Équations d'évolution

Diaraf SECK

Cédric VILLANI

Département Mathématiques de la Décision, Faculté des Sciences Économiques et de Gestion, Université Cheikh Anta Diop, Sénégal

E-mail address: diaraf.seck@ucad.edu.sn, dseck@ucad.sn

Université de Lyon & Institut Henri Poincaré (CNRS/UPMC), 11 rue Pierre et Marie Curie, F-75231 Paris Cedex 05, France

E-mail address: villani@ihp.fr, villani@math.univ-lyon1.fr

Table des matières

Résumé. Dans ces notes on trouvera une introduction aux concepts et techniques des équations d'évolution. Chacune des trois parties constitue un cours en soi : la première est consacrée aux équations différentielles, les deux suivantes aux équations aux dérivées partielles. On se concentrera sur l'analyse mathématique, tout en évoquant au passage la modélisation et la simulation numérique.

Introduction : Modèles et équations

La démarche scientifique cherche à décrire, comprendre et agir sur le monde qui nous entoure. Dans de nombreux cas, cela passe par la **prédiction** de phénomènes naturels variés : l'orbite d'une planète, la trajectoire d'un boulet de canon ou d'une fusée, l'évaluation de la température ambiante dans 48 heures, le courant qui va s'établir dans un cours d'eau, le champ électrique au sein d'une expérience de physique, etc. La mathématique joue un rôle fondamental dans le traitement de ces questions, qui en ont constitué une motivation majeure.

En mathématique, la phase de description passe par l'établissement d'un **modèle mathématique**, idéalisation ou reflet mathématique abstrait plus ou moins complexe d'un certain phénomène. Même si cette description est abstraite, le phénomène représenté est souvent bien réel ; et les applications de la démarche peuvent être, elles aussi, très concrètes. En fait, l'impact de la mathématique dans l'économie mondiale est de plus en plus important au fur et à mesure que se développe l'informatique, caisse de résonance sans pareil pour les techniques mathématiques.

L'ordre des étapes telles que nous les avons indiquées plus haut (décrire ; comprendre ; agir) est le plus naturel, mais en pratique les choses se passent rarement ainsi... on constate plutôt des aller-retour incessants entre ces trois étapes. En outre la phase de compréhension, particulièrement chère aux mathématiciens, peut être prise à des degrés très variés d'exigence. Pour prendre un cas frappant parmi d'innombrables exemples : le défibrillateur cardiaque sauve des vies tous les jours, mais personne ne sait vraiment "pourquoi ça marche"... Et pour rester sur cette même thématique vitale : malgré des modélisations sophistiquées mélangeant mathématique, physique et biologie, on ne sait toujours pas expliquer de manière satisfaisante la mise en place des battements cardiaques.

L'art de la prédiction a été révolutionné par un événement majeur de l'histoire des sciences, survenu au milieu du 17ème siècle : la naissance des **équations différentielles**, et plus généralement des **équations d'évolution**. Ces équations cherchent à prédire mathématiquement l'évolution de phénomènes en étudiant leurs "tendances" ou "variations infinitésimales". Il est remarquable, mais pas étonnant, que ces équations soient apparues dans la foulée de la naissance du calcul différentiel. Les hérauts de cette révolution étaient tout à la fois mathématiciens et physiciens, voire philosophes : Newton, Leibniz, Huygens... Avec eux on put maîtriser la trajectoire des boulets de canon, des planètes, et de bien d'autres systèmes mécaniques. Dès la fin du 17ème siècle, les célèbres frères Bernoulli enseignaient des cours d'équations différentielles dans diverses villes d'Europe.

L'un des premiers succès pratiques du calcul différentiel fut la solution par Huygens du problème des "oscillations tautochrones" (oscillations de période invariante). Le mathématicien néerlandais montra en effet qu'une bille roulant sans frottement sur une courbe cycloïde était soumise à des oscillations dont la période était indépendante aussi bien de la masse de la bille que de son énergie ; indépendante, en particulier, de l'amplitude des oscillations. En s'inspirant de cette découverte, il découvrit le moyen de fabriquer des horloges d'une précision inimaginable jusqu'alors.

Un peu moins d'un siècle après l'invention des équations différentielles, survint une seconde révolution conceptuelle et scientifique : l'apparition d'équations d'évolution portant sur tout un champ, toute une fonction inconnue. Avec Euler, D'Alembert, puis Lagrange, Laplace et d'autres, on se prit ainsi à rêver de prédire le mouvement insaisissable des fluides. Cette approche remportera des succès considérables tout au long du 19ème siècle, avec entre autres : les équations de Fourier qui régissent les transferts de chaleur ; les équations de Navier–Stokes qui constituent aujourd'hui la base de toute la simulation des fluides ; les équations de Maxwell qui régissent l'électromagnétisme. Les communications transatlantiques n'auraient pu être réalisées sans une profonde compréhension des équations aux dérivées partielles... En 1890, dans un article visionnaire, Henri Poincaré commençait à parler de classifier les grandes équations de la physique mathématique, et anticipa le développement extraordinaire de la théorie des équations aux dérivées partielles au 20ème siècle.

La troisième révolution scientifique qui ponctue l'histoire des équations d'évolution, c'est le développement de l'informatique à partir des années 1950. De nombreuses équations, qui jusqu'alors ne pouvaient être étudiées que qualitativement ou dans des cas particuliers, se sont alors retrouvées accessibles au calcul approché par les ordinateurs, avec une précision qui ne cessa de croître. La simulation numérique s'affirma alors comme une composante majeure de la science et de l'industrie, en même temps que se développait l'analyse numérique, interface entre théorie mathématique et le calcul.

Aujourd'hui le domaine des équations d'évolution reste en expansion rapide, et il s'est imposé dans notre univers : ces équations ont envahi toutes les sciences et toute l'industrie, et elles alimentent également les nouvelles spéculations de la physique théorique.

Ce cours présente une introduction au sujet, ou plutôt une initiation, tant la matière est vaste. Nous allons commencer par décrire quelques domaines d'application, puis consacrer du temps, d'abord à

la théorie des équations différentielles dites “ordinaires”, puis à celle, bien plus vaste et complexe, des équations aux dérivées partielles.

Faisons quelques remarques finales avant d'entrer plus dans le vif du sujet.

Première remarque : le sujet est d'une complexité extrême, qui se mesure par exemple au temps très long de résolution des problèmes fondamentaux. La stabilité du système solaire préoccupait déjà Newton au 17ème siècle, mais ce n'est qu'après plus de trois siècles, plusieurs théories mathématiques (Laplace–Lagrange, Poincaré, Kolmogorov–Arnold–Moser, Laskar–Tremaine) et l'exploitation intensive des ordinateurs que l'on pense aujourd'hui avoir compris ce problème. L'existence de solutions aux équations de la mécanique des fluides, mentionnée par Euler au milieu du 18ème siècle comme une difficulté à traiter, est toujours aujourd'hui un problème ouvert formidable. Quant aux simulations informatiques, si puissantes soient-elles, elles restent complètement inefficaces pour traiter des sujets d'avant-garde physique comme les tentatives de fusion nucléaire... Mais, si difficile soit-elle, la théorie comprend aussi de nombreux principes simples et marquants.

Seconde remarque : il importe d'appréhender le sujet “à différents niveaux”. Face à un problème physique complexe, il sera important de faire la différence entre ce qui est bien compris et mathématiquement démontré, ce qui est conjecturé, et ce qui semble découler des essais, des expériences ou des simulations numériques. De même, une bonne compréhension de l'analyse numérique peut être vitale à l'obtention de simulations réalistes : il ne faut pas croire que l'ordinateur va “de lui-même” réaliser les bonnes approximations. Si l'on fait une erreur d'appréciation sur l'algorithme par lequel on lui demande de calculer les solutions d'une équation, on pourra aboutir, de bonne foi, à des résultats complètement aberrants...

Troisième remarque : on peut se poser des questions sur l'ambition que l'on peut avoir face à un sujet qui est devenu aussi varié et multiforme que les équations d'évolution. Un ouvrage de référence sur le sujet devrait compter au moins 50 000 pages, et nul n'a l'ambition de l'écrire ! On trouve de beaux ouvrages synthétiques et assez complets dont la taille avoisine “seulement” les 500 pages... Cependant, le plus important est de bien comprendre la démarche, certains exemples emblématiques, et certaines théories qui font partie du “fonds commun” ; ensuite on pourra, au cas par cas, se plonger dans la littérature correspondante. Nous donnerons au fur et à mesure quelques références bibliographiques.

Ces remarques faites, nous allons passer en revue quelques exemples frappants où les équations d'évolution ont pleinement joué leur rôle.

Les notions ne seront pas toutes expliquées en grand détail, il s'agit dans ces exemples de gagner un peu en "culture générale" et d'avoir une idée plus précise du type de questions qui a pu se poser. Une fois cela effectué, on reprendra une approche bien plus progressive et systématique.

1. Exemple : La mécanique céleste

Vers 1665, Isaac Newton fait une découverte capitale, emblématique : la loi de la gravitation universelle. Selon cette loi, deux objets massifs, quels qu'ils soient, s'attirent l'un l'autre ; et la force qui s'exerce entre deux objets "ponctuels", de masses respectives m et m', éloignés d'une distance r, est proportionnelle à mm'/r^2. Ainsi, par exemple, le Soleil attire la Terre (et réciproquement) ; et comme la distance Terre-Soleil est très grande devant la taille de la Terre (d'un facteur 23000 environ) et même devant la taille du Soleil (d'un facteur 100 environ), nous pouvons, en première approximation, négliger la géométrie de ces corps célestes pour calculer l'amplitude de la force d'attraction. Cette dernière sera donc fort bien approchée par $\mathcal{G}Mm/r^2$, où $\mathcal{G}$ est la constante de gravitation de Newton, M la masse du Soleil, m la masse de la Terre, et r la distance Terre-Soleil.

L'idée d'une attraction de la Terre sur les objets qui nous entourent n'est pas neuve : cette force d'attraction est en fait l'un des principaux facteurs de stabilité de notre univers quotidien. Mais ce qui est nouveau, c'est l'idée que cette propriété soit commune à *tous* les corps massifs, et s'effectue selon une loi universelle.

Un premier succès de Newton fut de retrouver, à partir de cette seule loi, les trois lois dégagées par Johannes Kepler au début du 17ème siècle. Kepler s'était notamment rendu célèbre en découvrant, à la suite des observations de Mars, que la trajectoire d'une planète autour du Soleil décrit une ellipse. Rappelons que si l'on se donne un plan P, deux points F et F' (les foyers) dans P, et un nombre $a \geq |F - F'|/2$, alors l'ellipse $\mathcal{E}$ de foyers F et F' et de rayon a dans le plan P est le lieu des points $x \in P$ tels que $|x - F| + |x - F'| = 2a$. Il est aisé de construire $\mathcal{E}$ géométriquement, et on montre que cette courbe, connue depuis l'Antiquité grecque, vérifie une équation algébrique du second degré. L'*excentricité* $e = |F - F'|/(2a)$ mesure la non-circularité de l'ellipse : ce paramètre est très petit pour la Terre (environ 0.017), mais moins pour Mars (environ 0.093), et va jusqu'à 0.25 dans le cas de Pluton.

En termes mathématiques, voici ce que prouve Newton, sous une forme très légèrement simplifiée. Soit $x = x(t)$ un point dépendant

du temps, vérifiant l'équation $\ddot{x}(t) = -Kx(t)/|x(t)|^3$, où $K > 0$ est une constante (l'accélération est toujours dirigée vers l'origine, et inversement proportionnelle au carré de la distance à l'origine). Ici $x(t)$ représente la position d'une planète en orbite autour du Soleil, supposé immobile et placé à l'origine ; l'accélération de la planète est donc proportionnelle à la force que le Soleil exerce sur elle. Alors on peut *démontrer* que $x(t)$ décrit une trajectoire elliptique, dont l'origine est un foyer (première loi de Kepler). En outre, si l'on se place dans le plan de cette ellipse, l'aire délimitée par (i) l'axe des foyers, (ii) l'axe $Ox(t)$ et (iii) l'ellipse croît comme une fonction linéaire du temps (deuxième loi de Kepler). Enfin, si l'on note a le demi-grand axe (c'est à dire, la moitié de la distance entre les deux points les plus écartés de l'ellipse ; ou tout simplement le rayon, si cette ellipse est un cercle), et T la période de révolution (la période que met la trajectoire $x(t)$ pour revenir à son point de départ), alors T^2/a^3 est une constante qui ne dépend que de K (troisième loi de Kepler).

Le théorème précédent est énoncé dans l'approximation d'un Soleil si lourd qu'il ne bouge pas du tout sous l'action de la gravitation ; cela n'est pas très éloigné de la vérité puisque le Soleil est environ 1000 fois plus massif que Jupiter, la plus grosse des planètes de notre système. Mais même si l'on tient compte de la correction due à ce que le Soleil est de masse finie, on aboutit aux mêmes conclusions, quitte à changer de référentiel.

Newton va ensuite plus loin : il peut écrire maintenant un système complet décrivant les mouvements des planètes, tenant compte des interactions entre planètes, si faibles soient-elles. Si l'on note x_0 la position du Soleil, $(x_j)_{1\leq j\leq N}$ les positions des planètes numérotées (par exemple x_1 la position de Mercure, x_2 la position de Vénus, etc.), et m_j les masses correspondantes (m_0 la masse du Soleil, m_1 la masse de Mercure, etc.), alors

$$\forall i, \qquad \ddot{x}_i(t) = -\mathcal{G} m_j \sum_{j\neq i} \frac{(x_i(t) - x_j(t))}{|x_i(t) - x_j(t)|^3}. \tag{1}$$

On peut compliquer le modèle et inclure dans (1), si on le souhaite, les satellites comme la Lune, des astéroïdes, etc. En fait il faudrait écrire le système complet avec tous les astres de l'Univers ! Mais en pratique, en incluant juste les planètes et la lune, on obtient une très bonne approximation de la réalité. En fait à la fin du 18ème siècle, les astronomes comme Laplace montrent que le système de Newton permet de retrouver *toutes* les observations astronomiques connues à l'époque. Des siècles de relevés de positions résumés en une seule équation : c'était

l'un des premiers triomphes des équations d'évolution. En fait il faudra attendre le milieu du dix-neuvième siècle pour que l'on observe une petite déviation au modèle de Newton : ce sera la fameuse "avance de périhélie" de Mercure, que la relativité générale d'Einstein permettra d'expliquer...

En même temps, la loi de Newton amène des difficultés conceptuelles profondes, qui montrent bien combien cela peut être subtil de "comprendre" un modèle. L'équation (1) est destinée à prédire l'évolution du système solaire : que peut-on en déduire sur le futur *à long terme* de ce système ? Et tout d'abord, est-ce que le système solaire restera tel que nous le connaissons ? Ou est-ce que dans un futur lointain, par exemple, Mars ne va pas venir se fracasser sur la Terre, forçant à un changement complet de modèle ? (en supposant qu'il y ait encore, quelque part dans l'Univers, quelqu'un qui soit intéressé à modéliser cela après la destruction de la Terre...) Ce cataclysme correspondrait à une situation où $x_3(t)$ et $x_4(t)$, disons, deviennent si proches que l'on ne peut plus considérer la Terre et Mars comme des masses ponctuelles bien séparées... Ou encore, est-ce que Mercure ne va pas s'effondrer sur le Soleil ? Avec les lois de Kepler, de tels cataclysmes sont impossibles ; mais ces lois ne sont que des lois approchées, et subissent de petites déviations prises en compte dans (1).

Nous sommes ainsi arrivés au problème de la **stabilité du système solaire**, l'un des plus anciens problèmes de la physique mathématique : étant donnés des conditions initiales (les positions et vitesses des planètes à notre époque, disons), peut-on montrer que les solutions du système (1) resteront stables, c'est à dire proches de l'existant ; ou qu'au contraire elles subiront des déviations importantes, menant possiblement à un événement catastrophique ? Et dans ce dernier cas, à quelle échelle de temps ? On peut, pour simplifier le problème, réduire le nombre N de planètes (par exemple en ne gardant que les effets les plus importants, comme celui qui est dû à Jupiter) ; supposer l'excentricité des planètes très faible (en pratique elle est effectivement assez faible, sauf bien sûr pour la planète naine Pluton) ; et supposer les orbites presque coplanaires (elles ne sont pas loin de l'être) : le problème reste pour autant très difficile dès que l'on a trois corps ou plus, c'est à dire dès que $N \geq 2$.

La déviation des lois de Kepler est due aux interactions planète-planète, qui sont proportionnellement faibles car la masse des planètes est environ 1000 fois plus faible que la masse du Soleil. Il n'empêche : si dans un problème on a une petite perturbation d'ordre $\varepsilon \simeq 10^{-3}$, il est naturel de penser qu'après un temps $1/\varepsilon$ cette petite perturbation a eu le temps de se transformer en une variation considérable (si votre

compte en banque perd 1% chaque jour, au bout de 100 jours il a très sensiblement diminué !). Or l'échelle de temps naturelle, dans le cas du système solaire ou tout au moins de la Terre, est l'année (période de révolution) : on pourrait donc s'attendre à une perturbation énorme au bout d'un millier d'années... Newton avait bien conscience de cette difficulté, et pensait qu'en conséquence les équations (1) n'étaient pas seules responsables de l'évolution des planètes sur le long terme : il invoquait donc, en sus des lois de la physique, un contrôle d'origine divine ; cela fut le départ d'une âpre controverse avec Leibniz, dont des échos moqueurs se retrouvent jusque dans les écrits de Voltaire au 18ème siècle...

À la fin du 18ème siècle cependant, Laplace et Lagrange, dans l'un des plus remarquables exemples de compétition-collaboration de l'époque, parviennent à résoudre le paradoxe sur lequel Newton a buté. Par une analyse fine de l'équation (1), ils prouvent que dans un régime de petite perturbation, il faut attendre un temps non pas $1/\varepsilon$, mais $1/\varepsilon^2$, pour que des déviations sensibles se produisent ! Cela est dû à un phénomène de compensations : si l'on considère par exemple l'influence de Jupiter sur la Terre, Jupiter sera parfois opposée à la Terre (par rapport au Soleil), et parfois dans l'alignement de la Terre, de sorte que Jupiter aura tendance parfois à rapprocher la Terre du Soleil, parfois à l'en éloigner... et que globalement les effets se compenseront presque "en moyenne". Ce raisonnement fonctionne bien, en fait, *sauf* si Jupiter et la Terre se retrouvent en "résonance", c'est à dire que l'année jupitérienne est un multiple rationnel de l'année terrestre, disons $T_{\text{Jupiter}} = (p/q)T_{\text{Terre}}$. En effet, si tel est le cas, toutes les p années, Jupiter et la Terre se retrouveraient (presque) dans la même configuration, et de petites variations en viendraient à s'amplifier rapidement, puisqu'elles se répéteraient toujours dans le même sens avec une périodicité de q années. Le théorème de Laplace–Lagrange s'applique donc en l'absence de résonances, c'est à dire si toutes les années planétaires sont rationnellement indépendantes. On trouve alors cette échelle de temps de stabilité, en $1/\varepsilon^2$, donc de l'ordre du million d'années.

On objectera qu'il est impossible de vérifier en pratique que des années planétaires sont rationnellement indépendantes, la précision de la mesure étant forcément limitée à un nombre fini de chiffres ! Le point important est que si les entiers en jeu (comme les entiers p et q ci-dessous) sont grands, alors on ne sentira pas ou presque pas les effets de cette dépendance.

Saluons ces astronomes de la fin du 18ème siècle qui avaient réussi à prédire qualitativement le devenir du système solaire sur une période

aussi incroyablement longue qu'un million d'années ! Leur solution allait même au-delà des observations : il semblait en effet que Jupiter se rapprochait lentement mais continûment du Soleil, tandis que Saturne s'en écartait inexorablement. Mais on put montrer que c'était un mouvement de durée "pas trop longue", qui s'inversait tous les 800 ans environ, et qui était dû à ce que Jupiter et Saturne ne sont pas loin d'être en résonance (le rapport de leurs périodes est d'environ $0,403$ qui est proche de $0,4 = 2/5$!).

C'est également à l'aide d'une analyse fine des équations de Newton, et l'observation du mouvement d'Uranus, que les astronomes Urbain Le Verrier et John Adams, indépendamment, prédirent par le calcul l'existence de la planète Neptune, qui fut effectivement observée plus tard. Une fois encore, les équations d'évolution avaient permis de découvrir un phénomène avant qu'il ne soit observé ! Mais cette fois sous un angle très intéressant : en recherchant un objet inconnu qui expliquerait, via le modèle, les observations. On parle dans ce cas de *problème inverse*.

L'histoire ne s'arrête pas là. L'analyse de Laplace et Lagrange se heurte à une impasse quand on cherche à la prolonger sur des temps bien plus longs que le million d'années : des corrections de plus en plus fines peuvent s'accumuler sur des temps énormes, conduisant à des situations similaires à des résonances. Cette difficulté majeure a empoisonné la vie des astronomes du 19ème siècle sous le nom de "petits diviseurs", car elle mène dans les équations à l'apparition de dénominateurs très petits. On sentait alors que le problème demanderait la mise au point de nouveaux outils conceptuels ; on ne fut pas déçu. Cette étude de stabilité du système solaire en temps très grand a en effet mené Henri Poincaré au phénomène d'intersection homocline, à la sensibilité aux conditions initiales, et partant de là à certaines des idées fondatrices de la théorie moderne des systèmes dynamiques, incluant la théorie du chaos et suggérant l'instabilité à long terme du système solaire. Cela mena aussi le mathématicien russe Andrei Kolmogorov à fonder une nouvelle théorie perturbative d'une puissance redoutable, suggérant au contraire l'instabilité de ce système. Et il fallut attendre les années 1990 pour que la thèse de l'instabilité soit vérifiée informatiquement par ordinateur, grâce aux calculs indépendants de Jacques Laskar et Scott Tremaine, et de nouvelles avancées audacieuses dans l'analyse numérique... Sur cette épopée, la lectrice pourra consulter avec profit le très bref et élémentaire article de revue d'Étienne Ghys [**16**].

Aux dernières nouvelles, (i) le système solaire est instable sur des périodes de temps de plus de 60 millions d'années ; (ii) la pire cause d'instabilité recensée dans le système solaire est liée à une quasi-résonance de deux astéroïdes, Ceres et Vesta, qui, bien que gros pour des astéroïdes, n'en sont pas moins microscopiques à l'échelle du Soleil, représentant environ un *milliardième* de la masse de ce dernier ! (Newton les aurait négligés sans le moindre état d'âme, et pourtant ils ont le pouvoir d'influer drastiquement sur la trajectoire de la Terre à horizon de 60 millions d'années environ...) ; (iii) le système solaire a une chance petite, mais pas si faible — de l'ordre de 1 ou 2% peut-être — de connaître une catastrophe majeure d'ici quelques milliards d'années. Autant de conclusions incroyables que l'on retrouvera dans les articles de synthèse de Laskar[**25, 26**], et qui montrent combien il est difficile de vraiment "comprendre" le système (1).

Cela dit, on n'avait pas attendu cela pour commencer à agir ! Au vingtième siècle, les équations de Newton ont joué un rôle capital dans la conquête de l'espace : elles régissent en effet les trajectoires des fusées, satellites, etc. – avec bien sûr les modifications liées à la prise en compte des forces de propulsion. Le fait que l'on ait pu envoyer un homme sur la Lune et le ramener sain et sauf ; et, plus récemment, que l'on ait pu envoyer un robot se poser sur une comète, représentent des triomphes de la théorie mathématique initiée par Newton.

En outre, et cela est fréquent en mathématique, les théories développées dans l'étude de ce problème (théorie perturbative de Laplace–Lagrange, théorie du chaos, théorie de Kolmogorov–Arnold–Moser...) sont venues enrichir un arsenal théorique dont les applications ont des conséquences bien plus vastes que la simple prédiction céleste !

2. Exemple : Mécanique des fluides

La maîtrise de l'évolution des fluides constitue une autre grande conquête de la science. Pendant des milliers d'années, on n'a pu décrire les flots autrement que par des analogies et des termes vagues ; mais à partir du milieu du 18ème siècle on a commencé à traduire ce mouvement en équations mathématiques. Si l'histoire commence encore avec les frères Jakob et Johann Bernoulli, c'est Leonard Euler, le plus célèbre élève de Johann et l'un des plus grands mathématiciens de tous les temps, qui a eu l'honneur d'écrire de manière explicite le plus ancien de ces modèles, **l'équation d'Euler incompressible** :

$$\begin{cases} \dfrac{\partial u}{\partial t} + u \cdot \nabla u + \nabla p = 0 \\ \nabla \cdot u = 0. \end{cases} \tag{2}$$

Ici $u = u(t, x)$ est le champ de vitesses du fluide, c'est une fonction à valeurs vectorielles dépendant du temps t et de la position x (qui appartient, disons, à l'espace $\mathbb{R}^3$) ; et p est une fonction à valeurs réelles, également dépendant de t et x, appelée pression. Le premier terme $\partial u/\partial t$ représente la variation infinitésimale (en t) de u ; et l'on utilise les abréviations bien commodes $(u \cdot \nabla u)_i = \sum_j (u_j \, \partial/\partial x_j) u_i$, $(\nabla p)_i = \partial p/\partial x_i$, $\nabla \cdot u = \sum \partial u_i/\partial x_i$. On note que l'on ne dit rien sur la variation de p, de sorte qu'il semble manquer une équation... en fait cela est caché dans la condition $\nabla \cdot u = 0$, qui serait impossible à satisfaire si l'on ne permettait pas à p de varier sans contrainte. Cette condition, dite condition d'incompressibilité, traduit le fait que le volume d'un ensemble de particules transportées par le fluide est préservé au cours du temps.

L'équation (2) est un exemple typique d'équation d'évolution : elle permet de prédire ce qu'il adviendra du champ de vitesses quand on imprime un mouvement au fluide puis qu'on le laisse évoluer à son gré... Ici on a supposé que le fluide occupe tout l'espace, ce qui est bien sûr irréaliste, sauf peut-être pour modéliser des eaux profondes ; pour plus de réalisme, il faudrait tenir compte du récipient qui contient le fluide, introduire des conditions au bord pour modéliser l'interaction entre le fluide et la paroi, etc.

L'histoire dit qu'Euler écrivit le système (2) pour aider à la réparation d'une fontaine défectueuse, que pour autant il ne réussit jamais à faire fonctionner... il faut dire que le modèle (2) ne résout pas tout, et apporte avec lui des difficultés considérables ! Tout d'abord, cette équation contient des variations infinitésimales, comme il se doit, mais pas seulement par rapport au temps ($\partial u/\partial t$) : il s'y trouve également des dérivées spatiales, c'est à dire par rapport aux variables x_i. On parle d'**équation aux dérivées partielles**, et la théorie en est bien plus délicate que celle des équations différentielles considérées jusqu'alors. Si délicate en fait que l'on ne sait toujours pas, en 2015, si l'on peut, pour le système (2), construire des solutions régulières ou non ! !

À ce sujet, on peut même dire que nous semblons régresser, car l'incertitude a tendance à croître. Il y a quelques décennies, la majorité des spécialistes auraient bien parié que les solutions de (2) ne sont pas régulières ; mais depuis lors, les simulations numériques se sont révélées incapables de fournir un seul exemple de solutions singulières, menant une proportion non négligeable de mathématiciens et physiciens à douter de leur intuition... quand d'autres disent que de toute façon il est difficile de tirer des conclusions qualitatives fines de l'étude numérique, bourrée de chausse-trappes et d'illusions (on a ainsi vu par le passé des

"conjectures numériques" de singularité s'écrouler spectaculairement face à un théorème mathématique).

Une autre difficulté du modèle (2) est qu'il mène à des conclusions paradoxales, voire incroyables. La plus notable de ces anomalies est le **paradoxe de D'Alembert :** partons pour simplifier d'un écoulement irrotationnel, c'est à dire sans aucun tourbillon. Alors un objet qui avance à vitesse constante dans cet écoulement ne sentira aucune résistance de la part du fluide. Cela est tout à fait contraire à notre expérience : quand on avance dans une piscine, on sent bien la résistance de l'eau ! En fait sans résistance fluide, bien des choses ne fonctionneraient pas, et pour commencer tous les avions et oiseaux s'écraseraient lourdement au sol...

La solution de ce paradoxe réside dans le fait que le modèle (2) est impuissant à décrire les phénomènes fins qui se produisent au contact entre l'objet et le fluide dans lequel il est plongé. Et pour le surmonter, il faut donc changer de modèle, par exemple en tenant compte de la *viscosité* du fluide, c'est à dire sa capacité à être le siège de frottements internes ; on obtient ainsi l'équation dite de **Navier–Stokes incompressible**,

$$\begin{cases} \dfrac{\partial u}{\partial t} + u \cdot \nabla u + \nabla p = \nu \Delta u \\ \nabla \cdot u = 0, \end{cases} \tag{3}$$

où $\nu > 0$ est le coefficient de viscosité, et $(\Delta u)_i = \sum_j \partial^2 u_i / \partial x_j^2$. Avec cette équation, il y a spontanément génération de tourbillons à petite échelle, au contact entre le solide et le fluide, et le paradoxe de D'Alembert n'est plus un problème !

En fait il existe de nombreux types d'écoulements fluides, d'interactions, de conditions au bord, etc. Cela fait toute une zoologie d'équations fluides : compressible ou incompressible ; visqueux ou non visqueux ; en dimension 1, 2 ou 3 ; newtoniens ou non newtoniens ; et ainsi de suite. On trouvera un survol dans l'épais *Handbook of Mathematical Fluid Mechanics* de Friedlander et Serre [**13**], et une introduction élémentaire dans l'ouvrage classique de Chorin et Marsden [**6**]. Les applications à l'industrie sont considérables, en particulier dans le secteur naval ou aéronautique, et bien sûr la météorologie qui repose fondamentalement sur la mécanique des fluides...

Les mystères de ce domaine sont encore bien plus considérables que ceux de la mécanique céleste. Par exemple, c'est un problème mis à prix pour 1 million de dollars que de savoir si les solutions de l'équation (3)

sont régulières... et l'on ne conseille pas vraiment cette énigme au lecteur soucieux de s'enrichir ! Encore plus mystérieux est le sujet de la turbulence [**14**] : peut-on prédire les propriétés statistiques des fluctuations d'un fluide peu visqueux subissant un forçage à grande échelle ? La recherche dans ce sujet fondamental est moins avancée que dans la mécanique quantique !

Pourtant, la mécanique des fluides continue à faire des progrès, en particulier là où on ne l'attend pas. Ces dernières années, on a montré comment construire des solutions très peu régulières de l'équation d'Euler (2), avec des propriétés paradoxales incroyables... Par exemple, en fonction du temps t, on peut obtenir le comportement "pathologique" que voici : pour $0 \leq t \leq 1$, le fluide est entièrement au repos ; pour $1 < t < 2$, le fluide est au-repos en-dehors de la boule $|x| \leq 1$ et sa vitesse $|u|$ est égale à 1 partout à l'intérieur de cette boule ; puis pour $t \geq 2$ le fluide retourne subitement dans un état de repos absolu, par la seule action du système (2). Euler lui-même n'en serait pas revenu ! On trouvera dans [**37**] des explications sur ce phénomène, le "paradoxe de Scheffer–Shnirelman", qui est de mieux en mieux compris de nos jours, au point que l'on commence à y entrevoir ce qui pourrait constituer une limitation intrinsèque à la régularité des solutions typiques de (2)...

Et en même temps, on trouve de plus en plus de modèles nouveaux, décrivant par exemple des fluides complexes, des mélanges, des aérosols, des interactions subtiles entre les flots et les dunes, avec des applications par exemple en écologie. La mécanique des fluides, c'est devenu une science en soi !

3. Exemple : L'équation de Boltzmann

Dans la section précédente nous avons évoqué les équations d'Euler, Navier–Stokes et autres... toutes ces équations sont dites **hydrodynamiques** : elles s'appliquent à des fluides suffisamment denses, comme de l'eau, mais pas aux fluides très dilués comme de l'air en atmosphère raréfiée.

Or il y a une différence majeure entre les fluides denses et les fluides dilués ; elle se comprend au niveau des interactions microscopiques, quand on se souvient de la structure atomique de la matière. Dans un fluide dense, les collisions entre particules sont incessantes, et les lois statistiques s'appliquent très bien : en particulier, la distribution des vitesses des particules dans un tel fluide est très bien décrite par une courbe gaussienne. Mais si le fluide est peu dilué, cette conclusion n'est plus forcément vraie, et cela change la physique ! On abandonne alors la modélisation hydrodynamique pour une description plus précise, dite

cinétique, qui tient compte de ce que les vitesses ne sont pas forcément distribuées de manière gaussienne.

Dans le formalisme cinétique on se préoccupe de la distribution des positions x et des vitesses v des particules : l'inconnue est donc une fonction $f(t,x,v)$ qui représente, pour chaque temps t, la densité de particules dans l'espace des positions x et des vitesses v. Ou encore : si A est un ensemble de positions et de vitesses, alors $\int_A f(t,x,v)\,dx\,dv$ indique quelle est la quantité de gaz dont les particules sont situées dans A au temps t. On appelle équation cinétique un modèle destiné à prédire l'évolution de la distribution $f(t,x,v)$ en fonction de la donnée initiale $f(0,x,v)$ et du temps t ; cette équation dépendra de la nature des interactions entre particules.

La théorie cinétique demande une certaine sophistication conceptuelle ; si l'on remonte à sa source, on trouve à nouveau un Bernoulli ! Il s'agit cette fois de Daniel, le fils de Johann. Mais ce n'est qu'avec James Clerk Maxwell et Ludwig Boltzmann, entre 1865 et 1875, que le sujet se développe spectaculairement ; et avec lui, une nouvelle révolution scientifique, celle de la physique statistique, dont le but est d'expliquer et prédire des propriétés complexes de la matière à notre échelle par le fait qu'elle est constituée de très nombreuses particules microscopiques. Ce formalisme peut s'appliquer aussi bien à un gaz constitué d'innombrables molécules, qu'à un plasma fait d'électrons et de noyaux atomiques, à une galaxie faite de dizaines de milliards d'étoiles, à un essaim de centaines d'oiseaux, ou à un embouteillage faisant intervenir des milliers de voitures...

Le plus célèbre modèle cinétique est l'**équation de Boltzmann**, mise au point implicitement par Maxwell vers 1865, avant que Boltzmann ne lui donne sa forme définitive quelques années plus tard. Dans ce modèle, la fonction de distribution f évolue sous l'effet des collisions chaotiques et désordonnées des molécules entre elles. L'équation s'écrit

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f = Q(f,f) \tag{4}$$

où $(v\cdot\nabla f) = \sum_i v_i\,\partial f/\partial x_i$ et

$$Q(f,f) = \int_{\mathbb{R}^3}\int_{S^2} B(v-v_*,\sigma)\big[f(t,x,v')f(t,x,v'_*) - f(t,x,v)f(t,x,v_*)\big]\,dv_*\,d\sigma.$$

Le premier terme $\partial f/\partial t$ est la variation temporelle de la densité, le second $v\cdot\nabla_x f$ est le terme de transport, qui traduit la tendance spontanée des particules à un mouvement rectiligne uniforme, et le dernier $Q(f,f)$ est le terme de collision, qui exprime l'effet des collisions d'une particule sur une autre. En outre, v et v_* peuvent être considérées comme les

vitesses après collision de deux particules qui avant de s'entrechoquer avaient pour vitesses respectives v' et v'_*, et σ est une paramétrisation du choc :

$$v' = \frac{v+v_*}{2} + \frac{|v-v_*|}{2}\,\sigma, \qquad v'_* = \frac{v+v_*}{2} - \frac{|v-v_*|}{2}\,\sigma.$$

Enfin, $B(v-v_*,\sigma)$, le "noyau de collision de Boltzmann", indique quels sont les paramètres σ qui sont favorisés par l'interaction microscopique.

On note que l'opérateur de collision Q est quadratique, ce qui traduit une propriété physique importante : la décorrélation des vitesses avant collision, aussi appelée hypothèse de *chaos moléculaire* de Boltzmann : c'est comme si "deux particules sur le point d'entrer en collision ne se connaissaient pas" ; la multiplication de f par elle-même est alors un écho de la propriété de multiplication des densités des variables aléatoires indépendantes. On imagine bien que la justification de cette propriété d'indépendance pose de douloureux problèmes conceptuels, et de fait il a fallu attendre un siècle avant qu'Oscar Lanford, vers 1973, montre que l'équation de Boltzmann est mathématiquement cohérente avec la physique de Newton ! Et encore, sa démonstration ne s'applique que sous certaines hypothèses restrictives...

L'équation (4) peut sembler effrayante au premier abord, et elle l'est effectivement... tout oppose le terme de transport au terme de collision : l'un est linéaire et différentiel, l'autre est quadratique et intégral ; l'un a des propriétés de mélange des variables de vitesse et de position, l'autre est indifférent à la variable de position. Ce sont là divers points saillants d'une théorie particulièrement délicate, dont on pourra trouver une présentation synthétique, pas encore trop obsolète, dans [**35**].

Mais au-delà, cette équation a permis l'une des avancées conceptuelles les plus importantes de tout le 19ème siècle. Boltzmann a en effet démontré que si l'on pose $H(f) = \int f(x,v)\log f(x,v)\,dx\,dv$, et que l'on laisse la distribution f évoluer au cours du temps selon son équation, modélisant un gaz enfermé dans un récipient, alors la quantité H décroît spontanément au cours du temps. Ce résultat, le *Théorème H*, est interprété en physique comme la croissance spontanée du désordre macroscopique, ou **entropie** (l'entropie est l'opposé de H) sous l'influence de phénomènes microscopiquement réversibles. Il a ainsi contribué à changer notre vision du monde et à expliquer une loi fondamentale de la thermodynamique, science des échanges d'énergie et de température. Plus encore, il a permis un progrès majeur dans l'une des plus fascinantes interrogations de la physique théorique : à quoi est due la sensation d'écoulement du temps ? Tout cela dans une équation d'évolution !

L'équation de Boltzmann a également été le point de départ de nombreux problèmes mathématiques : montrer qu'un gaz isolé a tendance à revenir spontanément vers un état d'équilibre ; montrer que dans un régime où les collisions sont nombreuses ce gaz est bien décrit par des équations hydrodynamiques ; étudier les échanges d'information au sein d'un gaz...

Et enfin, l'équation de Boltzmann a permis des progrès technologiques importants dans des disciplines où les équations hydrodynamiques sont insuffisantes, comme en aéronautique. Il s'agit là encore d'un outil à la fois très fondamental et très pratique.

4. Exemple : Le flot de Ricci

Les équations d'évolution ne servent pas seulement à résoudre des problèmes issus du monde physique ; souvent elles sont aussi utiles dans le monde mathématique abstrait. Elles peuvent en effet servir à faire évoluer des objets mathématiques selon des règles que l'on prescrit, en fonction d'un paramètre que l'on appellera "temps" par convention. Cette démarche est très fructueuse dans toutes les branches mathématiques, en particulier en analyse, probabilité et géométrie.

L'exemple le plus populaire est certainement l'action de la diffusion. Dans le monde physique, la diffusion est le phénomène qui mélange des systèmes statistiques comme des gaz : par exemple, si l'on émet un jet de parfum dans une pièce fermée, au départ le parfum sera concentré au voisinage de l'endroit d'émission ; mais si l'on attend suffisamment longtemps le parfum sera répandu dans toute la pièce, en concentration à peu près uniforme. La diffusion est, physiquement, causée par les mouvements incessants des particules transportant avec elles de la matière ou de l'énergie ; elle s'applique aussi bien pour de la chaleur qui se répand dans un métal que pour une goutte de lait que l'on verse dans du thé. Les équations qui la régissent sont assez variées, parfois linéaires et parfois non linéaires, avec des traits communs qui les font regrouper par les spécialistes dans la famille des "équations de diffusion"... De loin le plus célèbre de ces modèles est l'équation de la chaleur,

$$\partial_t f = \Delta_x f \tag{5}$$

qui agit sur une quantité $f = f(t, x)$ en l'homogénéisant par rapport à la variable x, au fur et à mesure que le temps passe. L'équation de la chaleur remonte aux travaux de Joseph Fourier au début du 19ème siècle ; elle jouit de nombreuses propriétés, mais elle a surtout une tendance à tout améliorer : elle rend les fonctions moins variables, plus lisses... Et dans de nombreux problèmes mathématiques, on fait agir

cette équation, ou ses variantes, sur des objets, par exemple pour les rendre plus réguliers.

L'une de ces variantes a connu récemment une célébrité folle quand elle a permis de résoudre l'un des problèmes mathématiques les plus célèbres de tous les temps. Pour expliquer un peu de quoi il s'agit, commençons d'abord par rappeler ce que c'est que la courbure : un objet mathématique remontant aux travaux des mathématiciens allemands Karl-Friedrich Gauss et Bernhard Riemann, qui permet de quantifier à quel point une géométrie est non euclidienne. Par exemple un plan est de courbure nulle ; une sphère est de courbure strictement positive et uniforme ; un visage (idéalisé) est fait de points où la courbure est positive (menton, bout du nez, front...) et de points où la courbure est négative (le creux entre les lèvres et le menton ; le sommet du nez ; etc.) La courbure est commode d'usage et permet d'exprimer de très nombreux théorèmes.

Pour des surfaces, objets bidimensionnels par nature, on utilise toujours la même notion de courbure, dite courbure de Gauss. Mais pour des géométries de dimension 3 et plus, il existe plusieurs notions de courbure, qui sont apparentées les unes aux autres. L'une d'entre elles, dite *courbure de Ricci*, joue un rôle fondamental dans la théorie de la relativité générale.

Une géométrie arbitraire sera associée à des variations, éventuellement fortes, de la courbure ; comment faire pour l'homogénéiser, la rendre "plus uniforme" ? La réponse consiste à faire agir sur la géométrie une variante non linéaire de l'équation de la chaleur, le **flot de Ricci**, qui s'écrit

$$\frac{\partial g}{\partial t} = -2\mathrm{Ric}_g, \tag{6}$$

où g, appelé tenseur métrique, permet de mesurer les longueurs, tandis que Ric_g est appelé tenseur de courbure de Ricci. Sans donner plus de détails, on peut exprimer (6) de manière imagée en disant qu'on étire la géométrie la où la courbure est négative, et qu'on la rétrécit là où la courbure est positive. Ce que l'on réalise ainsi revient à effectuer une diffusion de la courbure !

À quoi cela peut-il servir ? À des procédés de déformation en traitement d'image, certainement. Mais aussi, le flot de Ricci a permis à Grigori Perelman de démontrer, au début du 21ème siècle, la conjecture de Poincaré, un énoncé de géométrie vieux de presque un siècle. Il s'agissait de savoir si une géométrie compacte, sans bords, en trois dimensions, simplement connexe (c'est à dire dans laquelle toute chemin fermé peut être continûment déformé en un point) peut se déformer continûment

en une 3-sphère (c'est à dire la géométrie d'une sphère dans l'espace de dimension 4). Cette question avait été introduite par Henri Poincaré tout à la fin de la série d'articles dans laquelle il jetait les bases de la topologie : dans cette optique, il s'agissait d'une première étape vers ce que l'on pourrait appeler la classification de toutes les formes que peut adopter une géométrie bornée de dimension 3. Notons qu'avant Poincaré, la réponse à la question était déjà connue en dimension 2 ; et au cours du vingtième siècle, la réponse a été apportée, au prix de maints efforts, en dimensions 4 et supérieures ; mais en dimension 3, le problème restait particulièrement rebelle malgré des contributions spectaculaires comme celle de William Thurston.

Il faut noter aussi que la question de Poincaré était de nature purement topologique, ne faisant aucunement intervenir la notion de courbure ! Pourtant, Richard Hamilton avait avancé l'idée que l'on pourrait y répondre en utilisant le flot de Ricci, homogénéisant la courbure pour "simplifier" la géométrie au bout d'un temps suffisamment long. Le programme de Hamilton, après des débuts prometteurs, avait buté sur des obstacles d'analyse non linéaire très subtile, et ce fut une surprise considérable quand Perelman réussit à le faire aboutir après sept années de travail solitaire [**30**].

Au delà de la solution de cette conjecture fameuse, la contribution de Perelman a (re)démontré la puissance que peuvent parfois avoir les équations d'évolution dans des problèmes mathématiques où ils n'ont a priori aucune raison d'intervenir. Elle a ainsi consolidé leur statut comme un outil important, non seulement dans les applications, mais aussi dans les branches plus abstraites et théoriques de la mathématique.

5. Sur la nature des équations d'évolution

Dans cette introduction, nous avons évoqué plusieurs équations tirées de domaines fort différents : mécanique céleste, mécanique des fluides, physique statistique, analyse géométrique... Tous ces modèles peuvent se regrouper en deux grandes catégories :

- les équations différentielles "ordinaires" (**EDO**), qui régissent l'évolution temporelle d'un nombre fini de paramètres inconnus : par exemple, en mécanique céleste, les positions et vitesses de toutes les planètes et du soleil. Avec 3 coordonnées pour chaque position et 3 composantes pour chaque vitesse, cela fait bien une soixantaine de de paramètres : c'est beaucoup, mais cela reste un nombre fini.

- les équations aux dérivées partielles (**EDP**), qui régissent l'évolution temporelle de fonctions tout entières, dont la description fait intervenir

une infinité de paramètres. Ainsi, en mécanique des fluides, les fonctions inconnues peuvent être le champ de vitesses et la température du fluide ; si l'on représente ces fonctions inconnues (par exemple dans un contexte périodique) par leurs coefficients de Fourier, cela représentera une infinité dénombrable de quantités inconnues, réelles ou complexes. Ces équations font presque toujours intervenir les dérivées partielles des fonctions inconnues par rapport à diverses variables, d'où leur nom.

La théorie des équations aux dérivées partielles pose tout de suite des difficultés supplémentaires, ne serait-ce que parce qu'il faut donner un sens à ces dérivées partielles, alors que l'on travaille avec des fonctions dont on ne sait pas a priori qu'elles sont différentiables – et dont on sait parfois qu'elles ne le sont pas !

Historiquement, les EDO sont nées avec Newton et ses contemporains, mais ce n'est qu'avec Poincaré, à la fin du 19ème siècle, que commence leur étude qualitative systématique. Quant aux EDP, leur développement a été plus laborieux : apparues au milieu du 18ème siècle, elles ont pris leur véritable essor après la Seconde Guerre Mondiale, en même temps que l'informatique permettait leur simulation numérique. Les deux types d'équations sont cruciaux dans toutes les branches des sciences, dans l'environnement, dans l'industrie. L'analyse des EDO et celle des EDP présentent des similitudes importantes, et certains principes généraux en commun ; mais elles ont aussi des différences fondamentales. En voici une importante : il est possible d'aborder la théorie moderne des EDO dans son ensemble, de maniere quelque peu unifiée ; alors que la théorie des EDP est fractionné en de nombreuses branches et sous-specialités qui ont chacune leurs recettes propres. Une autre différence majeure est que la théorie moderne des équations différentielle est fortement empreinte de géométrie (et réciproquement, la géométrie différentielle repose fondamentalement sur des équations différentielles), alors que dans les théories des équations aux dérivées partielles c'est souvent l'analyse qui domine.

Ces notes sont divisées en trois parties principales : la première sera consacrée aux EDO, les deux suivantes aux EDP. Chaque partie fournit la matière d'un cours en soi.

Première partie

Équations Différentielles Ordinaires

On aborde dans cette partie l'étude générale des équations différentielles (EDO), en distinguant théorie locale (temps petits) et théorie globale (temps long). On passera en revue plusieurs équations et comportements emblématiques ; on terminera avec une introduction aux omniprésents systèmes hamiltoniens.

CHAPITRE 1

Un bon départ

1.1. Mise en place

Une équation différentielle est un modèle destiné à prédire l'évolution, au cours du temps, d'un système physique ou abstrait décrit par un nombre fini de paramètres. On appelle espace des états du système l'ensemble dans lequel cet état est a priori autorisé à varier, et dans lequel on le recherche ; cet espace est supposé de dimension finie en un certain sens. Ce pourra être, par exemple, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $\mathbb{C}^n$ ou tout espace vectoriel de dimension finie ; ou une variété de dimension finie ; ou encore tout sous-ensemble ouvert de l'un de ces espaces. L'inconnue sera donc une fonction, dépendant d'un paramètre temps, à valeurs dans l'espace des états.

L'équation différentielle en elle-même est une relation entre dérivées de la fonction inconnue $t \longmapsto x(t)$. Voici quelques exemples :

- modélisation de la décroissance d'une quantité de matière radioactive : espace des états $X = \mathbb{R}_+$, EDO $x'(t) = -ax(t)$;

- position d'une bille roulant dans un bol sans frottement : X sera la surface du bol (modélisée comme une surface ouverte dans $\mathbb{R}^3$), et l'EDO sera $mx''(t) = -mge_y + R(x(t))$, où m est la masse de la bille, g la force de la gravité, e_y le vecteur vertical, $R(x)$ la réaction exercée par le bol sur la bille ;

- position d'une bille qui roule tout près du fond du bol : l'EDO est inchangée, mais l'espace X sera un petit ouvert centré autour du fond du bol ; on parlera alors d'étude *locale*.

Les choses sont un peu moins simples qu'il n'y paraît : dans l'exemple de la bille qui roule, on a écrit $x''(t)$ pour la dérivée seconde *calculée dans l'espace* $\mathbb{R}^3$, qui est plus large que l'espace des états. En géométrie des surfaces, on distingue l'accélération "intrinsèque", toujours tangente à la surface (dérivée seconde covariante), et l'accélération "extrinsèque" qui ne l'est pas forcément. Si l'on avait décidé d'utiliser l'approche intrinsèque, il aurait fallu remplacer la force de gravité verticale par sa projection orthogonale sur l'espace tangent au bol, et il n'y aurait alors pas eu à considérer de force de réaction. Pour commencer, et dans presque tout le cours, l'espace des états sera un ouvert de $\mathbb{R}^n$,

et nous n'aurons pas à nous préoccuper de ces subtilités car il n'y aura pas d'ambiguïté sur la notion de dérivée seconde. La lectrice qui souhaite cependant approfondir le formalisme de la géométrie différentielle en liaison avec la modélisation physique peut consulter l'ouvrage très pédagogique d'Arnold [**2**] ou celui, un peu plus intransigeant, de Thirring [**34**].

On notera par convention t la variable de temps, que l'on omettra souvent pour alléger les notations : ainsi x sera une abréviation de $x(t)$, x' une abréviation de $x'(t)$, etc. En outre, selon une convention qui remonte au 17ème siècle, on utilisera souvent un point pour dénoter l'opération de dérivation par rapport à t : ainsi $\dot{x} = x'$, $\ddot{x} = x''$, etc. On notera parfois aussi $x^{(2)} = x''$, $x^{(3)} = x'''$, etc.

Le temps t dans une équation différentielle est une variable réelle, qui varie toujours dans un *intervalle*, appelé intervalle de temps. En effet, si l'on ignore l'état du système à un certain temps, on ne peut espérer en prédire les états ultérieurs ; la prédiction n'est donc envisagée que s'il y a continuité des temps. Réciproquement, on cherche à travailler avec des équations vérifiant une propriété de déterminisme :

Définition 1 (Propriété de Déterminisme). Une équation est dite déterministe si toute solution $x(t)$ de cette équation, définie sur un intervalle de temps $I =]a, b[$, est entièrement déterminée par ses valeurs sur un intervalle de temps $]a, a + \varepsilon[$, avec ε arbitrairement petit.

En d'autres termes, si x et $\widetilde{x}$ sont deux solutions, définies sur $]a, b[$, telles que $x(t) = \widetilde{x}(t)$ pour tout $t \in]a, a + \varepsilon[$, alors $x(t) = \widetilde{x}(t)$ pour tout $t \in]a, b[$. Cette propriété indique que le système est entièrement prédictible à partir de son observation initiale. Pour prendre des exemples très simples : l'équation $x' = -ax$ est déterministe, alors que l'équation $|\dot{x}| \leq 1$ ne l'est pas.

Ce dernier exemple, et l'expérience, nous mènent à restreindre la forme des équations différentielles. Les équations vues en introduction étaient de la forme $\dot{y} = f(t, y)$ ou $\ddot{y} = g(t, y, \dot{y})$. De telles équations sont dites *explicites* car la dérivée de plus haut degré y est donnée comme une fonction des dérivées de degré inférieur. Cependant, une définition plus générale autorise des équations *implicites*.

Définition 2. Une EDO est une équation de la forme $F(t, y, \dot{y}, \ldots, y^{(m)}) = 0$, où m est un entier. Une EDO explicite est une EDO de la forme $y^{(m)} = G(t, y, \ldots, y^{(m-1)})$. On appelle m l'ordre de l'équation différentielle.

On note que l'exemple "non déterministe" $|\dot{x}| \leq 1$ est de la forme $F(\dot{x}) = 0$ si F est la fonction discontinue $F(y) = 1_{|y| \leq 1}$; cet exemple sera exclu par des conditions de régularité sur F.

On dira naturellement qu'une EDO est linéaire si elle est définie par une fonction F linéaire ; et polynomiale si elle est définie par une fonction F polynomiale. Voici deux équations polynomiales non linéaires :

$\ddot{y} = \dot{y}^3 + y^2 + 1$ (équation explicite)

$\ddot{y}^2 y + \dot{y}^2 = 1$ (équation implicite)

Les EDO implicites ne sont pas commodes à étudier ; en fait le premier réflexe du mathématicien, face à une EDO implicite, est de la transformer en une EDO explicite, soit par application d'un théorème des fonctions implicites, soit par résolution directe. Ainsi, à partir du dernier exemple, on écrira $\ddot{y} = \pm\sqrt{1-\dot{y}^2}$, après avoir noté la condition $|\dot{y}| \leq 1$; puis on considérera séparément les deux possibilités $\ddot{y} = \sqrt{1-\dot{y}^2}$ et $\ddot{y} = -\sqrt{1-\dot{y}^2}$.

Plusieurs difficultés peuvent surgir au cours de ce processus. Par exemple, si l'on part de l'équation implicite $t^2\dot{y} = y$, on aboutit à l'équation explicite $\dot{y} = y/t^2$, qui est singulière près de $t = 0$. Ou encore, si l'on part de l'équation $\dot{y}^3 = y^2$, on trouve l'équation $\dot{y} = \pm|y|^{2/3}$; et on note alors que la fonction définissant l'EDO n'est pas régulière, car $y \longmapsto |y|^{3/2}$ est non lisse en $y = 0$. Mais laissons toutes ces subtilités de côté pour l'instant...

Cela peut paraître surprenant au premier abord, mais *toute EDO peut être réécrite comme une EDO d'ordre 1.* La recette de cette réécriture est à la fois simple et importante : si l'on se donne une équation $F(t, y, \dot{y}, \ldots, y^{(m)}) = 0$, on définit la nouvelle fonction inconnue $z = (y, \dot{y}, \ldots, y^{(m-1)})$. Par exemple, si y varie dans l'espace des états $Y = \mathbb{R}$, alors z varie dans l'espace $Z = \mathbb{R}^m$. L'idée est que les composantes de z donneront accès à toutes les dérivées d'ordre au plus $m-1$, et qu'il suffira de dériver z une seule fois pour avoir accès à la dérivée d'ordre m. Ainsi, la nouvelle équation prendra la forme $F(t, z_1, z_2, \ldots, z_{m-1}, \dot{z_{m-1}}) = 0$.

Exemple 3. Soit l'équation (polynomiale explicite) $\ddot{y} = y^3 - ay^2 + 2t\dot{y}$, avec $y \in \mathbb{R}$. On pose $z = (y, \dot{y}) =: (z_1, z_2)$. Alors, par définition $\dot{z}_1 = z_2$, et par ailleurs

$$\dot{z}_2 = \ddot{y} = y^3 - ay^2 + 2t\dot{y} = z_1^3 - az_1^2 + 2tz_2.$$

On définira donc $f(t, z_1, z_2) = (z_2, z_1^3 - az_1^2 + 2tz_2)$, et l'équation prendra la forme $\dot{z}(t) = f(t, z(t))$.

Bien sûr, dans cet exemple on a perdu quelque chose au change : si l'ordre de l'équation a été réduit de 2 à 1, la dimension de l'espace des états est passée de 1 à 2. C'est un phénomène général !

1.2. Premières notions

Considérons une équation différentielle explicite du premier ordre, $\dot{y}(t) = f(t, y(t))$. Pour chaque état z, le vecteur $f(t, z)$ indique la variation de la solution "quand elle prend la valeur z". Ainsi $z \longmapsto f(t, z)$ est une fonction à valeurs vectorielles, ou **champ de vecteurs** ; on peut se la représenter comme une famille de vecteurs, dépendant du temps t, telle que $f(t, z)$ a son origine en z. Imposer une régularité sur f revient à parler de la régularité de la variation de ce vecteur en fonction de z ou de t. Réciproquement, un champ de vecteurs définit une EDO.

Définition 4. Soit I un intervalle de $\mathbb{R}$, et U un ouvert de $\mathbb{R}^n$; soit $(t, z) \longmapsto f(t, z) \in \mathbb{R}^n$ un champ de vecteurs ; il définit alors une EDO : $\dot{z} = f(t, z)$. On dit que les solutions de l'EDO sont les courbes intégrales du champ de vecteurs.

Remarque 5. 1. On peut poser la même définition si U est un ouvert d'une variété V différentiable ; un champ de vecteurs f doit alors vérifier $f(t, z) \in T_z U$ (condition de tangence à V).

2. Si $f(t, z)$ est un champ de vecteurs, une courbe intégrale $y(t)$ est, en chaque temps t, *tangente* à ce champ de vecteurs. La résolution d'une EDO est ainsi étroitement liée à un problème que l'on peut formuler en termes purement géométriques.

La notion de champ de vecteurs est une simple reformulation, mais elle est cruciale dans l'intuition des EDO. Avec cela en tête, continuons la classification des équations différentielles au moyen de quelques notions simples.

Définition 6. Si $f(t, z)$ est un champ de vecteurs défini dans $I \times U$, on appelle équilibre associé à f tout point z tel que $f(t, z) = 0$ pour tout t.

Les équilibres sont donc les états z invariants par l'évolution du système : c'est à dire que si le système est au départ dans l'état z, il y reste pour toujours.

Définition 7. Une EDO est dite autonome si elle est définie par un champ de vecteurs stationnaire, c'est à dire indépendant de la variable t ; en d'autres termes, si elle prend la forme $\dot{x} = f(x)$.

On se rappelle que toute équation différentielle, quelle que soit son ordre, est équivalente à une équation différentielle d'ordre 1, obtenue en élargissant éventuellement l'espace des états. De même, toute équation différentielle est équivalente à une équation différentielle autonome.

Cela se montre en introduisant un "temps auxiliaire", que nous allons noter s, en considérant la nouvelle variable comme partie des données et en faisant en sorte qu'elle coïncide automatiquement, *le long de la solution*, avec le temps naturel t. En effet, soit $f = f(t, x)$ un champ de vecteurs défini sur $I \times U$. On introduit une nouvelle variable que l'on note s, et on note $y = (x, s)$ la nouvelle inconnue ; explicitement, ce sera une fonction $(x(t), s(t))$. On définit alors la nouvelle EDO par

$$\begin{cases} \dot{x} = f(s, x) \\ \dot{s} = 1 \end{cases} \tag{7}$$

La deuxième équation entraîne $s(t) = t + c$, où c est une constante. Si l'on impose alors une *condition* $s(t_0) = t_0$, où t_0 est un temps quelconque dans l'intervalle I, alors s et t coïncideront automatiquement : pour toute solution, $s(t) = t$. En posant donc $F(x, s) = (f(s, x), 1)$ on aura $\dot{x} = f(t, x) \Longrightarrow \dot{y} = F(y)$. Cela vient donc avec : une dimension supplémentaire dans l'espace des états ; et une condition supplémentaire $s(t_0) = t_0$.

Si toutes les équations peuvent se transformer en équations autonomes du premier ordre, il faut bien s'attendre à ce que ces dernières ne soient en général pas résolubles "explicitement". En fait, comme nous le verrons, même les équations d'ordre 1, explicites, autonomes, en dimension 1, ne sont pas résolubles explicitement : résoudre $\dot{x} = f(x)$ se ramène en effet, en dimension 1, à calculer une primitive de $1/f$, ce qui est en général impossible à faire explicitement...

Il existe cependant plusieurs exceptions notables : des situations dans lesquelles un calcul plutôt explicite est possible. La plus importante est le cas des **équations linéaires à coefficients constants**, c'est à dire des EDO vectorielles prenant la forme $\dot{y} = Ay$.

PROPOSITION 8. *Soit A une matrice réelle $n \times n$, et soit $t_0 \in \mathbb{R}$. Alors l'équation différentielle $\dot{y} = Ay$, posée dans $\mathbb{R} \times \mathbb{R}^n$, se résout en*

$$y(t) = e^{(t-t_0)A} y(t_0).$$

En particulier, la solution est déterminée par sa valeur en t_0, et l'espace des solutions est de dimension exactement n.

La fonction e^M qui apparaît ci-dessus, l'exponentielle matricielle, est un concept fondamental qu'il convient de maîtriser. On rappelle que (i) sa définition est similaire à celle de l'exponentielle des nombres réels ou complexes :

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!};$$

(ii) Si la matrice A est triangulaire inférieure, alors e^A est aussi triangulaire inférieure, et les coefficients de la diagonale sont les exponentielles des coefficients de la diagonale de A ; et (iii) Le déterminant de l'exponentielle coïncide avec l'exponentielle de la trace :

$$\det(e^A) = e^{\mathrm{tr}\ A}.$$

(Pour retrouver cette formule en cas d'oubli, il suffit de penser au cas où la matrice est triangulaire ; on a alors $\det(e^A) = \prod(e^{\lambda_i}) = e^{\sum \lambda_i}$, où les λ_i, coefficients de la diagonale de A, en sont aussi les valeurs propres.

EXEMPLE 9. (i) L'équation $\dot{y} = \lambda y$ se résout en $y(t) = e^{\lambda t} y(0)$.

(ii) L'équation $\ddot{y} = -y$ se résout en $y(t) = y(0)\cos t + \dot{y}(0)\sin t$. Pour inclure cela dans le formalisme de l'exponentielle de matrice, on commence par réécrire

$$z = \begin{pmatrix} y \\ \dot{y} \end{pmatrix}, \qquad \dot{z} = Mz, \qquad M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

et partant de là $z(t) = e^{tM} z(0)$, ce qui redonne la formule précédente. (Attention à la convention : le vecteur est ici rangé "verticalement" ; si on l'avait écrit "horizontalement" il aurait fallu remplacer M par sa transposée.)

On peut aller plus loin pour certaines classes d'équations linéaires, en liaison avec des sujets tels que les polynômes orthogonaux, ou certaines fonctions spéciales, qui sont parfois définies comme des solutions d'équations linéaires (par exemple la fonction hypergéométrique). Un certain nombre de "recettes exactes" se trouvent dans des ouvrages de référence [**18, 29, 39**].

Dans le cas général, en l'absence de solution explicite, on se retrouve forcé de faire une étude qualitative des solutions ; cette approche, systématiquement développée à partir de Henri Poincaré, connaît l'une de ses plus belles expositions dans les ouvrages de synthèse de Vladimir Arnold [**1, 3, 2**].

1.3. Le problème de Cauchy et l'espace des phases

Quel que soit le sujet mathématique, ce n'est pas la forme particulière de l'équation qui compte, ce sont ses solutions ; d'ailleurs il existe souvent de très nombreuses manières de changer l'équation sans changer les solutions. Dans l'optique des équations d'évolution, on observe initialement l'état du système, par exemple en un temps t_0, et on cherche à prédire le futur, disons l'état en un temps $t > t_0$ (ou parfois à reconstituer le passé, soit retrouver l'état en un temps $t < t_0$).

Ce processus va s'effectuer, par exemple, selon les lois de la physique ; et ce n'est pas toujours conceptuellement très simple. Nous allons évoquer le formalisme lagrangien, qui régit beaucoup de systèmes physiques. Un système est dit lagrangien à la condition suivante : appelant x l'état du système, il existe une fonction $L(x, v)$ (le Lagrangien du système), où v est la variable de vitesse, telle que pour tout intervalle de temps, disons $[0, T]$, si l'on se donne un état x_0 au temps 0 et un état x_T au temps T, alors la trajectoire est solution du problème de minimisation, aussi appelé principe de moindre action,

$$\min \left\{ \int_0^T L(\gamma(t), \dot{\gamma}(t))\, dt; \quad \gamma(0) = x_0, \gamma(T) = x_T \right\}, \tag{8}$$

où l'on recherche le minimum parmi toutes les courbes $\gamma \in C([0, T]; X) \cap C^1((0, T); X)$. Ici X est l'espace des états, et v (variable de vitesse) est un vecteur tangent à l'espace X.

Les exemples les plus importants sont $L(x, v) = |v|$ dans $\mathbb{R}^n$ (alors $\int_0^T L(\gamma, \dot{\gamma}) = \int |\dot{\gamma}|$ est la longueur de la courbe) et $L(x, v) = |v|^2/2$ (alors $\int_0^T L(\gamma, \dot{\gamma}) = \int |\dot{\gamma}|^2/2$ est l'intégrale de l'énergie cinétique, que l'on appelle l'action en mécanique classique). Dans les deux cas, les trajectoires minimisantes dessinent des courbes droites !

Cette approche a rencontré des succès remarquables en physique, en commençant par le principe de moindre action de Fermat en optique : la lumière se propage selon des courbes de moindre action, avec un lagrangien défini par $L(x, v) = \|v\|$, où la norme dépend des propriétés du milieu, et peut d'ailleurs varier d'un point à l'autre. De nos jours, certains des modèles les plus élaborés de physique fondamentale sont encore posés à l'aide de lagrangiens...

Mais le formalisme lagrangien pose aussi un problème conceptuel non négligeable : à travers lui, c'est quand on se donne un état initial et un état final que l'on trouve la trajectoire... alors qu'en pratique, on se donne un état initial et on veut déterminer un état final ! Pour ce faire, il faudra transformer le principe lagrangien en équation différentielle, avec une valeur prédictive. Il y a une recette systématique pour cela, dont nous aurons l'occasion de reparler... Pour l'heure, nous allons simplement noter qu'il y a au moins deux façons populaires de formuler une loi d'évolution :

- connaissant la condition initiale et la condition finale, on détermine la trajectoire par un principe de minimisation : c'est alors un problème lagrangien.

- connaissant la condition initiale, on détermine la trajectoire par la résolution d'une équation différentielle : on dit que c'est un **problème de Cauchy**.

Dans les chapitres qui suivent, on se concentrera sur le problème de Cauchy, et ce n'est qu'à la fin du cours que l'on reviendra sur le problème lagrangien.

Passons maintenant à la notion d'espace des phases. Considérons le mouvement d'une balle qui roule sans frottement sur une table plane : on l'assimile à un point matériel qui se déplace selon la loi de Galilée, c'est à dire en mouvement rectiligne uniforme. De manière équivalente, le mouvement a une accélération nulle, quand on écrit l'équation dans le plan de la table. Il est évident dans ce cas que la position de la balle au temps t ne dépend pas seulement de la position de la balle au temps t_0 : la vitesse initiale a aussi son mot à dire. En revanche, à partir de la position et de la vitesse initiales, on peut déduire la position en tous les temps ultérieurs ; et aussi la vitesse (constante) à tous les temps ultérieurs.

Revenons sur cet exemple : même si l'équation vérifie un principe de déterminisme au sens de la Définition 1, la connaissance de l'état (la position) en un temps initial ne suffit pas à prédire l'avenir du système (les positions futures). En revanche, si l'on reformule le système dans l'espace plus large des positions et des vitesses, alors la connaissance de l'état initial permet de prédire l'avenir du système. On appelle cet espace élargi **l'espace des phases** : c'est l'espace le plus "économique" dans lequel l'évolution du système peut entièrement se prédire en fonction de l'état initial. Ici, "économique" n'est pas un mot bien défini, mais, par exemple, nul besoin d'élargir encore l'espace des états en y ajoutant les accélérations (dans cet exemple tout simple, l'accélération est toujours nulle et n'apporte donc aucune nouvelle information).

Cette discussion rappelle celle que nous avons déjà eue pour transformer les équations d'ordre m en équations du premier ordre. En effet, en général, l'espace des phases est aussi l'espace le plus économique dans lequel on peut reformuler le système comme une EDO du premier ordre, disons $\dot{x}(t) = f(t, x(t))$. Cela se comprend intuitivement : partant de x_0, on va faire évoluer le système dans la direction $f(t, x_0)$, et partant de là, chaque fois que l'on est en $x(t)$, le faire évoluer dans la direction $f(t, x(t))$... au contraire, avec une équation du second ordre, quand on part de x_0 on ne sait pas où aller !

En invoquant la technique de réduction des équations différentielles à l'ordre 1, on aboutit ainsi tout naturellement à la définition explicite qui suit :

DÉFINITION 10 (espace des phases). Soit $y^{(m)} = F(t, y, \dots, y^{(m-1)})$ une EDO explicite d'ordre m, où y est à valeurs dans l'espace d'états $\mathbb{R}^n$; alors l'espace des phases associé à l'EDO est

$$X = \big\{(y, y_1, \dots, y_{m-1})\big\} \subset \mathbb{R}^{mn}.$$

Cette définition simple est à prendre avec quelques précautions. D'abord, l'équation est peut-être définie sur un ouvert de $\mathbb{R}^n$ plutôt que sur $\mathbb{R}^n$, et peut-être qu'on aura alors intérêt à définir l'espace des phases comme un ouvert de $\mathbb{R}^{mn}$ si l'on sait, pour une raison ou une autre, que cela sera suffisant pour l'étude. Ensuite, durant l'exercice de modélisation, il n'est pas forcément évident de déterminer les variables de l'espace des phases. À titre d'exemple, revenons sur le modèle de la balle qui se déplace sur une table. On a vu que si la balle est assimilée à un point, l'espace des phases est fait des positions (2 variables) et des vitesses (2 variables), soit 4 variables en tout. Mais s'il y a un dessin sur la balle (penser à la marque d'une balle de ping-pong) et que l'on veut aussi prédire l'évolution de l'orientation de ce dessin, alors il faut modéliser la balle comme une sphère en contact avec la table, et ajouter aux inconnues l'orientation de cette sphère : cela fait trois variables de plus (deux pour placer le point de contact, une pour éventuellement changer l'orientation de la sphère autour de l'axe vertical) ; on a donc envie de dire que l'espace des phases a 7 dimensions... mais cela ne suffit toujours pas, car l'évolution de l'orientation dépendra aussi de l'effet de la balle, c'est à dire le fait qu'elle puisse, en plus de sa vitesse par rapport à la table, avoir une composante de rotation (dans un sens ou dans l'autre) autour de l'axe orthogonal à la table... il faudra donc ajouter encore une variable, et on conclut que l'espace des phases associé aura 8 dimensions.

On voit sur cet exemple que l'espace des phases peut être subtil. Pour autant, dans la majorité des exemples, il sera facile d'appliquer la Définition 10.

1.4. Changement de variables et Flot

Soit une équation différentielle ; pour exprimer les composantes de ses solutions, on se place dans un système de coordonnées, et cela influe aussi sur la forme de l'équation : si l'on change de système de coordonnées, l'équation est transformée en conséquence. Cette opération est appelée changement de variables. Supposons que le changement de coordonnées soit défini par un difféomorphisme (c'est à dire une application bijective différentiable, de réciproque différentiable, définie d'un ouvert dans un autre ouvert) ; soit $\psi = \phi^{-1}$ l'application réciproque

de ϕ : on pourra écrire, x étant l'inconnue dans les variables originales, $X = \phi(x)$, $x = \psi(X)$. Comment cela affecte-t-il une équation différentielle ?

Écrivons cette équation sous la forme $\dot{x} = f(t, x)$; alors

$$\dot{X}(t) = \frac{d}{dt}\phi(x(t)) = d\phi(x(t))\,\frac{dx}{dt} \\ = d\phi(x(t))\,f(t, x(t)) = d\phi\big(\psi(X(t))\big) f\big(t, \psi(X(t))\big). \tag{9}$$

Pour t fixé, le membre de droite est l'image du vecteur f par l'application linéaire $d\phi$, ces objets étant évalués en $(t, \psi(X(t)))$. Nous avons donc via (9) une *nouvelle équation différentielle* et un nouveau champ de vecteurs ; on l'appelle champ de vecteurs image de f par ϕ.

Définition 11 (image d'un champ de vecteurs). Soit f un champ de vecteurs défini sur un ouvert U de $\mathbb{R}^n$, et ϕ un difféomorphisme de U sur un ouvert V ; alors on définit $f_*\phi$, champ de vecteurs image de f par ϕ, via la formule

$$\phi_* f(X) = d\phi(\phi^{-1}(X)) f(\phi^{-1}(X)).$$

En pratique, pour déterminer la nouvelle forme d'une EDO après changement de variables, il sera plus commode de refaire le calcul précédent, basé sur la formule de différentiation des fonctions composées, que d'appliquer directement la formule ci-dessus. Le plus souvent, on utilisera des difféomorphismes de classe C^1.

Un changement de variables transforme tout à la fois le champ de vecteurs, l'équation et les solutions : si $x(t)$ est une solution du système initial, alors $\phi(x(t))$ est solution du système image (ou encore courbe intégrale du champ de vecteurs image). Résoudre l'équation dans le jeu de variables initial, c'est exactement équivalent à la résoudre dans le jeu de variables image. Mais cela peut être, en pratique, bien plus simple dans un système que dans un autre !

Exemple 12. Revenons à l'équation toute simple $\ddot{x} = 0$ dans $\mathbb{R}$. On introduit $E = \{(x, v)\} = \mathbb{R}^2$ l'espace des phases, où l'équation se réécrit $\dot{x} = v$, $\dot{v} = 0$. Considérons le changement de variables $\phi(x, v) = (x + v, x^3) =: (\phi_1, \phi_2)$. Tant que l'on ne s'approche pas de $x = 0$, ϕ définit bien un difféomorphisme. Écrivons x et v en fonction de ϕ :

$$x = \phi_2^{1/3}, \qquad v = \phi_1 - v = \phi_1 - \phi_2^{1/3}.$$

Pour déterminer comment se transforme l'équation de départ, on calcule :

$$\frac{d}{dt}(x + v) = \dot{x} + \ddot{x} = \dot{x} = \phi_1 - \phi_2^{1/3},$$

$$\frac{d}{dt}x^3 = 3x^2\dot{x} = 3\phi_2^{2/3}(\phi_1 - \phi_2^{1/3}).$$

Dans ces nouvelles variables, le système s'écrit donc

$$\dot{\phi} = \big(\phi_1 - \phi_2^{1/3}, 3\phi_2^{2/3}(\phi_1 - \phi_2^{1/3})\big).$$

Cela est équivalent au système initial $\ddot{x} = 0$, certes... mais le lecteur avouera que cela ne saute pas aux yeux !

Les changements de variables jouent un rôle fondamental en géométrie : c'est par un choix de coordonnées (une "carte") que l'on ramène un problème posé dans une géométrie non euclidienne en n dimensions à un problème posé dans un ouvert de $\mathbb{R}^n$; mais ce choix de coordonnées n'est pas intrinsèque, et l'on passe d'un choix à un autre par changement de variables. Pour revenir à l'un de nos exemples initiaux : pour décrire une bille roulant dans un bol, il vaudra mieux formuler l'équation dans un morceau de $\mathbb{R}^2$ que dans l'espace $\mathbb{R}^3$.

Soit maintenant une EDO, et une condition initiale dans l'espace des phases : il lui correspond toute une trajectoire qui décrit l'évolution du système. Cette correspondance, qui à la condition initiale associe la trajectoire, s'appelle le **flot** de l'EDO. Il dépend bien sûr du choix de temps initial, disons t_0. Voici une définition :

Définition 13 (flot). Soit (E) une équation différentielle définie sur un intervalle de temps I, et soit X l'espace des phases associé. Soit $t_0 \in I$, appelé temps initial. On appelle flot associé à (E) l'application

$$\Phi : x_0 \longmapsto (x(t))_{t\in I}, \quad \text{solution de (E) telle que } x(t_0) = x_0.$$

On note immédiatement plusieurs subtilités associées à cette définition. D'abord, il sera bon de rappeler dans la notation la dépendance du flot par rapport à t_0 : on écrira par exemple

$$x(t) = \Phi_{t_0,t}(x_0),$$

que l'on pourra lire "l'état au temps t du système qui au temps t_0 est en x_0". On pourra dire, au choix, que $\Phi_{t_0,t}$ est l'application qui à la position initiale x_0 associe la position au temps t ; ou que Φ_{t_0} est l'application qui à la position initiale x_0 associe toute la trajectoire $(x(t))_{t\in I}$.

Ensuite, tel que nous l'avons écrit, le flot n'est bien défini que si les solutions de (E) sont définies sur tout l'intervalle I... Or en général il est *impossible* d'espérer un résultat aussi fort ; on est donc amené à travailler avec des flots qui ne sont que partiellement définis ; et avec la notion de "flot maximal", qui est en quelque sorte le flot défini sur un intervalle de temps aussi grand que possible. Nous reparlerons de cela ultérieurement.

Enfin et surtout, il convient d'insister fortement sur la différence entre la trajectoire et le flot ! La trajectoire, c'est une fonction $t \longmapsto x(t)$ qui décrit les positions successives du système ; alors que le flot c'est l'application qui à une condition initiale associe la trajectoire tout entière. En outre, le flot est toujours défini sur l'espace des phases, alors que la trajectoire peut prendre ses valeurs dans un espace d'états plus restreint que l'espace des phases (par exemple, pour une équation du second ordre, on peut très bien étudier la trajectoire dans l'espace des positions, alors que cela n'a pas de sens de parler du flot dans l'espace des positions).

Souvent le temps t_0 est fixé une fois pour toutes ; par exemple $t_0 = 0$. On omet alors de le rappeler dans la notation. En outre, il existe un cas important où l'on n'a de toute façon pas besoin d'indiquer t_0 : si l'équation différentielle est *autonome*, alors $\Phi_{t_0,t}$ ne dépend que de $t - t_0$, et l'équation est inchangée quand on modifie l'origine des temps. On peut alors, sans perte de généralité, supposer que $t_0 = 0$, et écrire simplement Φ_t. Dans ce cas, l'état du système au temps t, partant de x_0 au temps t_0, sera $\Phi_{t-t_0}(x_0)$.

Exemple 14. Reprenons encore le mouvement libre d'une bille ponctuelle sur une table horizontale sans frottement : $\ddot{x} = 0$, l'espace des phases est de dimension 4. Notons x_0 la position initiale et v_0 la vitesse initiale, alors $\Phi_{t_0,t}(x_0, v_0) = (x_0 + (t - t_0)v_0, v_0) = \Phi_{t-t_0}(x_0, v_0)$.

Le flot jouit d'une importante **propriété de composition**, ou propriété de semigroupe :

$$\Phi_{t_1,t_2} \circ \Phi_{t_0,t_1} = \Phi_{t_0,t_2}.$$

Cette propriété, quasiment tautologique, exprime le fait que si l'on part de x_0 en t_0 pour aboutir à x_1 en t_1, et que partant de x_1 en t_1 on aboutit à x_2 en t_2, alors finalement, en partant de x_0 en t_0 on a abouti à x_2 en t_2... En conséquence, le flot est toujours inversible, puisque

$$\Phi_{t,t_0} \circ \Phi_{t_0,t} = \Phi_{t,t} = Id.$$

Dans le cas d'une équation autonome, ces relations s'écrivent simplement

$$\Phi_s \circ \Phi_t = \Phi_{t+s}; \qquad \Phi_t \circ \Phi_{-t} = \Phi_0 = \mathrm{Id}\,.$$

Pourquoi prendre tant de précautions pour parler du flot qui n'est, somme toutes, qu'un jeu de réécriture conceptuelle ? C'est parce que ce changement de point de vue est d'une richesse considérable. Avec la notion de flot il ne s'agit pas seulement de résoudre l'équation à partir d'une donnée initiale, il s'agit de vraiment comprendre comment le choix de la condition initiale influe sur la solution tout entière. Avec

la notion de flot on peut aborder des questions telles que : *Si je commets une erreur sur la condition initiale, est-ce que cela engendrera une erreur importante sur la solution ? L'erreur aura-t-elle tendance à diminuer avec le temps, ou à s'amplifier ? Est-ce que la réponse à telle question dépend crucialement du choix de la donnée initiale, ou pas trop ?*

La notion de flot est d'usage constant en géométrie ; on y parle sans cesse du *flot géodésique*, dont les trajectoires sont les courbes géodésiques, qui, entre deux positions pas trop éloignées, minimisent l'action $\int |\dot\gamma|^2$ (ce sont aussi, à reparamétrisation près, les courbes qui minimisent la longueur). On peut penser aux trajectoires des avions transatlantiques, qui cherchent à économiser le carburant et donc la distance parcourue : ce sont des approximations de courbes géodésiques à la surface de la Terre.

On apprend en géométrie riemannienne (non euclidienne) à calculer les équations des courbes géodésiques, dans un jeu de coordonnées quelconque :

$$\ddot x^k + \Gamma^k_{ij} \dot x^i \dot x^k = 0; \tag{10}$$

où l'on note par convention les composantes avec des indices en haut plutôt qu'en bas, et où les fonctions Γ^k_{ij}, appelées symboles de Christoffel, se calculent en fonction des propriétés de la géométrie. Cette équation est du second ordre, ce qui nous rappelle l'équation des droites dans l'espace euclidien, $\ddot x = 0$. Et l'application qui à une position et une vitesse initiale (x_0, v_0) associe la courbe géodésique, sous la forme $(x(t), \dot x(t))$, avec $x(0) = x_0$, $\dot x(0) = v_0$, est appelée tout naturellement le flot géodésique ; cette application contient en elle toute la géométrie. Pour tous ces concepts on renvoie par exemple à Do Carmo [**10**].

Finalement, remarquons que la notion de flot s'applique dès que l'on a une notion d'évolution, et déborde donc du cadre des équations différentielles. Par exemple, donnons-nous une loi de transformation discrète, disons une application ϕ qui à une configuration initiale x_0 associe une configuration $x_1 = \phi(x_0)$, puis une configuration $x_2 = \phi(x_1)$, etc. On pourra alors définir le flot associé à la transformation comme l'application $x_0 \longmapsto (x_0, x_1, x_2, x_3, \ldots)$. On parlera de flot discret, car c'est comme si l'on travaillait avec un ensemble discret de temps $t = 0, 1, 2, 3, \ldots$.

1.5. Lois de conservation et fonction de Lyapunov

Il est souvent intéressant d'étudier les équations différentielles au moyen de fonctions définies sur l'espace des phases ; ces fonctions n'apparaissent pas forcément dans l'équation elle-même, mais pourront

fournir des informations sur ses solutions. Parfois on pourra les construire par des procédés systématiques, parfois elles seront suggérées par l'intuition...

DÉFINITION 15 (loi de conservation). Soit (E) une EDO, et X l'espace des phases associé. On appelle invariant, ou loi de conservation, de (E), toute fonction $h : X \to \mathbb{R}$ qui reste constante le long des solutions de (E).

Autrement dit, un invariant est une fonction h telle que pour toute solution x de (E), définie sur l'espace des phases X, on aura

$$\frac{d}{dt}h(x(t)) = 0,$$

i.e. $t \mapsto h(x(t))$ est constante sur l'intervalle I où est définie x.

Les invariants permettent de restreindre l'espace des possibles, et donc de préciser et simplifier l'étude du système. La recherche des invariants est souvent l'une des premières tâches que l'on se fixe dans l'étude d'une équation différentielle, que ce soit en mécanique classique ou en physique contemporaine des hautes énergies.

EXEMPLE 16. L'exemple archétypal d'invariant remonte au 17ème siècle : c'est la fonction énergie d'un système mécanique classique. Soit en effet un système de la forme $\ddot{x} = F(x)$, où x est à valeur dans $\mathbb{R}^d$ par exemple (penser à F comme une force, la masse étant supposée fixée à 1). En général ce système n'admettra pas d'invariant non trivial. Mais considérons maintenant le cas où F dérive d'une fonction $V : \mathbb{R}^d \to \mathbb{R}$, c'est à dire que

$$F = -\nabla V,$$

ou encore

$$F_i(x) = -\frac{\partial V(x)}{\partial x_i}.$$

La fonction V, que l'on supposera par exemple de classe C^1, est appelée potentiel, et on peut lui ajouter une constante arbitraire sans changer l'équation associée $\ddot{x} = -\nabla V(x)$. Définissons $E : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ par la formule

$$E(x, v) = \frac{|v|^2}{2} + V(x). \tag{11}$$

Le premier terme $|v|^2/2$ est appelé énergie cinétique, et le second énergie potentielle ; on appelle souvent E l'énergie totale. Soit maintenant une solution $x(t)$, que l'on réécrit dans l'espace des phases

$$\dot{x}(t) = v(t), \quad \dot{v}(t) = -\nabla V(x(t));$$

on trouve

$$\frac{d}{dt}E(x(t),\dot{x}(t)) = v(t)\cdot\dot{v}(t) + \nabla V(x(t))\cdot\dot{x}(t) \\ = -v(t)\cdot\nabla V(x(t)) + \nabla V(x(t))\cdot v(t) = 0. \tag{12}$$

La fonction E est donc un invariant du mouvement : cette conclusion est appelée *loi de conservation de l'énergie*.

L'invariance nous donne une information sur la solution, en la restreignant. Pour aller au-delà, on peut s'intéresser au comportement de la fonction E sur l'espace des phases. Par exemple, supposons que $d = 1$ (mouvement monodimensionnel), alors E est définie sur le plan $\mathbb{R}^2$. Si V est convexe, alors E sera une fonction convexe de deux variables, et on peut tracer, au moins qualitativement, l'allure de ses lignes de niveau – des sortes d'ellipses imbriquées, typiquement. La propriété d'invariance signifie que les solutions prennent leurs valeurs dans ces lignes de niveau ; autrement dit, étant donnée une condition initiale, on n'a plus qu'un degré de liberté et non deux.

Si le mouvement a lieu dans les courbes de niveau, la vitesse est forcément tangente à ces mêmes courbes ; cela revient aussi à dire que le champ de vecteurs définissant l'équation est orthogonal au gradient de la fonction E, ce que l'on avait bien vu dans le calcul (12).

On retient donc l'*interprétation géométrique de l'invariance* pour une équation définie par un champ de vecteurs f : la fonction E est invariante si $\nabla E \perp f$ en tout point ; ou de manière équivalente, si f est, en tout point, tangent à la courbe de niveau de E.

Passons maintenant à la notion de fonction de Lyapunov. Pour la motiver, rappelons-nous que la conservation de l'énergie modélise une situation idéale, sans frottements : dans la réalité, l'énergie diminue plus ou moins vite. C'est le cas par exemple d'une bille que l'on laisse osciller dans un bol : après un temps plus ou moins long, la bille finira par s'immobiliser, après avoir dissipé toute son énergie en frottements. En fait ces frottements jouent un rôle fondamental de stabilisation à notre échelle ; on ne s'attend certainement pas, en revanche, à voir la bille accélérer et quitter le bol sous l'effet d'une énergie débordante ! La constatation importante est que l'effet des frottements va toujours dans la direction d'un "ralentissement".

Les forces de frottement sont parfois très complexes à exprimer ; elles ont toujours tendance à s'opposer au mouvement, et elles sont souvent (mais pas toujours !) d'autant plus importantes que la vitesse est élevée. Pour simplifier, prenons donc le modèle d'une force de frottement linéaire en la vitesse : $F(x,\dot{x}) = -\lambda\dot{x}$, où $\lambda > 0$ est un coefficient

de frottement. L'équation devient

$$\ddot{x} = -\nabla V(x) - \lambda \dot{x}. \tag{13}$$

Le calcul (12) doit en conséquence être modifié : on trouve

$$\frac{dE}{dt} = v \cdot (-\nabla V(x) - \lambda \dot{x}) + \nabla V(x) \cdot v = -\lambda v \cdot \dot{x} = -\lambda |v|^2,$$

ce qui est toujours négatif ou nul. Autrement dit, le long des solutions de (13), on a une décroissance systématique de E : on dit que E est une **fonctionnelle de Lyapunov**.

DÉFINITION 17 (fonction de Lyapunov). Soit (E) une equation differentielle, et X l'espace des phases associé. On appelle fonction de Lyapunov une fonction $h : X \to \mathbb{R}$ qui est soit toujours décroissante le long des solutions de (E), soit toujours croissante le long de ces solutions.

Les fonctions de Lyapunov ne permettent pas de réduire l'espace des possibles aussi bien que les invariants, mais elles permettent quand même de le restreindre, ou plutôt de le confiner. Si l'on connaît la valeur initiale de la fonction de Lyapunov h décroissante, alors on sait que le système aux temps ultérieurs restera dans l'ensemble de sous-niveau $\{h \leq h(t_0)\}$; en particulier il ne peut pas aller divaguer trop loin de l'état de repos...

On peut également traduire géométriquement la propriété de Lyapunov : en effet, supposons que h décroît le long des solutions, alors, dans l'espace des phases on doit avoir $(d/dt)h(x(t)) = \nabla h(x) \cdot \dot{x} \leq 0$; ce qui veut dire que *le gradient de h forme toujours un angle obtus avec le champ de vecteurs f associé à l'équation* ; ou encore que "le champ de vecteurs f pointe vers l'intérieur des courbes de niveau de h".

Le cas le plus drastique de fonction de Lyapunov se produit quand le champ de vecteurs pointe dans la direction opposée au gradient de l'énergie : c'est le cas en particulier pour les **flots gradients** :

DÉFINITION 18 (flot gradient). Soit E une fonction différentiable définie dans $\mathbb{R}^n$; on appelle flot gradient de E le flot associé à l'EDO $\dot{y} = -\nabla E(y)$.

On peut y penser comme à un flot dans lequel la fonction énergie décroît, en quelque sorte, le plus efficacement possible... De telles équations jouent un rôle essentiel en physique et en mathématique, où elles apparaissent très fréquemment, soit qu'elles existent dans la nature, soit qu'on les introduise pour les fins d'une démonstration.

Terminons ce chapitre avec un exemple célèbre de fonction de Lyapunov, qui nous demandera un calcul plus élaboré. En physique statistique, on s'intéresse à l'évolution des statistiques de particules ; soit donc un ensemble de très nombreuses particules qui peuvent se répartir selon m états différents (on les supposera quantiques pour simplifier, avec des niveaux d'énergie discrets). L'état du système sera modélisé par un vecteur de probabilité $p = (p_1, \dots, p_m)$: chaque p_i indique la proportion de particules qui sont dans l'état i. On suppose également que les particules passent, de manière aléatoire et indépendante, d'un état à un autre, avec une probabilité instantanée de transition de l'état i à l'état j qui vaut $K_{ij} \geq 0$ (c'est à dire que sur un intervalle de temps infinitésimal, τ, la probabilité de réaliser la transition vaut τK_{ij}). On suppose que

$$\forall i, \quad \sum_j K_{ij} = 1, \qquad \forall j, \quad \sum_i K_{ij} = 1. \tag{14}$$

La théorie des processus aléatoires nous indique l'équation différentielle régissant l'évolution de p en fonction du temps :

$$\frac{dp_i}{dt} = \sum_j K_{ji}\, p_j - \sum_j K_{ij}\, p_i. \tag{15}$$

Faisons un petit aparté pour expliquer informellement comment on établit (15). Sur un intervalle de temps infinitésimal, entre t et $t + dt$, faisons le bilan des particules qui sont dans l'état i. Appelons N le nombre total, très grand, de particules. Pour chaque état $j \neq i$, une proportion $K_{ij}\, dt$ des $Np_i(t)$ particules qui étaient dans l'état i sont passées dans l'état j, et donc ne sont plus dans l'état i ; mais une proportion $K_{ji}\, dt$ des $Np_j(t)$ particules qui étaient dans l'état j sont, en guise de compensation, passées dans l'état i. Au total, au temps $t + dt$, le nombre de particules dans l'état i est donc

$$Np_i(t) - \sum_{j \neq i} Np_i(t) K_{ij}\, dt + \sum_{j \neq i} Np_j(t) K_{ji}\, dt.$$

Ajouter $Np_{ii}(t) K_{ii}\, dt$ dans chacune des deux sommes ne changera rien ; finalement

$$Np_i(t + dt) = Np_i(t) - \sum_j Np_i(t) K_{ij}\, dt + \sum_j Np_j(t) K_{ji}\, dt.$$

Il suffit alors de retrancher $Np_i(t)$ de part et d'autre, et de tout diviser par N pour obtenir (15).

Revenons maintenant à l'étude de l'équation (15). C'est visiblement une EDO du premier ordre, portant sur le vecteur p, qui compte r

composantes. Elle est de plus linéaire et à coefficients constants, de sorte que l'on pourrait écrire $p(t) = e^{Ct}p(0)$, où C est une certaine matrice (laquelle ?). Cependant, cela ne nous informera guère sur le comportement de p en fonction de t ; cherchons donc des informations qualitatives à ce sujet.

Pour commencer, vérifions que la fonction $M(p) = \sum p_i$ est un invariant de cette équation :

$$\begin{aligned}\frac{d}{dt}M(p(t)) &= \sum_i \frac{dp_i}{dt} \\ &= \sum_{ij} K_{ji}\, p_j - \sum_{ij} K_{ij}\, p_i \quad = \sum_{ij} K_{ji}\, p_j - \sum_{ji} K_{ji}\, p_j = 0,\end{aligned}$$

où l'on a échangé les indices i et j dans la dernière somme. En fait, comme p_i représente la fraction de particules dans l'état i, le vecteur p doit être un vecteur de fréquences, et donc la somme de ses composantes doit être égale à 1. La conservation de M n'a donc rien d'étonnant, mais encore fallait-il vérifier qu'elle était compatible avec l'équation (15) : si M n'avait pas été invariante par (15), cela aurait été l'indice d'une modélisation défectueuse. (Notons cependant qu'il existe dans la physique des équations d'évolution de fréquences, portant sur un nombre infini d'états, pour lesquelles la somme des fréquences n'est pas conservée ; cela traduit par exemple l'apparition de particules dont l'énergie, à l'échelle de modélisation, doit être considérée comme infinie.)

Et pour être complétement cohérents avec notre modélisation, nous devrions aussi vérifier que les composantes de p restent positives si elles ont cette propriété initialement ! C'est un peu plus subtil : fixons i et écrivons

$$\frac{dp_i}{dt} + \big(\sum_{j\neq i} K_{ij}\big)p_i = \sum_{j\neq i} K_{ji}\, p_j \geq 0,$$

donc si l'on pose $A_i = \sum_{j\neq i} K_{ij}$, on aura

$$\dot{p_i} + A_i\, p_i \geq 0,$$

soit $(d/dt)(p_i(t)e^{A_i t}) \geq 0$; cela prouve que

$$p_i(t) \geq p_i(0)e^{-A_i t} \geq 0, \tag{16}$$

comme il se doit.

Souvent les coefficients K_{ij} vérifient une hypothèse supplémentaire, appelée *microréversibilité de la dynamique* :

$$\forall i\ \forall j, \qquad K_{ji} = K_{ij}. \tag{17}$$

Physiquement parlant, cela veut dire qu'il est aussi aisé à une particule de passer de l'état i à l'état j, que de faire le chemin inverse. Alors l'équation (15) se réécrit sous la forme plus simple

$$\dot{p}_i = \sum_j K_{ij}(p_j - p_i). \tag{18}$$

Nous allons maintenant prouver une propriété absolument remarquable. Suivant les travaux de Ludwig Boltzmann, on introduit la fonction S, ou entropie, qui mesure en un certain sens le "désordre" associé à la distribution p :

$$S(p) = -\sum_i p_i \ln p_i.$$

Une rapide étude de fonction montre que S est minimale quand p est un vecteur dont toutes les composantes valent 0 sauf une (qui vaut alors 1, c'est à dire que toutes les particules sont bien ordonnées dans un seul état), et que S est maximale quand tous les p_i sont égaux (c'est à dire que toutes les particules sont réparties de manière aussi équitable que désordonnée dans tous les états possibles).

Calculons maintenant l'évolution de S en fonction du temps. La fonction $p \ln p$ est différentiable pour tout $p \neq 0$, et sa dérivée vaut alors $\ln p + p/p = \ln p + 1$. En revanche cette fonction n'est pas différentiable en $p = 0$, ce qui peut causer des soucis quand on calcule la dérivée de S... mais supposons que tous les p_i soient strictement positifs au temps initial ; alors (16) montre que les p_i restent strictement positifs pour tous les temps, et l'on peut utiliser la dérivée sans états d'âme.

On trouve alors

$$\begin{aligned}
\frac{dS(p(t))}{dt} &= -\sum_i (\ln p_i + 1)\frac{p_i}{dt} \\
&= -\sum_i \ln p_i \frac{dp_i}{dt} \\
&= -\sum_i \ln p_i \sum_j (K_{ji}\, p_j - K_{ij}\, p_i) \\
&= -\sum_{ij} K_{ij}\, p_i \ln p_i - K_{ji}\, p_j \ln p_i \\
&= \sum_{ij} K_{ij}(p_i \ln p_i - p_i \ln p_j) \\
&= \sum_{ij} K_{ij}\, p_i \ln\left(\frac{p_i}{p_j}\right) \\
&= \sum_{ij} K_{ij}\, p_j \left(\frac{p_i}{p_j}\right) \ln\left(\frac{p_i}{p_j}\right) \\
&\geq \sum_{ij} K_{ij}\, p_j \left(\frac{p_i}{p_j} - 1\right) \\
&= \sum_{ij} K_{ij}\, p_i - \sum_{ij} K_{ij}\, p_j \\
&= \sum_i p_i - \sum_j p_j = 0,
\end{aligned}$$

où la seule inégalité dans le calcul ci-dessus provient de l'inégalité de convexité $q \log q \geq q - 1$, et la formule (14) a été utilisée pour aboutir à la dernière ligne. La conclusion est $dS/dt \geq 0$: dans ce processus aléatoire, linéaire, *l'entropie augmente toujours* au fur et à mesure que le temps s'écoule ; la fonction S est donc une fonctionnelle de Lyapunov pour l'équation (18). Cette information fondamentale est un cas particulier de la **seconde loi de la thermodynamique** : dans un système conservatif, dissipatif et isolé, l'entropie augmente spontanément.

Il est intéressant de noter qu'une autre preuve existe quand les coefficients K_{ij} vérifient l'hypothèse de microréversibilité (17) à la place

de (14) :

$$\begin{aligned}\frac{dS(p(t))}{dt} &= -\sum_i (\ln p_i + 1)\frac{p_i}{dt} \\ &= -\sum_i \ln p_i \frac{dp_i}{dt} \\ &= -\sum_i \ln p_i \sum_j K_{ij}(p_j - p_i) \\ &= \sum_{ij} K_{ij}(p_i \ln p_i - p_i \ln p_j) \\ &= \sum_{ij} K_{ij}\, p_i \ln\left(\frac{p_i}{p_j}\right) \\ &= \frac{1}{2}\sum_{ij} K_{ij}(p_i - p_j)\ln\left(\frac{p_i}{p_j}\right),\end{aligned}$$

où l'on a utilisé l'échange $i \leftrightarrow j$ une nouvelle fois, ainsi que l'identité $\ln p_i/p_j = -\ln p_j/p_i$. Le résultat du calcul précédent est alors positif, pour la simple raison que $(a-b)\ln(a/b) = (a-b)(\ln a - \ln b) \geq 0$ pour toutes valeurs de a et b. Cette preuve, valable seulement dans le cas où l'équation est microréversible, est une version très simplifiée du célèbre Théorème H de Boltzmann.

La lectrice attentive pourra rétorquer que l'on a oublié d'envisager le cas où l'une des composantes de p s'annule au temps initial ! Voici une façon de résoudre cette difficulté ; elle illustrera au passage certains avantages que l'on peut trouver à la notion de flot. La formule $p(t) = e^{Bt}p(0)$ montre que $p(t)$ dépend continûment de $p(0)$ (on dit que le flot est continu) ; comme la fonction S est continue sur $\mathbb{R}^m_+$, la fonction $S(p(t))$ dépendra continûment, elle aussi, de la valeur initiale $p(0)$. Et si $p(0)$ a une ou plusieurs composantes nulles, on peut toujours trouver une famille p_0^δ de vecteurs de probabilité dont toutes les composantes sont strictement positives, telle que p_0^δ converge vers $p(0)$ quand $\delta \to 0$. À chaque p_0^δ est associé une solution $p^\delta(t) = e^{Ct}(p_0^\delta)$, et notre calcul différentiel s'appliquera à cette solution, d'où

$$S(p^\delta(t)) \geq S(p_0^\delta).$$

Il suffit alors de faire tendre δ vers 0 pour conclure

$$S(p(t)) \geq S(p(0)),$$

ce qui prouve bien que S est croissante le long de la solution.

Avec un peu plus de travail, on peut approfondir la conclusion et prouver, sous l'hypothèse supplémentaire que tous les coefficients K_{ij} sont strictement positifs, que

$$p(t) \underset{t\to+\infty}{\longrightarrow} p_\infty = \left(\frac{1}{m}, \dots, \frac{1}{m}\right).$$

En d'autres termes, le système converge vers l'état d'entropie maximale ; on peut aussi prouver que cette convergence est exponentiellement rapide.

Enfin cette propriété de convergence exponentielle demeure vraie dès que l'on peut trouver un entier $r \geq 1$ tel que tous les coefficients de la matrice K^r soient strictement positifs ; on peut dire alors que le système est *ergodique*, et cela est lié à la possibilité pour les particules de passer de n'importe quel état à n'importe quel autre...

1.6. Modélisation et simulation numérique

L'étude de l'EDO n'est le plus souvent qu'une étape dans un processus qui commence par la **modélisation** d'un problème et qui se termine par la **simulation numérique** des solutions de l'EDO (après quoi viennent bien souvent d'autres étapes encore : réinterprétation, action, etc. sans qu'il y ait de règles générales à ce sujet).

Il existe sans doute des dizaines ou des centaines de milliers d'équations différentielles dans la littérature scientifique ; elles sont issues de modélisations extrêmement variées, ou parfois d'autres problèmes mathématiques. La démarche de modélisation combine de nombreux ingrédients : démarche déductive ou inductive, simplifications, analogies, prise en compte séparée de différents phénomènes, expérience gagnées sur d'autres situations, utilisation "phénoménologiques" d'équations choisies pour leur comportement, etc. Illustrons cela sur un problème très simple de dynamique des populations.

Soit une population animale qui s'accroît exponentiellement vite du fait de la reproduction ; notons a le "taux de reproduction" dû à la différence entre fécondité et mortalité. Les naissances sont des événements discrets, mais nous allons les remplacer par un processus continu, soit que cela provienne d'une démarche d'analogie, soit que l'on se place à une échelle de temps assez long, où les naissances apparaissent comme des événements qui ont lieu en continu. On se retrouve avec l'EDO

$$\dot{x}(t) = a\, x(t),$$

où x est la taille de la population.

On constate tout de suite que ce modèle prédit une croissance de la population jusqu'à l'infini. On peut

- soit considérer cela comme la fin en soi du modèle (conçu, comme tous les modèles, avec certaines limitations) qui doit alors "passer le relais" à une autre équation,

- soit modifier le modèle en prenant en compte une limitation, comme la population maximale que le milieu ambiant peut accueillir (du fait, par exemple, de la limitation des ressources disponibles). Dans cette optique une façon très simple de procéder est d'introduire de force cette population maximale.

Voici donc un autre modèle prenant en compte une population maximale K :

$$\dot{x}(t) = a\,x(t)\left(1 - \frac{x(t)}{K}\right). \tag{19}$$

Cette nouvelle équation autorise une croissance de type exponentiel quand x est petit par rapport à K, mais interdit à x de dépasser la valeur K.

On reviendra plus tard sur ce type d'équations, mais on peut déjà noter que (19) comprend deux équilibres, celui (stable) où la population est de taille K, et celui (instable) où la population est inexistante. Le premier équilibre est stable, car si, partant d'une population égale à K, on subit une petite baisse d'effectifs ou une petite hausse d'effectifs due à une cause inconnue, l'équation (19) entraînera rapidement un retour vers cet effectif K. Le second équilibre est instable, car si l'on introduit juste quelques individus, la population va croître très vite ! Bien sûr, cette conclusion est irréaliste par certains côtés : on n'a jamais vu une espèce complètement éteinte augmenter ses effectifs...

On voit à travers ces quelques considérations simples l'aller-retour qui peut s'effectuer entre l'étude de l'équation et l'interprétation du modèle. On peut ensuite compliquer (enrichir) le modèle à volonté, tenant compte de questions de prédation, de l'effet de l'environnement, etc. Il existe quantité de variantes de ces équations.

Souvent, les conclusions ne pourront être obtenues par un pur raisonnement déductif, mais devront être observées par résolution numérique. cela nous mène à l'autre interface naturelle des équations différentielles : son calcul approché, en pratique par des ordinateurs convenablement programmés. La science de cette simulation est l'analyse numérique des équation différentielles ; elle passe toujours par une discrétisation des variables et des inconnues.

Pour comprendre les problématiques qui se posent, réfléchissons au calcul pratique de la solution de $\dot{x} = f(t, x(t))$, avec la donnée initiale

$x(0) = x_0$. On se donne un pas de temps h, qui servira à la discrétisation temporelle.

On réalise une approximation en partant de la formule de Taylor :

$$x(h) = x(0) + \int_0^h f(s, x(s))\, ds \simeq x_0 + hf(0, x_0).$$

Appelons x_1 le membre de droite, qui est donc une approximation de $x(h)$. On recommence alors :

$$x_2 := x_1 + hf(h, x_1) \simeq x(2h),$$

$$\dots x_n := x_{n-1} + f((n-1)h, x_{n-1}) \simeq x(nh).$$

Ce processus, très simple, est appelé **schéma d'Euler**. Lui sont associés deux types d'erreur. D'abord, à chaque fois on a remplacé une intégrale exacte par une fonction affine :

$$\int_{kh}^{(k+1)h} f(s, x(s))\, ds \simeq h\, f(kh, x(kh));$$

l'erreur associée doit être en $o(h)$ pour garantir une bonne qualité d'approximation. Le deuxième type d'erreur est plus subtil : quand on calcule x_2, on ne part pas de $x(h)$, mais de x_1, qui est déjà une erreur par rapport à $x(h)$; il y a donc possibilité d'amplification des erreurs, d'autant plus que l'on va réaliser de très nombreux pas si h est petit (pour un intervalle $[0, 1]$, il faudra environ $O(1/h)$ pas). On dit que le schéma est **consistant** si l'on peut contrôler le premier type d'erreur ; et **stable** si l'on peut contrôler le second. L'un des principes les plus importants de l'analyse numérique peut se résumer en une phrase : si un schéma numérique est à la fois consistent et stable, alors il **converge**, ce qui pour notre exemple signifiera

$$\max_{k \in [0, T/h]} |x_k - x(kh)| \xrightarrow[h \to 0]{} 0.$$

Il y a de nombreuses méthodes numériques pour résoudre les EDO, avec des erreurs de consistence en $o(h)$, voire $O(h^s)$ pour $s > 1$; on parle de s comme de l'**ordre** du schéma. Si s est "grand", la méthode sera précise ; en général cela n'est possible que si f est assez régulière. On connaît toute une zoologie de schémas numériques, même pour des équations très simples. Reprenons l'équation $x_{k+1} = x_k + hf(kh, x_k)$, et écrivons $y(kh) = x_k$ (de sorte que y est une fonction, définie sur des temps discrets, qui approche x). Cela peut se réécrire

$$\frac{y((k+1)h) - y(kh)}{h} = f\big(kh, y(kh)\big); \tag{20}$$

et on voit bien alors que cela revient à remplacer une dérivée, ou accroissement infinitésimal, dy/dt, par un vrai accroissement fini, ou dérivée discrète, calculée sur l'intervalle de temps $[kh, (k+1)h]$. Pour ce faire, on a évalué le champ de vecteurs f au temps $t = kh$.

Mais en y repensant, on constate que l'on aurait pu choisir d'autres valeurs pour la dérivée discrète : par exemple

$$\frac{y(kh) - y((k-1)h)}{h},$$

ou encore

$$\frac{y((k+1)h) - y((k-1)h)}{2h},$$

qui est symétrique entre le passé et l'avenir.

On aurait pu aussi, au lieu du membre de droite dans (20), inscrire $f((k+1)h, y(k+1)h)$: alors $Y = y((k+1)h)$ aurait été donné par la solution d'une **équation implicite**

$$Y - h\, f(kh, Y) = y(kh).$$

On pourrait aussi changer le pas de temps, au lieu de toujours choisir h ; peut-être que certaines valeurs demandent un calcul plus détaillé, par exemple si la fonction f commence à varier de manière plus importante...

On trouve toutes ces possibilités, et de nombreuses variantes, dans la littérature : schéma d'Euler implicite ou explicite, schémas de Runge–Kutta d'ordres variés, etc. Les études d'analyse numérique des EDO, les algorithmes des calculatrices et des logiciels de calcul, font grand usage de toutes ces variantes. Ainsi, le solveur du logiciel libre SciLab utilise un schéma de Runge–Kutta d'ordre 4 avec pas variable.

L'analyse numérique dépasse le cadre de ce cours ; pour une introduction complète à la discipline, on renvoie aux ouvrages de Crouzeix & Mignot [**7**] et Delabrière & Postel [**8**].

1.7. Résoudre ?

On parle couramment de la *résolution* d'une équation ; mais que signifie "résoudre" une EDO ?

Première remarque : avant tout, il s'agit de savoir s'il y a une solution, ou plusieurs, ou aucune... Même pour des équations dont l'inconnue est un nombre réel, il est très fréquent qu'il y ait plusieurs solutions (comme $x^2 - 1 = 0$), ou aucune solution (comme $x^2 + 1 = 0$).

Deuxième remarque : pour des EDO, quel que soit le sens que l'on donne au mot "résoudre", on a d'habitude avantage à le formuler en termes de flot.

Avec ces remarques en tête, résoudre peut signifier, selon le contexte :

• déterminer une expression de la solution (expression analytique, formule exacte) ; cela n'est possible d'habitude que dans des cas très particuliers, comme des équations linéaires.

• prouver que l'équation admet un flot bien défini, unique et de régularité suffisante. On pourra alors établir certaines propriétés qualitatives de la solution, et dresser un **portrait qualitatif** du flot, donnant accès à l'"allure" des solutions.

• expliciter une recette d'approximation numérique du flot, voire effectuer cette résolution par un programme informatique.

Par exemple, considérons l'équation

$$\ddot{x} = -x^3 - \dot{x}.$$

Il n'existe pas de "formule" pour l'équation, de sorte que la résolution au sens premier du terme est impossible ; mais on peut prouver l'existence d'un flot bien défini, de classe C^∞, et montrer que toute solution converge exponentiellement vite vers le point d'équilibre stable $(0,0)$; on peut même préciser encore cette description qualitative. Pour ce qui est de la résolution numérique, on pourra par exemple entreprendre un schéma d'Euler implicite : on approchera la solution $(x, \dot{x})$ (écrite dans l'espace des phases) par la solution approchée $(\widetilde{x}, \widetilde{v})$ définie par

$$\frac{\widetilde{v}((k+1)\tau) - \widetilde{v}(k\tau)}{\tau} = -\widetilde{x}((k+1)\tau)^3 - \widetilde{v}((k+1)\tau)$$
$$\frac{\widetilde{x}((k+1)\tau) - \widetilde{x}(k\tau)}{\tau} = \widetilde{v}((k+1)\tau);$$

en reportant la seconde équation dans la première, on trouvera une équation de degré 3 sur $\widetilde{x}((k+1)\tau)$, que l'on pourra résoudre numériquement avec un solveur d'équations réelles ; on montrera alors que cette solution converge uniformémement en temps. Pour cela, il sera utile d'avoir réalisé l'étape précédente (existence, unicité et régularité du flot !) De manière générale, une bonne compréhension du flot permet souvent d'éviter que le schéma numérique adopte un comportement aberrant...

Nous rencontrerons plus tard des exemples où l'on ne peut définir le flot : les solutions, tout simplement, n'existent pas ! Avant cela, nous allons considérer une situation où l'on peut résoudre le flot "presque explicitement" : les équations autonomes d'ordre 1 en dimension 1, c'est à dire de la forme $\dot{x} = f(x)$, où x est une variable rélle.

Soit donc $f : \mathbb{R} \to \mathbb{R}$ une fonction continue, et $x_0 \in \mathbb{R}$. Si $f(x_0) = 0$, alors c'est un point d'équilibre, et la solution issue de x_0 est triviale. Sinon, on peut se placer dans un voisinage où f ne s'annule pas, de

sorte que l'équation deviendra, après division,

$$\frac{\dot{x}}{f(x)} = 1. \tag{21}$$

Soit G une primitive de $1/f$: on reconnaît alors au membre de gauche de (21) la dérivée $(d/dt)G(x(t))$; une intégration en temps donnera donc

$$G(x(t)) - G(x_0) = t - t_0,$$

dès que la solution est bien définie sur $[t, t_0]$. Cela implique

$$x(t) = G^{-1}\big(t - t_0 + G(x_0)\big).$$

Ce calcul peut souvent être fait en pratique assez simplement, modulo bien sûr une opération de primitivation qui n'est que rarement explicite.

Cependant, on peut rencontrer des obstructions analytiques. Voici deux exemples importants :

(a) $\dot{x} = x^2$;

(b) $\dot{x} = x^{2/3}$.

Commençons par résoudre (a) : de $\dot{x}/x^2 = 1$ on tire $1/x_0 - 1/x = t - t_0$, en notant $x_0 = x(t_0)$. Il s'ensuit

$$x(t) = \frac{1}{1/x_0 - (t - t_0)} = \frac{x_0}{1 - x_0\, t}.$$

Cela est l'expression du flot $\Phi_t(x_0)$. (L'équation est autonome et on peut donc se contenter d'une seule variable de temps.) Le problème semble complètement résolu... mais en fait $\Phi_t(x_0)$ n'est pas défini en $t = 1/x_0$! La solution *diverge* quand $t \to t^* = 1/x_0$. (Noter que ce n'est pas aberrant d'avoir un temps homogène à $1/x_0$: on peut penser à x comme une variable de position, mais dans l'EDO $\dot{x} = x^2$ on a alors implicitement une constante qui a la dimension espace/temps, et cette constante se retrouvera dans les calculs.)

Il est important de comprendre que l'on ne peut pas définir le flot pour les temps $t > t^*$; en effet, la notion de flot n'a de sens que sur un intervalle de temps. Une fois que l'on perd la trace de la solution, il n'y a plus d'espoir de lui redonner une valeur. Pour ce qui est de définir la solution issue de x_0 en $t = 0$, on en est réduit à travailler seulement sur l'intervalle $[0, 1/x_0[$; on parle de *flot local* :

$$\Phi_t(x_0) = x(t) \quad \text{pour } 0 \le t < 1/x_0.$$

Si $|x_0| < \varepsilon$, cela donne un flot défini sur un intervalle de temps au moins égal à $[0, 1/\varepsilon[$, grand quand $\varepsilon \to 0$.

Pour l'équation (b), nous allons aussi rencontrer un problème, mais fort différent, cette fois près de la valeur $x = 0$. Il s'agit bien évidemment d'un point d'équilibre, donc $x(t) = 0$ est solution. Pourtant, on peut

vérifier que $x(t) = t^3/27$ est également solution de l'équation ! Autrement dit, même si l'équation est du premier ordre, la connaissance de l'état initial ne détermine pas la trajectoire. Il n'y a pas déterminisme, même au sens large où nous l'avons défini : ainsi les fonctions $x(t) = 0$ et $y(t) = t_+^3/27$ ($t_+ = \max(t, 0)$ étant la partie positive de la variable de temps) sont toutes deux solutions de l'équation, alors qu'elles coïncident sur l'intervalle de temps $[-1, 0]$, et même sur $\mathbb{R}_-$ tout entier. Pour cette équation, "le futur n'est pas déterminé par le passé". Ce comportement "déplaisant" n'est possible que parce que la fonction $x \longmapsto x^{2/3}$ n'est pas régulière, en fait pas différentiable. Nous verrons dans le chapitre suivant comment utiliser la différentiabilité pour démontrer des théorèmes généraux d'existence et unicité des solutions et du flot.

CHAPITRE 2

Théorèmes locaux

2.1. Théorème de Cauchy–Lipschitz

Cette section est consacrée au plus important théorème local de résolution des EDO ; "local" veut dire que la résolution se fait sur un petit intervalle de temps, au voisinage de la condition initiale.

Ce type d'énoncé apparaît au 19ème siècle avec Cauchy, qui cherche un procédé itératif général pour l'approximation des solutions des EDO. Le plus classique de ces résultats est maintenant appelé **théorème de Cauchy–Lipschitz**. Il permet de prouver, sous des conditions très générales, l'existence d'un flot assez régulier. Dans sa formulation moderne il a été mis au point au début du 20ème siècle, à l'aide du théorème de point fixe de Picard.

THEORÈME 19 (Théorème de Cauchy–Lipschitz). *Soient U un ouvert de $\mathbb{R}^n$ (espace des phases), I un intervalle de $\mathbb{R}$ (intervalle de temps), et $f = f(t,x)$ un champ de vecteurs défini sur $I \times U$, de classe C^1. On se donne en outre un temps initial $t_0 \in I$, et un état $x_* \in U$. Alors il existe $\varepsilon > 0$ tel que pour tout $x_0 \in U$ vérifiant $|x_0 - x_*| \leq \varepsilon$, on peut définir un unique flot $(\Phi_{t_0,t}(x_0))$, défini pour tous les temps t tels que $|t - t_0| \leq \varepsilon$, et de classe C^1 par rapport à t et x_0.*

De plus, si f est de régularité C^r ($r \geq 1$), alors Φ est aussi de régularité C^r dans toutes les variables.

REMARQUES 20. 1. Le Théorème de Cauchy–Lipschitz construit un flot seulement pour des temps t proches de t_0 ; il ne permet jamais de définir un flot pour des temps arbitrairement grands, même si $I = \mathbb{R}$.

2. Le Théorème 19 affirme l'existence et la régularité du flot. Cela contient la solution du problème de Cauchy : partant de n'importe quel x_0 proche de x_* on trouve une solution unique à l'équation. Mais la conclusion est beaucoup plus forte : elle contient aussi le fait que la solution dépend de manière régulière de x_0. La solution est donc une fonction continûment différentiable de la condition initiale ! Au cours de la preuve du théorème, nous apprendrons comment calculer la différentielle du flot, en résolvant le flot de l'équation linéarisée : cela est important en théorie comme en pratique.

3. Comme $\Phi_{t_0,t}$ est inversible (d'inverse Φ_{t,t_0}), le flot est en fait un C^1-difféomorphisme ; voire un C^r-difféomorphisme si f est de classe C^r.

4. Quitte à réduire encore ε, le flot Φ dépend aussi continûment de t_0 et même de f (mesuré dans une topologie C^1).

5. Ce théorème, étant local, fonctionne tout aussi bien sur une variété (que l'on peut paramétrer localement par un morceau d'espace euclidien).

6. L'hypothèse de régularité C^1 en t et x peut être remplacée par une condition plus faible : "localement lipschitzienne en x, uniformément en t, et continue en t". Le flot est alors lipschitzien mais pas forcément de classe C^1. En revanche, la régularité Hölder est insuffisante pour définir un flot, même irrégulier.

Au sujet de la dernière remarque, on rappelle les définitions des fonctions lipschitziennes et höldériennes :

Définition 21 (fonction lipschitzienne). Si O est un ouvert de $\mathbb{R}^n$, une fonction $g : O \to \mathbb{R}^d$ est dite lipschitzienne s'il existe une constante $L > 0$ telle que $|g(y)-g(z)| \leq L|y-z|$ pour tous $y, z \in O$. La meilleure constante L admissible est dite constante de Lipschitz de f.

Si U est un ouvert de $\mathbb{R}^n$, une fonction $g : U \to \mathbb{R}^d$ est dite localement lipschitzienne si elle est lipschitzienne dans toute boule ouverte B telle que $\overline{B} \subset U$.

Définition 22 (fonction höldérienne). Si O est un ouvert de $\mathbb{R}^n$ et $\alpha \in]0,1[$, une fonction $g : O \to \mathbb{R}^d$ est dite α-höldérienne s'il existe une constante $C > 0$ telle que $|g(y) - g(z)| \leq C|y - z|^\alpha$ pour tous $y, z \in O$.

Si U est un ouvert de $\mathbb{R}^n$, une fonction $g : U \to \mathbb{R}^d$ est dite localement α-höldérienne si elle est α-höldérienne dans toute boule ouverte B telle que $\overline{B} \subset U$.

Exemple 23. Toute fonction $f \in C^1(U; \mathbb{R}^d)$ est localement lipschitzienne : en effet, si O est un ouvert convexe borné avec $\overline{O} \subset U$, alors la formule de Taylor montre que $|f(x) - f(y)| \leq (\max_{\overline{O}} |df|)\,|x - y|$, donc f est L-lipschitzienne sur O, avec $L = \max_{\overline{O}} |df|$. On peut montrer en exercice que L est la meilleure constante admissible. (Si O n'est pas convexe, on montre que f est lipschitzienne sur O en recouvrant O par un nombre fini de boules ouvertes.)

Exemple 24. La fonction valeur absolue, ou plus généralement la norme euclidienne sur $\mathbb{R}^n$, est lipschitzienne (de constante 1) sans être différentiable. Il en est de même pour la fonction distance sur une surface.

Exemple 25. La fonction $x \longmapsto \sqrt{x}$, définie sur $]0, +\infty[$, est de classe C^∞, et en particulier localement lipschitzienne ; mais elle n'est

pas lipschitzienne (sa dérivée n'est pas bornée près de 0). En revanche, elle est 1/2-höldérienne. Si maintenant on considère la fonction $g(x) = \sqrt{|x|}$, définie sur $\mathbb{R}$ tout entier, alors g n'est pas localement lipschitzienne, ni dérivable en 0 ; mais elle reste 1/2-höldérienne.

L'intérêt de la condition qui apparaît dans la Définition 21, introduite par Lipschitz au 19ème siècle, est précisément de permettre la généralisation de nombreux énoncés à des fonctions qui ne sont pas différentiables. En fait la régularité lipschitzienne apparaît comme une **régularité critique** pour le théorème de Cauchy–Lipschitz : c'est le minimum qui permet de construire quelque chose. À partir d'un champ de vecteurs f höldérien, on ne réussira pas à construire un flot décent ; en revanche, dès que le champ de vecteurs est lipschitzien, non seulement on aura un flot bien défini, mais toute régularité supplémentaire sera transmise par f au flot (si f est C^1 alors le flot est C^1 ; si la différentielle de f est α-Hölder alors la différentielle du flot sera α-Hölder, etc.)

Remarque 26. Le théorème de Cauchy–Lipschitz associe à chaque condition initiale une solution, et garantit que cette correspondance est C^1. Mais si l'on se donne une solution, on retrouve la condition initiale en évaluant tout simplement la solution au temps t_0 ; cette opération aussi est de classe C^1 si l'on équipe les solutions de la topologie C^1. Autrement dit, l'application qui met en correspondance la condition initiale x_0 et la solution $x(t) = \Phi_{t_0,t}(x_0)$ est un difféomorphisme. C'est un principe important, lié au déterminisme : même si les solutions sont a priori recherchées dans un espace de dimension infinie, elles forment en fait une structure (une variété !) de dimension finie.

Remarque 27. Le Théorème de Cauchy–Lipschitz a une traduction géométrique, que l'on verra dans une section ultérieure : si un champ de vecteurs f de classe C^1 ne s'annule pas en x_*, alors au voisinage de x_* on peut *redresser* f, c'est à dire trouver un changement de coordonnées dans lequel ce champ est constant.

Voyons maintenant comment réinterpréter une difficulté déjà mentionnée. Soit l'équation $\dot{y}^3 = \pm y^2$, qui se résout par $\dot{y} = \pm|y|^{2/3}$. (Ici on note $\pm$ pour le signe de y.) La fonction $f(y) = \pm|y|^{2/3}$ est C^∞ sur $\mathbb{R} \setminus \{0\}$, mais elle n'est pas différentiable en $y = 0$: au voisinage de 0 elle est continue mais pas différentiable, ni d'ailleurs lipschitzienne. Le Théorème de Cauchy–Lipschitz peut donc s'appliquer à toute condition initiale *non nulle*, et échouera à dire quoi que ce soit pour une condition initiale nulle. Or on se souvient que $t^3/27$ est une solution exacte valant 0 en $t = 0$: il n'y a donc justement pas d'unicité, et pas

de flot bien défini, quand on passe par la valeur particulière 0. Ainsi le contre-exemple de la fin du Chapitre 1 s'interprète par le caractère non-lipschitzien du champ de vecteurs (qui est en fait 2/3-höldérien).

Pour de telles équations, l'existence de solutions est toujours garantie par le **Théorème de Cauchy–Peano** :

THEORÈME 28 (Cauchy–Peano). *Soit $f = f(t,x)$ un champ de vecteurs continu défini dans un ouvert O et un intervalle de temps I. Soient $x_* \in O$ et $t_0 \in I$. Alors pour tout x_0 assez proche de x_* on peut définir une solution au problème de Cauchy :*

$$\frac{dx}{dt} = f(t,x), \qquad x(t_0) = x_0.$$

Cette solution existe pour $|t - t_0| \leq \varepsilon$, où $\varepsilon > 0$ est assez petit.

Le théorème de Cauchy–Peano permet de prouver l'existence des solutions de l'EDO, sous des conditions bien plus générales que Cauchy–Lipschitz ; mais il ne fournit ni l'unicité des solutions, ni l'existence d'un flot ! En pratique, on l'utilise rarement. Nous ne le démontrerons pas ici.

En rapport avec l'unicité, on va énoncer une conséquence géométrique capitale du Théorème de Cauchy–Lipschitz.

PROPOSITION 29. *Soient $(x(t))$ et $(y(t))$ deux solutions d'une EDO $\dot{z} = f(t, z(t))$ définie par un champ de vecteurs f de classe C^1. Alors les fonctions x et y, si elles sont distinctes, ne peuvent jamais se rencontrer. En d'autres termes, s'il existe t tel que $x(t) = y(t)$, alors x et y coïncident pour tous les temps (où elles sont bien définies).*

COROLLAIRE 30. *Les trajectoires d'une EDO autonome, définie par un champ de vecteurs de classe C^1, ne s'intersectent jamais.*

DÉMONSTRATION. Soit t_0 tel que $x(t_0) = y(t_0)$; on note x_0 cette valeur commune. En appliquant Cauchy–Lipschitz, on trouve que $x(t)$ et $y(t)$ sont tous deux égaux à $\Phi_{t,t_0}(x_0)$, donc coïncident en tout temps. Cela prouve la Proposition 29. Le Corollaire 30 en découle : soient deux trajectoires distinctes d'une EDO autonome, disons x et y. S'il existe t_1 et t_2 tels que $x(t_1) = y(t_2)$, alors définissons $\widetilde{y}(t) = y(t + t_2 - t_1)$: comme l'EDO est autonome, $\widetilde{y}$ en est aussi solution (pour une EDO autonome, le choix de l'origine des temps n'a pas d'importance). Or $\widetilde{y}$ coïncide avec x en t_1, donc pour tous les temps par la Proposition 29, ce qui est impossible par hypothèse. □

Ces dernières propriétés illustrent encore une fois le fait que les solutions forment un espace très structuré.

2.2. Preuve de Cauchy–Lipschitz : Existence, unicité

Le Théorème 19 peut se prouver de différentes manières. Une démonstration utilise le théorème des fonctions implicites, appliqué à la fonctionnelle

$$\Phi(x) = \Big(x(t) - \int_0^t f(x,s)\,ds\Big)_{t\in[t_0-\tau,t_0+\tau]},$$

où τ est assez petit. Une autre preuve est basée sur le théorème du point fixe. Ces deux démonstrations sont cependant plus ou moins équivalentes, car le théorème des fonctions implicites repose lui aussi sur le théorème de point fixe.

THEORÈME 31 (Théorème du point fixe de Picard). *Soit E un espace de Banach (espace vectoriel normé complet), et soit $\phi : E \to E$ une fonction k-lipschitzienne, $0 < k < 1$. Alors l'équation de point fixe $\phi(x) = x$ admet une unique solution dans E.*

Ce théorème, vrai en dimension finie ou infinie, est à la base de nombreux théorèmes d'existence et d'unicité.

PREUVE DU THÉORÈME 31. On note d'abord qu'il ne peut y avoir plus d'un point fixe : en effet, si x et y sont deux points fixes distincts, alors $x - y = \phi(x) - \phi(y)$ implique $|x - y| = |\phi(x) - \phi(y)| \leq k|x - y|$, et en simplifiant par $|x - y|$ on trouve $1 \leq k$, impossible.

Pour prouver l'existence d'un point fixe, on choisit $x_0 \in E$ quelconque, puis on pose $x_1 = \phi(x_0)$, et plus généralement $x_n = \phi(x_{n-1})$ $(n \geq 1)$. On a donc

$$|x_{n+1} - x_n| = |\phi(x_n) - \phi(x_{n-1})| \leq k|x_n - x_{n-1}|,$$

et par récurrence $|x_{n+1} - x_n| \leq k^n|x_1 - x_0|$. On en déduit que la série $\sum(x_{n+1} - x_n)$ est une suite de Cauchy ; plus précisément

$$|x_{n+q} - x_n| \leq k^n(1 + ... + k^{q-1})|x_1 - x_0| \leq \frac{k^n}{1-k}|x_1 - x_0|.$$

La suite (x_n) est donc de Cauchy, et converge donc vers une limite $y \in E$. En passant à la limite dans l'équation $x_{n+1} = \phi(x_n)$, on trouve $y = \phi(y)$, c'est à dire que y est effectivement un point fixe. □

Le Théorème 31 a l'inconvénient de se placer dans un espace vectoriel normé E tout entier, alors que souvent l'on doit se limiter à une boule de E, pour localiser par exemple. Voici deux énoncés, très proches, qui permettent de répondre à ces besoins. On notera $B[x, R]$ la boule fermée de centre x et de rayon R.

THEORÈME 32 (Théorème du point fixe dans une boule, version I). *Soit x_* un élément d'un espace de Banach E, soit $R > 0$, et soit*

ϕ une fonction k-lipschitzienne définie sur $B[x_*, R]$, à valeurs dans $B[x_*, R]$. On suppose que $k < 1$; alors F admet un unique point fixe dans $B[x_*, R]$.

THEORÈME 33 (Théorème du point fixe dans une boule, version II). *Soit x_* un élément d'un espace de Banach E, soit $R > 0$, et soit ϕ une fonction k-lipschitzienne définie sur $B[x_*, R]$, à valeurs dans E. On suppose que $k < 1$ et $|\phi(x_*) - x_*| < R(1-k)$; alors F admet un unique point fixe dans $B(x_*, R)$.*

La différence entre les deux théorèmes est mince : tous deux reposent fondamentalement sur la k-lipschitzianité, mais dans le premier cas on utilise aussi un contrôle des valeurs de la fonction ϕ sur toute la boule (l'information selon laquelle $\phi(x)$ appartient à la boule), alors que dans le second cas on se contente d'un contrôle sur $\phi(x_*)$. Dans certaines situations, le second énoncé est plus précis.

PREUVE DU THÉORÈME 32. La démonstration est entièrement similaire à celle du Théorème 31, en utilisant le fait que la limite des x_n, a priori définie dans E, est en fait un élément de $B[x_*, R]$. (De manière plus intrinsèque, on peut dire que la boule $B[x_*, R]$ est un espace métrique complet, en tant que fermé d'un espace complet ; et la démonstration s'applique à tout espace métrique complet.) □

PREUVE DU THÉORÈME 33. La preuve d'unicité est la même que dans le Théorème 31. La preuve d'existence suit le même schéma, mais demande un peu plus d'attention. On choisit cette fois $x_0 = x_*$, et l'on définit une suite récurrente $(x_n)_{n\in\mathbb{N}}$ par $x_n = \phi(x_{n-1})$; cela a un sens tant que les x_n restent dans $B(x_*, R)$. (A priori la fonction ϕ n'est même pas définie en dehors de cette boule !) Si l'on peut trouver $R' < R$ tel que $|x_n - x_*| \leq R'$ pour tout n, alors la convergence se prouve selon le même procédé, dans la boule fermée $B[x_*, R]$ (qui est complet comme fermé d'un espace complet).

Tout ce qui compte c'est donc de prouver par récurrence que $|x_n - x_*| \leq R' < R$. Calculons donc

$$\begin{aligned} |x_n - x_*| &= |\phi^n(x_*) - x_*| \\ &\leq |\phi^n(x_*) - \phi^{n-1}(x_*)| + ...|\phi(x_*) - x_*| \\ &\leq (k^{n-1} + ... + k + 1)\,|\phi(x_*) - x_*| \\ &\leq \frac{1}{1-k}\,|\phi(x_*) - x_*|. \end{aligned}$$

Posons $R' = |\phi(x_*) - x_*|/(1-k)$: le calcul ci-dessous prouve que $|x_n - x_*| \leq R'$; et en passant à la limite on trouve $|y - x_*| \leq R'$, et par hypothèse $R' < R$, ce qui conclut la preuve. □

Nous pouvons maintenant entamer la démonstration du théorème de Cauchy–Lipschitz.

Preuve du Théorème 19 : Existence et unicité. On abrègera $\Phi_{t_0,t}(x_0)$ en $x(t)$ par simplicité. On peut rassembler l'équation $\dot{x} = f(t, x(t))$ et la condition initiale $x(t_0) = x_0$ en une seule formule :

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds.$$

Cette *reformulation intégrale* est naturellement écrite sous forme d'un point fixe : il faut que ce soit la même fonction x au membre de gauche et au membre de droite. Posons donc

$$F(x) = \left\{ t \longmapsto x_0 + \int_{t_0}^{t} f(s, x(s))\, ds \right\}. \tag{22}$$

(Attention, x dans le membre de gauche est une fonction dépendant du temps.) Le problème est de prouver que F admet un point fixe unique !

La fonctionnelle F dépend bien sûr de x_0 et de t_0. Par ailleurs elle n'a de sens que si l'on précise l'intervalle de temps sur lequel est défini x, disons $[t_0 - \varepsilon, t_0 + \varepsilon]$; et si x reste proche de x_* (a priori $f(t,x)$ n'est même pas défini si x s'écarte trop...). On supposera donc une borne $|x(s) - x_*| \leq r$, avec r assez petit ; et l'on supposera aussi que ε est plus petit qu'un certain θ tel que $f(t,x)$ est bien défini pour $t \in [t_0 - \theta, t_0 + \theta]$ et $|x - x_*| \leq r$.

Pour appliquer à F un théorème de point fixe, il faut préciser l'espace de Banach sur lequel cette fonctionnelle est définie ; en particulier, préciser la norme. Appelons donc E l'espace des fonctions continues à valeurs dans la boule fermée $B[x_*, r]$. La norme de la convergence uniforme fait de E un espace complet. On définit

$$F : C\big([t_0 - \tau, t_0 + \tau]; B[x_*, r]\big) \to C\big([t_0 - \tau, t_0 + \tau]; \mathbb{R}^n\big),$$

au moyen de la formule (22). On note que l'espace de définition est une boule $B[x_*, r]$ dans la topologie uniforme, si l'on identifie x_* à la fonction constante égale à x_*.

Pour appliquer le Théorème 32, il faut vérifier que, si τ est assez petit,

(a) F est lipschitzienne avec une constante de Lipschitz $k < 1$;

(b) $\|F(x) - x_*\| \leq r$

Notons L la constante de Lipschitz de f dans le domaine $\{|t - t_0| \leq \theta,\ |x - x_*| \leq r\}$, et M le supremum de $|f|$ dans ce même domaine.

Pour prouver la propriété (a) on calcule

$$F(x) - F(\widetilde{x}) = \int_{t_0}^{t} \big[f(s, x(s))ds - f(s, \widetilde{x}(s)) \big] \, ds.$$

On passe aux normes, et on applique la propriété de lipschitzianité : on trouve qu'au temps t

$$\begin{aligned} |F(x)(t) - F(\widetilde{x})(t)| &\leq \int_{t_0}^{t} |f(s, x(s)) - f(s, \widetilde{x}(s))| \, ds \\ &\leq L \int_{t_0}^{t} |x(s) - \widetilde{x}(s)| \, ds \\ &\leq L \, \varepsilon \, \|x - \widetilde{x}\|. \end{aligned}$$

(Au niveau typographique, on a utilisé la notation $|a|$ pour la norme de a dans $\mathbb{R}^n$ et la notation $\|v\|$ pour la norme d'une fonction dans la topologie continue.) Cette série d'inégalités montre que F est $(L\varepsilon)$-lipschitzienne, donc k-lipschitzienne si l'on pose $L\varepsilon = k$. Si $\varepsilon < 1/L$, alors $k < 1$, ce qui était la première condition à vérifier.

Pour prouver la propriété (b), on écrit

$$\begin{aligned} |F(x)(t) - x_*| &= \left| x_0 - x_* + \int_{t_0}^{t} |f(s, x(s))| \, ds \right| \\ &\leq |x_0 - x_*| + \int_{t_0}^{t} |f(s, x(s))| \, ds \\ &\leq |x_0 - x_*| + \varepsilon M. \end{aligned}$$

On suppose $|x_0 - x_*| \leq r/2$, et $\varepsilon < r/(2M)$; alors la propriété (b) est vérifiée. L'application du Théorème 32 prouve alors l'existence et l'unicité de la solution passant par x_0 au temps t_0, construite comme point fixe de (22). Cela est obtenu pour $|x_0 - x_*| \leq r/2$ et $|t - t_0| \leq \varepsilon < r/(2M)$; pour obtenir l'énoncé précis du théorème il reste à choisir $\varepsilon < \min(r/2, r/(2M))$.

A priori cette solution est seulement continue, mais par *bootstrap* on peut dire mieux. Repartons en effet de l'expression $x(t) = x_0 + \int_{t_0}^{t} f(s, x(s)) \, ds$. Comme x est continu et f aussi, le membre de droite est continûment différentiable en t. Il s'ensuit que le membre de gauche aussi : la fonction x est donc de classe C^1 dans la variable t, et l'on peut alors dériver les deux membres de (22) pour trouver l'équation $\dot{x} = f(t, x(t))$, vérifiée au sens classique. □

Remarque 34. Dans le Théorème 19 comme dans la preuve, on a pris soin de distinguer x_* (point de référence au voisinage duquel on travaille) et x_0 (condition initiale choisie près de x_*). Cependant,

pour les besoins de la preuve on aurait pu se contenter de travailler dans une boule centrée en x_0, et noter que les estimations obtenues sur ε sont explicites et varient continûment en fonction de x_0, donc sont uniformes pour x_0 dans un voisinage de x_*.

REMARQUE 35 (Variante de la preuve). Si l'on applique le Théorème 33 plutôt que le Théorème 32, on aboutit aussi au résultat, avec des estimations légèrement différentes sur ε. La condition (b) à vérifier est maintenant $|F(x_*) - x_*| < r(1-k)$, et l'estimation devient

$$|F(x_*) - x_*| = \left| x_0 - x_* + \int_{t_0}^{t} |f(s, x_*)|\, ds \right| \leq |x_0 - x_*| + \varepsilon M_*,$$

où $M_* = \sup\{|f(s, x_*)|;\ |s - t_0| \leq \theta\}$. Pour conclure il suffit donc de vérifier que $|x_0 - x_*| + \varepsilon M < r(1 - L\varepsilon)$, ce qui est vrai si $|x_0 - x_*|$ et ε sont assez petit.

À ce stade nous avons prouvé l'existence d'un flot local continu. Il nous reste seulement à prouver que le flot est non seulement continu, mais aussi (continûment) différentiable...

2.3. Preuve de Cauchy–Lipschitz : Régularité du flot

FIN DE LA PREUVE DU THÉORÈME 19. Dans la section précédente nous avons prouvé que le flot est bien défini ; nous allons maintenant montrer qu'il est régulier, et plus précisément de classe C^1. Cela ne veut pas dire que la solution est de classe C^1 (nous l'avons déjà prouvé), mais que la solution, vue comme fonction de la donnée initiale, est continûment différentiable.

Commençons par établir la continuité du flot. Soit (x_0^k) une suite de conditions initiales, convergeant vers x_0 quand $k \to \infty$. Pour chaque x_0^k on peut trouver une solution $x^k(t)$ de l'EDO, telle que $x^k(t_0) = x_0^k$. Comme les fonctions x^k sont uniformément bornées et de dérivée uniformément bornée, on peut appliquer le théorème de compacité d'Ascoli : quitte à extraire une sous-suite, $x^k(t)$ converge, uniformément en t, vers une fonction $(y(t))$. En passant alors à la limite dans la formulation intégrale de l'équation,

$$x^k(t) = x_0^k + \int_{t_0}^{t} f(s, x^k(s))\, ds,$$

on trouve que y est aussi solution de l'équation :

$$y(t) = x_0 + \int_{t_0}^{t} f(s, y(s))\, ds;$$

c'est donc l'unique solution passant par x_0 en t_0. Cela prouve que $\Phi_{t_0,t}(x_0^k) \longrightarrow \Phi_{t_0,t}(x_0)$, c'est à dire que le flot est continu.

Pour prouver la différentiabilité de $x(t)$ par rapport à x_0, on va calculer la différentielle ; et pour cela, en pratique on va différentier l'équation. Écrivons

$$\dot{x}(t) = f(t, x(t)), \qquad x(t_0) = x_0 \tag{23}$$

et dérivons par rapport à x_0 : on trouve

$$\frac{d}{dt}\left(\frac{\partial x(t)}{\partial x_0}\right) = df(t, x(t)) \circ \frac{\partial x(t)}{\partial x_0}, \qquad \frac{\partial x(t_0)}{\partial x_0} = \mathrm{Id}\,.$$

Ici on se souvient que $x(t)$, abréviation de $\Phi_{t_0,t}(x_0)$, dépend de x_0 ; et on a appliqué la dérivation des fonctions composées pour calculer la dérivée de $f(t, x(t))$ par rapport à la condition initiale.

Nous avons donc deviné la formule recherchée : si X est la différentielle du flot par rapport à la condition initiale,

$$X(t) = \frac{\partial x(t)}{\partial x_0} = \frac{\partial \Phi_{t_0,t}(x_0)}{\partial x_0},$$

alors X vérifie

$$\begin{cases} \dot{X}(t) = df(t, x(t))X(t) \\ X(t_0) = \mathrm{Id}\,. \end{cases} \tag{24}$$

C'est une équation différentielle *linéaire, non-autonome, matricielle* ; en effet l'inconnue X est une application linéaire, que l'on peut représenter par une matrice $n \times n$.

Remarques 36. 1. On pourrait avoir l'impression, à première vue, que l'équation (24), étant linéaire, pourrait être résolue explicitement, plus facilement que l'équation non linéaire (23) ; mais ce n'est pas le cas puisque la matrice $df(t, x(t))$ dépend du temps, et en fait de toute la solution de (23).

2. Une façon de reformuler ce résultat consiste à dire que la différentielle du flot $\Phi_{t_0,t}$ n'est autre que le flot $\Phi^L_{t_0,t}$ associé à l'équation différentielle linéarisée.

Reprenons le cours de la preuve. Le Théorème de Cauchy–Lipschitz (du moins la partie que nous avons déjà démontrée) nous permet de définir une solution unique à (24), sur un intervalle de temps assez petit. Reste à vérifier que c'est bien la différentielle du flot. Pour cela repartons de la formulation intégrale de ce flot :

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

ou plus rigoureusement

$$\Phi_{t_0,t}(x_0) = x_0 + \int_{t_0}^{t} f(s, \Phi_{t_0,s}(x))\, ds. \tag{25}$$

Écrivons aussi l'équation sur X sous forme intégrale :

$$X(t) = \mathrm{Id} + \int_{t_0}^{t} df(s, \Phi_{t_0,s}(x_0))X(s)ds. \tag{26}$$

On forme ensuite

(27)

$$\delta(t, x_0, h) = \Phi_{t_0,t}(x_0 + h) - \Phi_{t_0,t}(x_0) - X(t)h$$

(28)

$$= \int_{t_0}^{t} \Big[f(s, \Phi_{t_0,s}(x_0 + h)) - f(s, \Phi_{t_0,s}(x_0)) - df(s, \Phi_{t_0,s}(x_0))X(s)h \Big]\, ds.$$

Par définition de δ, l'argument de la première occurrence de f vaut

$$\Phi_{t_0,s}(x_0 + h) = \Phi_{t_0,s}(x_0) + X(s)h + \delta(s, x_0, h);$$

donc l'expression entre crochets vaut (avec le raccourci $x(s) = \Phi_{t_0,s}(x_0)$)

$$f(x + Xh + \delta) - f(x) - df(x)Xh,$$

ce qui vaut, grâce à la définition de la différentiabilité, $df(x)\delta + o(|h|) + o(|\delta|)$, ou, plus explicitement,

$$df(x)\delta + \zeta(|h|)|h| + \zeta(|\delta|)|\delta|, \tag{29}$$

ζ étant une fonction telle que $\zeta(r) \to 0$ quand $r \to 0$.

On majore la norme de $|df|$ par L, et on majore brutalement $\zeta(|\delta|)$ par une constante C, de sorte que l'expression (29) est bornée en norme par

$$(L + C)|\delta| + \zeta(|h|)|h|.$$

En reportant dans (28) on obtient

$$|\delta| \leq (L + C)\,|t_0 - t||\delta| + \zeta(|h|)\,|h| \leq (L + C)\varepsilon|\delta| + \zeta(|h|)\,|h|.$$

On suppose alors $\varepsilon < 1/(L + C)$, et on déduit de cette inégalité

$$|\delta| \leq \frac{\zeta(|h|)\,|h|}{1 - (L + C)\varepsilon} = o(|h|).$$

Cela signifie exactement que $\Phi_{t_0,t}(x_0)$ est une fonction différentiable de x_0, et que sa dérivée est $X(t)$.

Une fois la différentiabilité du flot prouvée, on montre la continuité de la dérivée avec le même raisonnement que précédemment.

Ensuite c'est "seulement" une question de courage que de vérifier que si f est de classe C^r alors le flot aussi ! □

Terminons avec quelques remarques techniques sur la preuve.

REMARQUE 37. La majoration $\varepsilon < 1/(L+C)$ est grossière ; on peut en fait, avec un peu plus de soin, se contenter de $\varepsilon < 1/L$. C'est un détail mais on va esquisser l'argument car il est intéressant. Soit $L' > L$ (aussi proche de L qu'on le souhaite) ; si σ est assez petit on sait que

$$|\delta| \leq \sigma \Longrightarrow \zeta(|\delta|) \leq (L' - L), \tag{30}$$

et alors l'expression entre crochets dans (28) sera bornée par $L'|\delta| + o(|h|)$; on en déduira

$$|\delta| \leq L' \, |t_0 - t||\delta| + \zeta(|h|) \, |h| \leq L'\varepsilon \, |\delta| + \zeta(|h|) \, |h|,$$

et donc, dès que $\varepsilon < 1/L'$,

$$|\delta| \leq \frac{\zeta(|h|) \, |h|}{1 - L'\varepsilon} = o(|h|). \tag{31}$$

Cependant, il semble qu'il y ait un cercle vicieux dans le raisonnement : pour prouver que $|\delta|$ est petit, on a fait l'hypothèse que $|\delta|$ est borné par le petit nombre σ ! Cependant, cette dernière borne est bien moins forte que la borne que l'on a obtenue ; en effet, si $|h|$ est assez petit, le membre de droite de (31) (appelons-le $\xi(|h|)$) est bien plus petit que σ. Cela nous laisse la possibilité de démontrer la conclusion rigoureusement, par un argument de continuité, dont le principe est souvent fort utile. Pour mettre cela en forme, supposons que $|h|$ est assez petit pour que $|\xi(h)| \leq \sigma/2$. Notons ensuite que δ est une fonction continue de t, et vaut 0 en $t = t_0$; donc pour t assez proche de t_0, on aura forcément $|\delta| \leq \sigma$. Mais cela impliquera alors, par nos estimations, $|\delta| \leq \xi(|h|) \leq \sigma/2$; donc, quand t continuera à s'écarter de t_0, pour un petit intervalle de temps on aura encore $|\delta| \leq \sigma$, et donc en fait $|\delta| \leq \xi(|h|) \leq \sigma/2$. Et ainsi de suite ; en fait, si jamais $|\delta|$ croissait jusqu'à atteindre la valeur σ, alors à cet instant il serait encore borné par $\xi(|h|)$, ce qui impliquerait qu'il est plus petit que $\sigma/2$, d'où contradiction. On conclut que $|\delta| \leq \sigma/2$ tout au long du processus, et on a bien finalement $|\delta| \leq \xi(|h|)$.

REMARQUE 38. On verra plus tard (Section 2.5) une preuve directe de ce que le flot est lipschitzien, et plus précisément $e^{|t-t_0|L}$-lipschitzien (avec L la constante de Lipschitz de f). Cette borne directe, obtenue par le Lemme de Gronwall, permet de retrouver la continuité du flot de manière quantitative, et sans passer par l'argument de compacité ci-dessus. Elle permet en outre de prouver la différentiabilité avec un plan encore légèrement différent :

- on prouve que $|X(t)|$ est borné, également par Gronwall ;

- de cela et du caractère lipschitzien du flot, on déduit de (27) que $|\delta| = O(|h|)$;

- on en déduit que $f(x + Xh + \delta) - f(x) - df(x)Xh = df(x)\delta + o(|\delta|) + o(|h|) = df(x)\delta + o(|h|)$, ce qui se majore en $L|\delta| + o(|h|)$. On conclut comme dans la preuve déjà effectuée, en choisissant $L\varepsilon < 1$.

2.4. Différentiation par rapport à un paramètre

En pratique, dans les applications, on peut oublier l'essentiel de la preuve du théorème de Cauchy–Lipschitz et se contenter d'en appliquer les conclusions. Cependant, le raisonnement qui nous a permis de calculer la différentielle du flot est extrêmement utile et il est très fréquent que l'on soit amené à le reproduire, pour étudier les variations du flot par rapport à tel ou tel paramètre.

Dans la section précédente, nous avons dérivé l'équation non linéaire $\dot{x} = f(t, x)$, par rapport à la condition initiale, pour obtenir l'équation (24). Le même calcul peut s'appliquer à n'importe quel paramètre, au delà de la condition initiale. Soit a un paramètre auxiliaire, par exemple réel, et une équation de la forme

$$\dot{x} = f(t, x, a). \tag{32}$$

Dérivons, formellement, par rapport au paramètre a ; en échangeant dans le membre de gauche les dérivées par rapport à t et a, on obtient

$$\frac{d}{dt}\left(\frac{\partial x}{\partial a}\right) = \frac{\partial f}{\partial a}(t, x, a) + \frac{\partial f}{\partial x}(t, x, a)\frac{\partial x}{\partial a}. \tag{33}$$

C'est une équation affine, inhomogène en temps, portant sur la fonction inconnue $\partial x/\partial a$; contrairement à l'équation (24) elle comporte un terme source (le premier terme du membre de droite). On l'appelle **équation linéarisée** par rapport au paramètre a.

La différence entre (24) et (33) n'est qu'apparente : en fait on peut inclure la dérivation par rapport à un paramètre et la dérivation par rapport à la donnée initiale dans un formalisme commun. Il suffit pour cela d'agrandir l'espace des phases pour qu'il contienne le paramètre a, de définir un nouvelle inconnue (x, a), et de remplacer l'équation (32) par

$$\frac{dx}{dt} = f(t, x, a), \qquad \frac{da}{dt} = 0.$$

Ainsi le champ de vecteurs f est changé en $(f, 0)$, et la donnée initiale doit être (x_0, a).

La dérivation par rapport à a devient donc la dérivation par rapport à la seconde composante de la donnée initiale. En particulier,

on peut appliquer le Théorème de Cauchy–Lipschitz pour prouver la différentiabilité C^1 de la solution de (32) par rapport à a.

EXEMPLE 39. Considérons l'équation (13) d'un point dans un potentiel V, avec un coefficient λ de friction linéaire. Admettons que les solutions en sont bien définies pour tout temps. Quelle est la dépendance de la solution par rapport à λ ? Pour le savoir, on dérive l'équation, soit sous la forme (13) du second degré, soit sous la forme équivalente d'un système d'ordre 1. Faisons-le sous cette dernière forme : à partir de

$$\frac{dx}{dt} = v(t), \qquad \frac{dv}{dt} = -\nabla V(x(t)) - \lambda v(t),$$

on trouve

$$\frac{d}{dt}\left(\frac{dx}{d\lambda}\right) = \frac{dv}{d\lambda}, \qquad \frac{d}{dt}\left(\frac{dv}{d\lambda}\right) - \nabla^2 V(x(t))\left(\frac{dx}{d\lambda}\right) - v(t) - \lambda\left(\frac{dv}{d\lambda}\right).$$

(Ici on a utilisé la notation $dv/d\lambda$ pour $\partial v/\partial\lambda$, c'est juste une question de conventions.) Nous avons ainsi un système de deux équations à deux inconnues, qui permet de calculer la dérivée de (x, v) par rapport à λ numériquement, disons.

EXEMPLE 40. Soit un champ de vecteurs $f(t, x; a)$, de classe C^1, où t est le temps, x l'état du système et a un paramètre. Supposons que f est périodique en t, de période T, et supposons que pour $a = 0$ on a une solution T-périodique, disons $y(t)$: partant de $y(0) = y_0$ elle y retourne, de sorte que $y(T) = y_0$, et plus généralement $y(kT) = y_0$. (En fait, si $y(T) = y_0$, alors en appliquant le théorème de Cauchy–Lipschitz autant de fois que nécessaire on peut montrer que $y(kT) = y_0$, par un argument d'unicité.) On se demande si pour $a \neq 0$ on trouvera encore des solutions T-périodiques. Face à un tel problème, on peut d'abord mettre le but en équation : la condition recherchée est $x(T) = x(0)$. Supposons que l'on ait un flot bien défini au voisinage de $(x_0; 0)$: pour tout a proche de 0, et tout x_0 proche de y_0, on peut résoudre l'équation différentielle $\dot{x} = f(t, x; a)$ partant de x_0 en $t = 0$, en une solution $x(t, x_0; a)$. L'équation de périodicité devient donc

$$x(T, x_0; a) - x_0 = 0. \tag{34}$$

Pour résoudre localement cette *équation implicite* au voisinage de $(x_0; 0)$ il suffit de prouver l'inversibilité de la dérivée par rapport à la variable x_0, en la valeur $a = 0$. Soit donc

$$Z = \frac{\partial x}{\partial x_0}$$

la dérivée du flot par rapport à x_0 : on sait que

$$\frac{dZ}{dt} = df\big(t, x(t;a);a\big)\, Z(t), \qquad Z(0) = \mathrm{Id}$$

Pour $a = 0$ cette équation est

$$\frac{dZ}{dt} = df\big(t, y(t);0\big)\, Z(t), \qquad Z(0) = \mathrm{Id}\,.$$

Notons Z_0 la solution de cette équation, évaluée le long de la solution particulière $y(t)$. Notre but est de savoir si $Z(T) - \mathrm{Id}$ est inversible ; autrement dit, si 1 est valeur propre de $Z(T)$.

On n'ira pas plus loin dans cet exemple, il faudrait ajouter des hypothèses ou considérer des formes particulières ; il s'agissait juste de voir que le concept de flot, et la dérivation par rapport à un paramètre, permet de ramener un problème local (existe-t-il une orbite périodique pour de petites valeurs de a, au voisinage de l'orbite particulière $y(t)$) à l'étude d'une propriété évaluée uniquement en la valeur particulière du paramètre (une propriété de l'orbite $y(t)$).

2.5. Lemme de Gronwall

Le Lemme de Gronwall est d'usage tellement fréquent dans la théorie des EDO, et plus généralement des équations d'évolution, qu'il mérite bien une section pour lui seul.

Lemme 41 (Lemme de Gronwall). *Soient $z(t)$, $a(t)$, $b(t)$ des fonctions à valeurs réelles définies sur un intervalle de temps $[t_0, t_1]$. On suppose que a et b sont bornées (mesurables), et que z est une fonction continue vérifiant l'inégalité*

$$\forall t \in [t_0, t_1], \quad \frac{d^+ z}{dt} \leq a(t)z(t) + b(t). \tag{35}$$

Soit

$$A(t) = \int_{t_0}^{t} a(s)\, ds$$

la primitive de a qui s'annule en t_0. Alors pour tout $t \in [t_0, t_1]$,

$$z(t) \leq e^{A(t)} z(t_0) + \int_{t_0}^{t} e^{A(t)-A(s)}\, b(s)\, ds.$$

Remarques 42. 1. De manière légèrement plus intrinsèque, on peut écrire

$$z(t) \leq e^{A(t)-A(t_0)} z(t_0) + \int_{t_0}^{t} e^{A(t)-A(s)} b(s)\, ds,$$

où A est n'importe quelle primitive de a.

2. Ici d^+/dt désigne la dérivée (supérieure) à droite :

$$\frac{d^+ f}{dt} := \limsup_{s \to t^+} \frac{f(s) - f(t)}{s - t},$$

qui est toujours définie, à valeurs dans $\mathbb{R} \cup \{\pm\infty\}$, et que l'on pourra à l'occasion, par abus de notation, désigner avec un point, comme une vraie dérivée. En pratique on applique souvent ce lemme à des fonctions qui ne sont pas forcément différentiables, comme des normes de fonctions vectorielles ; de sorte que, même quand on travaille avec des solutions lisses, on est régulièrement amené à utiliser ces dérivées supérieures pour le Lemme de Gronwall. Cela ne pose guère de problème, grâce à l'inégalité très commode qui suit : si g est une fonction différentiable de t, à valeurs dans $\mathbb{R}^n$, et $\| \cdot \|$ une norme, alors

$$\frac{d^+ \|g\|}{dt} \leq \left\| \frac{dg}{dt} \right\|.$$

3. Il existe aussi une version entièrement intégrale du lemme de Gronwall, où l'on n'écrit aucune dérivée. On pourra la démontrer en exercice.

Preuve du Lemme de Gronwall. Réécrivons (35) sous la forme

$$\frac{d^+ z}{dt} - a(t) z(t) \leq b(t). \tag{36}$$

On multiplie le membre de droite par $e^{-A(t)}$ pour le transformer en une dérivée, en l'occurrence celle de $e^{-A(t)} z(t)$. En effet,

$$\frac{d^+}{dt}\left[e^{-A(t)} z(t)\right] = e^{-A(t)} \left[-a(t) z(t) + \frac{d^+ z}{dt}\right];$$

donc (36) est équivalent à

$$\frac{d^+}{dt}\left[e^{-A(t)} z(t)\right] \leq e^{-A(t)} b(t).$$

On intègre ensuite entre t_0 et t, en utilisant l'inégalité $F(b) - F(a) \leq \int_a^b (d^+ F/dt)$: on trouve

$$e^{-A(t)} z(t) - e^{A(t_0)} z(t_0) \leq \int_{t_0}^t e^{-A(s)} b(s)\, ds.$$

Après multiplication par $e^{A(t)}$ on obtient la conclusion souhaitée. □

Remarque 43. Dans cette preuve on a utilisé sans justification le fait que $dA/dt = a$, ce qui est vrai pour tout temps t si a est continue, et pour presque tout temps t si a est simplement mesurable bornée. (Noter que la fonction $e^{-A(t)} z(t)$ est continue par hypothèse.) En pratique,

presque toujours, la fonction a est continue et il n'y a pas à se soucier de la différentiabilité de A.

En application du Lemme de Gronwall, nous allons démontrer la

PROPOSITION 44. *Soient U un ouvert de $\mathbb{R}^n$, I un intervalle de temps, et $f(t,x)$ un champ de vecteurs $I \times U \to \mathbb{R}^n$, L-lipschitzien en x (uniformément en t). Soient aussi $t_0 \in I$ et $x_0 \in U$. Alors l'équation $\dot{x} = f(t,x)$ admet au plus une solution $(x(t))_{t\in I}$ telle que $x(t_0) = x_0$.*

REMARQUE 45. Dans le théorème de Cauchy–Lipschitz on avait aussi prouvé l'unicité, mais le théorème dans son ensemble était énoncé pour un petit intervalle de temps.

PREUVE DE LA PROPOSITION 44. Soient $(x(t))$ et $(y(t))$ deux solutions définies sur I, telles que $y(t_0) = x(t_0) = x_0$; on va montrer que $x(t_1) = y(t_1)$ pour tout $t_1 \in I$. Sans perte de généralité (quitte à renverser le sens du temps et le signe de l'équation), on peut supposer que $t_1 > t_0$. On va alors appliquer le Lemme de Gronwall à l'intervalle $[t_0, t_1] \subset I$.

Par hypothèse,

$$\dot{x}(t) = f(t, x(t)), \qquad \dot{y}(t) = f(t, y(t)).$$

Alors $x(t)$ et $y(t)$ sont des fonctions dérivables, et

$$\frac{d^+}{dt}\|x - y\| \leq \left\| \frac{dx}{dt} - \frac{dy}{dt} \right\| = \|f(t, x(t)) - f(t, y(t))\| \leq L\|x(t) - y(t)\|,$$

où L est la constante de Lipschitz de f. On peut alors appliquer le Lemme de Gronwall avec $a(t) = L$, et on trouve, pour tout $t \in [t_0, t_1]$,

$$\|x(t) - y(t)\| \leq \|x(t_0) - y(t_0)\| e^{L(t-t_0)}. \tag{37}$$

Comme par hypothèse, $x(t_0) = y(t_0)$, on en déduit effectivement que $x(t_1) = y(t_1)$. □

REMARQUE 46. L'estimation (37) va au-delà de la propriété d'unicité : ele montre que si, dans la résolution de l'EDO, on réalise une petite erreur sur la donnée initiale, alors l'erreur résultant au temps t est contrôlée par l'erreur au temps t_0, avec un facteur multiplicatif exponentiel. On parle d'estimation de **stabilité**.

REMARQUE 47. Notons $F(t,x) = a(t)x + b(t)$, où $x \in \mathbb{R}_+$. Le Lemme de Gronwall repose sur une *comparaison* entre les solutions de l'inégalité différentielle $\dot{z} \leq F(t,z)$ et les solutions de l'équation différentielle $\dot{y} = F(t,y)$. Il est facile de prouver un principe de comparaison dès que la fonction $F(t,x)$ est *croissante* en x : en effet, tant que $z \leq y$, on aura $\dot{z} \leq F(t,z) \leq F(t,y) = \dot{y}$, la différence entre x et

y sera donc toujours croissante. Mais cela ne couvre pas dans toute sa généralité le Lemme de Gronwall, qui est vrai aussi bien pour $a(t) > 0$ que pour $a(t) < 0$ (et dans ce dernier cas la fonction $a(t)x + b(t)$ est décroissante en x).

2.6. Étude locale et divergence

Nous avons appris à dériver par rapport aux conditions initiales, mais quelle information peut-on en tirer ? Nous allons en déduire l'un des outils les plus importants de l'analyse locale des EDO : la **divergence**.

Pour commencer, réécrivons à nouveau l'équation de la dérivée par rapport à la condition initiale :

$$\frac{dX}{dt} = df(t, x(t))X(t) \qquad X(t_0) = \mathrm{Id}\,. \tag{38}$$

Pour $t > t_0$, la matrice $X(t)$ indique comment le flot a fait varier l'espace des états entre le temps t_0 et le temps t. En particulier, les valeurs propres de $X(t)$ nous donnent des informations précieuses sur la façon dont les distances ont été distordues par le flot – contractées (valeurs propres inférieures à 1 en module) ou étirées (valeurs propres supérieures à 1 en module).

Si l'on a une valeur propre $\lambda \in \mathbb{C}$, alors l'exponentielle matricielle a une valeur propre $e^{t\lambda}$, dont le module est égal à $e^{t\mathcal{R}e\,\lambda}$ Ce sont donc les *parties réelles des valeurs propres de la matrice df* qui nous donnent ces informations sur la distorsion :

- partie réelle strictement positive, tendance à la croissance exponentielle ;

- partie réelle strictement négative, tendance à la décroissance exponentielle

On peut aussi s'intéresser à la façon dont les *volumes* sont transformés par le flot. Considérons les solutions issues d'un ensemble de configurations initiales au temps t_0 : quand on arrivera au temps t est-ce que les configurations seront moins nombreuses ou plus nombreuses ?

C'est le *déterminant jacobien* $\det(dx(t)/dx_0) = \det X(t)$ qui nous donnera cette information. On se souvient en effet de l'interprétation du déterminant jacobien : si l'on fait agir la transformation linéaire $X(t)$ sur un petit cube infinitésimal, on obtient un parallélépipède infinitésimal ; et

- d'une part, le volume de ce parallélépipède est le volume initial multiplié par le déterminant ;

- d'autre part, le parallélépipède est, à peu de choses près (à des infiniments petits d'ordre supérieur près) l'image du petit cube par le flot.

En résumeé, $\det(d\phi)$ est le coefficient par lequel la transformation ϕ multiplie les volumes infinitésimaux.

Commençons par le cas d'une équation différentielle linéaire à coefficients constants, $\dot{z} = Az$. La formule $\det(e^{tA}) = e^{t(\operatorname{tr} A)}$ nous laisse trois possibilités :

- $\operatorname{tr} A > 0$: le déterminant jacobien croît exponentiellement vite ;
- $\operatorname{tr} A < 0$: le déterminant jacobien décroît exponentiellement vite ;
- $\operatorname{tr} A = 0$: le déterminant jacobien reste constant, égal à 1.

Exemple 48. Soit $\ddot{x} = -x - \lambda\dot{x}$ l'équation de l'oscillateur harmonique avec frottement ($\lambda > 0$) dans $\mathbb{R}^n$. On réduit l'équation à une EDO du premier ordre en se plaçant dans l'espace des phases (x, v) : on trouve un champ de vecteurs $f(x, v) = (v, -x - \lambda v) = A(x, v)$ où

$$A = \begin{pmatrix} 0 & I \\ -I & -\lambda I \end{pmatrix}.$$

La trace de A vaut $-\lambda n$, donc le déterminant jacobien est égal à $e^{-\lambda n t}$. Ce déterminant est exponentiellement décroissant : cela traduit l'effet stabilisant du frottement, qui réduit rapidement l'espace des configurations que le système peut occuper : il y a *contraction* dans l'espace des phases.

Que reste-t-il de ce calcul quand on s'intéresse à des équations qui ne sont pas linéaires à coefficients constants ? Il se peut que certaines configurations entraînent une contraction, et d'autres une dilatation de l'ensemble des configurations.

Pour cela nous allons chercher une équation différentielle sur le déterminant jacobien, que nous appellerons $\phi(t)$. Ce calcul utilise la différentielle du déterminant jacobien, que nous rappelons : le déterminant est une application qui va de l'ensemble des matrices $n \times n$ dans $\mathbb{R}$, et sa différentielle en M est l'application linéaire définie par

$$d_M \det : H \longmapsto (\det M) \operatorname{tr}(M^{-1}H). \tag{39}$$

On note que cette formule n'est valable que si M est inversible.

En utilisant successivement la dérivation des fonctions composées, l'équation (39) avec $H = X(t)$, et l'équation (38), on obtient

$$\frac{d}{dt}(\det X(t)) = (\det X(t))\ \mathrm{tr}\,(X(t)^{-1}\,\dot{X}(t))$$
$$= (\det X(t))\ \mathrm{tr}\,\Big(X(t)^{-1}\,df(t,x(t))\,X(t)\Big).$$

On rappelle en effet que $X(t)$ est inversible, car différentielle d'un difféomorphisme. Mais la trace est invariante par changement de base, c'est à dire par conjugaison par une application linéaire inversible :

$$\mathrm{tr}\,\Big(X(t)^{-1}\,df(t,x(t))\,X(t)\Big) = \mathrm{tr}\,(df(t,x(t))).$$

En conclusion, on a démontré que

$$\frac{d\phi}{dt} = \phi(t)\ \mathrm{tr}\,(df(t,x(t))). \tag{40}$$

Cette dernière quantité, la trace de la différentielle, s'appelle la **divergence** du champ de vecteurs f, et on la note $\mathrm{div}\, f$ ou encore $\nabla \cdot f$. On rappelle au passage la très importante formule d'intégration par parties (variante de la formule de Stokes) en plusieurs variables : si f est un champ de vecteurs et u une fonction, alors sous de bonnes hypothèses de régularité, $\int f \cdot \nabla u\, dx = -\int (\nabla \cdot f)\, u\, dx$.

La formule (40) se réécrit donc

$$\frac{d}{dt}\log \phi(t) = (\nabla \cdot f)(t,x(t)). \tag{41}$$

On rappelle que $x(t)$ est la solution de l'équation différentielle, et $\phi(t)$ le déterminant jacobien du flot le long de cette solution. On peut exprimer (41) en mots : *La dérivée logarithmique du volume infinitésimal des conditions, le long du flot, c'est la divergence du point que l'on est en train de visiter.* On note au passage que le déterminant jacobien ϕ ne peut s'annuler : c'est normal, puisque le flot est toujours un difféomorphisme, donc sa jacobienne est de déterminant non nul.

Remarque 49. La divergence porte bien son nom : elle mesure à quel point le champ de vecteurs diverge, c'est à dire a tendance à s'écarter. Il faut imaginer de petites particules infinitésimales transportées par le flot : dans un champ divergent (divergence positive), les particules vont s'écarter les unes des autres, et la densité de particules va diminuer ; dans un champ convergent (divergence négative), les particules vont se concentrer et leur densité va augmenter.

Un cas particulier très important est celui où la divergence est nulle : le déterminant jacobien reste alors constant !

EXEMPLE 50. Soit $\ddot{x} = -\nabla V(x)$ une équation newtonienne définie par un gradient ; alors $f(x, v) = (v, -\nabla V(x))$, et les coefficients diagonaux de df sont tous nuls, sa trace est donc bien évidemment nulle.

2.7. Complément : théorème de redressement

Cette section est consacrée à l'équivalence entre le Théorème de Cauchy–Lipschitz (Théorème 19) et le Théorème de Redressement qui suit :

THEORÈME 51 (Théorème de Redressement). *Soit $g \in C^1(O; \mathbb{R}^n)$ un champ de vecteurs défini dans un ouvert O de $\mathbb{R}^n$, et soit $y_* \in O$ tel que $g(y_*) \neq 0$. Alors il existe un difféomorphisme ϕ, défini sur un voisinage de y_*, tel que $\phi_* g = \xi$ est un champ de vecteurs constant.*

Pour la commodité, rappelons l'énoncé du Théorème de Cauchy–Lipschitz :

THEORÈME 52. *Soit $f \in C^1(U \times I; \mathbb{R}^n)$, où U est un ouvert de $\mathbb{R}^n$ et I un intervalle ouvert de $\mathbb{R}$. Soient $x_* \in U$ et $t_0 \in I$. Alors à l'équation différentielle $\dot{x} = f(t, x)$ on peut associer un flot C^1, $\Phi_{t_0,t}$, défini pour x_0 proche de x_* et sur un intervalle de temps $|t - t_0| \leq \tau$.*

Ces deux théorèmes sont vrais et peuvent se démontrer, indépendamment l'un de l'autre, par des méthodes d'analyse réelle. Mais quand on les qualifie d'*équivalents*, on veut dire que l'on peut déduire le Théorème 52 du Théorème 51, par un raisonnement assez direct ; et réciproquement, on peut déduire simplement le Théorème 51 du Théorème 52.

PREUVE DE L'ÉQUIVALENCE. On va montrer d'abord que le redressement implique Cauchy–Lipschitz ; ensuite on montrera le contraire.

Pour démontrer le Théorème de Cauchy–Lipschitz à partir du Théorème de redressement, l'idée est que l'on va d'abord redresser le champ de vecteurs, après quoi ce sera un jeu d'enfant que de résoudre l'équation !

On part donc d'une équation $\dot{x} = f(t, x)$, que l'on ramène à une équation autonome par la technique habituelle : on pose $s = t - t_0$, et la nouvelle inconnue $y = (x, s)$ vérifie l'EDO $dy/dt = (f(s, x), 1)$.

Sur $U \times I$, le champ de vecteurs $\widetilde{f} = (f, 1)$ ne s'annule jamais, puisque sa seconde composante est non nulle. D'après le Théorème 51, on peut donc le redresser, c'est à dire construire un difféomorphisme ψ, défini au voisinage de (x_*, t_0), tel que $\psi_* \widetilde{f} = \psi_*(f, 1) = \xi$ est un vecteur constant. On applique alors le changement de variables ψ à l'EDO : on trouve que $d\psi(y)/dt = d\psi(dy/dt) = d\psi(f, 1) = \xi$; donc $\psi(y(t)) = \psi(y_0) + \xi(t - t_0)$; et il suffit d'inverser ψ pour avoir une solution $y(t) = \psi^{-1}(\psi(y_0) + \xi(t - t_0))$, unique et régulière. Le flot que

l'on a construit dans ce nouveau jeu de variables définit aussi un flot pour le système originel, et la conclusion du Théorème 52 en découle.

La réciproque (prouver le théorème de redressement à partir de Cauchy–Lipschitz) est plus subtile. Une première idée consiste à construire le difféomorphisme comme solution d'une EDO ; c'est une recette utile pour nombre de problèmes. Mais surtout, l'idée principale est que les trajectoires associées à f ne peuvent jamais se couper, et vont *fibrer* l'espace autour du point x_* : le découper en fibres qui sont les trajectoires. Redresser le champ de vecteurs revient alors à redresser ces fibres, c'est à dire trouver un jeu de coordonnées où les fibres ont toutes leurs coordonnées fixes, sauf une variable, qui joue le rôle de variable de temps.

Construisons pour commencer cette coordonnée qui repère la fibre. Soit H un petit ouvert d'un hyperplan passant par x_* et transverse aux fibres, disons orthogonal à $f(x_*)$. Ici l'hypothèse $f(x_*) \neq 0$ est capitale ! On choisit un repère orthonormal de $\mathbb{R}^n$ dont les $(n-1)$ premiers vecteurs forment une base orthonormale de H, et dont le dernier est dirigé selon $f(x_*)$. Tout élément de H peut être vu comme un élément de $\mathbb{R}^n$, via $x_0 \longmapsto (x_0, 0)$.

Pour tout $x_0 \in H$, en faisant agir le flot pendant un temps t on obtient un certain point $x = \Phi_t(x_0)$. Prouvons que l'application $H \times I \to \mathbb{R}^n$ qui à (x_0, t) associe x est un difféomorphisme au voisinage de (x_*, t_0). Pour cela on applique le théorème d'inversion locale, rappelé ci-dessous :

THEORÈME 53 (Théorème d'inversion locale). *Si une fonction $\psi \in C^1(U; \mathbb{R}^n)$ a sa différentielle inversible en un point, alors au voisinage de ce point ψ définit un difféomorphisme.*

Considérons donc $(x_0, t) \longmapsto \Phi_t(x_0)$. Que vaut la matrice jacobienne en (x_*, t) ? Il y a $n-1$ composantes correspondant à x_0, et une composante pour la variable t ; on va les considérer séparément. On commence par dériver par rapport à la variable x_0, en fixant $t = 0$: or $\Phi_0(x_0) = x_0 \in H$; ou plutôt, si l'on considère le membre de droite comme un élément de U, $\Phi_0(x_0) = (x_0, 0)$. La jacobienne comprend donc un bloc carré I_{n-1} (la matrice identité $(n-1) \times (n-1)$) et une colonne nulle adjacente à ce bloc. Puis, si l'on dérive par rapport à t, en fixant $x_0 = x_*$, on obtient $(d/dt)\Phi_t(x_*)$, qui n'est autre que $f(x_*)$, dont seule la dernière composante est non nulle ; la ligne manquante de la jacobienne a donc un seul coefficient non nul, sur la diagonale. Finalement la matrice jacobienne est diagonale, avec tous ses coefficients diagonaux non nuls ; par le Théorème d'inversion locale, l'application définit donc un difféomorphisme local.

Donc, pour tout x dans une boule $B(x_*, r)$, avec r assez petit, on peut trouver $x_0 \in H$ et $t \in \mathbb{R}$ tel que $x = \Phi_t(x_0)$. Dans ces nouvelles coordonnées, le flot a pour expression $(x_0, t_0) \longmapsto (x_0, t)$; et donc le flot se déplace en ligne droite, à vitesse constante. Ce flot est donc associé à un champ de vecteurs constant ! (L'important était bien de trouver le jeu de coordonnées pertinent.) □

Cette belle équivalence entre les Théorèmes 52 et 51 illustre bien le lien profond entre les EDO et la géométrie. Cependant, il ne faut pas surestimer l'importance de ce résultat, qui souffre de plusieurs limitations :

- Il est rare que l'on ait avantage à se placer dans le jeu de coordonnées fourni par le Théorème de redressement ;
- Le Théorème de redressement est un énoncé purement local, car il est très rare que l'on puisse redresser *globalement* un champ de vecteurs, même s'il ne s'annule jamais. Ainsi le flot géodésique à la surface d'un anneau est globalement défini, mais ne peut se redresser globalement. A contrario, pour de nombreuses équations on peut résoudre le problème de Cauchy globalement : il en est ainsi, par exemple, du flot géodésique sur l'anneau.
- Le Théorème de redressement ne s'adapte pas à la dimension infinie et aux équations aux dérivées partielles.

2.8. Complément : EDO non régulières

Que faire pour résoudre une équation quand la régularité du champ de vecteurs n'est pas suffisante pour appliquer le théorème de Cauchy–Lipschitz ? C'est un problème que l'on peut rencontrer en pratique dans de nombreuses situations, en physique par exemple. D'ailleurs, certains soupçonnent que le champ de vitesse d'un fluide "typique" n'est pas lipschitzien...

Deux approches principales ont été développées pour travailler avec des notions affaiblies de régularité. L'une comme l'autre sont considérablement plus élaborées que la théorie de Cauchy–Lipschitz, et dépassent le cadre de ce cours ; on les mentionnera pour mémoire.

La première approche est issue de la théorie des équations de transport, une certaine catégorie d'équations aux dérivées partielles. Rendue célèbre par Ron DiPerna et Pierre-Louis Lions à la fin des années 1980, cette théorie remplace la régularité lipschitzienne par des notions affaiblies de régularité, vraies en moyenne en un certain sens. Ainsi DiPerna et Lions ont prouvé, par exemple, que si le champ de vecteurs f est dérivable au sens de Sobolev (disons dérivable presque partout avec une dérivée appartenant à un espace de Lebesgue L^p convenable),

alors on peut définir un flot presque partout. Ces travaux ont repris de la vigueur dans les années 2000, avec une série de contributeurs, tout particulièrement Luigi Ambrosio. Un séminaire Bourbaki de Camillo De Lellis [**9**] faisait le point il y a quelques années sur ce sujet, qui continue à connaître des développements et applications de grand intérêt.

La seconde approche est issue de la théorie des probabilités, et plus particulièrement des processus aléatoires. Pour la motiver, reprenons le champ de vitesses $f(x) = \pm|x|^{2/3}$, dont la dérivée vaut $(2/3)|x|^{-1/3}$. Quand une solution s'approche de l'origine, elle peut soit traverser l'origine, soit rester stationnaire en l'origine un temps arbitraire, puis repartir à un moment arbitraire. Cela fait une infinité de solutions possibles ! Pour les départager, on peut dire que, une fois qu'on a atteint l'origine, on tire au hasard la suite du processus, comme si l'on lançait une pièce pour déterminer la date à laquelle on va passer de l'autre côté de l'origine.

Cette idée a suggéré de *résoudre aléatoirement des équations déterministes non régulières*, une voie de recherche qui a été explorée en particulier par Yves Le Jan et Olivier Raimond dans les années 2000. On peut modéliser cela avec une famille de *noyaux de transition* $K_t(x, dy)$, qui représentent la probabilité, partant de x, d'aller en y après un temps t. Pour que ce soit cohérent, il faut un substitut à la propriété de composition du flot : c'est la relation

$$\int_y K_t(x, dy)\, K_s(y, dz) = K_{t+s}(x, dz).$$

(Pour aller de x à z en un temps $t + s$, on commence par aller en un point intermédiaire y en un temps t.)

Ensuite, comment exprimer dans ce formalisme la condition qui définit l'EDO ? Ici la discussion deviendra plus délicate. Dans une optique probabiliste, il faut intégrer et raisonner sur un ensemble d'états possibles. Soit donc ζ une fonction lisse (une fonction coordonnée par exemple), on cherchera donc à donner un sens à $\int \zeta(y_t)\, \mu(dx)$, où x est la condition initiale et y_t est l'état au temps t. Pour cela il faut prendre en compte tous les états y_t possibles, et on aboutit donc à la formule

$$\int \zeta(\Phi_t(x))\, \mu(dx) \simeq \int_x \int_y \zeta(y)\, K_t(x, dy)\, \mu(dx). \tag{42}$$

(Le membre de gauche dans (42) n'a pas vraiment de sens, mais c'est moralement ainsi que l'on pense au membre de droite qui, lui, est bien défini.) La dérivée du membre de droite de (42) vaut, s'il n'y a pas de

problème à dériver sous l'intégrale,

$$\int_x \int_y \zeta(y)\, \frac{d}{dt} K_t(x, dy)\, \mu(dx). \tag{43}$$

Quant à la dérivée du membre de gauche de (42), si tout était bien défini, ce serait

$$\int_x \nabla\zeta(\Phi_t(x)) \cdot \left(\frac{d}{dt} \Phi_t(x) \right) \mu(dx) = \int_x \nabla\zeta(\Phi_t(x)) \cdot f(\Phi_t(x))\, \mu(dx);$$

pour donner un sens à cette dernière expression, on remplace à nouveau le flot par sa variante probabiliste, et l'on trouve

$$\int \nabla\zeta(\Phi_t(x)) \cdot f(\Phi_t(x))\, \mu(dx) \simeq \int_x \int_y \nabla\zeta(y) \cdot f(y) K_t(x, dy)\, \mu(dx). \tag{44}$$

Le membre de droite a bien un sens, et on peut réaliser, au moins sous de bonnes hypothèses de régularité, une intégration par parties en y :

$$\int_x \int_y \nabla\zeta(y) f(y) K_t(x, dy)\, \mu(dx) = - \int_x \int_y \zeta(y) \nabla_y \cdot \big(f(y) K_t(x, dy) \big)\, \mu(dx).$$

Cette dernière quantité représente donc la dérivée du membre de gauche de (42) ; en l'identifiant à (43) on obtient

$$\int_x \int_y \zeta(y)\, \frac{d}{dt} K_t(x, dy)\, \mu(dx) = - \int_x \int_y \zeta(y) \nabla_y \cdot \big(f(y) K_t(x, dy) \big)\, \mu(dx).$$

Ici μ et ζ sont arbitraires, et finalement cette identité implique une équation ponctuelle sur K_t :

$$\partial_t K_t(x, dy) + \nabla_y \cdot \big(f(y)\, K_t(x, dy) \big) = 0. \tag{45}$$

On reconnaît ici une forme de l'**équation de continuité**, fondamentale en mécanique des fluides, sous la forme

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0 \tag{46}$$

(ρ la densité de particules advectée par le champ de vitesses u).

On trouve, par exemple dans les travaux de Le Jan et Raimond, des résultats généraux garantissant l'existence de familles de noyaux de transition $K_t(x, dy)$ vérifiant toutes ces conditions ; mais c'est bien trop technique pour ce cours... Dans toute la suite, nous nous concentrerons sur la résolution déterministe des EDO.

On note au passage que l'équation de continuité (46) joue aussi un rôle central dans l'approche de DiPerna et Lions, et que nous la retrouverons dans le cours d'équations aux dérivées partielles.

CHAPITRE 3

Étude globale

3.1. Exemples et contre-exemples

Pour entamer ce chapitre consacré à l'étude globale des équations différentielles ordinaires, on va dresser une liste d'exemples et contre-exemples.

Le chapitre précédent se concentrait sur l'étude *locale* des équations, et le théorème de Cauchy–Lipschitz qui permet de définir un flot en temps petit. Mais la plupart des phénomènes intéressants se produisent en temps non petit, voire en temps grand... Un certain nombre d'obstacles peuvent cependant se mettre en travers du chemin quand on cherche à saisir le comportement en temps grand.

Un bon réflexe est de chercher, dès le début de l'étude, les équilibres de l'équation : cela fournit des solutions particulières importantes. Mais certains équilibres sont stables, d'autres instables, et bien sûr ils ne nous disent pas tout sur le comportement en temps grand.

Premier exemple de comportement : Il est possible que le flot ne soit simplement pas bien défini en temps grand : on parle alors d'explosion. On a déjà vu un exemple en une dimension d'espace :

$$\dot{x} = x^2. \tag{47}$$

Dans ce cas la solution tend vers $+\infty$ en temps fini, dès que $x(0) > 0$; et une fois que la solution est partie à l'infini, on perd toute possibilité de la prolonger... Dans le cas de l'équation (47) on a abouti à la conclusion par un calcul explicite, mais il faut pouvoir reconnaître ce cas de figure dans des situations où le calcul ne sera pas possible. Considérons par exemple

$$\dot{x} = x^3 + x^2 + \sin x.$$

Comment montrer que le flot n'est pas bien défini et que la solution explose en temps fini ?

Deuxième exemple de comportement : Il se peut que le flot soit bien défini pour tous les temps et que les solutions convergent vers une valeur d'équilibre bien déterminée. Considérons par exemple

l'équation toute simple de désintégration radioactive

$$\dot{x} = -\lambda x, \tag{48}$$

où $\lambda > 0$. La solution est alors unique, c'est $x(t) = x_0\, e^{-\lambda t}$, qui converge vers 0. Ici c'est facile, car on a une solution explicite... mais peut-on en dire autant, par exemple, de

$$\dot{x} = -\sin x - \mu x?$$

Quand x est proche de 0 on aimerait bien dire que c'est comme $\dot{x} = -(1+\mu)x$, et donc conclure à un comportement qualitativement similaire à celui de (48) ; mais comment le prouver ?

Troisième exemple de comportement : Le système peut se retrouver pris dans une sorte de boucle périodique, repassant toujours par les mêmes états. C'est bien sûr le cas pour

$$\ddot{x} = -x, \tag{49}$$

qui se résout en $x(t) = x_0 \cos t + x_1 \sin t$. De tels cycles apparaissent dans de nombreuses équations importantes, et la plupart du temps on ne peut pas déterminer leur équation exacte. L'un des exemples les plus célèbres est celui du système proie-prédateur : quand les prédateurs sont nombreux, ils prélèvent de nombreuses proies ; alors les effectifs de proies diminuent, et les prédateurs n'ont plus assez à manger, et ils meurent de faim ; les effectifs des prédateurs décroissent, et ils prélèvent donc moins de proies, ce qui permet aux effectifs des proies de remonter ; quand les proies sont à nouveau assez nombreuses, les prédateurs peuvent en profiter et leurs effectifs remontent... et ainsi de suite ! De manière remarquable, le système aboutira à des oscillations incessantes plutôt qu'à un équilibre où les populations de proies et de prédateurs seraient tout juste adaptées l'une à l'autre. Nous développerons cet exemple au Chapitre 5.

Quatrième exemple de comportement : Le système peut errer de manière apparemment imprévisible, sans qu'un comportement bien identifié se dégage. Un cas d'école simple est le système du double pendule (deux pendules rigides accrochés l'un à l'autre, que l'on laisse osciller). La solution est toujours bien définie, mais les solutions ne convergent vers rien, et on ne voit aucune régularité se dégager des observations...

Pour de tels systèmes erratiques, au delà de la difficulté de prédire le comportement d'une solution, on peut parfois dégager des régularités statistiques. Par exemple, est-ce que le pendule inférieur penchera aussi souvent à droite qu'à gauche, en moyenne ? On parle de chaos déterministe quand le flot vérifie deux conditions :

- imprédictibilité du futur des trajectoires individuelles (c'est à dire que l'on est incapable de prévoir, en fonction d'informations simples sur la condition initiale, comment se comportera la solution au bout d'un temps très grand)

- distribution d'états statistiquement prévisible en fonction du temps (c'est à dire que l'on sait prévoir de manière probabiliste la répartition des états au cours du temps, éventuellement pour un ensemble de solutions plutôt que pour une solution particulière).

Cinquième exemple de comportement : Le système peut converger vers un sous-ensemble bien précis, en rester très proche, et pourtant ne pas rester prédictible pour autant. Dans le système de Lorenz par exemple, on s'approche irrésistiblement de l'attracteur de Lorenz, mais ensuite on se déplace erratiquement au voisinage de cet attracteur ; il y a donc à la fois prédictibilité et imprédictiblité. Nous aurons l'occasion de reparler de ce système en fin de chapitre.

Voilà donc passés en revue cinq comportements remarquables : explosion ; convergence vers un équilibre ; convergence vers un cycle ; voyage erratique dans l'espace des configurations ; voyage erratique au voisinage d'un attracteur. Mais un système différentiel peut aussi adopter plusieurs de ces comportements, en fonction des régions de l'espace des phases. Par exemple, soit l'équation

$$\dot{x} = x^2 - x \tag{50}$$

dans $\mathbb{R}$. Supposons que x est petit : on s'attend alors à ce que x^2 soit négligeable par rapport à x, auquel cas l'équation (50) devrait se comporter comme $\dot{x} = -x$; si l'on part d'une valeur initiale petite, on doit donc converger vers 0 en temps grand, et le système doit rester petit pour tous les temps (ce qui est cohérent avec l'approximation par $\dot{x} = -x$). Mais supposons maintenant que x est grand (positif) : on s'attend alors à ce que x soit négligeable devant x^2, et que l'équation soit bien approchée par $\dot{x} = x^2$, qui explose en temps fini. Ces considérations nous laissent deviner que le flot est bien défini, pour tous les temps, quand la condition initiale $x(0)$ est petite ; et qu'il y a au contraire explosion en temps fini quand $x(0)$ est suffisamment grand.

En outre on remarque l'existence d'un point fixe $x = 1$, équilibre. On soupçonne alors que c'est une valeur séparatrice, et on peut même conjecturer que pour $x < 1$ il y a convergence vers 0, alors que pour $x > 1$ il y a explosion en temps fini. Avec de l'expérience on apprend ainsi à faire des raisonnements qualitatifs préliminaires en combinant des théorèmes et des analogies avec des équations représentatives.

3.2. Théorème de prolongement

Le Théorème de Cauchy–Lipschitz fournit un flot local ; pour l'étendre autant que possible, on utilisera le théorème de prolongement, qui nous permet de continuer le flot jusqu'à l'explosion.

DÉFINITION 54 (Flot maximal). Soient O un ouvert de $\mathbb{R}^n$, I un intervalle de $\mathbb{R}$, et f une application $I \times O \to \mathbb{R}$. On considère l'équation différentielle $\dot{x} = f(t,x)$ et on fixe $t_0 \in I$. On appelle flot maximal associé à l'équation une application $\Phi_{t_0,t}(x_0)$, définie pour tout (x_0,t) dans un sous-ensemble D de $O \times I$, appelé domaine du flot, et vérifiant les propriétés suivantes :

(i) pour tout $x_0 \in O$, $I(x_0) = \{t \in I;\ (x_0,t) \in D\}$ est un intervalle ouvert, sur lequel $x(t) = \Phi_{t_0,t}(x_0)$ est solution de $\dot{x}(t) = f(t,x(t))$;

(ii) il n'existe aucune application $\Phi_{t_0,t}(x_0)$ vérifiant les mêmes propriétés et définie sur un domaine strictement plus grand.

REMARQUE 55. Ici on choisit de considérer des flots définis aussi bien vers les temps passés que vers les temps futurs. C'est une question de convention : on pourrait aussi préférer se limiter aux temps futurs par exemple ; il faudrait alors modifier légèrement la définition.

THEORÈME 56 (Théorème de prolongement). *Soient O un ouvert borné de $\mathbb{R}^n$, I un intervalle borné de temps, $t_0 \in I$ et $f = f(t,x)$ un champ de vecteurs dans O, dépendant du temps, de classe C^1. Alors il existe un unique flot maximal $(\Phi_{t_0,t}(x_0))_{t \in I(x_0)}$ pour l'équation $\dot{x} = f(t,x)$. Il a en outre une régularité C^1, voire C^r si f est de classe C^r ; et sa différentielle est donnée par la formule* (24).

En outre, soit $x_0 \in O$ et soit t^* une extrémité de l'intervalle $I(x_0)$, de sorte que (x_0,t^*) est un point frontière du domaine D ; alors il existe un point d'accumulation y^* de $\Phi_{t_0,t}(x_0)$ quand $t \to t^*$, tel que $(y^*,t^*) \in \partial(O \times I)$.

REMARQUES 57. 1. Les extrémités de l'intervalle $I(x_0)$, ou "temps maximaux d'existence", correspondent donc à une sortie du domaine, soit dans la variable de temps (bord de I), soit dans la variable d'état (bord de O).

2. La "brique élémentaire locale" sur laquelle repose le Théorème 56 est le théorème de Cauchy–Lipschitz, qui est invariant par changement de variables. En conséquence, l'énoncé du Théorème 56 s'applique encore, sans aucune modification, quand O est un ouvert borné d'un espace topologique que l'on peut localement paramétrer par $\mathbb{R}^n$; c'est à dire quand O est un ouvert borné d'une variété C^1-différentiable de dimension n.

3. La fin de l'énoncé signifie moralement que (x^*, t^*) appartient à la frontière de $O \times I$, où x^* est la valeur de la solution au temps maximal t^*. Mais rien ne garantit que cette solution soit bien définie au temps maximal : il se pourrait qu'elle oscille de plus en plus vite en approchant de la frontière. Cela justifie d'introduire, dans la conclusion, un point d'accumulation y^*, c'est à dire un état y^* tel que l'on puisse trouver une suite $t_k \to t^*$ telle que $\Phi_{t_0,t_k}(x_0) \to y^*$. Si la limite de $\Phi_{t_0,t}(x_0)$ existe quand $t \to t^*$, alors y^* est forcément égal à cette limite, que l'on peut noter $\Phi_{t_0,t^*}(x_0)$.

4. Si l'on est seulement intéressé aux temps $t \geq t_0$, on peut remplacer les intervalles ouverts $I(x_0)$ par des intervalles semi-ouverts de la forme $[t_0, t^*[$; l'énoncé est inchangé et les preuves sont faciles à adapter.

Avant d'aborder la preuve de ce théorème, il faut bien en comprendre l'esprit : partant d'une condition initiale x_0, on cherche à construire une solution qui dure aussi longtemps que possible. Pour cela on résout l'équation différentielle $\dot{x} = f(t, x)$ en temps petit par Cauchy–Lipschitz ; cela nous mène à une position $x(t_1)$; on applique alors de nouveau Cauchy–Lipschitz en un temps petit, et on continue autant que possible... jusqu'à ce que l'on arrive à la fin de l'intervalle de temps, ou que "quelque chose de catastrophique" se produise, qui empêche la continuation. Ce que dit le Théorème 56 c'est que cette catastrophe est forcément l'arrivée à la frontière du domaine O.

Dans le Théorème 56 on a supposé O borné. Que se passe-t-il si l'on résout l'équation dans un ouvert non borné, comme l'espace $\mathbb{R}^n$ tout entier ? On peut toujours écrire $\mathbb{R}^n$ comme une union de boules de plus en plus grandes, et résoudre l'équation dans l'une de ces grandes boules... Par exemple une boule $B(0, R_1)$ pour les temps $t \in [0, 1]$, puis une boule $B(0, R_2)$ pour les temps $t \in [1, 2]$, avec $R_2 > R_1$, etc. la seule chose qui pourrait nous empêcher de faire cela, ce serait que le système ne soit confiné dans aucune boule $B(0, R)$ en temps fini ; autrement dit qu'il "parte à l'infini en temps fini".

En conclusion, si $f(t, x)$ est un champ de vecteurs défini sur $\mathbb{R}^n$ et dépendant du temps t, alors la seule chose qui puisse empêche le flot maximal d'être défini sur $\mathbb{R}^n \times \mathbb{R}_+$, c'est l'*explosion en temps fini*, c'est à dire

$$\limsup_{t \to t^*} |x(t)| = +\infty.$$

De même, si I est un intervalle borné, la seule chose qui puisse empêcher le flot d'être défini pour tous les temps $t \in I$, c'est la possibilité d'explosion en un temps $t^* \in I$.

Plus généralement, il suffira de retenir que l'*on peut toujours prolonger le flot local en un flot maximal, et ce flot ne s'interrompt pas tant que la solution reste à l'intérieur de l'ouvert où l'équation est définie.*

Le cas particulier où $I = \mathbb{R}$ et $O = \mathbb{R}^n$ est suffisamment important pour mériter un énoncé explicite :

COROLLAIRE 58 (Un flot localement borné sur $\mathbb{R}^n$ est globalement défini). *Soit $f = f(t,x)$ un champ de vecteurs de classe C^1, défini sur $\mathbb{R} \times \mathbb{R}^n$. Soit Φ le flot maximal associé à l'EDO $\dot{x} = f(t,x)$, partant de $t_0 = 0$. Si l'on sait que pour tout (x_0, t) dans le domaine de Φ,*

$$|t| \leq T,\ |x_0| \leq R \Longrightarrow |\Phi_{t_0,t}(x_0)| \leq C(T,R) < +\infty,$$

alors le flot est défini pour tous les temps et toutes les conditions initiales.

La démonstration du Théorème 56 est moins technique que celle du Théorème de Cauchy–Lipschitz ; elle consiste principalement à appliquer ce dernier théorème à bon escient. Le lemme suivant jouera un rôle clé.

LEMME 59 (Lemme de compatibilité). *Soit $f(t,x)$ un champ de vecteurs $I \times O \to \mathbb{R}^n$ de classe C^1. Soient $(x(t))_{t\in J}$ et $(\widetilde{x}(t))_{t\in \widetilde{J}}$ deux courbes intégrales de f, définies respectivement sur des intervalles ouverts J et $\widetilde{J}$, telles que $x(t_0) = \widetilde{x}(t_0)$. Alors $x(t) = \widetilde{x}(t)$ pour tout $t \in J \cap \widetilde{J}$.*

DÉMONSTRATION. Posons

$$K = \Big\{ t \in J \cap \widetilde{J};\quad x(t) = \widetilde{x}(t) \Big\}.$$

Bien sur K est non vide, puisqu'il contient t_0. Si K est égal à $J \cap \widetilde{J}$, le lemme est démontré. Sinon, c'est que K est inclus strictement dans $J \cap \widetilde{J}$; quitte à changer t en $-t$, on peut toujours supposer que $\sup K < \sup(J \cap \widetilde{J})$. Soit alors

$$t_* = \inf\big\{ t \geq t_0;\ x(t) \neq \widetilde{x}(t) \big\}.$$

Nous avons alors deux cas à envisager :

(a) $x(t_*) \neq \widetilde{x}(t_*)$;

(b) $x(t_*) = \widetilde{x}(t_*)$ mais t_* est limite d'une suite de temps $(t_k)_{k\in\mathbb{N}}$ où x et $\widetilde{x}$ diffèrent :

$$x(t_k) \neq \widetilde{x}(t_k), \qquad t_k \to t. \tag{51}$$

Dans le cas (a), forcément $t_* > t_0$. Par continuité, on a aussi $x(t) \neq \widetilde{x}(t)$ pour $t < t_*$ assez proche de t_* ; cela contredit la définition de l'infimum.

Dans le cas (b), appliquons le théorème de Cauchy–Lipschitz au nouveau "temps initial" t_*, avec la nouvelle "condition initiale" $x(t_*)$: on en déduit l'existence d'un intervalle $]t_* - \epsilon, t_* + \epsilon[$ sur lequel le flot est uniquement défini ; donc $x(t)$ et $\widetilde{x}(t)$ coïncident sur $]t_* - \epsilon, t_* + \epsilon[$, ce qui contredit (51). □

COROLLAIRE 60. *Avec les mêmes hypothèses que le Lemme 59, si $x(t)$ est une courbe intégrale définie sur J, $\widetilde{x}(t)$ une courbe intégrale définie sur un intervalle $\widetilde{J}$ contenant J, et t_* une extrémité de J, alors $x(t) \to \widetilde{x}(t_*)$ quand $t \to t_*$.*

On peut maintenant aborder la preuve du théorème de prolongement.

PREUVE DU THÉORÈME 56. On commence par se ramener au cas où I est un intervalle borné. Pour cela on intersecte I, si besoin, par un intervalle $]-T, T[$; Les conclusions générales s'en déduiront facilement.

Soit x_0 une condition initiale. On considère alors toutes les solutions $(x(t))$ définies sur tous les intervalles possibles $J \subset I$. On définit $J_{\max}$ comme l'union de tous ces intervalles ; c'est une union d'intervalles ouverts qui ont tous un point commun, donc c'est un intervalle ouvert. On peut rappeler explicitement sa dépendance en x_0 en le notant $J_{\max}(x_0)$.

Pour s dans $J_{\max}$, on définit $X(t)$ comme suit : si $s \in J_{\max}$, il existe un intervalle $J \subset J_{\max}$ et une solution $x(t)$ définie sur J ; on définit alors $X(t) = x(t)$. Par le Lemme 59 cela ne dépend pas du choix de l'intervalle J. La solution $X(t)$ ainsi définie résout l'équation sur l'intervalle $J_{\max}$ tout entier : c'est donc la solution maximale.

On peut effectuer cette opération pour toute condition initiale x_0, et obtenir ainsi une solution $X(t; x_0)$.

Soit x_0 donné, et $J_{\max}$ l'intervalle maximal correspondant ; supposons que $\sup J_{\max} = t^*$. Il y a trois possibilités :

- soit $t^* = \sup I$;
- soit $t^* \neq \sup I$ et $\liminf_{t \to t^*} d(X(t), \partial O) = 0$;
- soit $t^* \neq \sup I$ et $\liminf_{t \to t^*} d(X(t), \partial O) \geq \delta > 0$.

Pour démontrer le théorème, il suffit de réfuter la troisième possibilité. Nous allons le faire par l'absurde, et donc supposer que cette troisième possibilité est vraie : en particulier, $X(t)$ reste à une distance au moins δ du bord quand t est proche de t^*. La fonction $t \longmapsto X(t)$ prend donc ses valeurs dans un ensemble compact K de O ; sur cet ensemble, $f(t, x)$ est borné ; donc $dX/dt = f(t, X(t))$ reste borné. Donc X est une fonction lipschitzienne, et par conséquent uniformément continue. Donc $X(t)$ admet une limite ξ quand $t \to t^*$, et $\xi \in O$.

Le temps t^* est par hypothèse intérieur à I, et ξ est intérieur à O. On peut donc appliquer le Théorème de Cauchy–Lipschitz et résoudre l'équation $\dot{y} = f(t, y)$ pour t dans un voisinage de t^*, avec condition initiale $y(t_*) = \xi$. Cette solution est unique, et d'après le Lemme 59, coïncide avec $X(t)$ pour $t^* - \varepsilon < t < t^*$.

Soit maintenant $t \longmapsto Y(t)$ la fonction définie par

$$\begin{cases} Y(t) = X(t) & \text{pour } t \in J_{\max} \\ Y(t) = y(t) & \text{pour } t \in]t^* - \varepsilon, t^* + \varepsilon[. \end{cases}$$

Cette fonction est bien définie sur $J_{\max} \cup]t^* - \varepsilon, t^* + \varepsilon[$, et de classe C^1 sur ces deux intervalles ouverts séparément, donc sur leur réunion.

Finalement, $Y(t)$ est une solution définie sur un intervalle plus grand que $J_{\max}$; mais cet intervalle était par construction le plus grand ! Il y a donc contradiction.

La fin de la démonstration des propriétés du flot n'est pas conceptuellement profonde, et peut être omise ; de toute façon elle ne sera pas formalisée de manière très rigoureuse. À ce stade on a prouvé l'existence du flot maximal, et on a caractérisé les frontières de son domaine de définition. Si l'on se donne (x_0, t) dans ce domaine de définition, on sait que t est intérieur à I et que la solution $(x(s))$ issue de x_0 reste séparée de ∂O d'une distance $\delta > 0$. En chaque $s \in]t_0, t[$ on peut appliquer le théorème (local) de Cauchy–Lipschitz et trouver une boule B_s centrée en $x(s)$ et un $\theta > 0$ tel que le flot issu du temps s dans la boule B_s est bien défini et continu sur un intervalle de temps ouvert I_s. Par un argument de compacité, on peut recouvrir la trajectoire par un nombre fini de telles boules B_s et de tels intervalles I_s, avec s variant dans un ensemble discret ; ensuite en ne gardant que les trajectoires qui appartiennent à la boule B_s sur chaque intervalle de temps I_s, on voit que l'on a défini un flot continu sur un petit voisinage de x_0, sur l'intervalle de temps $[t_0, t[$ complet.

En particulier, on peut écrire $\Phi_{t_0,t} = \Phi_{t_m,t} \circ \Phi_{t_{m-1},t_m} \circ \ldots \circ \Phi_{t_0,t_1}$, où les différences successives des temps $t_0, \ldots, t_m, t$ sont suffisamment petites pour que le théorème de Cauchy–Lipschitz s'applique à chacun des flots $\Phi_{t_j,t_{j+1}}$ ainsi qu'à $\Phi_{t_m,t}$. Le flot $\Phi_{t_0,t}$ est ainsi C^1 comme composition de fonctions C^1 ; et même C^r comme composition de fonctions C^r si f est de classe C^r. En outre la différentielle de $\Phi_{t_0,t}$ est obtenue en composant les différentielles des flots Φ_{t_0,t_1}, etc. La différentielle du flot Φ_{t_0,t_1} coïncide, par la Section 2.3, avec le flot $\Phi^L_{t_0,t_1}$ de l'équation linéarisée (24) entre le temps initial t_0 et le temps final t_1 ; en composant, on trouve que la différentielle du flot $\Phi_{t_0,t}$ coïncide avec le flot

$\Phi^L_{t_0,t}$: l'équation (24) est toujours valable, mais on la considère sur tout l'intervalle de temps $[t_0, t]$. La preuve du théorème est ainsi achevée. □

REMARQUE 61. La fin de la preuve montre que $I(x_0)$ est une fonction continue de x_0 : si $x_0^k \to x_0$, alors $\lim I(x_0^k) = I(x_0)$. (La convergence des intervalles revient à la convergence des bornes supérieures et inférieures.) À titre d'exemple, on pourra calculer la fonction $I(x_0)$ en fonction de x_0 pour l'équation $\dot{x} = x^3$, qui explose en temps fini pour toutes les conditions initiales non nulles.

3.3. Critère de compacité

Cette section est une longue paraphrase des résultats de la section précédente, agrémentée de quelques exemples typiques et de recettes importantes.

Nous savons maintenant que l'on peut toujours définir le flot maximal associé à un champ de vecteurs de classe C^1 sur un ouvert borné de $\mathbb{R}^n$, ou plus généralement d'une variété de dimension n ; et nous savons aussi que ce flot est bien défini tant que l'on reste à l'intérieur de l'ouvert.

En corollaire, si l'on sait, pour une raison ou une autre, que les solutions restent toujours à l'intérieur de l'ouvert, sans jamais s'approcher du bord de cet ouvert, alors on sait aussi que le flot est défini globalement ! Ici "globalement" veut dire "pour tous les temps", et s'oppose à "localement".

La plus simple de ces situations survient quand $\partial O = \emptyset$. Bien sûr, un ouvert borné de $\mathbb{R}^n$ a forcément une frontière ; mais si l'on travaille sur une variété différentielle M compacte, alors on peut définir O comme étant la variété M tout entière, et le flot sera automatiquement défini globalement.

EXEMPLE 62. Soit $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ le tore à d dimensions : un élément de $\mathbb{T}^d$ est un vecteur $(x_1, \dots, x_d)$ où chaque x_i est défini modulo 1. On peut aussi l'écrire sous la forme $(\theta_1/(2\pi), \dots, \theta_d/(2\pi))$, où chaque θ_i est un angle, donc défini modulo 2π. Si l'on se donne $x^* \in \mathbb{T}^d$, alors une petite boule de $\mathbb{T}^d$ autour de x^* est difféomorphe à une petite boule de $\mathbb{R}^d$: le tore à d dimensions est donc une variété de dimension d. Un champ de vecteurs f défini sur $\mathbb{T}^d$ correspond à un champ de vecteurs défini sur $\mathbb{R}^d$ et $\mathbb{Z}^d$-périodique, c'est à dire tel que

$$f(x+z) = f(x) \qquad \forall x \in \mathbb{R}^d, \quad \forall z \in \mathbb{Z}^d.$$

Un tel champ de vecteurs, défini sur $\mathbb{R}^d$, engendre un flot $\mathbb{Z}^d$-périodique, c'est à dire tel que $\Phi_t(x+z) = \Phi_t(x) + z$. En passant au quotient modulo $\mathbb{Z}^d$, on obtient un flot sur $\mathbb{T}^d$. On parle donc de champ de

vecteurs sur $\mathbb{T}^d$ et de flot sur $\mathbb{T}^d$. Le Théorème 56, appliqué dans l'ouvert $O = \mathbb{T}^d$, montre alors que ce flot est *global*, c'est à dire défini pour tous les temps. (On peut aussi, alternativement, démontrer ce théorème en reprenant la preuve du théorème de prolongement sur $\mathbb{R}^d$ avec des fonctions périodiques.)

Réécrivons explicitement l'exemple précédent sous une forme équivalente avec des angles : soit donnée une équation différentielle dans l'espace des angles,

$$\dot{\theta}_i = f_i(\theta_1, \dots, \theta_d),$$

où chaque θ_i est défini modulo 2π ; alors le flot associé est globalement défini pour tous les temps. Les fonctions f_i peuvent être absolument quelconques, pourvu qu'elles soient de classe C^1 ! Il en sera de même pour n'importe quelle équation différentielle définie sur un espace compact. (Attention : dans tous nos théorèmes on a défini une ODE sur un domaine ouvert ; quand on dit "espace compact", cela signifie donc un ouvert compact, comme $\mathbb{T}^d$.)

Si l'équation est définie sur un ouvert O non compact, le raisonnement précédent ne s'applique plus ; il faut alors trouver une raison pour laquelle les solutions (hypothétiques) resteraient confinées. Donnons-nous un ensemble C_0 de conditions initiales. Supposons que l'on trouve un compact $K \subset O$ tel que *les solutions issues de C_0 restent confinées dans K, tant qu'elles existent*. Le Théorème de Prolongement prouve justement que les solutions sont bien définies tant qu'elles restent dans K ; on en déduit que ces solutions sont bien définies pour tous les temps. Autrement dit, le flot est bien défini, globalement en temps, sur C_0.

On en déduit la très importante **méthode des estimations a priori** : on raisonne *comme si* les solutions existaient pour tous les temps, et l'on cherche une borne qui confine ces solutions. Si l'on trouve cette borne, cela prouve tout à la fois que les solutions existent, qu'elles sont confinées, et que le flot est bien défini.

Exemple 63. Soit l'EDO $\dot{x} = f(t, x)$ avec f bornée. Soit une solution $x(t)$: tant qu'elle est bien définie, on a

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

donc

$$|x(t)| \leq |x_0| + |t - t_0|\, \|f\|_\infty,$$

où $\|f\|_\infty = \sup |f|$. Pour $|t - t_0| \leq C$, on a donc $|x(t)| \leq |x_0| + C\|f\|_\infty$, et le flot reste contrôlé : la solution ne peut exploser en temps fini. En conséquence, le flot est défini pour tous les temps.

EXEMPLE 64. Soit, sur $\mathbb{R}^n$, l'EDO $\dot{x} = f(t,x)$ avec f bornée sur la tranche $(x = 0)$ et lipschitzienne en x. Alors on a $|f(t,x)| \leq |f(t,0)| + L\,|x|$, où L est la constante de Lipschitz de f dans la variable x. On en déduit

$$|x(t)| \leq |x_0| + \int_{t_0}^{t} |f(x,s)|\,ds \leq |x_0| + \int_{t_0}^{t} (A + L|x(s)|)\,ds.$$

Tant que $|t - t_0| \leq T$, on en déduit (en supposant le flot bien défini)

$$|x(t)| \leq |x_0| + AT + L\int_{t_0}^{t} |x(s)|\,ds.$$

Par le Lemme de Gronwall cela donne

$$|x(t)| \leq (|x_0| + AT)e^{LT} \leq (|x_0| + AT)e^{LT}.$$

Cette borne montre que le flot est effectivement bien défini sur l'intervalle $|t - t_0| \leq T$. Comme T est arbitraire, le flot est en fait bien défini pour tous les temps. Il en serait de même si $f(t,0)$ était supposée lipschitzienne et non bornée ; et même si $f(t,0)$ était simplement suppos'ee localement bornée en t.

EXEMPLE 65. Considérons deux pendules rigides pesants accrochés l'un à l'autre, et évoluant selon les équations de Newton. L'espace des phases est de dimension 4 : aux positions des pendules, repérées par deux angles, il faut ajouter les deux vitesses angulaires. Le système agit sous l'effet de la gravité : il y a donc conservation de l'énergie totale E, qui est la somme des deux énergies cinétiques des pendules et des deux énergies potentielles. Chacune de ces énergies potentielles est bornée en valeur absolue, disons par un certain nombre positif P, puisque la différence de hauteur est bornée. (L'énergie potentielle est définie à une constante additive près ; on peut supposer que la position de repos verticale correspond à une énergie nulle.) Notons m_1 et m_2 les masses des pendules, et θ_1 et θ_2 leurs positions : on a donc

$$m_1\,\frac{\dot{\theta}_1^2}{2} + m_2\,\frac{\dot{\theta}_2^2}{2} \leq E + 2P.$$

Cela implique bien sûr des bornes sur $\dot{\theta}_1$ et $\dot{\theta}_2$: par exemple

$$|\dot{\theta}_1| \leq \sqrt{\frac{2(E+2P)}{m_1}} =: K_1, \quad |\dot{\theta}_2| \leq \sqrt{\frac{2(E+2P)}{m_2}} =: K_2.$$

Comme θ_1 et θ_2 prennent leurs valeurs dans l'ensemble compact $\mathbb{R}/(2\pi\mathbb{Z})$ (l'espace des angles), on en déduit finalement que la solution, écrite

dans l'espace des phases, $(\theta_1, \theta_2, \dot\theta_1, \dot\theta_2)$, est a priori confinée dans le compact

$$K = \mathbb{T} \times \mathbb{T} \times K_1 \times K_2.$$

On en déduit, *sans aucun calcul*, que le flot est bien défini pour tous les temps. Cela est d'autant plus remarquable que ce double pendule est un modèle classique de comportement "erratique", imprédictible, utilisé en théorie du chaos.

S'il existe une fonction de Lyapunov $\psi = \psi(x)$ continue telle que $(d/dt)\psi(x(t)) \leq 0$ pour $x(t)$ solution de l'équation, et $\psi \to +\infty$ à l'infini, alors $x(t)$ est confinée dans le sous-ensemble compact $\{\psi(x) \leq \psi(x_0)\}$, et il s'ensuit que le flot est bien défini pour tous les temps. C'est la très importante **méthode de la fonction de Lyapunov**. Elle s'applique de même si l'on généralise la notion de fonction de Lyapunov en autorisant une dépendance par rapport au temps (de sorte que ce serait une fonction $\psi(x,t)$, tendant vers $+\infty$ quand la variable x tend vers l'infini, qui vérifierait $(d/dt)\psi(x(t),t) \leq 0$).

Remarque 66. La fonction ψ ci-dessus n'est pas forcément continue : ce qui compte c'est que ses ensembles de sous-niveau soient compacts. En fait, en dimension infinie (dans l'étude des EDP par exemple), il est classique de travailler avec des fonctions de Lyapunov semi-continues inférieurement.

Exemple 67. On a déjà vu que $\dot x = x^2$ mène à une explosion en temps fini quand $x(0) > 0$. On pourrait refaire le raisonnement pour une équation comme $\dot x = x^2 + x^4$: on aboutirait à la même conclusion au terme d'un calcul plus fastidieux. En revanche on ne pourra faire aucun calcul si l'on considère, disons, l'EDO $\dot x = a(t)x^2 + b(t)x^4$, où $a(t)$ et $b(t)$ sont des fonctions quelconques, de classe C^1, toutes deux comprises entre 1 et 2. Montrons cependant qu'il y a explosion pour $x(0) > 0$. Tant que la solution est bien définie, on a $\dot x \geq x^2$; donc cette solution reste *plus grande* que la solution de $\dot y = y^2$. (C'est intuitif mais prenons un moment pour le montrer : soit $\varepsilon > 0$ arbitrairement petit, soit $y_\varepsilon = x(0) - \varepsilon$, et soit $y(t)$ la solution de $\dot y = y^2$ partant de $y(0) = y_\varepsilon$: au temps initial on a $x(0) > y(0)$, et tant que $x(t) > y(t)$ on a $\dot x \geq x^2 > y^2 = \dot y$, donc $x - y$ ne peut qu'augmenter au cours du temps ; donc $x - y$ ne peut jamais reprendre la valeur 0, restant donc strictement positif tant que ces deux solutions sont bien définies. On obtient la conclusion en faisant tendre ε vers 0.) On conclut qu'il y a effectivement explosion en temps fini. Cela peut aussi être vu comme une conséquence du principe de comparaison de la Remarque 47 (voir aussi l'exercice 7, semaine 3).

Exemple 68. Considérons maintenant l'équation

$$\dot{x} = a(t)x^2 - b(t)x^4,$$

avec les mêmes conditions sur a et b. Pour x grand, on s'attend à ce que le second terme, négatif, domine, et que donc il n'y ait pas d'explosion. Mais aucun calcul n'est possible... à la place, on va confiner la solution a priori. Pour cela on va supposer par exemple que $x_0 > 0$; on note qu'alors $x(t)$ reste strictement positif pour tous les temps. En effet, sinon la trajectoire $x(t)$ croiserait en un certain temps t la trajectoire constante égale à 0 ; mais cela est impossible par théorème de Cauchy–Lipschitz. Il reste à vérifier que $x(t)$ ne peut pas devenir trop grand. Soit donc une solution régulière $(x(t))$: elle vérifie

$$\dot{x} \leq 2x(t)^2 - x(t)^4.$$

Le membre de droite est négatif ou nul dès que $x^2 \geq 2$, soit $x \geq \sqrt{2}$. Donc, si à un moment donné $x(t) \geq \sqrt{2}$, alors $x(t)$ ne peut que diminuer ; on en déduit que $x(t)$ ne peut jamais passer au-dessus de cette valeur. (En effet, sinon, soit t tel que $x(t) > \sqrt{2}$, et soit t_* le plus grand $t < t_*$ tel que $x(t) = \sqrt{2}$, qui existe par continuité. Alors $\dot{x}(s) \leq 0$ pour tout $s \in [t_*, t]$, donc $x(s)$ est décroissante sur cet intervalle, et $x(t_*) \geq x(t)$, ce qui est faux.) Tout cela aboutit à la borne a priori

$$0 < x(t) \leq \max(x_0, \sqrt{2}),$$

qui prouve que $x(t)$ est défini pour tous les temps.

Exemple 69. Nous allons retrouver la conclusion de l'exemple précédent avec une fonction de Lyapunov, que l'on cherchera sous la forme

$$F(x,t) = \frac{x^2}{2} - C\,t.$$

On calcule :

$$\begin{aligned}\frac{dF}{dt} &= x(t)\,\dot{x}(t) - C \\ &= x(t)\,\big(a(t)\,x(t)\big) - x(t)\,b(t)\,x(t)^3 - C \\ &\leq 2\,x(t)^2 - x(t)^4 - C.\end{aligned}$$

Si l'on choisit C assez grand, la fonction $x \longmapsto x^4 - 2x^2 + C$ est positive, et l'on aura donc

$$\frac{d}{dt}F(x(t),t) \leq 0.$$

Comme la fonction $F(x,t)$ tend vers $+\infty$ quand $|x| \to \infty$, on peut appliquer le critère de compacité, et en déduire que le flot est défini pour tous les temps.

Les exemples de cette section l'ont illustré : au-delà des théorèmes généraux, l'analyse globale des EDO repose souvent sur des arguments de comparaison et de continuité, et des estimations élémentaires. Nous allons maintenant aborder des outils plus sophistiqués.

3.4. Stabilité locale en spectre négatif

Cette section est consacrée à un théorème de convergence du flot près d'un équilibre stable ; on va présenter un critère de stabilité important et général, via l'opérateur linéarisé.

La **linéarisation** consiste à remplacer une équation non linéaire $\dot{x} = f(x)$, près d'un équilibre x^*, par une équation linéaire ; cette dernière décrit les petites variations de la solution quand la condition initiale est une petite variation de l'équilibre. On obtient cette équation en considérant la dérivée du flot, comme cela est fait dans la démonstration du Théorème de Cauchy–Lipschitz. On se limite à une équation autonome $\dot{x} = f(x)$, l'équation linéarisée près de l'équilibre x^* sera donc

$$\dot{h}(t) = df(x^*)\, h(t). \tag{52}$$

Cette équation linéaire décrira le comportement de l'équation non linéaire, avec une bonne précision, sur un intervalle de temps fini, à condition que x_0 soit proche de x^* ; plus précisément, $x^* + h(t)$ sera proche de la solution non linéaire $x(t)$ sur un intervalle $[0, T]$ fixé à l'avance. Plus généralement, on s'attend à ce que l'équation (52) soit une bonne approximation de la solution tant que celle-ci reste très proche de x^*.

Posons-nous la question : *Partant d'une condition initiale très proche de l'équilibre x^*, est-il vrai que la solution $x(t)$ converge vers x^* quand $t \to +\infty$?* Si cela est vrai, on parlera de stabilité asymptotique, ou tout simplement de stabilité. Formalisons cela un peu plus précisément.

Définition 70. Soit $f = f(t, x)$ un champ de vecteurs défini pour $t \in [t_0, +\infty[$ et x dans un ouvert O de $\mathbb{R}^n$. Soit $x^* \in O$ un équilibre de l'équation différentielle associée à f. On dit que x^* est un équilibre asymptotiquement stable, ou simplement un équilibre stable, s'il existe un voisinage V de x^* tel que pour toute condition initiale x_0 appartenant à V, la solution $x(t)$ de l'EDO, partant de x_0 en temps t_0, est bien définie pour tout temps $t \geq t_0$, et converge vers x^* quand $t \to +\infty$.

Cela veut dire que x^*, en tant qu'équilibre de l'EDO, a un *bassin d'attraction* contenant tout un voisinage de x^*.

Commençons par examiner des équations linéaires près d'un équilibre. Considérons, en dimension 1, l'équation $\dot{x} = ax$, avec $a \neq 0$. (Si $a = 0$,

aucune évolution n'a lieu.) L'équilibre est $x = 0$, la solution partant de x_0 en $t = 0$ s'écrit

$$x(t) = x_0\, e^{at}.$$

Si $a > 0$ la solution diverge quand $t \to +\infty$, alors que si $a < 0$ il y a convergence vers l'équilibre $x = 0$. La condition de stabilité est donc $a < 0$.

Soit maintenant une équation linéaire vectorielle, définie par une matrice diagonale :

$$\dot{x}_i = a_i\, x_i.$$

Alors la convergence n'aura lieu que si l'on a $a_i < 0$ pour tout i, c'est à dire si tous les coefficients diagonaux sont strictement négatifs.

Considérons ensuite le cas d'une matrice diagonalisable M : l'équation sera $\dot{x} = Mx$. Par un changement de base, on peut se ramener au cas d'une matrice diagonale, dont les coefficients seront les valeurs propres de la matrice M, et on pourra appliquer le résultat précédent *si et seulement si toutes ces valeurs propres sont strictement négatives.*

Le but est de généraliser ce résultat. Les matrices que l'on rencontre en pratique ne sont pas forcément diagonalisables ; en revanche on sait que l'on peut toujours les *trigonaliser* sur le corps des nombres *complexes.* Autrement dit, si M est une matrice $n \times n$ à coefficients réels, on peut trouver une matrice inversible P et une matrice triangulaire T, toutes deux de taille $n \times n$ et à coefficients complexes, telles que $M = PTP^{-1}$. Les coefficients diagonaux de la matrice T sont alors les valeurs propres complexes de M, c'est à dire ses valeurs propres sur le corps $\mathbb{C}$.

Proposition 71 (Critère de stabilité linéaire). *Soit $\dot{x} = Ax$ une équation différentielle linéaire à coefficients constants sur $\mathbb{R}^n$. Alors 0 est un équilibre asymptotiquement stable si et seulement si toutes les valeurs propres complexes de A ont leur partie réelle strictement négative.*

Theorème 72 (Théorème de stabilité non linéaire). *Soit $\dot{x} = f(x)$ une équation différentielle définie dans un ouvert U de $\mathbb{R}^n$, et soit x^* un équilibre. Alors x^* est asymptotiquement stable si toutes les valeurs propres de $df(x^*)$ ont leur partie réelle strictement négative.*

Remarque 73. Comme x^* est un équilibre, on a $f(x^*) = 0$; donc $f(x) = df(x^*)(x - x^*) + Q(x - x^*)$, où Q est en $o(|x - x^*|)$, et même en $O(|x - x^*|^2)$ si f est de classe C^2. En particulier $df(x^*)$ est l'approximation d'ordre 1 de f au voisinage de x^*.

En résumé : *pour prouver la stabilité d'un équilibre, on linéarise l'équation au voisinage de cet équilibre, et on cherche à montrer que*

toutes les valeurs propres de l'opérateur linéarisé sont de partie réelle strictement négative.

REMARQUES 74. 1. Si l'on compare la Proposition 71 et le Théorème 72, on comprend que la question est de savoir quand la stabilité linéaire implique la stabilité non linéaire. La réponse à cette question, que l'on rencontre souvent dans les équations d'évolution, n'est pas automatique.

2. La Proposition 71 est une équivalence, alors que le Théorème 72 est seulement une implication. Même si le linéarisé a une valeur propre strictement positive, l'instabilité n'en découle pas forcément. En pratique, les systèmes qui ont des valeurs propres de partie réelle strictement positive ne sont que rarement stables. On verra des énoncés plus précis dans la suite du cours.

On va esquisser la démonstration de la Proposition 71.

PREUVE DE LA PROPOSITION 71. Le plus simple est de résoudre l'équation : pour cela il faut être familier avec la décomposition de Jordan de la matrice A. On peut en effet décomposer l'espace $\mathbb{R}^n$ en somme directe de sous-espaces vectoriels E_i, chacun de dimension n_i, stable par l'action de la matrice A ; et dans ce sous-espace la matrice A prend la forme $A_i = \lambda_i\, I + N_i$, où I est la matrice identité (ici de taille n_i) et N_i une matrice triangulaire nilpotente, avec des 0 sur la diagonale.

Mais les matrices $\lambda_i I$ et N_i commutent, donc

$$e^{tA_i} = e^{\lambda_i t} e^{tN_i} = e^{\lambda_i t}\Big(I + tN_i + \frac{t^2\, N_i^2}{2} + \dots\Big).$$

Comme N_i est nilpotente, la somme entre parenthèses est faite d'un nombre fini de termes ; l'expression e^{tA_i} tout entière est en fait égale à $e^{\lambda_i t}$ multiplié par un polynôme non nul en tN_i. Attention, λ_i est maintenant un nombre complexe ! Comme $|e^{\lambda_i} t| = e^{(\mathcal{R}e\,\lambda_i)t}$, on a stabilité sur le sous-espace E_i si et seulement si $\mathcal{R}e\,\lambda_i < 0$.

Ce raisonnement vaut pour tout i, et donc on a stabilité si et seulement si tous les $\mathcal{R}e\,\lambda_i$ sont strictement négatifs. □

La section suivante sera consacrée à la preuve du Théorème 72.

3.5. Preuve du Théorème de stabilité locale non linéaire

On entreprend maintenant la preuve du Théorème 72. On rappelle que l'on considère l'équation différentielle

$$\dot{x} = f(x), \tag{53}$$

et que x^* est un équilibre, c'est à dire $f(x^*) = 0$.

On notera A la matrice associée à l'application linéaire $df(x^*)$, et R la partie non linéaire de f au voisinage de x^*, de sorte que

(54)
$$f(x) = A(x - x_*) + R(x - x_*), \qquad |R(h)| \leq \eta(h)\,|h|, \qquad \eta(h) \xrightarrow[h\to 0]{} 0.$$

On note que si f est de classe C^2, alors on peut choisir $\eta(h) = O(|h|)$, soit $|R(h)| = O(|h|^2)$. On va diviser la preuve en trois étapes.

Première étape : *Trigonalisation de A.* On considère A comme la matrice d'un opérateur $\mathbb{C}$-linéaire, via $A(x + iy) = Ax + iAy$; et l'on peut trouver un changement de base qui transforme la matrice A en une matrice Δ triangulaire (disons triangulaire supérieure), dont la diagonale est constituée des valeurs propres complexes $\lambda_1, \ldots, \lambda_n$ de A. La matrice P de changement de base est complexe ; on ne sait rien des coefficients hors diagonale de Δ. On note $(e_1, \ldots, e_n)$ cette base de trigonalisation : l'idée est de travailler dans cette base.

Bien sûr, une équation différentielle sur $\mathbb{R}^n$ est définie de manière intrinsèque, indépendamment du choix de base ; on peut donc choisir la plus commode pour effectuer les calculs. Cependant, ici le changement de base n'est permis que si l'on se place sur $\mathbb{C}^n$ plutôt que $\mathbb{R}^n$: pour appliquer le raisonnement il faut commencer par étendre l'équation différentielle à $\mathbb{C}^n$. Cela peut se faire de manière très naïve, en posant $f(x+iy) = f(x) + if(y)$; en fait, si $z = x + iy$, l'équation $\dot{z} = f(z)$ est alors équivalente à l'équation de départ $\dot{x} = f(x)$ (on le voit en séparant partie réelle et partie imaginaire de l'équation). En outre l'hypothèse faite sur le reste R est inchangée.

Pour résumer : quitte à dédoubler les variables pour se placer sur $\mathbb{C}^n$, et à choisir une base convenable, on peut supposer que A est triangulaire supérieure. On a alors

$$A\big(\sum u_i e_i\big) = \sum_{ij} a_{ij} u_j\, e_i, \tag{55}$$

où $a_{ij} \in \mathbb{C}$, $a_{ij} = 0$ pour $i > j$.

Deuxième étape : *"Diagonalisation approchée" de A.* Nous allons maintenant changer une nouvelle fois la base pour simplifier encore la matrice A. Bien sûr on ne peut pas, a priori, diagonaliser A : il y a une obstruction algébrique en général. En revanche on peut chercher à s'en approcher, via l'analyse. Pour cela on va modifier les vecteurs de la base en utilisant des "changements d'échelle". Soit en effet ε un paramètre à fixer ultérieurement, $0 < \varepsilon < 1$. Changeons la base $(e_1, e_2, \ldots, e_n)$ en $(e'_1, e'_2, \ldots, e'_n) = (e_1, \varepsilon e_2, \ldots, \varepsilon^{n-1} e_n)$. (La matrice de changement de

base est donc diagonale, avec des coefficients qui sont des puissances successives de ε.)

Calculons les coefficients de A après changement de base : de (55) on déduit

$$\begin{aligned} A\big(\sum u_i\,(\varepsilon^{i-1}e_i)\big) &= \sum_{ij} a_{ij}\,u_j\,\varepsilon^{j-1}\,e_i \qquad (i\le j)\\ &= \sum_{ij}(\varepsilon^{j-i}\,a_{ij})\,u_j\,(\varepsilon^{i-1}e_i). \end{aligned}$$

Donc la nouvelle matrice est une matrice triangulaire supérieure Δ_ε, dont le coefficient (i,j) est $a^\varepsilon_{ij} = \varepsilon^{j-i}a_{ij}$:

- sur la diagonale, les coefficients sont inchangés et égaux à $a_{ii} = \lambda_i$
- au-dessus de la diagonale, les coefficients sont multipliés par ε, ε^2, ... ε^{n-1} en fonction de leur position par rapport à la diagonale. Tous les coefficients hors diagonale sont donc en $O(\varepsilon)$, et tendent vers 0 quand $\varepsilon \to 0$. En fait, Δ_ε converge vers la matrice *diagonale* dont les coefficients sont les valeurs propres complexes de A. L'idée sera maintenant de fixer $\varepsilon > 0$ suffisamment petit et de faire les calculs dans la base correspondante.

Étape 3 : *Usage de la norme comme fonction de Lyapunov*

On introduit maintenant la norme ℓ^2 (norme euclidienne) associée au produit scalaire dans la base $(e'_1, e'_2, \dots, e'_n)$. Autrement dit, si $x = \sum v_i\,e'_i$, on pose

$$Q(x) = \frac{1}{2}\left(\sum_i |v_i|^2\right),$$

où $|v_i|$ est le module du nombre complexe v_i. Montrons maintenant que Q, carré de la norme, agit comme une fonction de Lyapunov pour l'équation non linéaire : non seulement Q va décroître avec le temps, mais en outre Q convergera exponentiellement vite vers 0.

En notant z^* le complexe conjugué de z, $(v_i(t))$ les composantes de $x(t)-x^*$, et $\langle x, y\rangle = \sum x_i\,y_i^*$, on trouve

$$\begin{aligned} \frac{d}{dt}Q\big(x(t)-x^*\big) &= \frac{1}{2}\frac{d}{dt}\sum_i v_i\,v_i^*\\ &= \frac{1}{2}\left(\sum_i v_i\,\dot v_i^* + \sum_i \dot v_i\,v_i^*\right)\\ &= \mathcal{R}e\sum v_i\,\dot v_i^* = \mathcal{R}e\,\big\langle f(x(t)), x(t)-x^*\big\rangle\\ &= \mathcal{R}e\,\big\langle A(x(t)-x^*), x(t)-x^*\big\rangle + \mathcal{R}e\,\big\langle R(x(t)-x^*), x(t)-x^*\big\rangle. \end{aligned}$$

On considère séparément les deux contributions du membre de droite ; et dans le premier on distinguera la contribution de la diagonale et celle des coefficients hors diagonale. On écrit donc

$$\begin{aligned}\mathcal{R}e\,\big\langle A(x(t)-x^*),x(t)-x^*\big\rangle &= \sum\big((\mathcal{R}e\,\lambda_i)|v_i|^2\big)+\mathcal{R}e\sum_{i<j}a_{ij}^{\varepsilon}v_i\,v_j^*\\ &\le (\sup\mathcal{R}e\,\lambda_i)(\sum|v_i|^2)+(\max_{i<j}|a_{ij}^{\varepsilon}|)(\sum_i|v_i|)^2\\ &\le (\sup\mathcal{R}e\,\lambda_i)(\sum|v_i|^2)+(\max_{i<j}|a_{ij}^{\varepsilon}|)n(\sum_i|v_i|^2),\end{aligned}$$

où l'on a utilisé l'inégalité de Cauchy–Schwarz sous la forme $(\sum_{i\le n}|\alpha_i|)\le\sqrt{n}\sqrt{\sum|\alpha_i|^2}$. En se rappelant la définition de Q, on conclut que

(56)
$$\mathcal{R}e\,\big\langle A(x(t)-x^*),x(t)-x^*\big\rangle \le 2\Big[(\sup\mathcal{R}e\,\lambda_i)+n\max_{i<j}|a_{ij}^{\varepsilon}|\Big]\,Q(x(t)-x^*).$$

À l'intérieur des crochets, on note que $\sup\mathcal{R}e\,\lambda_i$ est strictement négatif, et que les coefficients $|a_{ij}^{\varepsilon}|=O(\varepsilon|a_{ij}|)$ sont arbitrairement petits ; de sorte que toute cette partie aura une contribution négative. Passons maintenant à la contribution du reste R : d'après (54),

$$\mathcal{R}e\,\big\langle R(x(t)-x^*),x(t)-x^*\big\rangle \le \eta(|x(t)-x^*|)\,\|x(t)-x^*\|^2 \tag{57}$$
$$\le 2\eta(|x(t)-x^*|)\,Q(x(t)-x^*). \tag{58}$$

Écrivons $\eta(r)=\delta(r^2/2)$ pour tout exprimer en termes de la fonction de Lyapunov $Q(x(t)-x^*)$. En combinant (56) et (57), on obtient une équation portant sur $q(t)=Q(x(t)-x^*)$:

$$\frac{dq(t)}{dt}\le 2\Big[(\max\mathcal{R}e\,\lambda_i)+n(\max_{i<j}|a_{ij}|)\,\varepsilon+\delta(q)\Big]q(t). \tag{59}$$

Si ε et $q(t)$ sont assez petits, alors $\delta(q)$ sera très petit également, et le terme prédominant entre crochets dans (59) sera le premier, $\max\mathcal{R}e\,\lambda_i$, qui est par hypothèse strictement négatif. Choisissons donc un nombre $\Lambda>0$ tel que

$$0<\Lambda<-\max\mathcal{R}e\,\lambda_i.$$

Puisque $\delta(q)\to 0$ quand $q\to 0$, on pourra trouver $\theta>0$ et $\overline{\varepsilon}>0$ tels que

$$\varepsilon\le\overline{\varepsilon},\ q\le\theta\Longrightarrow n(\max_{i<j}|a_{ij}|)\varepsilon+\delta(q)\le(-\max\mathcal{R}e\,\lambda_i)-\Lambda.$$

En conclusion, tant que $q(t)\le\theta$ et $\varepsilon\le\varepsilon_0$,

$$\frac{dq(t)}{dt}\le -2\Lambda q(t).$$

Supposons que $q(0) \leq \theta$: alors tant que $q(t) \leq \theta$ on a $q(t) \leq q(0)e^{-2\Lambda t}$, en particulier $q(t) < \theta$. Il est donc impossible à $q(t)$ de dépasser la valeur θ (toujours le même argument de continuité qui nous a déjà rendu service plusieurs fois...). Finalement, on a la borne

$$q(t) \leq q(0)\, e^{-2\Lambda t}$$

pour tous les temps positifs. Comme $q(t)$ est le carré d'une norme de $x(t) - x^*$, on peut conclure que

$$\|x(t) - x^*\| \leq C\, \|x(0) - x^*\|\, e^{-\Lambda t},$$

pour une certaine constante C qui dépend du choix de la base et de la valeur de ε.

Remarques 75. 1. L'exemple linéaire montre que le "vrai" taux de convergence n'est pas forcément égal à $\sup \mathcal{R}e\, \lambda_i$. En effet, même si f est une application trigonalisable, le taux de convergence sera typiquement en $t^k\, e^{(\sup \mathcal{R}e\, \lambda_i)t}$, où $k \leq n-1$.

2. La fonction de Lyapunov utilisée dans cet argument pour prouver la convergence n'utilise la propriété de décroissance que dans un voisinage de l'équilibre ; dans sa construction même on a utilisé l'hypothèse de parties réelles strictement négatives.

3.6. Variétés stable et centrale

On a étudié le cas d'une EDO autonome $\dot{x} = f(x)$ au voisinage d'un équilibre x^* sous une hypothèse de stabilité linéarisée, c'est à dire si toute les valeurs propres de la matrice différentielle $df(x^*)$ ont leur partie réelle strictement négative. Peut-on aller au-delà de ce critère de stabilité ?

On se donnnera encore f un champ de vecteurs de classe C^1 dans un ouvert O de $\mathbb{R}^n$, et on se donne un équilibre x^* (on dit aussi que x^* est un point critique de f). L'équation linéarisée s'écrit $\dot{h} = df(x^*)h$. On dit que x^* est un point critique *hyperbolique* de f si *aucune valeur propre de $df(x^*)$ n'est de partie réelle nulle.* On peut classer les points critiques hyperboliques en trois catégories, selon les signes des parties réelles des valeurs propres de $df(x^*)$:

- puits (stable) : quand toutes les parties réelles sont strictement négatives ;

- source (instable) : quand toutes les parties réelles sont strictement positives ;

- point selle : si au moins une partie réelle est strictement positive, et une autre strictement négative.

EXEMPLE 76. Déterminer la nature des équilibres du système $\dot{x} = x^2 - y^2 - 1$, $\dot{y} = 2y$.

Avant d'énoncer un théorème, on va rappeler deux notions importantes, dont la première a déjà été mentionnée plusieurs fois en remarque.

DÉFINITION 77. Une variété différentielle de dimension n, de classe C^k est un espace topologique que l'on peut localement paramétrer par un ouvert de $\mathbb{R}^n$, avec des changements de variables de classe C^k.

On renvoie à un cours de géométrie différentielle [**10, 15**] pour plus de détails. Souvent les variétés sont définies comme des *sous-variétés* de l'espace euclidien : ce sont des sous-ensembles de $\mathbb{R}^N$ définis par une équation, au moyen d'un théorème de fonctions implicites par exemple.

DÉFINITION 78. Soit $(\Phi_t)_{t\in\mathbb{R}}$ le flot associé à un champ de vecteurs autonome de classe C^1. Un ensemble E est dit invariant pour le flot si

$$\forall x_0 \in E, \quad \forall t \in \mathbb{R}, \quad \Phi_t(x_0) \in E.$$

On dit aussi que E est positivement invariant si cette propriété est vraie pour tout $t \geq 0$, et négativement invariant si elle est vraie pour tout $t \leq 0$.

L'idée est de construire des *variétés invariantes* avec des propriétés de stabilité.

THEORÈME 79 (Théorème de la variété stable). *Soient O un ouvert de $\mathbb{R}^n$, f un champ de vecteurs C^1 dans O et Φ_t le flot maximal associé. Soit x^* un point critique hyperbolique de f. On suppose que $df(x^*)$ a k valeurs propres de partie réelle strictement négative, associées à un sous-espace propre E^S ; et $n-k$ valeurs propres de partie réelle strictement positive, associées à un sous-espace propre E^U. (Ici les valeurs propres sont comptées avec leur multiplicité.) Alors il existe une variété différentiable S, de classe C^1 et de dimension k, tangente à E^S, positivement stable, telle que pour tout $x_0 \in S$, $\Phi_t(x_0)$ converge vers x^* quand $t \to +\infty$. Il existe aussi une variété différentiable U, de classe C^1 et de dimension $n-k$, tangente à E^U, négativement stable, telle que pour tout $x_0 \in S$, $\Phi_t(x_0)$ converge vers x^* quand $t \to -\infty$.*

On appelle S la variété stable et U la variété instable. Dans un cas comme dans l'autre, les taux de convergence sont déterminés (comme on peut s'y attendre) par les plus petites valeurs absolues des parties réelles des valeurs propres associées.

REMARQUE 80. On n'a pas fait d'hypothèse a priori sur le flot ; le théorème contient donc implicitement la propriété selon laquelle $\Phi_t(x_0)$

est bien défini pour $t \to +\infty$ quand $x_0 \in S$ et bien défini pour $t \to -\infty$ quand $x_0 \in U$.

THEORÈME 81 (Théorème de Hartman–Grobman). *Avec les mêmes notations que dans le Théorème 79, le flot non linéaire et le flot linéarisé sont* localement équivalents*, c'est à dire que l'on peut trouver un homéomorphisme ψ au voisinage de x^* qui transforme localement les trajectoires de l'un en les trajectoires de l'autre : pour $|t|$ assez petit,*

$$\psi^{-1}\big(\Phi_t(\psi(x_0))\big) = e^{tA}x_0.$$

Si l'on suppose en outre f de classe C^2, alors on peut imposer à ψ d'être un C^1-difféomorphisme.

À venir : quelques mots sur les preuves, dans une version future ?

Les Théorèmes de la Variété Stable et de Hartman–Grobman ne s'appliquent qu'aux équilibres hyperboliques. Le Théorème de la Variété Centrale, en revanche, donne des informations quand certaines valeurs propres sont nulles.

THEORÈME 82 (Théorème de la Variété Centrale). *Soient O un ouvert de $\mathbb{R}^n$, f un champ de vecteurs C^1 dans O et Φ_t le flot associé. Soit x^* un point critique hyperbolique de f. On suppose que $df(x^*)$ a*

- *k valeurs propres de partie réelle strictement négative, associées à un sous-espace propre E^S ;*
- *j valeurs propres de partie réelle strictement positive, associées à un sous-espace propre E^U ;*
- *$m = n - k - j$ valeurs propres de partie réelle nulle, associées à un sous-espace propre E^C.*

(Toutes ces valeurs propres sont encore comptées avec multiplicité.)

Alors il existe une variété différentiable C, de classe C^1 et de dimension m, tangente à E^C, invariante sous l'action du flot. Elle est localement unique (uniquement déterminée au voisinage de x^*) et on l'appelle variété centrale du système.

EXEMPLE 83. Le système $\dot{x}_1 = x_1^2$, $\dot{x}_2 = -x_2$ admet-il une variété centrale, et si oui laquelle ?

3.7. Complément : Théorème de Poincaré–Bendixson

Le Théorème de Poincaré–Bendixson est, tout à la fois, conceptuellement important, techniquement difficile, et limité dans son application. Il ne s'applique qu'aux équations différentielles *à deux degrés de liberté*, c'est à dire posées dans un espace des phases de dimension 2.

Comme l'étude des équations en dimension 1 est conceptuellement facile, la dimension 2 est la plus petite dans laquelle on peut commencer à observer des phénomènes riches ; le Théorème de Poincaré–Bendixson, démontré en 1901, les classifie.

Ce théorème s'intéresse à décrire l'ensemble l'ensemble attracteur, dont voici une définition très générale :

DÉFINITION 84 (ensemble ω-limite). On appelle ensemble ω-limite d'une trajectoire $(x(t))$ l'ensemble de tous ses points d'accumulation quand $t \to \infty$, soit

$$\Omega := \Big\{ y;\ \exists (t_n) \to \infty;\ x(t_n) \to y \Big\}.$$

En d'autres termes, c'est l'ensemble de toutes les limites possibles de positions successives de la trajectoire évaluées en des temps qui tendent vers l'infini.

Dans le cadre des EDO, nous avons déjà rencontré quelques exemples de structures limite notables :

- un équilibre est un point fixe du flot, donc c'est évidemment un ensemble ω-limite ; il peut être asymptotiquement stable (partant près de l'équilibre, on converge vers cet équilibre quand $t \to +\infty$), ou orbitalement stable (partant près de l'équilibre, on en reste proche pour tous les temps ultérieurs), ou encore instable (partant de l'équilibre, on peut s'en retrouver loin à un temps ultérieur).

- un cycle est une trajectoire fermée, périodique en temps.

Introduisons maintenant une autre structure intéressante. Pour la discussion, on supposera toujours que l'EDO est de classe C^1. Les équilibres peuvent être connectés par des trajectoires du flot : on part d'un équilibre (forcément instable), et on aboutit à un autre équilibre (stable ou instable). Si cela se fait, c'est forcément *en temps infini* : c'est à dire que la trajectoire tendra vers un équilibre (instable) pour $t \to -\infty$, et tendra également vers un équilibre pour $t \to +\infty$. (En effet, si la trajectoire atteignait un équilibre, disons x^*, en temps fini, et puisque $t \mapsto x^*$ est solution de $\dot{x} = f(x)$, le Théorème de Cauchy–Lipschitz, sous la forme de la Proposition 44 de non-croisement, impliquerait que la trajectoire est identiquement égale à x^*, ce qui est absurde.)

Rien n'empêche d'ailleurs ces deux équilibres d'être identiques ; on a une terminologie spéciale pour ce cas de figure.

DÉFINITION 85 (Trajectoires reliant des équilibres). Soit $(x(t))_{t\in\mathbb{R}}$ une trajectoire d'une EDO, reliant deux équilibres :

$$\lim_{t\to-\infty} x(t) = x^*_-, \qquad \lim_{t\to+\infty} x(t) = x^*_+.$$

On dit que x est une trajectoire hétérocline si $x^*_+ \neq x^*_-$, et homocline si $x^*_+ = x^*_-$.

En d'autres termes, une trajectoire homocline est issue d'un équilibre instable, et revient vers ce même équilibre. Nous pouvons maintenant énoncer le théorème de Poincaré–Bendixson, de manière légèrement informelle.

THEORÈME 86 (Théorème de Poincaré–Bendixson). *Considérons une EDO autonome d'ordre 1, de classe C^1, posée dans un domaine ouvert O du plan $\mathbb{R}^2$. Alors, dans tout sous-domaine compact de O contenant un nombre fini d'équilibres, l'ensemble ω-limite d'une orbite de l'équation ne peut être que l'un des trois objets suivants :*

- un équilibre ;

- un cycle ;

- un graphe fait d'un nombre fini d'équilibre instables connectés par des trajectoires (homoclines ou hétéroclines) du flot. En outre, deux équilibres distincts sont connectés par au plus deux trajectoires hétéroclines, une dans chaque direction du temps. (Ce graphe peut, en revanche, comporter un nombre arbitraire, voire une infinité dénombrable, de trajectoires homoclines issues d'un équilibre donné.)

REMARQUE 87. Un équilibre est un cycle particulier, donc on pourrait simplifier encore l'énoncé ; mais les équilibres sont si importants qu'il est légitime de les considérer à part.

La démonstration de ce théorème est difficile, nous n'allons pas en parler ! C'est seulement la morale qu'il faut retenir : en deux dimensions, un système différentiel ne peut avoir, asymptotiquement, qu'un petit nombre de comportements possibles.

REMARQUE 88. Un exercice intéressant montre que si une trajectoire est compacte, elle est forcément périodique et ce même en dimension supérieure à 2 (ainsi on a équivalence, pour des trajectoires d'EDO, de deux sens différents du mot "fermé"). L'hypothèse est très forte, puisque l'on suppose que la trajectoire est fermée, sans prendre l'adhérence comme dans la définition de l'ensemble ω-limite. La conclusion est aussi très forte, montrant que la trajectoire se confond avec son comportement asymptotique, qui est toujours cyclique. Pour une preuve, voir l'exercice 12 de la semaine 3.

3.8. Complément : Chaos déterministe, attracteurs étranges

Pour terminer cette semaine, nous allons évoquer l'un des plus célèbres développements mathématiques du vingtième siècle : la théorie

du *chaos déterministe*, ou tout simplement théorie du chaos. Dans toute la discussion, on se limite à des équations de classe au moins C^1, autonomes et du premier ordre.

Le Théorème de Poincaré–Bendixson montre que le flot d'une équation différentielle en dimension 2 ne peut pas faire "n'importe quoi" : quand on l'examine en temps long, ne subsistent que des équilibres, des cycles, des trajectoires homoclines et hétéroclines.

À partir de là, on pourrait imaginer qu'en dimension 3 on peut prouver des théorèmes de ce genre, avec des objets plus compliqués ? Mais la réalité est tout autre : dès la dimension 3, des comportements radicalement différents peuvent se produire, et l'on ne peut décrire la dynamique en temps grand par des objets "simples". Cela est très bien illustré par le célèbre *système de Lorenz*.

Edward Lorenz, mathématicien et météorologiste américain, travaille dans les années 50 et 60 sur les équations de la météorologie. Conscient des complications considérables que les non-linéarités peuvent entraîner, il s'attache à détecter des comportements qualitativement complexes dans ces équations. En 1963 il publie des résultats surprenants : il a remplacé les équations aux dérivées partielles par un système simplifié à l'extrême, comportant seulement trois degrés de régularité, et avec la régularité la plus simple que l'on puisse imaginer – quadratique.

Le système de Lorenz s'écrit donc avec 3 inconnues réelles (disons x, y, z), et 3 paramètres strictement positifs (σ, ρ, β) :

$$\begin{cases} \dot{x} = \sigma\,(y - x) \\ \dot{y} = \rho\, x - y - xz \\ \dot{z} = -\beta\, z + xy. \end{cases} \tag{60}$$

Pour en arriver là, Lorenz a travaillé dans une géométrie périodique et n'a retenu que les premiers modes de Fourier (c'est à dire, les composantes qui oscillent les plus lentement) dans l'équation de Boussinesq, un modèle populaire de mécanique des fluides. Il faut imaginer que l'on garde seulement une direction horizontale et une direction verticale ; et que l'on modélise un fluide qui est chauffé par le bas et refroidi par le haut, sujet donc à des phénomènes de convection. Du point de vue physique,

- σ est le rapport entre la viscosité (propension du fluide à la friction) et la diffusivité thermique (capacité de la chaleur à se propager)
- ρ est le nombre de Rayleigh adimensionné (paramètre mesurant si le transfert de chaleur se fait plutôt par convection ou plutôt par conduction)

- β est un paramètre lié à la géométrie de l'écoulement.

En outre x est le mode de Fourier $(1,1)$ de la fonction potentiel du fluide, et y, z sont les modes de Fourier $(1,1)$ et $(0,2)$ de la fonction température.

En résumé, le système (60) est un modèle ultra-simplifié de convection décrivant l'interaction entre l'écoulement d'un fluide et l'évolution de la température en son sein. L'allure du flot dépend des valeurs des paramètres, bien sûr : dans certains régimes c'est un flot tranquille, dans d'autres ce sera un flot turbulent. Mais la complexité est si réduite que l'on peut s'attendre à un comportement plutôt facile à étudier. Cependant, quand on examine l'ensemble asymptotique (ω-limite), on découvre que les trajectoires asymptotiques forment une figure mystérieuse, avec deux boucles... on peut y voir un symbole infini, ou une sorte de papillon, ce qui est ironique car c'est à la suite de ces travaux que l'on a inventé l'expression d'"effet papillon".

Lorenz avait choisi les paramètres $\beta = 8/3$, $\sigma = 10$, $\rho = 28$ (un régime assez turbulent). L'ensemble n'est pas un ruban solide, mais une structure très fine, pleine de trous, qui évoque lointainement les anneaux de Saturne. Cet objet, que l'on appelle un **attracteur étrange**, s'apparente à un fractal, et sa dimension est nettement plus grande que 1 : on l'a estimée à environ 2,06.

Les équations de Lorenz sont extrêmement proches des équations écrites par Tsuneji Rikitake quelques années avant lui, dans le cadre de l'étude de la polarisation magnétique du globe terrestre (l'effet dynamo). Mais le grand mérite de Lorenz est d'avoir su dégager de son modèle, de manière limpide, quelques principes caractéristiques qui constituent la base de la **théorie du chaos** :

(1) *Sensibilité aux conditions initiales*, et plus précisément instabilité exponentielle : si l'on commet une erreur minuscule sur l'état de départ, évalué au temps t_0, on obtient une erreur éventuellement importante sur l'état en un temps $t > t_0$; et en fait l'imprécision croît exponentiellement vite au fur et à mesure que le temps passe. Dans le cas du système de Lorenz cela peut paraître paradoxal, puisque l'on est en train d'étudier une structure limite, a priori stable ! Mais en fait on a, en chaque point de cet attracteur,

- la direction du flot, qui fait avancer le système le long de l'attracteur ;

- une direction stable : un écart dans cette direction tend à s'amenuiser aux temps ultérieurs ;

- une direction instable : un écart dans cette direction tend à s'amplifier aux temps ultérieurs.

De la sorte, on ne peut pas prédire à l'avance l'allure de la trajectoire, *bien que le système soit déterministe*. En effet, on peut considérer deux conditions initiales, presque identiques – disons identiques jusqu'à la 30e décimale – telles qu'en un temps $t > t_0$ assez grand la première solution sera située dans la boucle de gauche, et la deuxième solution sera située dans la boucle de droite, bien loin de la première. Lorenz introduit la terminologie d'*effet papillon* : en extrapolant ces conclusions au système météorologique initial, on peut penser qu'un effet infime comme un battement d'aile de papillon résultera en un changement radical – tempête ou pas tempête – à quelques semaines et quelques milliers de kilomètres de distance.

(2) *Imprédictibilité du comportement global* : Une fois l'attracteur pris en compte, à peu près toutes les trajectoires sont possibles... Par exemple, décidons de couper l'attracteur en deux avec une boucle gauche (G) et une boucle droite (D), et enregistrons une trajectoire, de manière très grossière, en fonction de la succession de boucles qu'elle emprunte : si elle commence par tourner selon la boucle de gauche avant de passer dans la boucle de droite, on notera (G,D), etc. Peut-être qu'ainsi la trajectoire sera repérée par (G,D,D,D,G,G,G,D,G,....) Et bien, pour n'importe quelle suite finie de lettres G et D, on peut trouver des solutions de l'équation qui suivent ce comportement ! Autrement dit, le système explore un peu tout ce qui est possible.

(3) *Ergodicité* : Cela veut dire que le comportement *statistique* du système au cours du temps est prédictible au moyen d'une probabilité dans l'espace des phases, et plus précisément sur l'attracteur : on peut prévoir statistiquement quelle proportion du temps le système restera à gauche, quelle proportion du temps il restera à droite... De manière générale, un système est dit μ-ergodique si pour presque toute condition initiale x_0 et pour toute fonction ζ (continue bornée) on a

$$\frac{1}{T}\int_0^T \zeta(\Phi_t(x))\,dt \xrightarrow[T\to\infty]{} \int \zeta(y)\,\mu(dy).$$

En d'autres termes, si l'on peut remplacer la moyenne temporelle le long du flot (a priori extrêmement complexe) par une moyenne sur l'espace des phases, plus simple conceptuellement. (C'est un problème considérable que de savoir reconnaître si un système est ergodique ; quand il l'est, on peut utiliser des moyennes sur les trajectoires pour calculer numériquement la mesure μ.)

Dans la vision de Lorenz, on a ainsi une combinaison d'*imprédictibilité* des trajectoires individuelles, et de *prédictibilité statistique* de ces mêmes trajectoires. Pour continuer l'analogie du papillon, Lorenz pense qu'il

est impossible de prédire si une tempête aura lieu à un certain moment, mais que l'on peut en revanche prédire la quantité de tempêtes qui se déchaîneront sur une longue période de temps.

Lorenz a ainsi dit, de manière frappante, *"J'avance qu'au fil des années les petites perturbations ne modifient pas la fréquence d'apparition des événements tels que les ouragans : la seule chose qu'elles peuvent faire, c'est modifier l'ordre dans lequel ces événements se produisent."*

Pour illustrer la sensibilité aux conditions initiales, Lorenz aimait réaliser l'expérience suivante : on lâche une feuille de papier à l'horizontale, plusieurs fois d'affilée, en prenant garde de la lâcher exactement dans le même état d'une fois sur l'autre ; et l'on note que les endroits où la feuille atterrit sont fort différents les uns des autres.

C'est toujours un problème débattu que de savoir si l'effet papillon s'applique vraiment à la météorologie. En effet, "dans la vraie vie", les équations de la mécanique des fluides font intervenir une quantité infinie de degrés de liberté. On pourrait penser que cela rend les choses encore plus compliquées, mais on peut aussi argumenter que cela va stabiliser le système, par moyenne. Au-delà de ce débat, la théorie de Lorenz montre que la simplification des équations peut avoir des effets incontrôlés, et que le déterminisme peut se coupler avec un comportement subtilement imprédictible.

La théorie du chaos rassemble un arsenal de théorèmes, concepts et recettes mathématiques en rapport avec ces différents aspects : instabilité, imprédictibilité, ergodicité... Ces outils appartiennent à divers champs mathématiques tels que la géométrie, la topologie et les systèmes dynamiques. L'un des concepts clés est la robustesse du comportement : on parle de *stabilité structurelle* si le comportement des solutions reste inchangé quand on perturbe (raisonnablement) la forme de l'équation. On étudie aussi la possibilité de remplacer un modèle différentiel par une dynamique abstraite, par exemple en temps discret, définie par une opération symbolique (comme le décalage de Bernoulli qui décale l'écriture binaire d'un nombre d'un cran vers la droite) ou une itération polynomiale (par exemple celle de Mandelbrot qui à un nombre complexe z associe z^2+c). Certains attracteurs issus de la dynamique complexe, tels que les ensembles fractals de Mandelbrot et de Julia, sont l'objet d'études mathématiques ainsi que d'expérimentations artistiques.

La théorie du chaos prend sa source lointaine dans les travaux de Henri Poincaré, à la fin du 19ème siècle, sur l'instabilité du système solaire ; et dans les oeuvres subséquentes de George D. Birkhoff. Lorenz n'avait pas conscience explicitement de ces travaux, mais avait

suivi les enseignements de Birkhoff ; c'est en partie sous cette influence qu'il a redécouvert heuristiquement certains des concepts les plus importants de la théorie de Poincaré et Birkhoff. Poincaré lui-même avait tout à fait la vision d'un mélange d'imprédictibilité individuelle et de prédictibilité statistique ; il pensait aussi que l'on peut obtenir des informations importantes sur le flot sans pour autant calculer ses solutions, et même sans connaître précisément la forme des équations : *"Vous me demandez de vous prédire les phénomènes qui vont se produire. Si, par malheur, je connaissais les lois de ces phénomènes, je ne pourrais y arriver que par des calculs inextricables et je devrais renoncer à vous répondre ; mais comme j'ai la chance de les ignorer, je vais vous répondre tout de suite. Et, ce qu'il y a de plus extraordinaire, c'est que ma réponse sera juste."*

Dans les années 1980, les astrophysiciens Jacques Laskar en France et Scott Tremaine au Canada ont montré, via de lourdes simulations informatiques, que la dynamique des planètes du système solaire est chaotique : les trajectoires des petites planètes sont imprédictibles sur des périodes d'une centaine de millions d'années, ce qui n'est pas long par rapport à l'âge du système solaire. Ils ont ainsi confirmé l'intuition de Poincaré. Cette instabilité est due à des phénomènes de quasi-résonance entre les orbites des planètes : une résonance vient avec une possibilité d'amplification des instabilités, un peu comme quand un pont suspendu casse si un régiment passe dessus en marchant au pas à une fréquence bien (ou plutôt mal) choisie.

Laskar estime ainsi que la distance entre deux objets que l'on lâche "au hasard" dans le système solaire croît exponentiellement vite, avec un facteur multiplicatif d'environ $10^{t/10}$, quand le temps t est exprimé en millions d'années. Laskar et ses collaborateurs ont même montré numériquement, il y a quelques années, que le comportement des planètes du système solaire est influencé de manière majeure par de minuscules variations des orbites de deux petits satellites, Ceres et Vesta, en orbite entre Mars et Jupiter. Il suffit qu'une cause inconnue affecte de quelques centimètres la position de ces satellites pour que cela change radicalement la position de la Terre par rapport au Soleil soixante millions d'années plus tard. L'effet papillon existe donc vraiment dans le système solaire !

La théorie du chaos est un sujet bien plus vaste et complexe que ce que nous pouvons traiter dans ce cours ; elle a été particulièrement populaire dans les années 70 et 80. Pour une introduction légère, on pourra consulter le DVD *Chaos* d'Aurélien Alvarez, Étienne Ghys et Jos Leys ; et pour une introduction plus approfondie, l'ouvrage de James

Gleick, *Chaos* [**17**]. On retrouvera dans ces sources tous les mots clés de cette section : instabilité, imprédictibilité trajectorielle, attracteurs étranges, fractals, prédictibilité statistique...

CHAPITRE 4

Le pendule pesant

Un pendule pesant est constituée d'une petite boule (la "masse") suspendue à un fil, lui-même accroché à un support fixe. Sous l'effet de la gravité, le pendule hors d'équilibre oscille, et l'on peut mettre son mouvement en équations.

Le modèle du pendule pesant est l'une des plus simples équations différentielles non linéaires qui soient. Son étude remonte au 17ème siècle, et a eu d'importantes répercussions technologiques sur la mesure précise du temps. En même temps, c'est une équation emblématique, qui permet d'apprécier la richesse des concepts généraux ; c'est un incontournable d'un cours sur les équations différentielles. Nous allons prendre notre temps pour l'étudier pas à pas.

4.1. Modélisation

Soit un pendule accroché par le haut, initialement au repos, auquel on imprime une petite poussée : le pendule se met alors à osciller. Il en va de même si l'on tire le pendule pour le déplacer légèrement de sa position d'équilibre, tout en gardant le fil tendu, et qu'on le lâche, avec ou sans vitesse initiale. On observe que

- le fil reste à peu près tendu et droit (il peut se déformer légèrement sous l'action de son propre poids) ;

- le mouvement s'effectue dans un plan vertical, défini par la position initiale du fil ;

- les oscillations sont régulières et à peu près constantes en amplitude et en durée.

Si l'on observe le pendule un peu plus longtemps, on constate que les frottements réduisent peu à peu l'ampleur des oscillations ; et si l'on veut entretenir ces dernières indéfiniment, il faut fournir un peu d'énergie au système, en agissant soit sur la masse, soit sur le fil (si l'on tient le pendule entre ses doigts, on peut entretenir les oscillations par d'infimes mouvements des doigts). Cependant, un pendule de bonne qualité peut osciller pendant très longtemps sans avoir besoin d'un apport d'énergie.

Si la masse du fil est complètement négligée, les seules forces qui s'exercent au sein du fil sont des forces de tension, qui se propagent le long de l'axe ; le fil reste donc tendu, tant que la valeur de la tension est strictement positive.

Si l'on imprime un mouvement de grande ampleur au pendule, il peut très bien se produire que la tension s'annule subitement quand le pendule occupe une position assez élevée ; il peut alors se faire que le fil perde d'un coup sa rigidité, de sorte que le pendule "tombe brusquement" sous l'effet de la seule gravité, avant de "rebondir" quand le fil se tend à nouveau. Pour éviter cette difficulté majeure, on va commencer par remplacer le fil par une *tige rigide*, que l'on supposera aussi de masse négligeable. On reviendra plus tard sur cette hypothèse.

À part la force de tension, le fil est soumis à la force exercée par le solide, et aux forces exercées par le point d'accroche. Si le pendule est accroché à un mur, ces forces sont faibles ; mais on sait qu'elles ont une influence, la preuve expérimentale en étant que si l'on accroche deux pendules à un mur et qu'on les fait osciller de manière indépendante, ils finissent par se synchroniser, du fait des petites vibrations transmises par le mur. Là encore, c'est un effet très subtil que nous allons négliger.

Assimilons la masse au bout de la tige à un point. La force de tension exercée par le fil et la force de gravité se combinent pour mettre la masse en mouvement. Appelons T l'intensité de la force de tension, g l'intensité de la gravité, m la valeur de la boule, et L la longueur du fil. On applique la relation de Newton : $ma = \sum F$, la masse multipliée par l'accélération est égale à la somme des forces. L'accélération ici est calculée dans un repère galiléen (on ne peut pas choisir un repère dirigé selon la tige !). On définit donc une direction verticale fixe, et une direction horizontale fixe, puis on écrit la relation de Newton dans ce repère, et ensuite on projette sur l'axe de la tige.

On dirige l'axe vertical vers le haut, et on introduit θ l'angle avec la verticale, mesuré avec l'orientation trigonométrique habituelle. La position de la boule est alors

$$x = L(\sin\theta, -\cos\theta),$$

et on en déduit la vitesse et l'accélération

$$\dot{x} = L(\cos\theta, \sin\theta)\,\dot{\theta}$$

$$\ddot{x} = L(-\sin\theta, \cos\theta)\,\dot{\theta}^2 + L(\cos\theta, \sin\theta)\,\ddot{\theta}.$$

Les forces en présences sont la gravité F_g et la tension F_T :

$$F_g = (0, -mg), \qquad F_T = T(\sin\theta, -\cos\theta).$$

On applique le bilan des forces : après avoir tout divisé par m, on obtient

$$\begin{cases} L\cos\theta\,\ddot{\theta} - L\sin\theta\,\dot{\theta}^2 = -\dfrac{T}{m}\sin\theta \\ \\ L\sin\theta\,\ddot{\theta} + L\cos\theta\,\dot{\theta}^2 = -g + \dfrac{T}{m}\cos\theta. \end{cases} \tag{61}$$

Projetons maintenant sur les vecteurs unitaires

$$e = (\sin\theta, -\cos\theta), \qquad e^{\perp} = (\cos\theta, \sin\theta).$$

Par exemple, la projection sur e revient à multiplier la première équation de (61) par $\sin\theta$, la seconde par $-\cos\theta$, et a àjouter les résultats. On trouve en fin de compte deux nouvelles équations :

$$\begin{cases} L\dot{\theta}^2 = -g\cos\theta + \dfrac{T}{m} \\ \\ L\ddot{\theta} = -g\sin\theta. \end{cases} \tag{62}$$

La deuxième équation se réécrit

$$\ddot{\theta} = -\frac{g}{L}\sin\theta. \tag{63}$$

C'est la célèbre **équation du pendule** : elle fournit l'évolution de l'angle θ en fonction du temps. Elle ne dépend que du paramètre $k = g/L$, et donc pas de la masse m. Cette observation remonte (au moins) au début du 17ème siècle, quand Galilée effectuait des expériences avec des chandeliers attachés à des cordes, qu'il chronométrait avec les battements de son cœur !

La première équation dans (62) donne la valeur de T en fonction de θ :

$$T = mL\dot{\theta}^2 + mg\cos\theta. \tag{64}$$

Laissons maintenant de côté pendant quelque temps le problème physique, et concentrons-nous sur l'équation du pendule (63), objet mathématique intéressant en soi. Notons qu'on l'appelle aussi "équation du pendule non linéaire", par opposition à l'équation linéaire $\ddot{\theta} = -k\theta$ qui décrit ses petites oscillations ; on l'appelle encore "équation du pendule pesant" pour rappeler que c'est la gravité, via le poids du pendule, qui gouverne les oscillations (même si, ironiquement, la masse a disparu des calculs !)

4.2. Résolution de l'équation du pendule

Essayons dans un premier temps de résoudre l'équation (63) de manière aussi explicite que possible.

On commence par ramener (63), équation du second ordre à une variable, à une équation du premier ordre à deux variables. Pour cela on pose $\psi = \dot\theta$, de sorte que

$$\dot\theta = \psi, \qquad \dot\psi = -k \sin\theta.$$

En général on ne sait pas "résoudre" un système à deux variables. Mais celui-ci admet une loi de conservation ; on peut le découvrir en jouant avec l'équation, ou en se laissant guider par l'intuition physique de la *conservation de l'énergie*.

L'énergie ici est constituée de deux contributions :

- énergie cinétique $E_c = mv^2/2$ (v la vitesse) ;
- énergie potentielle gravitationnelle $E_P = mgh$ (h la hauteur).

Dans notre jeu de variables, cela prend la forme

$$E_c = \left(\frac{m}{2}\right) L^2\dot\theta^2, \qquad E_P = mgL\,(1-\cos\theta).$$

(On a normalisé pour que l'énergie associée à la position de repos, $\theta = 0$, soit nulle ; c'est juste une question de convention, l'énergie potentielle étant définie à une constante près.)

On définit l'énergie totale comme la somme de l'énergie cinétique et de l'énergie potentielle :

$$E(\theta,\dot\theta) = \left(\frac{m}{2}\right) L^2\dot\theta^2 + mgL\,(1-\cos\theta). \tag{65}$$

Vérifions que cette énergie est effectivement préservée par l'évolution temporelle : grâce à (63) on trouve

$$\begin{aligned}\frac{dE}{dt} &= mL^2\dot\theta\,\ddot\theta + mgL\sin\theta\,\dot\theta \\ &= -mL^2\dot\theta\left(\frac{g}{L}\right)\sin\theta + mgL\sin\theta\,\dot\theta = 0.\end{aligned}$$

Voici une autre façon d'effectuer le calcul : partant de l'équation $\ddot\theta = -(g/L)\sin\theta$, on multiplie les deux membres par $\dot\theta$, et on trouve $\dot\theta\ddot\theta = -(g/L)\dot\theta\sin\theta$. Le membre de gauche est la dérivée de $\dot\theta^2/2$, et le membre de droite est la dérivée de $(g/L)\cos\theta$; on en déduit que $\dot\theta^2/2 - (g/L)\cos\theta$ est constante. Quitte à ajouter une constante, et multiplier par une autre constante, c'est l'énergie E.

Il sera commode, pour alléger les calculs, de définir

$$\mathcal{E}(\theta,\dot\theta) = \frac{E}{mL^2} = \frac{\dot\theta^2}{2} + k(1-\cos\theta), \qquad k = \frac{g}{L}. \tag{66}$$

Nos calculs ont prouvé que $\mathcal{E}$ est un *invariant* du mouvement.

L'invariant $\mathcal{E}$ réduit la dynamique de deux degrés de liberté à un seul : en effet, le mouvement doit s'effectuer dans les lignes de niveau de l'invariant. Si donc $\mathcal{E}_0$ est la valeur initiale de l'énergie (cela dépend de l'angle initial, et de la vitesse qu'on lui imprime au départ), alors on aura, par conservation de l'énergie

$$\dot{\theta}^2 = 2\big[\mathcal{E}_0 - k(1-\cos\theta)\big],$$

d'où

$$\dot{\theta} = \pm\sqrt{2\big[\mathcal{E}_0 - k(1-\cos\theta)\big]}.$$

Les variations angulaires sont donc directement liées aux variations temporelles :

$$\frac{\pm d\theta}{\sqrt{2\big[\mathcal{E}_0 - k(1-\cos\theta)\big]}} = dt. \tag{67}$$

À partir de là, il faut distinguer selon que $\dot{\theta} > 0$ (déplacement du pendule vers la droite) ou $\dot{\theta} < 0$ (déplacement du pendule vers la gauche). Par exemple, dans le cas d'un pendule se déplaçant vers la droite, on posera

$$F(x) = \int \frac{dy}{\sqrt{2\left[\mathcal{E}_0 - k(1-\cos y)\right]}},$$

et l'équation (67) s'intègrera en

$$F(\theta(t)) - F(\theta(t_0)) = t - t_0, \tag{68}$$

équation qui fournit $\theta(t)$ *implicitement* en fonction de t et de $\mathcal{E}_0$.

À dire vrai, cette formule n'est pas, pour l'heure, d'une grande utilité : on ne peut pas raisonnablement prétendre qu'elle nous fait mieux comprendre les propriétés mathématiques et physiques du système. Il sera plus instructif de tracer le portrait de phases de l'équation, et de l'étudier qualitativement, réservant la formule (68) pour certaines questions subtiles.

4.3. Portrait de phase

On va maintenant étudier l'allure géométrique des solutions de l'équation

$$\ddot{\theta} = -k\sin\theta$$

dans l'espace des phases. Chaque point dans l'espace des phases donne deux informations : position (angle), et vitesse (angulaire). La seconde

variable, que l'on a appelée ψ plus haut, correspondra à la vitesse angulaire, et souvent on l'appellera donc $\dot\theta$ par léger abus de notation. On parlera donc de l'espace des phases comme de l'ensemble des couples $(\theta, \dot\theta)$.

Le système est périodique de période 2π dans la variable θ : la figure sera donc invariante par translation horizontale de 2π.

On rappelle que

$$\mathcal{E}(\theta, \dot\theta) = \frac{\dot\theta^2}{2} + k(1 - \cos\theta).$$

est constante le long des trajectoires. Le mouvement du système a donc entièrement lieu dans les lignes de niveau de la fonction : traçons ces lignes de niveau. Pour ce faire, on se donne un niveau d'énergie $\overline{e}$, et on écrit la courbe $\mathcal{E} = \overline{e}$ comme une union de graphes $\dot\theta = f(\theta)$. On trouve facilement

$$\dot\theta = \pm\sqrt{2[\overline{e} - k(1 - \cos\theta)]}.$$

Il y a deux branches : l'une positive, l'autre négative, symétriques l'une de l'autre.

Étudions la branche positive. La fonction $k(1 - \cos\theta)$ prend toutes les valeurs entre 0 et $2k$. Si $\overline{e} > 2k$, la fonction sous la racine carrée reste strictement positive. Si $\overline{e} < 2k$, elle s'annule pour un certain angle θ^0 ; et près de cet angle elle se comporte à peu près comme une fonction linéaire, donc la racine carrée se comporte à peu près comme la racine carrée d'une fonction linéaire, avec une pente verticale. Enfin si $\overline{e} = 2k$, la fonction sous la racine carrée vaut $2k(1 + \cos\theta)$ et s'annule précisément pour $\theta = \pi$, où elle a un comportement quadratique en $k(\theta - \pi)^2$; le graphe se comporte donc comme $\sqrt{k}\,|\theta - \pi|$ au voisinage de π.

On note en outre que pour $\overline{e} \simeq 0$, on a forcément $\theta \simeq 0$, le graphe est donc approché par $\sqrt{2\,[\overline{e} - k\theta^2/2]}$, ce qui est l'équation d'une ellipse dont les axes sont en proportion 1 et $\sqrt{1/k}$. (Pour $k = 1$ on a donc une figure presque circulaire.)

À ce stade on a assez d'informations pour tracer un dessin approché (voir les dessins dans le cours vidéo ou sur les animations !) :

- des ellipses déformées, concentriques, centrées sur l'axe horizontal, au niveau de l'origine et plus généralement des multiples pairs de π ; ces "ellipses" sont de plus en plus grandes et traversent l'axe horizontal avec une tangente verticale ;

- elles sont encerclées par des "yeux" qui joignent, sur l'axe vertical, les multiples impairs consécutifs de π ; ces yeux aboutissent avec des pentes $\pm\sqrt{k}$ et forment comme une longue chaîne ;

- des ondulations périodiques au-dessus et en-dessous des yeux, de plus en plus écartées de l'axe horizontal.

Une fois ce graphe tracé, on a un découpage de l'espace des phases en "tranches courbes", sur lequel on peut lire l'évolution en utilisant la propriété de conservation de l'énergie. Dans le demi-plan supérieur $\dot{\theta} > 0$, θ va bien sûr augmenter, donc au-dessus de l'axe horizontal toutes les courbes peuvent être orientées de la gauche vers la droite ; symétriquement, au-dessous de l'axe horizontal toutes les courbes peuvent être orientées de la droite vers la gauche.

On a deux positions d'équilibre : $\theta = 0$ et $\theta = \pi$, correspondant aux annulations de $\sin\theta$. La position $\theta = 0$ est la position "de repos" où le pendule est dirigé vers le bas, la position $\theta = \pi$ est la position "inversée" où le pendule est dirigé vers le haut. (Ici il est vital de travailler avec une tige rigide plutôt qu'avec un fil !) Intuitivement, la position de repos est stable, la position inversée est instable : il suffit de lâcher le pendule extrêmement près de cette position pour le voir effectuer une grande oscillation.

La stabilité de la position de repos se voit facilement sur le portrait de phases : la solution est en effet "coincée" par les courbes environnantes, de sorte que si on part près de l'origine on reste près de l'origine. Noter que ce n'est pas la stabilité asymptotique (le pendule ne retourne pas vers l'équilibre dans ce modèle !) mais la stabilité orbitale (l'orbite du pendule reste proche de l'équilibre). On voit également sur ce portrait de phases la régularité des oscillations quand leur amplitude n'est pas trop grande ; avec des oscillations tantôt vers la droite (demi-plan supérieur dans l'espace des phases), tantôt vers la gauche (demi-plan inférieur dans l'espace des phases).

L'instabilité de la position inversée se voit bien également : partant de l'équilibre, il suffit de perturber légèrement pour se retrouver pris dans des trajectoires de grande ampleur. On voit aussi qu'il y a des trajectoires particulières qui joignent cet équilibre à lui-même : en suivant le contour d'un œil, sur la figure, on va de $\theta = -\pi$ à $\theta = \pi$ (mais comme tout est modulo 2π, $\theta = \pi$ et $\theta = -\pi$ représentent la même position d'équilibre !). Ces trajectoires sont des *courbes homoclines*, comme nous les avions définies dans la Section 3.7. On voit tout de suite sur le portrait de phases du pendule le rôle qu'elles jouent pour structurer ce portrait. On rappelle que le long de ces trajectoires, le pendule ne parvient jamais à l'équilibre (cela violerait la règle de non-croisement de trajectoires) : il s'en approche quand $t \to +\infty$, et quand $t \to -\infty$.

En conclusion : outre les deux équilibres (stable $\theta = 0$, instable $\theta = \pi$), on lit sur le portrait de phases trois régimes hors équilibre,

la transition de l'un à l'autre se faisant en fonction de la valeur de l'énergie :

- si l'énergie est assez petite ($0 < \mathcal{E} < 2k$), le pendule oscille entre deux angles $-\theta_{\max}$ et $+\theta_{\max}$, le mouvement est périodique ;

- si l'énergie est assez grande ($\mathcal{E} > 2k$), le pendule tourne toujours dans la même direction (soit toujours dans le sens des aiguilles d'une montre, soit toujours dans le sens opposé), le mouvement est également périodique ;

- si l'énergie est critique ($\mathcal{E} = 2k$), le pendule converge vers la position d'équilibre instable, aussi bien quand $t \to +\infty$ que quand $t \to -\infty$; et cela lui prend un temps infini.

4.4. Étude des équilibres

Étudions le régime de petites oscillations au voisinage de l'équilibre stable. On réécrit l'équation dans l'espace des phases :

$$\dot\theta = \psi, \qquad \dot\psi = -k \sin\theta.$$

L'équilibre stable correspond à $\theta = 0$, $\psi = 0$.

Le théorème de régularité du flot (régularité par rapport à la condition initiale), et l'équation 24 établie dans la Section 2.3, montrent que la première variation du flot Φ est fournie par le flot linéarisé, c'est à dire le flot Φ^L de l'équation linéarisée. Plus précisément, pour tout intervalle $[-T, T]$ donné, on aura

$$\sup_{t\in[-T,T]} \left| \Phi_t^L(\theta_0, \dot\theta_0) - \Phi_t(\theta_0, \dot\theta_0) \right| = o(|\theta_0| + |\dot\theta_0|). \tag{69}$$

On a donc, pour de petites valeurs des oscillations et des temps pas trop longs, une très bonne approximation du flot non linéaire par le flot linéarisé. Calculons ce dernier : pour cela on linéarise $\sin\theta$ en θ, et on obtient

$$\dot\theta = \psi, \qquad \dot\psi = -k\theta.$$

Ou, de manière équivalente,

$$\ddot\theta = -k\theta. \tag{70}$$

(On peut linéariser puis passer dans l'espace des phases, ou passer d'abord dans l'espace des phases et ensuite linéariser, c'est équivalent.)

L'équation se met donc sous la forme matricielle $\dot z = Mz$ avec $z = (\theta, \psi)$, et

$$M = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix}.$$

Les valeurs propres sont $\pm i\sqrt{k}$: imaginaires pures, elles ne disent rien, en elles-mêmes, de la stabilité de l'équation non linéaire. Quant à l'équation linéaire, on peut la résoudre explicitement :

$$\theta^L(t) = \theta_0 \cos(\sqrt{k}t) + \dot{\theta}_0 \frac{\sin(\sqrt{k}t)}{\sqrt{k}}. \tag{71}$$

(Cette discussion peut paraître un peu pédante pour une équation si élémentaire ; mais elle permet de mettre en place une méthodologie qui est précieuse dans des problèmes plus complexes !)

En combinant (71) et (69), on conclut que, avec une très bonne approximation pour de petites valeurs de l'énergie, le pendule effectue des oscillations sinusoïdales régulières de période

$$T^* = \frac{2\pi}{\sqrt{k}}. \tag{72}$$

Cette conclusion est toute simple, mais pourtant contre-intuitive : on pourrait croire que dans un régime de très faibles oscillations la période du pendule est très brève ; or nous venons de montrer qu'elle est à peu près fixée et tend vers une valeur non nulle, qui ne dépend que du rapport $k = g/L$.

Nous avons ainsi retrouvé une autre observation expérimentale de Galilée : *Dans un régime de petites oscillations, la période des oscillations est presque indépendante de l'amplitude de ces oscillations.*

Remarques 89. 1. L'observation de Galilée, bien que motivée par la curiosité (et par des considérations philosophiques quelque peu obscures), n'avait rien d'anecdotique : ce fut le point de départ de tout un champ scientifico-technologique visant à mesurer le temps au moyen de systèmes oscillants. Apparemment Galilée a cru que la propriété d'isochronie restait vraie pour de grandes oscillations, alors qu'il n'en est rien. De fait, Huygens allait faire entrer ce domaine dans une phase décisive en calculant la courbe (non circulaire) qu'un pendule de longueur variable doit suivre pour que la période des oscillations soit *strictement* indépendante de l'amplitude. Avant Huygens les horloges avaient une précision relative d'environ 10^{-2} dans la mesure du temps ; grâce à lui on put obtenir 10^{-4} ; un siècle plus tard, en lointain développement, on parviendrait à la précision remarquable de 10^{-6}. (Aujourd'hui on obtient des précisions relatives hallucinantes de 10^{-16} en tirant partie des oscillations d'atomes froids...)

2. En pratique, les expériences de pendule oscillant sont faites dans un environnement (laboratoire, bâtiment) où le champ de gravité peut être considéré comme fixé ; la période d'oscillation dépend alors uniquement de la longueur du pendule. Cependant, c'est bien le rapport

g/L qui détermine la période, et on peut aussi, en principe, mesurer des variations de g en étudiant la période d'un pendule en fonction de l'endroit où il oscille. De fait, l'astronome Jean Richer avait remarqué en 1672 que la période d'oscillation d'un pendule est plus longue à Cayenne qu'à Paris, de sorte que le pendule retardait de 2,5 minutes par jour (soit un écart de moins de 1/500) ; Richer put corriger son pendule, qui mesurait environ 1m et battait la seconde, en le raccourcissant d'un peu plus de 3mm... Ces observations étaient le signe d'une variation du champ gravitationnel entre Paris et Cayenne ; elles permirent à Newton et Huygens de confirmer la théorie (développée par Newton) selon laquelle la Terre est légèrement aplatie aux pôles ! Il est remarquable que cet écart à la forme sphérique, de quelques millièmes seulement, puisse s'observer sur les oscillations d'un pendule...

3. La variation de la période en fonction de la longueur du pendule est à la base des jolies expériences d'"ondes pendulaires" (*pendulum waves*) où l'on observe les oscillations simultanées de pendules de longueurs variables ; voir par exemple

`www.youtube.com/watch?v=yVkdfJ9PkRQ` .

4. L'équation (70) est appelée l'équation de l'**oscillateur harmonique**. C'est aussi l'équation du mouvement d'une masse rappelée à l'origine par un ressort linéaire parfait. Elle joue un rôle majeur en physique et mathématique.

Passons maintenant à l'étude locale de l'équilibre instable : $(\pi, 0)$ dans l'espace des phases. Écrivons l'approximation linéaire :

$$\Phi_t(\theta_0, \dot\theta_0) = (\pi, 0) + \Phi_t^L(\theta_0 - \pi, \dot\theta_0) + o(|\theta_0 - \pi|) + o(|\dot\theta_0|).$$

Résolvons cette équation, qui porte sur les accroissements $\theta - \pi$ et $\dot\theta$. Si l'on pose $\phi = \theta - \pi$, $\psi = \dot\theta$, l'équation non linéaire se réécrit

$$\dot\phi = \dot\theta = \psi, \qquad \dot\psi = -k \sin\theta = -k \sin(\pi + \phi) = k \sin\phi,$$

et la linéarisation donne donc

$$\dot\phi = \psi, \qquad \dot\psi = k\phi.$$

La matrice associée est

$$\begin{pmatrix} 0 & 1 \\ k & 0 \end{pmatrix},$$

dont les valeurs propres sont $\sqrt{k}$ et $-\sqrt{k}$. Une valeur propre positive, une valeur propre négative : le linéarisé est instable, et admet une

direction stable et une direction instable. La solution de l'équation linéarisée s'écrit

$$\phi(t) = \phi_0 \cosh(\sqrt{k}t) + \psi_0 \frac{\sinh(\sqrt{k}t)}{\sqrt{k}}.$$

Au voisinage de l'équilibre instable, les courbes homoclines viennent former le dessin d'une courbe stable et d'une courbe instable (stable là où le mouvement converge vers l'équilibre, instable là où il s'en éloigne). Cela est en accord avec la conclusion du théorème de la variété stable.

L'analyse linéarisée donne des informations sur un intervalle de temps fixé ; pour aller au-delà, il faut se fier au portrait de phases. Par rapport à l'équilibre inversé (instable),

- soit on fait partir le système avec une énergie légèrement inférieure à l'énergie critique, et alors le pendule fera de grandes oscillations, s'éloignant autant que possible de l'équilibre avant de s'en rapprocher au bout d'une période, puis de s'en éloigner à nouveau, etc.

- soit on le fait partir avec une énergie lègèrement supérieure à l'énergie critique, et il tournera sans cesse, s'éloignant de l'équilibre avant d'y repasser, puis de s'en éloigner à nouveau, etc.

- soit on le fait partir avec *exactement* l'énergie critique, et on va alors, soit se rapprocher lentement de cet équilibre jusqu'à la fin des temps, soit s'en écarter complètement avant de s'en rapprocher. Dans les deux cas, on convergera vers l'équilibre en temps infini.

Notons que si l'on choisit la position du pendule "au hasard" au voisinage de l'équilibre instable, on se retrouvera toujours dans l'un deux premiers cas de figure : le scénario homocline où l'on tend vers l'équilibre en temps infini correspond à un ensemble de conditions initiales de mesure nulle. Il est d'ailleurs très délicat de le capturer numériquement !

4.5. Compléments : période ; pendule souple

Nous avons calculé la période du pendule dans le régime de petites oscillations ; mais qu'en est-il pour de grandes oscillations ? On suppose que l'énergie est inférieure à l'énergie critique, de sorte que le pendule effectue des oscillations de gauche à droite, de droite à gauche, etc. La solution est faite de deux branches de durée égale ; le temps pour aller de $-\theta_{\max}$ à $+\theta_{\max}$ est donc une demi-période, soit $P/2$. Repartant de la fin de la Section 4.2 on trouve que

$$\frac{P}{2} = \int_{-\theta_{\max}}^{\theta_{\max}} \frac{d\tau}{\sqrt{2\big[\mathcal{E} - k(1-\cos\tau)\big]}}.$$

En outre, le calcul montre que les valeurs $\pm\theta_{\max}$ sont celles qui annulent le dénominateur.

Il n'est pas du tout évident, sur l'expression précédente, que la période tend vers la période du système linéarisé ! Pour le retrouver, coupons l'intégrale en deux : par parité,

$$\frac{P}{4} = \int_0^{\theta_{\max}} \frac{d\tau}{\sqrt{2\big[\mathcal{E} - k(1-\cos\tau)\big]}}.$$

On effectue alors un changement de variables en choisissant tout ce qui est sous la racine carrée comme nouvelle inconnue :

$$s := 2\big[\mathcal{E} - k(1-\cos\tau)\big].$$

Ainsi $ds = -2k\sin\tau\, d\tau$; $s = 0$ en $\tau = \theta_{\max}$, $s = 2\mathcal{E}$ en $\tau = 0$ (les bornes sont inversées, ce qui est cohérent avec le fait que $ds/d\tau < 0$). En outre

$$\cos\tau = 1 - \left(\frac{\mathcal{E} - s/2}{k}\right), \qquad \sin\tau = \sqrt{1-\cos\tau^2}.$$

D'où

$$\frac{P}{4} = \frac{1}{2k}\int_0^{2\mathcal{E}} \frac{ds}{\sqrt{2s}\sqrt{\dfrac{2(\mathcal{E}-s/2)}{k} - \dfrac{(\mathcal{E}-s/2)^2}{k^2}}}.$$

Pour $\mathcal{E}$ petit, négligeons, au dénominateur, le terme $(\mathcal{E}-s/2)^2$ devant $(\mathcal{E}-s/2)$: on obtient

$$\frac{P}{4} \simeq \frac{1}{2k}\int_0^{2\mathcal{E}} \frac{ds}{\sqrt{2s}\sqrt{\dfrac{2(\mathcal{E}-s/2)}{k}}} = \frac{1}{2\sqrt{k}}\int_0^{2\mathcal{E}} \frac{ds}{\sqrt{s}\,\sqrt{2\mathcal{E}-s}}.$$

On commence à y voir plus clair : en effectuant le changement de variable $u = s/(2\mathcal{E})$, on voit que l'intégrale finale vaut

$$\frac{1}{2\sqrt{k}}\int_0^1 \frac{du}{\sqrt{u(1-u)}} = \frac{\pi}{2\sqrt{k}},$$

et on retrouve

$$P \simeq \frac{2\pi}{\sqrt{k}},$$

ce qui est cohérent avec l'étude de l'équation linéarisée !

En raffinant le calcul, on peut faire l'étude de la période à des ordres plus élevés. On peut aussi remplacer la fonction sinus par un développement limité d'ordre plus élevé en θ, et effectuer les calculs

sur ce système approché. Une formule populaire en la matière est la *formule de Borda* :

$$P \simeq \frac{2\pi}{\sqrt{k}} \left(1 + \frac{\theta_{\max}^2}{16}\right),$$

qui reste vraie à 3% près jusqu'à des amplitudes $\theta_{\max} \simeq \pi/2$.

On peut aussi montrer que la période est une fonction strictement croissante de l'énergie, jusqu'à diverger vers $+\infty$ quand on s'approche de l'énergie critique ; et ensuite, pour des valeurs de l'énergie plus grandes, c'est une fonction strictement décroissante de l'énergie, tendant vers 0 quand l'énergie tend vers l'infini.

En choisissant une énergie suffisamment petite, on peut garantir que la période est égale à la période T^* de l'équation linéarisée, disons à 1/1000 près. Cependant, même dans ce cas, l'accord entre l'équation linéarisée et l'équation non linéaire ne restera pas valable sur des intervalles de temps arbitrairement longs ! Supposons en effet que les deux systèmes partent en phase à partir de l'angle $\theta_{\max}$, et que la période P soit égale exactement à 1,001 fois la période T^* du linéarisé. Au bout d'un temps $500P$, le système non linéaire aura connu 500 oscillations, et le système linéarisé en aura connu 500,5 : les deux systèmes seront donc à cet instant en opposition de phase, avec l'un en position $\theta_{\max}$ et l'autre en position $-\theta_{\max}$. À cet instant, l'équation linéarisée ne fournira pas d'approximation plus précise que $\theta \simeq 0$! On retrouve l'avertissement déjà mentionné plusieurs fois : en général la linéarisation ne donne de résultats précis que sur un intervalle de temps fixé a priori, et l'étude de la stabilité non linéaire passe aussi par l'étude du portrait de phases complet.

Terminons par un développement bien distinct qui visera à réexaminer l'hypothèse de rigidité du pendule. Dans la "vraie vie", les pendules sont le plus souvent accrochés par un fil souple, et le fil peut très bien perdre sa rigidité (ne serait-ce que dans le cas le plus évident où l'on prépare le pendule au repos dans la position d'équilibre inversé). Cherchons à déterminer exactement quand se produit ce "déferlement" du pendule : il s'agit de trouver l'annulation de la tension. Jusqu'à présent nous n'avons utilisé que l'équation (63), mais maintenant nous pouvons employer l'équation compagnon (64) : la quantité $T = L\dot{\theta}^2 + g\cos\theta$ devient négative ou nulle si

$$\dot{\theta}^2 + k\cos\theta \leq 0.$$

Cela correspond, dans l'espace des phases, à l'intérieur de courbes qui ressemblent à des ellipses déformées, centrées autour de l'équilibre instable et de ses copies. Comment se comportent ces courbes ? Pour

$\dot\theta = 0$, on trouve $\theta = \pi/2$, soit une énergie $\mathcal{E} = k$. La valeur maximale admissible de $\dot\theta$ correspondra à $\cos\theta = -1$, soit $\dot\theta^2 = k$, soit $\dot\theta = \pm\sqrt{k}$, avec une énergie égale à $\mathcal{E} = 5k/2$.

En conclusion, avec un pendule "souple", on a trois régimes principaux :

- énergie $\mathcal{E} \leq k$: régime d'oscillations ;
- énergie $\mathcal{E} \geq 5k/2$: régime tournoyant ;
- énergie $\mathcal{E}$ comprise entre k et $5k/2$: régime où le pendule finit par s'écrouler, et où notre modèle ne suffira plus à décrire son évolution.

En particulier, les trajectoires homoclines (énergie égale à $2k$) ne s'observent pas avec un pendule souple ! Cela restreint leur intérêt pratique, mais il n'empêche que ces trajectoires sont des objets mathématiques remarquables.

On pourra utiliser les simulations numériques pour comparer le comportement du pendule rigide et du pendule souple.

4.6. Complément : Méthodes d'Aubry–Mather–Mañé

L'équation du pendule, $\ddot\theta = -k\sin\theta$, appartient à certaines classes importantes d'équations qui ont été extrêmement étudiées, et que nous reverrons dans le chapitre final du cours.

Pour commencer, c'est un **système lagrangien** : cela veut dire que ses trajectoires peuvent être reconstituées à l'aide d'un problème variationnel basé sur une fonction, définie sur l'espace des phases et appelée le lagrangien. En l'occurrence, un lagrangien pour le pendule est la fonction

$$L(\theta, \dot\theta) = \frac{\dot\theta^2}{2} - k(1 - \cos\theta), \tag{73}$$

le dernier terme, $V(\theta) = k(1-\cos\theta)$ étant l'énergie potentielle. Sur de petits intervalles de temps les trajectoires de ce système, entre un état initial θ_1 et un état final θ_2, sont les solutions du problème variationnel *de moindre action* :

$$\min\left\{\int_{t_1}^{t_2} L(\theta(s), \dot\theta(s))\, ds; \quad \theta(t_1) = \theta_1,\ \theta(t_2) = \theta_2\right\}.$$

La fonctionnelle ci-dessus, c'est à dire l'intégrale temporelle du lagrangien, est appelée une action.

Remarques 90. 1. Par extension, on appelle parfois systèmes lagrangiens des systèmes dont les trajectoires sont des points critiques de la fonctionnelle d'action, pas forcément des minimiseurs.

2. La restriction "pour des temps assez proches" est importante ici. Elle signifie qu'entre des temps t_1 et t_2 tels que $|t_2 - t_1| \leq \delta$, les solutions de l'équation du pendule sont solutions du problème ci-dessus ; mais elle ne dit rien pour des temps t_1 et t_2 quelconques ; en outre la taille de δ n'est pas connue a priori, elle dépend de la solution. Prenons un exemple : soit une trajectoire effectuant des révolutions de grande ampleur, passant par l'équilibre stable au temps 0 et par l'équilibre instable au temps T. La trajectoire la plus économique entre le temps 0 et le temps T consiste à effectuer une demi-révolution sur l'intervalle $[0, T]$, de sorte que la période soit égale à $P = 2T$. (En fait il y a deux trajectoires possibles : soit vers la droite, soit vers la gauche.) Mais la période d'un pendule tournoyant est une fonction strictement décroissante de l'énergie ; on peut donc trouver des solutions qui auront une période exactement $2T/3$. Alors, sur l'intervalle de temps $[0, T]$ le pendule aura le temps de faire trois demi-révolutions, et joindra encore l'équilibre stable à l'équilibre instable. Nous avons donc ainsi une solution alternative du système, qui ne minimise pas l'action sur l'intervalle de temps $[0, T]$; en revanche elle minimise l'action sur chacun des intervalles $[0, T/3]$, $[T/3, 2T/3]$ et $[2T/3, T]$.

Il existe une recette générale pour calculer l'équation différentielle associée à un principe de moindre action. Nous la justifierons plus tard, et pour cette discussion nous nous contenterons de l'énoncer : à un lagrangien $L(\theta, \dot{\theta})$ correspond l'équation différentielle

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = \frac{\partial L}{\partial \theta}. \tag{74}$$

Dans notre exemple, $\partial L/\partial\dot{\theta}$ est simplement $\dot{\theta}$, donc le membre de gauche de (74) vaut $\ddot{\theta}$; quand au membre de droite c'est $\partial L/\partial\theta$ qui vaut $-k \sin\theta$. On retrouve donc effectivement l'équation du pendule.

Le pendule est aussi un **système hamiltonien** : si l'on pose

$$H(\theta, p) = \frac{p^2}{2} + V(\theta), \tag{75}$$

alors l'équation se réécrit

$$\begin{cases} \dot{\theta} = \dfrac{\partial H}{\partial p} \\ \dot{p} = -\dfrac{\partial H}{\partial \theta}. \end{cases}$$

On remarque que, sous l'identification $p = \dot{\theta}$, la fonction H n'est autre que l'énergie totale du système, dont on sait qu'elle est conservée. En fait la plupart des systèmes lagrangiens peuvent aussi se réécrire

comme des systèmes hamiltoniens, modulo une recette simple que nous étudierons plus tard. Les systèmes hamiltoniens constituent une classe très générale et importante d'équations différentielles conservatives, source d'innombrables travaux. De nombreux auteurs ont travaillé à mettre au point des théories visant à mieux comprendre le comportement des systèmes hamiltoniens : parmi eux, citons Serge Aubry, John Mather (les pères de la célèbre "théorie d'Aubry–Mather") et Ricardo Mañé. Voyons comment ces méthodes élégantes se comportent dans le cas particulier du pendule.

La *méthode de Mather* consiste à chercher des mesures invariantes vérifiant une propriété de minimisation : ainsi on cherchera à résoudre

$$c = \min \left\{ \iint L(\theta, \dot\theta)\, \mu(d\theta\, d\dot\theta); \quad \mu \text{ mesure invariante} \right\}. \tag{76}$$

Ici μ est cherchée parmi les mesures de probabilité invariantes, c'est à dire qu'elle doit rester inchangée par l'action du flot Φ_t :

Définition 91 (mesure invariante). Étant donné un flot Φ_t défini sur un espace X, on appelle mesure Φ-invariante une mesure μ vérifiant

$$\int \zeta \circ \Phi_t \, d\mu = \int \zeta \, d\mu.$$

pour toute fonction test ζ, μ-intégrable.

Il faut imaginer que si μ est un assemblage de grains de matière dans l'espace des phases, se déplaçant selon le flot, alors μ est inchangée au cours du temps. La plus simple des mesures de probabilité invariantes est une masse de Dirac portée par un équilibre ; et juste un peu plus compliquée, la mesure uniforme μ_γ portée par une trajectoire fermée γ :

$$\int \zeta \, d\mu_\gamma = \frac{1}{T} \int_0^T \zeta(\gamma(t), \dot\gamma(t))\, dt,$$

où T est la période de la trajectoire. Une mesure de probabilité invariante est donc une sorte de généralisation des trajectoires fermées ; elle donne des indications précieuses sur la dynamique.

Plus une mesure invariante est concentrée, plus elle est utile : elle permet ainsi de bien distinguer certaines zones de l'espace de phases. Le problème de minimisation de Mather est une façon de sélectionner des mesures "assez concentrées". On appelle ensemble de Mather l'adhérence de l'union de tous les supports de mesures de probabilité invariantes minimisantes :

$$\mathcal{M} = \overline{\left\{ \bigcup \mathrm{Spt}\mu;\ \mu \text{ solution de (76).} \right\}}$$

Si on applique la définition au pendule, on trouve que cet ensemble est réduit à un point : l'équilibre instable $(\pi, 0)$. Mather démontre en fait que pour de larges classes de systèmes lagrangiens, l'ensemble $\mathcal{M}$ est un *graphe lipschitzien*, donc quelque chose d'assez concentré ! En dimension 2, c'est donc soit un point, soit une courbe.

Très liée à la méthode de Mather, et en quelque sorte duale, nous avons la *méthode de Mañé*, qui consiste à étudier les solutions de l'équation aux dérivées partielles de Hamilton–Jacobi :

$$H(x, \nabla\psi) = h, \tag{77}$$

où h est un nombre réel. Il peut paraître surprenant que l'on cherche à résoudre une EDO au moyen d'EDP, a priori bien plus complexes ; pourtant c'est une approche efficace, également employée en physique pour divers calculs explicites [**23**].

Dans l'équation (77), il faut a priori considérer toutes les solutions, et elles peuvent dépendre de h... cependant, on peut montrer que cette équation n'est soluble que pour (au plus) une seule valeur de h ! On appelle cette valeur le hamiltonien effectif ; il se trouve que c'est aussi la valeur de c dans le problème (76). On définit alors l'ensemble d'Aubry comme l'intersection de tous les graphes des gradients des solutions de cette équation :

$$\mathcal{A} = \bigcap \big\{ \text{Graphe}(\nabla\psi); \quad \psi \text{ solution de (77)} \big\}.$$

Cet ensemble est également très intéressant pour l'étude de systèmes dynamiques un peu généraux ; quand on applique cette méthode au pendule, on retrouve la position d'équilibre instable.

Même en ayant conscience de la simplicité du système sur lequel nous les testons, c'est un peu décevant de ne retrouver que l'équilibre instable avec ces méthodes ingénieuses. Pour aller plus loin dans l'analyse, on peut faire varier le lagrangien : en effet, une équation donnée peut être associée à divers lagrangiens. Soit ω un nombre réel quelconque, posons

$$L_\omega(\theta, \dot\theta) = \frac{\dot\theta^2}{2} + \omega\,\dot\theta - k(1 - \cos\theta).$$

Alors l'équation découlant de L_ω par (74) est

$$\frac{d}{dt}(\dot\theta + \omega) = -k \sin\theta,$$

ce qui revient à $\ddot\theta = -k \sin\theta$. L'équation est donc inchangée, mais les ensembles d'Aubry et de Mather vont changer ! En fait il existe une valeur critique ω_c telle que

- si $|\omega| < \omega_c$, alors $\mathcal{M} = \mathcal{A}$ correspond à l'équilibre instable ;

- si $|\omega| > \omega_c$, alors $\mathcal{M} = \mathcal{A}$ est concentré sur une révolution du pendule ;

- si $|\omega| = \omega_c$, alors $\mathcal{M}$ est l'équilibre instable, et $\mathcal{A}$ la réunion de l'équilibre instable et des trajectoires homoclines (ce qui peut, au passage, être vu comme l'union de deux graphes lipschitziens).

Nous voyons ainsi qu'en appliquant la théorie générale d'Aubry–Mather–Mañé et en tirant parti des symétries du lagrangien, nous avons pu retrouver les caractéristiques les plus notables de l'équation du pendule rigide : les trajectoires homoclines et la transition entre le régime d'oscillations et le régime de révolutions. Dans des situations plus complexes où le portrait de phases est difficile à explorer, ces méthodes seront capables de détecter certaines structures remarquables.

Ce domaine — l'exploration de la dynamique via des problèmes de minimisation et des équations aux dérivées partielles — est le sujet de la théorie "KAM faible", où les solutions des équations de Hamilton–Jacobi jouent un rôle clé. Cette théorie est truffée de chausse-trappes conceptuelles : par exemple il faut être vigilant au sens que l'on donne à l'équation (77) car ses solutions ne sont en général pas différentiables... Elle a été développée au cours des vingt dernières années par Albert Fathi ; au-delà de son ouvrage technique [**11**], on pourra consulter les introductions de Karl Siburg [**33**] et du second auteur [**36**, fin du Chapitre 8].

CHAPITRE 5

Cycles

Ce chapitre est consacré à des phénomènes *cycliques* apparaissant dans les équations différentielles. Un cycle, c'est un phénomène qui se reproduit à l'identique, ou quasiment à l'identique, périodiquement. Ce qui nous intéressera particulièrement, c'est la situation dans laquelle une équation différentielle amène spontanément un système à évoluer sous forme de cycles, éventuellement après une période transitoire.

Avec les équilibres, les cycles sont parmi les phénomènes les plus importants que l'on puisse identifier dans le portrait de phases d'une équation différentielle. Ils sont cependant difficiles à capturer numériquement : comment détecter, malgré les erreurs numériques, que la solution est revenue exactement au point de départ ? Outre la précision de calcul, cela nécessite souvent l'usage de méthodes bien choisies.

Nous allons examiner quelques exemples emblématiques de cycles, issus d'équations qui apparaissent dans des branches aussi diverses que l'écologie et l'électronique.

5.1. Le modèle de Lotka–Volterra

Le système de Lotka–Volterra, aussi appelé système proie-prédateur, est l'une des plus célèbres équations différentielles qui soient. Il date des années 1910–1920, à l'époque où l'on s'efforçait de développer l'étude mathématique des phénomènes chimiques et biologiques, et d'aller au-delà des modèles les plus élémentaires de réactions chimiques ou de reproduction des espèces. Lotka avait travaillé sur la dynamique de réactions chimiques autocatalytiques ; quant à Volterra, qui était déjà un mathématicien reconnu, il souhaitait expliquer qualitativement les fluctuations de stocks de poissons dans la Mer Adriatique. L'analyse de Volterra était partie d'une observation paradoxale : pendant la première guerre mondiale, malgré la réduction de l'effort de pêche, on avait assisté à la décroissance de certains effectifs de poissons bons pour consommation.

Pour établir le modèle, considérons des populations de proies et de prédateurs. Pour fixer les idées, selon un usage assez bien établi, nous supposerons que ce sont des lapins et des renards. Appelons x

l'effectif de proies (lapins), et y l'effectif de prédateurs (renards). Les unités ne sont pas bien spécifiées ici, mais on va supposer que les populations sont assez nombreuses pour que l'on puisse appliquer un formalisme continu : par exemple, si la population est mesuré en milliers d'individus, la valeur de x et y pourra être fractionnaire, et en fait on considérera ces quantités comme des variables réelles. Quelles équations écrire alors dans les variables x et y pour modéliser les interactions entre proies et prédateurs ?

Des observations innombrables montrent que (a) si l'on éradique le prédateur, les proies se mettent à pulluler ; (b) si l'on éradique les proies, les prédateurs eux-mêmes vont voir leur population décliner. Pour tenir compte de cela, nous supposerons que

- en l'absence de prédateurs, les proies croissent selon un taux de reproduction $\alpha > 0$;

- en l'absence de proies, les prédateurs déclinent selon un taux de mortalité $\delta > 0$.

Passons maintenant à l'interaction entre les deux. Les renards chassent les lapins et en bénéficient ; la quantité de prédation est proportionnelle au nombre de rencontres proies-prédateurs ; si ces rencontres sont dictées par le hasard et réparties de manière uniforme dans l'aire géographique considérée, ce nombre de rencontres est lui-même proportionnel au produit xy.

Enfin cette biomasse consommée se retrouve au niveau de la population de prédateurs. Bien sûr, il y a une déperdition, mais on peut modéliser cela avec un nouveau coefficient.

On en arrive ainsi à la modélisation simple :

$$\dot{x} = \alpha\, x - \beta\, xy, \qquad \dot{y} = \gamma\, xy - \delta\, y. \tag{78}$$

On note le rôle des quatre coefficients, tous strictement positifs : reproduction des lapins (α) ; mortalité des proies par prédation (β) ; croissance des prédateurs par prédation (γ) ; mortalité des prédateurs (δ).

La nature simpliste de ces équations saute aux yeux : on a oublié le taux de reproduction des prédateurs, le taux de mortalité naturelle pour les proies, les effets coopératifs, les ressources environnementales... Également on peut protester contre l'absence de prise en compte des ressources de l'environnement (les lapins aussi ont besoin de se nourrir, et s'ils sont trop nombreux cela pourra être un problème), etc. En outre il y a souvent, dans la nature, une compétition acharnée entre plusieurs espèces de prédateurs qui se disputent l'accès aux proies ; et les proies sont elles-mêmes en compétition pour l'accès aux ressources naturelles.

On renvoie au cours de Steve Baigent [**4**] pour de nombreuses variantes plus sophistiquées et certains effets subtils.

Mais pour simple qu'il soit, le système proie–prédateur mérite toute notre attention. D'abord il mène à des comportements intéressants, et ses prédictions ont été qualitativement vérifiées dans des données réelles de biologie des populations. Ensuite, on le retrouve, sous une forme ou une autre, dans de nombreux phénomènes, qui vont de la biologie aux algorithmes de finance mathématique moderne en passant par l'économie... et la chimie, puisque des réactions chimiques auto-oscillantes ont effectivement été découvertes dans la nature (réaction de Belusov-Zhabotinski, qui a d'abord suscité un scepticisme considérable !). C'est donc l'un des quelques modèles types qu'il faut absolument connaître quand on s'intéresse aux équations d'évolution.

5.2. Résolution du système proie–prédateur

Oublions maintenant la modélisation, et concentrons-nous sur l'étude du système mathématique à deux inconnues

$$\dot{x} = \alpha x - \beta\, xy = x(\alpha - \beta y), \qquad \dot{y} = \gamma\, xy - \delta y = y(\gamma x - \delta). \tag{79}$$

A priori x, y sont des nombres réels ; cela dit, ils n'auront de signification, en tant qu'effectifs de population, que s'ils sont positifs !

Le système (79) est du premier degré, et donc l'espace des phases est simplement l'espace à deux dimensions des couples (x, y) : chaque point de l'espace des phases donne à la fois l'effectif des lapins, et l'effectif des renards.

Après quelques essais on découvre qu'il y a une loi de conservation associée au système : si l'on pose

$$I(x, y) = \beta y + \gamma x - \alpha \ln y - \delta \ln x, \tag{80}$$

alors le long des trajectoires du système,

$$\begin{aligned} \frac{dI}{dt} &= \beta\dot{y} + \gamma\dot{x} - \frac{\alpha\,\dot{y}}{y} - \frac{\delta\,\dot{x}}{x} \\ &= \beta\,(\gamma\, xy - \delta\, y) + \gamma\,(\alpha\, x - \beta\, xy) - \alpha\,(\gamma\, x - \delta) - \delta\,(\alpha - \beta\, y) \\ &= 0. \end{aligned}$$

La fonction I est donc un invariant de l'équation ! On en déduit un découpage de l'espace des phases, le mouvement ayant lieu dans les lignes de niveau de I. On en déduit aussi une estimation a priori sur les solutions : en effet, si x ou y tend vers l'infini, alors I tend vers l'infini ; par contraposée, si I est borné cela confine la solution dans un espace

borné. Par critère de compacité, on a donc existence d'un flot global pour le modèle proie–prédateur.

Il est facile d'analyser l'équation pour trouver les équilibres : la première équation donne $x = 0$ ou $y = \alpha/\beta$; si $x = 0$ alors forcément la deuxième équation donne $y = 0$. Si en revanche $y = \alpha/\beta$, alors la deuxième équation donne $x = \delta/\gamma$. Nous avons donc deux équilibres : $(0, 0)$ et $(\delta/\gamma, \alpha/\beta)$.

Cherchons également des solutions particulières : si $x = 0$ on a $\dot{y} = -\delta y$, et si $y = 0$ on a $\dot{x} = \alpha x$. Cela correspond aux situations que nous avons envisagées au début de la modélisation : une population de renards sans lapins, qui meurt de faim ; une population de lapins sans renards, qui explose. Ces trajectoires particulières correspondent au axes $(x = 0)$ et $(y = 0)$.

Dans l'espace des phases, l'équation est autonome et du premier degré ; par Cauchy–Lipschitz, les trajectoires ne peuvent jamais se croiser. On en déduit que x et y restent strictement positifs s'ils le sont au départ : en effet, si une trajectoire passait par $(x = 0)$ ou $(y = 0)$ elle croiserait les solutions particulières portées par les axes.

On peut aussi déduire la positivité stricte de la loi de conservation I : si $x = 0$ ou $y = 0$ cela correspond, formellement, à $I = \infty$. Une valeur finie de I exclut donc des valeurs nulles de x ou de y.

En conclusion temporaire, le modèle proie–prédateur ne fournit pas de conclusions "évidemment absurdes" telles que des populations négatives !

Ensuite, pour tracer le portrait de phases, le plus simple est de tirer parti de la loi de conservation pour restreindre le champ des possibles. On peut tracer les lignes de niveau de I, et l'on sait que le mouvement a lieu dans ces lignes de niveau : on voit ainsi que la solution effectue des *cycles* où les populations respectives de proies et de prédateurs oscillent entre des valeurs tour à tour faibles et fortes ; et aussi qu'il y a en gros décalage de phase de $\pi/2$ entre les variations des proies et les variations des prédateurs. Cette conclusion est très importante : la population des proies ne répond pas "en temps réel" à l'évolution de la population des prédateurs, il y a un décalage temporel. Cela est bien observée dans des données réelles. Il y a donc alternance de différentes phases qui se chevauchent : augmentation du nombre de lapins ; augmentation du nombre de renards ; diminution du nombre de lapins ; diminution du nombre de renards ; etc.

On note qu'avec ce modèle *il n'y a pas d'approche d'un équilibre* ; ce serait le cas avec d'autres modèles de dynamique des populations.

Cependant, on peut répéter que pour certaines interactions proies–prédateurs, les cycles s'observent effectivement dans la nature ; l'absence de convergence vers un équilibre ne doit donc pas être considérée en soi comme un défaut du modèle.

D'autres failles du modèle, en revanche, apparaissent dans les conclusions. Quelles que soient les valeurs des paramètres $\alpha, \beta, \gamma, \delta$ on a abouti à des cycles limites, ce que l'on peut considérer comme une forme d'"équilibre dynamique" ; alors que dans "la vraie vie", on s'attend à la disparition pure et simple des proies si le taux de reproduction est trop faible par rapport au taux de prédation.

Pour conclure ce bref aperçu, nous allons présenter une variante du système proie–prédateur, décrivant de manière à la fois abstraite et générale une situation de compétition. Cette discussion est tirée de l'ouvrage de Hirsch & Smale [**20**]. On note x et y deux populations de prédateurs en compétition pour l'accès aux proies ; on écrit alors

$$\dot{x} = Q(x, y)\, x; \qquad \dot{y} = R(x, y)\, y,$$

où Q et R sont les taux de croissance effectifs de ces espèces. On pose les hypothèses suivantes :

(i) si l'une des espèces croît, le taux de croissance décroît (cela traduit le fait que les deux espèces se gênent mutuellement). En d'autres termes,

$$\frac{\partial Q}{\partial y} < 0, \qquad \frac{\partial R}{\partial x} < 0.$$

(ii) Si x ou y est très importante, aucune espèce ne pourra se multiplier car les ressources seront insuffisantes : on suppose donc l'existence d'un seuil $K > 0$ tel que si $x > K$ ou $y > K$ alors Q et R sont tous deux négatifs ou nuls.

(iii) Si une seule des espèces est présente, sa croissance est positive jusqu'à une certaine taille de population, et négative au-delà. Il existe donc des seuils $a > 0$, $b > 0$ tels que

$$x < a \Longrightarrow Q(x, 0) > 0; \qquad x > b \Longrightarrow Q(x, 0) < 0;$$

$$y < a \Longrightarrow R(0, y) > 0; \qquad y > b \Longrightarrow R(0, y) < 0.$$

Ces trois hypothèses sont très générales, mais suffisent à étudier la stabilité des points critiques et à tracer des portraits de phases qualitatifs, comme on pourra le voir dans Hirsch & Smale [**20**].

5.3. Équation de Van der Pol

D'innombrables systèmes complexes, vivants ou artificiels, reposent sur des variations de courant et de potentiel électrique. On peut penser à un cerveau comme à un circuit d'ordinateur. Ces phénomènes se modélisent aussi par des équations différentielles.

Dans ce domaine, on peut citer en tout premier le célèbre circuit RLC, cas d'école bien connu des cours d'électricité. Il comprend

- une résistance (ou résistor) : c'est un composant électrique qui dissipe de l'énergie ; la différence de potentiel à ses bornes est proportionnelle au courant électrique, selon la loi d'Ohm $U = RI$ (R la valeur de la résistance). Cette loi simple est en fait une approximation d'une loi "légèrement" non linéaire.

- une inductance, c'est à dire une bobine qui, traversée par un courant électrique, engendre un champ magnétique qui a tendance à "contrer" l'action du courant ; la loi correspondante est alors $U = L(dI/dt)$, où L est la valeur de l'inductance.

- un condensateur, qui stocke de l'énergie électrostatique, selon l'équation $U = Q/C$, où Q est la charge électrique emmagasinée et C la capacité du condensateur. Le courant électrique I est égal à la variation temporelle de la charge ($dQ/dt = I$).

Dans le circuit RLC, les trois composants sont montés en série, et on impose une différence de potentiel E aux bornes de l'ensemble, que l'on supposera constante. L'équation est alors

$$E = RI + L\frac{dI}{dt} + \frac{Q}{C},$$

que l'on peut dériver par rapport au temps pour trouver

$$0 = R\,\frac{dI}{dt} + L\,\frac{d^2I}{dt^2} + \frac{I}{C}.$$

Soit encore

$$\ddot{I} + \left(\frac{R}{L}\right)\dot{I} + \frac{I}{LC} = 0. \tag{81}$$

Si $R = 0$ c'est l'équation d'un oscillateur harmonique ; elle montre que le courant subit des *oscillations régulières*, sinusoïdales, de période $T = 2\pi\sqrt{LC}$. Le courant passe dans un sens, stockant de l'énergie électrostatique dans le condensateur, et l'inductance agit comme une "force contraire" qui finit par faire repartir le courant dans le sens opposé.

Si $R > 0$, se superpose à cette oscillation régulière une déperdition de courant due à l'action de la résistance ; on obtient donc des *oscillations amorties*, exactement comme dans le pendule linéaire avec frottement.

Nous allons maintenant considérer un modèle plus subtil. Dans les années 1920, l'ingénieur électricien Balthasar Van der Pol considérait un circuit où la résistance R (dite "passive" car son effet est toujours proportionnel au courant) est remplacée par un constituant "actif" qui réagit de manière différente en fonction du courant, et "privilégie" certaines valeurs de l'intensité. Ce composant peut, en pratique, être obtenu par un assemblage de diodes, ou un composant semiconducteur alimenté par une source de courant externe... La relation $U = RI$ sera ainsi remplacée par $U = \phi(I)$, où ϕ est une fonction non linéaire.

Considérons donc une non-linéarité cubique de la forme

$$\phi(I) = \lambda I(I^2 - a), \qquad \lambda > 0, \quad a > 0.$$

Ainsi le composant exerce une forte résistance quand le courant est grand ; aucune résistance quand le courant a une intensité $\sqrt{a}$; et une résistance négative (tendant à accroître le courant) quand l'intensité est petite. On note que la fonction ϕ est impaire, ce qui traduit le fait que le composant est insensible au sens dans lequel passe le courant.

Avec un tel composé l'équation de la tension devient

$$E = \lambda I(I^2 - a) + L\,\frac{dI}{dt} + \frac{Q}{C},$$

et quand on dérive on trouve

$$0 = 3\lambda I^2 \dot{I} - a\lambda \dot{I} + L\ddot{I} + \frac{I}{C},$$

soit

$$LC\,\ddot{I} - C\lambda(a - 3I^2)\dot{I} + I = 0.$$

Réduisons le nombre de paramètres par un choix soigneux des unités de référence pour le temps et le courant électrique : on pose

$$x(t) = \frac{I(\nu t)}{I_0},$$

où ν est un paramètre à déterminer. On trouve, en factorisant par I_0,

$$\left(\frac{LC}{\nu^2}\right)\ddot{x} - \left(\frac{C\lambda}{\nu}\right)(a - 3I_0^2 x^2)\,\dot{x} + x = 0.$$

Choisissons $I_0 = \sqrt{a/3}$ et $\nu = \sqrt{LC}$: cela revient à se placer dans l'échelle de temps des oscillations spontanées du système LC (non

amorti). L'équation devient

$$\ddot{x} - \left(\frac{C\lambda a}{\sqrt{LC}}\right)(1-x^2)\,\dot{x} + x = 0.$$

Cette équation ne dépend plus que du paramètre $\mu = C\lambda a/\sqrt{LC}$, et nous avons finalement abouti à l'**équation de Van der Pol** :

$$\ddot{x} - \mu(1-x^2)\,\dot{x} + x = 0. \tag{82}$$

Ici $\mu > 0$ est un paramètre strictement positif, et x représente donc l'intensité du courant dans un circuit comportant un composant actif "cubique", une inductance et un condensateur.

L'équation (82) a acquis une grande popularité dans la modélisation de certaines classes de phénomènes non linéaires. Avant d'approfondir son étude, essayons de deviner le comportement de ses solutions. La symétrie du système, et le fait que $(x = 0)$ est évidemment un équilibre, nous incitent à prendre garde à la position de l'équilibre par rapport à 0.

- si x est petit (proche de 0), alors le terme en μ est à peu près $-\mu\dot{x}$; cela correspond à un terme d'amortissement linéaire... avec le mauvais signe ! Comme une friction qui apporterait de l'énergie au système au lieu de lui en extraire. On s'attend donc que les petites oscillations soient amplifiées par le système.

- si x est grand, alors le terme en μ est à peu près $(\mu x^2)\dot{x}$, ce qui correspond bien à une friction (le signe est bon cette fois !), mais extrêmement forte (car μx^2 est très grand), et d'autant plus forte que x est grand. On s'attend donc à ce que les mouvements de grande amplitude soient très vite amortis.

Une première conclusion : si le système est amorti extrêmement vite pour de grandes valeurs de x, on s'attend à ce qu'il soit impossible au système de partir à l'infini en temps fini, et donc que les solutions soient globalement définies (pour tous les temps).

Par ailleurs, si les petites oscillations sont spontanément amplifiées et que les grandes oscillations sont spontanément atténuées, peut-être existe-t-il une taille d'oscillation qui se met en place d'elle-même ?

5.4. Cycle limite de l'équation de Van der Pol

Commençons maintenant à tracer le portrait de phases de l'équation de Van der Pol (82). Dans l'espace des phases, l'équation se réécrit

$$\begin{cases} \dot{x} = y \\ \dot{y} = \mu\,(1-x^2)y - x. \end{cases} \tag{83}$$

On peut noter cela de manière compacte $\dot{z} = f(z)$, avec $z = (x, y)$ et $f(x, y) = (y, \mu(1 - x^2)y - x)$.

On constate immédiatement qu'il n'y a qu'un seul équilibre : $x = 0, y = 0$.

Étudions la dynamique près de l'équilibre : la linéarisation fournit

$$\ddot{x} = -x + \mu\dot{x}$$

dans l'espace des positions, ou de manière équivalente $\dot{z} = df(0, 0)z$, l'application linéaire $df(0, 0)$ se représentant sous forme matricielle

$$\begin{pmatrix} 0 & 1 \\ -1 & \mu \end{pmatrix}.$$

La trace vaut μ, le déterminant vaut 1 : les valeurs propres valent donc λ et λ^{-1}, avec $\lambda + \lambda^{-1} = \mu$. D'où l'on distingue trois cas en fonction de μ :

- si $\mu = 2$, on a seulement la valeur propre 1 ;

- si $\mu > 2$, il y a deux valeurs propres réelles distinctes : $\mu/2 \pm \sqrt{\mu^2 - 4}/2$;

- si $0 < \mu < 2$, il y a deux valeurs propres complexes distinctes : $\mu/2 \pm i\sqrt{4 - \mu^2}/2$; et la partie réelle vaut alors $\mu/2$.

Dans tous les cas, les deux valeurs propres ont leur partie réelle strictement positive, ce qui correspond à une instabilité, et même un point source. Partant d'une valeur proche de l'équilibre (mais différente de l'équilibre), on a tendance à s'en éloigner en se déplaçant dans l'espace des phases par une spirale tournant dans le sens des aiguilles d'une montre.

On peut aussi tracer le portrait des phases, de manière qualitative, pour de grandes valeurs de $|x|$. Le champ de vecteurs f dans l'espace des phases peut être examiné pour différentes valeurs particulières :

- sur l'axe des abscisses, il vaut $(0, -x)$, un profil linéaire vertical (pointant vers le bas dans la partie droite de l'espace des phases, vers le haut dans la partie gauche)

- sur l'axe des ordonnées, il vaut $(y, \mu y)$, un profil linéaire oblique (pointant en haut à droite dans la partie droite, en bas à gauche dans la partie gauche)

- sur l'axe $y = x$, on trouve à peu près $(x, -\mu x^3)$, qui pointe très fortement vers l'axe des abscisses, et légèrement en oblique ;

- sur l'axe $y = -x$, on trouve à peu près $(-x, \mu x^3)$, qui pointe très fortement vers l'axe des abscisses, et légèrement en oblique.

Tout cela suggère un système qui, partant loin de l'origine, s'approche rapidement de l'axe des abscisses, et a tendance à tournoyer

autour de l'origine selon une trajectoire alternativement montante et descendante.

Comment connecter les deux comportements – pour $|x|$ grand et $|x|$ petit ?

Le théorème de Poincaré–Bendixson recense les comportements asymptotiques possibles pour ce système de dimension 2 : soit des équilibres, soit des cycles, soit des trajectoires reliant des équilibres. Comme nous n'avons ici qu'un équilibre, et qu'il est un point source, le comportement asymptotique partant de n'importe quelle autre position de l'espace des phases ne peut être décrit que par un cycle limite.

De fait, la propriété la plus remarquable de l'équation de Van der Pol est d'admettre un *unique cycle limite* qui attire toutes les solutions (à la seule exception de l'équlibre !). Ce cycle limite n'a pas d'équation explicite, mais nous allons tâcher de le cerner sans faire appel au marteau-pilon du théorème de Poincaré–Bendixson. Notre discussion suivra celle présentée dans le cours de Knill [**21**], elle-même inspirée d'autres sources.

Quand le système est placé (dans l'espace des phases) au-dessus de l'axe des abscisses, x augmente ; et quand il est placé en-dessous de cet axe, x diminue. Un cycle limite alterne les phases d'augmentation et de diminution de x, et doit donc nécessairement être en partie au-dessus et en partie en-dessous de l'axe des abscisses ; et bien sûr, il doit croiser cet axe, parfois en descendant et parfois en montant. L'examen du champ de vecteurs sur l'axe des abscisses montre qu'un croisement descendant n'est possible qu'à droite de l'axe des ordonnées, et qu'un croisement montant n'est possible qu'à gauche de ce même axe. On en conclut qu'un cycle tourne forcément autour de l'origine, traversant successivement les zones $(x > 0, y > 0)$, $(x > 0, y < 0)$, $(x < 0, y < 0)$ et $(x < 0, y > 0)$.

Pour trouver un cycle, on peut donc partir d'un certain $(0, y_0)$ avec $y_0 > 0$, suivre le système, et vérifier que l'on revient à $(0, y_0)$ au bout d'un certain temps.

On peut simplifier le problème davantage, en remarquant qu'un cycle est forcément à symétrie centrale autour de 0 : cela découle de la symétrie impaire du champ de vecteurs $f(x, y) = (y, \mu(1 - x)^2 y - x)$. En effet, cette symétrie impaire montre que si $\mathcal{C}$ est un cycle, alors $-\mathcal{C}$ en est un aussi (c'est une courbe fermée bien sûr, et par imparité elle vérifie aussi les équations). Mais si $\mathcal{C}$ est une courbe tournant autour de l'origine, non symétrique par rapport à l'origine, alors un argument de continuité montre que $\mathcal{C}$ et $-\mathcal{C}$ ont au moins deux intersections. Comme ce sont les trajectoires d'un champ de vecteurs autonome, il s'ensuit

(par Cauchy–Lipschitz) que $\mathcal{C} = -\mathcal{C}$, ce qui contredit l'hypothèse de non-symétrie.

Si $\mathcal{C}$ est un cycle pour l'équation de Van der Pol, partant de $(0, y_0)$, il viendra donc intersecter à nouveau l'axe vertical en $(0, -y_0)$. Et réciproquement, si l'on a une courbe qui part de $(0, y_0)$ pour aboutir en $(0, -y_0)$, sa continuation reviendra, par imparité, en $(0, y_0)$ et l'on aura donc un cycle.

Récapitulons : trouver un cycle est équivalent à trouver une trajectoire qui, issue d'un point $(0, y_0)$, avec $y_0 > 0$, vient couper à nouveau l'axe vertical en $(0, -y_0)$.

Pour avancer, nous aurons besoin de tirer parti de la forme de l'équation. On rappelle que dans (82) les deux premiers termes $\ddot{x} - \mu(1-x^2)\dot{x}$ forment une dérivée temporelle (c'est bien comme cela qu'on les a obtenus !). Posons

$$\phi(x) = \mu\left(\frac{x^3}{3} - x\right), \tag{84}$$

de sorte que $\phi'(x) = -\mu(1 - x^2)$. L'équation (82) se réécrit alors

$$\frac{d}{dt}(\dot{x} + \phi(x)) + x = 0. \tag{85}$$

Sans le terme $\phi(x)$ on aurait l'équation de l'oscillateur harmonique, $\ddot{x} + x = 0$; on sait que pour cette dernière on a une fonction d'énergie naturelle, $(\dot{x}^2 + x^2)/2$. Une méthode astucieuse due à Lienard s'inspire de cette analogie pour définir une "énergie", en remplaçant le terme en $\dot{x}$ par ce qui est "dans la dérivée" en (85) :

$$E(x, \dot{x}) = \frac{\big(\dot{x} + \phi(x)\big)^2 + x^2}{2}. \tag{86}$$

Une première remarque est que $E(0, y_0) = y_0^2/2$, de sorte qu'il est équivalent de démontrer que le système recoupe l'axe vertical en $-y_0$, ou de montrer que la valeur de E est la même au départ et en ce nouveau point d'intersection.

Calculons la dérivée temporelle de E le long du flot de Van der Pol : en utilisant (86) et (85) on trouve

$$\begin{aligned} \frac{dE}{dt} &= (\dot{x} + \phi(x))\,\frac{d}{dt}(\dot{x} + \phi(x)) + x\dot{x} \\ &= (\dot{x} + \phi(x))(-x) + x\dot{x} \\ &= -x\phi(x). \end{aligned} \tag{87}$$

Cette dérivée ne dépend donc que de x !

La fonction $-x\phi(x) = x^2(1 - x^2/3)$ présente une double bosse symétrique, qui s'annule en 0 et en $\pm\sqrt{3}$, et qui est maximale en ± 1 où elle vaut $2/3$. Au vu de (87), on en déduit le comportement de $E(t) := E(x(t), \dot{x}(t))$ en fonction de $x(t)$:

- tant que $0 < |x(t)| < \sqrt{3}$, la fonction $E(t)$ est strictement croissante ;

- dès que $|x(t)| > \sqrt{3}$, la fonction $E(t)$ est strictement décroissante.

Par ailleurs, la fonction dE/dt est bornée par $2\mu/3$, donc $E(t)$ reste contrôlé pour tous les temps, ne serait-ce que par la borne $E(t) \leq E(0) + 2\mu t/3$. Cela donne des bornes sur $x(t)$ et $\dot{x}(t)$, et par critère de compacité le flot de l'équation de Van der Pol est défini globalement.

Dans la suite, on fixe y_0 et une trajectoire issue de $(0, y_0)$. Appelons T le premier temps $T > 0$ tel que la trajectoire du système recoupe l'axe vertical : le problème est de montrer que $E(0, y(T)) = E(0, y_0)$; ou encore que

$$\Delta(y_0) := E(0, y(T)) - E(0, y_0)$$

s'annule. La formule de la moyenne implique

$$\Delta(y_0) = \int_0^T \frac{dE}{dt}\, dt = \int_0^T \big(-x(t)\, \phi(x(t))\big)\, dt. \tag{88}$$

L'annulation de cette intégrale se fait forcément par compensation entre valeurs positives et négatives : en effet, dE/dt ne vaut 0 que si x vaut 0 ou $\sqrt{3}$. Or, si $(x = 0)$ est un équilibre, ce n'est pas le cycle que nous cherchons ; et $(x = \sqrt{3})$ n'est pas un équilibre.

Appelons x_1 la coordonnée de la première intersection de la trajectoire avec l'axe horizontal ; c'est aussi l'abscisse maximale de la trajectoire (car l'abscisse augmente pour $y > 0$ et diminue pour $y < 0$). On note que x_1 est une fonction croissante de y_0 : en effet, si l'on considère deux trajectoires, l'une issue de $(0, y_0)$ et l'autre de $(0, y_0')$ avec $y_0' > y_0$, alors la seconde trajectoire reste forcément "au-dessus et à droite" de la première, car les deux trajectoires ne peuvent se croiser. On note également que $x_1 \to 0$ quand $y_0 \to 0$.

L'application $y_0 \longmapsto x_1$ est un difféomorphisme de $]0, +\infty[$ sur lui-même : l'injectivité découle encore une fois de la propriété de non-croisement, la différentiabilité est conséquence de Cauchy–Lipschitz, et pour avoir la surjectivité il suffit de résoudre l'équation "à l'envers" en partant de $(0, x_1)$ et en remontant jusqu'à $(0, y_0)$. (Cette courbe croise forcément l'axe des ordonnées : en effet, s'il existait x_1 tel que la courbe "remontante" ne croise pas l'axe des ordonnées, alors on considèrerait l'infimum de ces x_1, on verrait que pour cet infimum la courbe correspondante est forcément asymptotique à l'axe des ordonnées, et cela est

impossible au vu des valeurs du champ de vecteurs f sur cet axe.) Il s'ensuit que $x_1 \to +\infty$ quand $y_0 \to +\infty$.

Nous pouvons maintenant aborder l'étude des variations de la fonction Δ.

Si $x_1 < \sqrt{3}$, alors $0 \leq x(t) < \sqrt{3}$ pour tout $t \in [0, T]$; donc la fonction E est strictement croissante entre 0 et T, et forcément $\Delta(y_0) > 0$.

Si $x_1 > \sqrt{3}$, la fonction E est successivement croissante puis décroissante. Examinons le régime où $x_1 \to \infty$. Le raisonnement n'est pas difficile mais sera un peu fastidieux. Montrons d'abord que la portion de la courbe sur laquelle x est compris entre 0 et $\sqrt{3}$ correspond à un intervalle de temps contrôlé. Pour $|x| < \sqrt{3}$, l'ordonnée y ne peut pas croître ou décroître trop vite : le champ de vecteurs $f(x, y)$ est contrôlé en norme par $2\mu|y| + \sqrt{3}$, et en appliquant Gronwall on en déduit une minoration de la forme $|y(t)| > |y_0|e^{-2\mu t} - bt$; ce qui prouve que pour tout $T > 0$,

$$\max_{t\in[0,T]} |x(t)| < \sqrt{3} \Longrightarrow \inf_{0\leq t\leq T} |y(t)| \to [y_0 \to \infty] + \infty.$$

Comme $\dot{x} = y$ il s'ensuit que $x(t)$ parcourt en un très bref intervalle de temps l'intervalle $[0, \sqrt{3}]$, et sur cet intervalle l'augmentation de E est forcément petite puisque $dE/dt \leq 2/3$. Il en est de même pour l'augmentation de E quand x passe de $\sqrt{3}$ à 0 aux temps ultérieurs.

En revanche, l'inclinaison du champ de vecteurs au voisinage de l'axe horizontal montre que, le système passe un temps au moins $O(1/\mu|x_1|^3)$ dans un voisinage de taille 1 autour de $x = x_1$, pour lequel dE/dt vaut environ $-x_1^4/3$; il s'ensuit que sur cet intervalle de temps la variation de E est négative et d'amplitude au moins $O(|x_1|/\mu)$, qui est arbitrairement grand quand $|x_1| \to \infty$.

La conclusion est que $\Delta(y_0) \to -\infty$ quand $x_1 \to +\infty$, donc quand $y_0 \to +\infty$. Si l'on ajoute à cela que $\Delta(y_0) \to 0$ quand $y_0 \to 0$, on peut déjà conclure, par continuité, qu'il existe y_0 tel que $\Delta(y_0) = 0$, et cela prouve l'existence d'un cycle.

Maintenant on cherche à prouver que ce cycle est unique, autrement dit que Δ s'annule une seule fois. C'est un peu plus sportif ! L'idée est de raffiner le calcul pour étudier les variations de Δ. À cette fin, on va troquer l'intégrale en temps pour une intégrale sur l'espace des phases : pour cela, on peut utiliser l'équation, en trouvant

$$dt = \frac{dx}{y} \qquad \text{quand } y \neq 0$$

$$dt = -\frac{dy}{\phi'(x)y + x} \qquad \text{quand } y \neq -\frac{x}{\phi'(x)}.$$

La première relation sera commode en début et en fin de courbe, quand y est bien distinct de 0 ; mais pas en milieu de courbe, en particulier quand on traverse l'axe des abscisses ; quant à la seconde relation, elle est peu commode dans tous les cas. C'est le moment de pousser plus avant la technique de Lienard et de considérer un nouveau jeu de variables :

$$(x, y) \longmapsto \big(x, z = y + \phi(x)\big).$$

Appelons-les "variables de Lienard". On trouve alors facilement

$$\begin{cases} \dot{x} = z - \phi(x) \\ \dot{z} = -x \end{cases} \tag{89}$$

(la première équation résulte de la définition du changement de variables et de la première équation dans (83), et la seconde équation résulte de la deuxième équation dans (83), réécrite selon (85)). En outre on peut réécrire l'énergie dans ce nouveau jeu de variables :

$$E = \frac{x^2 + z^2}{2}. \tag{90}$$

Par (89) on trouve

$$dt = \frac{dx}{z - \phi(x)}, \qquad dt = -\frac{dz}{x}.$$

La deuxième relation est simple et sera commode pour des valeurs de x éloignées de l'origine.

Fixons donc une valeur $a > 0$, à spécifier plus tard, et découpons l'intégrale (88) en trois morceaux : on a ainsi $\Delta = (I) + (II) + (III)$, avec

- Premier morceau (I) : x croissant de 0 à a ; pour ce morceau on change de variable par $dt = dx/y$, pour trouver

$$(I) = \int_0^a (-x\phi(x))\,\frac{dx}{y} = \int_0^a \left(\frac{-x\phi(x)}{y}\right) dx.$$

Ici $y = y_+(x; y_0)$ est l'ordonnée du point d'abscisse x sur la trajectoire issue de $(0, y_0)$, dans le demi-plan $y > 0$.

- Second morceau (II) : x croissant de a à x_1, puis décroissant de x_1 à a ; pour ce morceau on change de variables en passant dans les variables de Lienard (z décroît de $z+- = y_a - \phi(a)$ à $z_- = -y_a - \phi(a)$)

puis en posant $dt = -dz/x$ (le signe négatif résulte en une interversion des bornes dans la variable z) ; on trouve ainsi

$$(II) = \int_{z_-}^{z_+} (-x\phi(x)) \frac{dz}{x} = \int_{z_-}^{z_+} (-\phi(x))\, dz,$$

où $x = x(z; y_0)$ est l'abscisse du point d'ordonnée z sur la trajectoire issue de $(0, z = y_0)$ dans les variables (x, z). (Noter que $y = z$ sur l'axe $x = 0$, de sorte que la valeur initiale de z est bien y_0.)

- Troisième morceau (III) : x décroissant de a à 0 ; pour ce morceau on change de variable comme dans le premier morceau, en prenant garde aux bornes, pour trouver

$$(III) = -\int_0^a (-x\phi(x))\, \frac{dx}{y} = -\int_0^a \left(\frac{-x\phi(x)}{y} \right) dx.$$

Ici $y = y_-(x; y_0) < 0$ est l'ordonnée du point d'abscisse x sur la trajectoire issue de $(0, y_0)$, dans le demi-plan $y < 0$. (Le plus difficile dans l'histoire est sans doute de ne pas se tromper avec les signes !)

Analysons maintenant les variations des trois morceaux : On commence par

$$(I) = \int_0^a \left(\frac{-x\phi(x)}{y_+(x; y_0)} \right) dx.$$

Quand y_0 augmente, toute la courbe est décalée vers le haut ; la fonction $y_+(x; y_0)$ est donc une fonction croissante de y_0. Si $a \leq \sqrt{3}$, la fonction $-x\phi(x)$ est strictement positive pour $0 < x < a$, et l'intégrande $-x\phi(x)/y_+$ est donc une fonction décroissante de y_0.

La même conclusion s'applique à

$$(III) = \int_0^a \left(\frac{-x\phi(x)}{-y_-(x; y_0)} \right) dx,$$

car y_- est une fonction décroissante négative de y_0, donc $-y_-$ est une fonction croissante de y_0.

Enfin considérons le terme intermédiaire

$$(II) = \int_{z_-}^{z_+} (-\phi(x(z; y_0))\, dz,$$

où $z_\pm$ est défini par $x = a$. Quand y_0 augmente, le graphe de la courbe se décale vers la droite (dans le plan x, z), donc x augmente, à z fixé. Si $a \geq 1$ la fonction $-\phi$ est une fonction décroissante de x ; donc l'intégrande est, pour tout z, une fonction décroissante de y_0. À ce stade il faut prendre garde : quand y_0 varie les bornes de l'intégrale varient aussi, l'intervalle d'intégration s'agrandit et de nouvelles valeurs de z sont à prendre en compte ; mais si $a \geq \sqrt{3}$ ces valeurs seront toujours négatives (car $-\phi(x) \leq -\phi(a) \leq 0$).

On choisit finalement $a = \sqrt{3}$, et on conclut que $\Delta(y_0)$ est la somme de trois termes qui sont tous des fonctions décroissantes de y_0. La conclusion est que dès que $x_1 > \sqrt{3}$, la fonction Δ est strictement décroissante ; comme elle a une valeur strictement positive pour $x_1 \leq \sqrt{3}$ et qu'elle tend vers $-\infty$ quand $x_1 \to +\infty$, on conclut qu'il existe un unique x_1 (et donc un unique y_0) tel que $\Delta(y_0) = 0$. Nous avons ainsi prouvé l'existence d'un unique cycle !

REMARQUE 92. La forme particulière de la fonction ϕ n'a joué aucun rôle : ce qui compte est que ϕ soit une fonction impaire, surlinéaire à l'infini, avec un maximum local et un minimum local...

5.5. Analyse qualitative du cycle de Van der Pol

À quoi ressemble le cycle de Van der Pol ? On peut le contempler dans des simulations numériques, et voir comment il change d'allure quand le paramètre μ varie entre 0 et l'infini. Sans que l'on ait de formule explicite pour une valeur donnée de μ, on peut aussi étudier mathématiquement l'allure de ce cycle dans les deux régimes asymptotiques $\mu \to 0$ et $\mu \to \infty$; c'est ce que nous allons faire dans cette section.

Commençons par le régime $\mu \to 0$. Si on prend la limite $\mu \to 0$ dans l'équation, on trouve l'équation de l'oscillateur harmonique

$$\ddot{x} + x = 0,$$

qui admet une infinité de cycles stables. Il s'agit maintenant de voir comment ce comportement est modifié quand μ est strictement positif et petit ; nous sommes donc face à un *problème perturbatif*, comme l'étaient les astronomes du 18ème siècle qui étudiaient l'effet des petites interactions planète-planète sur les trajectoires dans le système solaire.

Partons d'un cycle de l'oscillateur non perturbé :

$$\begin{cases} X(t) = x_0 \cos t + y_0 \sin t \\ Y(t) = y_0 \cos t - x_0 \sin t. \end{cases}$$

Si μ est petit et non nul, $X(t)$ et $Y(t)$ ci-dessus ne sont plus des solutions exactes ; cependant leur variation est faible. Le théorème de variation du flot par rapport à un paramètre nous permet d'estimer la variation par rapport à μ :

$$x(t) = X(t) + \mu\xi(t) + O(\mu^2), \qquad y(t) = Y(t) + \mu\eta(t) + O(\mu^2),$$

où (ξ, η) est solution du système linéarisé

$$\begin{cases} \dot{\xi} - \eta = 0 \\ \dot{\eta} + \xi = (1 - X^2(t))Y(t). \end{cases}$$

On résout par la méthode de la "variation de la constante" : tous calculs faits, on trouve

$$\begin{cases} \xi(t) = \xi_0(t)\cos t + \eta_0(t)\sin t \\ \eta(t) = \eta_0(t)\cos t - \xi_0(t)\sin t \end{cases}$$

avec

$$\begin{cases} \dot{\xi}_0(t) = -[1 - X(t)^2]\, Y(t)\, \sin t \\ \dot{\eta}_0(t) = [1 - X(t)^2]\, Y(t)\, \cos t. \end{cases}$$

Au premier ordre en μ, la variation Δx de $x(t)$, sur une période, est d'environ

$$\mu \int_0^{2\pi} -\big[1 - (x_0 \cos t + y_0 \sin t)^2\big](y_0 \cos t - x_0 \sin t)\sin t\, dt; \tag{91}$$

en utilisant les identités

$$\frac{1}{2\pi}\int \cos^2 t\, dt = \frac{1}{2}, \qquad \frac{1}{2\pi}\int \cos^4 t\, dt = \frac{3}{8}, \qquad \frac{1}{2\pi}\int \cos^2 t\, \sin^2 t\, dt = \frac{1}{8},$$

on trouve finalement

$$\Delta x = \frac{2\pi\mu}{8}\, x_0\big(4 - (x_0^2 + y_0^2)\big) + O(\mu^2). \tag{92}$$

Un calcul similaire montre que la variation de y sur $[0, 2\pi]$ vaut

$$\Delta y = \frac{2\pi\mu}{8}\, y_0\big(4 - (x_0^2 + y_0^2)\big) + O(\mu^2). \tag{93}$$

Sur le cycle, cette variation est forcément nulle à l'ordre μ, ce qui entraîne

$$x_0^2 + y_0^2 = 4 + O(\mu). \tag{94}$$

En conclusion, le cycle de Van der Pol est asymptotique, dans la limite $\mu \to 0$, au cercle centré en l'origine de rayon $r = 2$.

Un autre calcul que l'on peut faire selon les mêmes lignes concerne la variation de $r = \sqrt{x^2 + y^2}$. Tous calculs faits, on trouve que la variation de r sur $[0, 2\pi]$ vaut

$$\Delta r = \frac{2\pi\mu}{8}\, r(4 - r^2) + O(\mu^2). \tag{95}$$

En prenant l'augmentation moyenne, on trouve une dynamique approchée

$$\dot{r} \simeq \frac{\mu}{8}\, r(4 - r^2).$$

C'est comme si le système évoluait selon un cercle de rayon lentement changeant : sur une échelle de temps $O(1/\mu)$, le rayon de ce cercle augmente si $r < 2$ et diminue sur $r > 2$, convergeant comme il se doit vers un cycle limite $r \simeq 2$. On peut ainsi remplacer, sur une échelle de temps assez grande, la dynamique de Van der Pol, en variables polaires, par

- une dynamique lente $\dot{r} = (\mu/8)r(4 - r^2)$
- une dynamique rapide $\dot{\theta} = 1$.

Remarque 93. La valeur $\mu = 0$ peut être vue comme une *valeur de bifurcation du système*. En effet, si l'on assigne à μ une valeur strictement négative, l'équilibre $(0, 0)$ est toujours stable ; alors que pour $\mu = 0$ tout l'espace des phases est fait de cycles, et que pour $\mu > 0$ l'équilibre $(0, 0)$ est devenu instable et se fait remplacer par un cycle stable. On qualifie de *bifurcation de Hopf* la situation dans laquelle la variation continue d'un paramètre amène au remplacement soudain d'un équilibre par un cycle limite.

Maintenant considérons l'asymptotique opposée où $\mu \to \infty$. Une étape clé est l'identification d'un jeu de variables commode. Posons

$$\psi(x) = \frac{\phi(x)}{\mu} = \frac{x^3}{3} - x.$$

Le graphe de ψ présente deux branches croissantes et une branche décroissante, les changements de variation se faisant en ± 1. La deuxième équation du système de Van der Pol s'écrit $\dot{y} = \mu\, \psi'(x)y - x$. En admettant que x ne s'éloigne pas trop à l'infini, on peut s'attendre à ce que le dernier terme s'efface devant le premier, et on escompte une dynamique approchée de la forme $\dot{y} = -\mu\psi'(x)y = (d/dt)(-\mu\psi(x))$. Cela suggère de s'intéresser à la quantité $z = y + \mu\psi(x)$; on retrouve ainsi le changement de variable de Lienard. La variable z vérifiera $\dot{z} = -x = O(1)$, de sorte que z ne varie pas trop vite ; en revanche, ses valeurs peuvent devenir très grandes, de sorte que l'on s'intéressera à $w = z/\mu = \psi(x) + \dot{x}/\mu$. L'équation $\dot{x} = y$ devient, dans ce jeu de variables, $\dot{x} = \mu(w - \psi(x))$. Pour récapituler : le changement de variables $w = \psi(x) + \dot{x}/\mu$ nous a fourni une nouvelle formulation de l'équation de Van der Pol :

$$\begin{cases} \dot{x} = \mu(w - \psi(x)) \\ \dot{w} = -\dfrac{x}{\mu}. \end{cases} \tag{96}$$

Jusqu'ici nous avons équivalence exacte avec (82). C'est dans la limite $\mu \to \infty$ que le système (96) va s'avérer particulièrement utile. En l'examinant, on voit en effet que

- w a toujours une dynamique très lente ;
- x a une dynamique très rapide dès que le système s'écarte du graphe $w = \psi(x)$.

La description qualitative du système est donc comme suit : le système monte et descend lentement le long du graphe $w = \psi(x)$, puis saute très rapidement d'une valeur de x à l'autre, sans changer la valeur de w.

On trace l'allure du champ de vecteurs correspondant : il est presque horizontal, pointant vers la droite pour toute la partie située "à gauche du graphe" et pointant vers la gauche pour toute la partie située "à droite du graphe".

À quoi ressemble la dynamique ? Tout d'abord, le système est très vite attiré par le graphe $w = \psi(x)$. Supposons que l'on parte d'un point (x, w) situé en dehors du graphe ; on subit donc l'action d'un champ de vecteurs quasiment horizontal et de grande amplitude, qui projette rapidement le système sur le graphe ou à son voisinage immédial. Pour savoir sur quelle partie du graphe le système va atterrir, on examine la droite horizontale qui passe par (x, w). Si cette droite coupe le graphe en un seul point (c'est à dire si $|w|$ est assez grand) on est très vite précipité vers ce point d'intersection. Si cette droite coupe le graphe en trois points (c'est à dire si $|w|$ est assez petit), alors cela dépend de la position par rapport au graphe : si on est complètement à gauche, on est précipité vers le brin gauche ; si l'on est complètement à droite, on est précipité vers le brin droit ; si l'on est "dans le coude", entre deux branches du graphe, alors on est précipité vers la branche gauche, quand on est en-dessous du graphe, et vers la branche droite, quand on est au-dessus. Seules les parties croissantes du graphe sont atteintes, la partie décroissante (centrale) ne l'est jamais. (Un dessin est aussi indispensable qu'éclairant.) Enfin, si la droite coupe le graphe en deux points, la discussion est un tout petit peu plus subtile : par exemple si l'on est au-dessus de l'axe des abscisses, exactement à la hauteur du maximum local, alors on va être projeté sur le brin droit si l'on est tout à droite de la courbe ou si l'on est à droite du maximum local ; et si l'on est à gauche de ce maximum, on est projeté sur le brin gauche, très légèrement en-dessous de la bosse correspondant au maximum local (car $\dot{w} < 0$).

Que se passe-t-il maintenant si l'on part du graphe ? La partie horizontale du champ de vecteurs s'annule sur le graphe, la partie verticale va donc jouer alors pleinement son rôle. L'origine est toujours un équilibre. Si l'on est sur la partie centrale (décroissante) du graphe, à droite de l'axe des ordonnées, alors on va légèrement descendre, et aussitôt être attiré par le brin gauche du graphe ; symétriquement, si l'on est sur la partie centrale et à gauche de l'axe des ordonnées, alors on va légèrement monter, et aussitôt être attiré par le brin droit du graphe. Autrement dit, la partie centrale du graphe est instable, et, à l'exception de l'origine, mène tout de suite aux parties externes, croissantes, du graphe.

Supposons maintenant que l'on parte de l'une des deux branches croissantes du graphe, par exemple celle de gauche. Tout départ significatif du graphe est impossible tant que l'on reste sur cette branche ; la variation de w étant positive, on monte (très lentement) le long de cette branche ; cela dure jusqu'à ce que l'on atteigne le sommet, c'est à dire le maximum local. De là une infime variation de l'ordonnée pousse le système au-dessus du graphe, et aussitôt il est attiré par la branche de droite. Mais là il va descendre lentement le long du graphe, jusqu'au minimum local, puis descendre encore, et être attiré par la branche de gauche. Et ainsi de suite...

Le cycle limite pour $\mu \to \infty$ est donc trouvé : il est fait de

- deux segments de branches croissantes du graphe, déterminées par $\psi_-^{-1}(\psi(1)) \leq x \leq -1$ et $1 \leq \psi_+^{-1}(\psi(-1))$, où $\psi_-^{-1}(\psi(1))$ est le plus petit x tel que $\psi(x) = \psi(1)$, et $\psi_+^{-1}(\psi(-1))$ est le plus grand x tel que $\psi(x) = \psi(-1)$;
- et deux segments horizontaux qui joignent ces deux branches.

L'ensemble dessine donc une sorte de parallélogramme avec deux côtés courbes. Le mouvement sur ce cycle se décompose ainsi :

- un mouvement lent le long des deux branches courbes : ce mouvement se fait à vitesse $O(1/\mu)$ et dure un temps $O(\mu)$;
- un mouvement rapide le long des deux segments horizontaux : sur la majeure partie de ces segments, le mouvement est à vitesse $O(\mu)$ et dure environ $O(1/\mu)$. (Mais la durée complète de parcours des segments est plus longue – environ $O(\ln \mu/\mu)$ – car il faut du temps pour s'arracher du graphe...)

On peut ensuite retrouver la forme asymptotique de l'attracteur dans les variables (x, y) en utilisant $y = \mu(w - \psi(x))$: en particulier, les segments horizontaux sont remplacés par de larges bosses qui sont essentiellement les images renversés des segments correspondants du

graphe ψ, dilatées d'un facteur μ. L'attracteur dessine ainsi deux "cornes" très pointues, une à droite dirigée vers le haut, et une à gauche dirigée vers le bas.

Ces conclusions sont très bien observées numériquement, et là encore ne dépendent guère de la forme exacte de ψ.

5.6. Complément : Équations de FitzHugh–Nagumo et autres modèles

L'équation de Van der Pol s'inscrit dans un ensemble de travaux destinés à modéliser des phénomènes d'oscillations et de réponses à des impulsions. Cela se produit dans des circuits électriques, mais aussi dans les organes vivants : propagation de courants à travers les cellules (la membrane agissant comme un condensateur), en particulier à travers les axones ; battements réguliers du cœur causés par des variations complexes de potentiel électrique à la surface du myocarde...

Dans la hiérarchie des équations utilisées à ces fins, l'équation de Van der Pol est la plus simple. Citons aussi :

- l'équation de FitzHugh–Nagumo, système de 2 équations à 2 inconnues :

$$\dot{x} = f(x) - w + I \qquad \dot{w} = a(bx - cw). \tag{97}$$

Si l'on choisit $c = 0$, $ab = 1$, $f(x) = \mu(x - x^3/3)$, on retrouve le modèle de Van der Pol. Le modèle de FitzHugh–Nagumo est populaire pour décrire les battements du cœur.

- le modèle de Hodgkin–Huxley, système de 4 équations à 4 inconnues [**28**]. C'est l'un des plus élaborés en la matière : sa mise au point était un tour de force qui a valu un Prix Nobel de médecine à ses auteurs. Il modélise l'évolution du potentiel électrique dans un axone, via

$$I = C\,\frac{dV}{dt} + \sum_{\ell} g_\ell(t)\,\big(V(t) - V_\ell\big),$$

où I est le courant injecté, et les g_ℓ représentent les contributions de différents facteurs : concentration en ions potassium ($\ell = \mathrm{K}$) et en ions sodium ($\ell = \mathrm{Na}$), pertes dues à la fuite des ions ($\ell = \mathrm{f}$). Les quantités g_{K}, g_{Na} et g_{f} dépendent d'une subtile modélisation électrochimique :

$$g_{\mathrm{K}} = \text{const.}n^4, \qquad g_{\mathrm{Na}} = \text{const.}m^3h, \qquad g_{\mathrm{f}} = \text{const.}$$
$$\dot{n} = \alpha_n(V)(1-n) - \beta_n(V)n, \qquad \dot{m} = \alpha_m(V)(1-m) - \beta_m(V)m,$$
$$\dot{h} = \alpha_h(V)(1-h) - \beta_h(V)h.$$

Ce modèle dépend donc de pas moins de 6 fonctions (α_n, α_m, α_h, β_n, β_m, β_h) modélisées par des fractions de fonctions exponentielles... Sa

complexité le rend difficile à manier en pratique, et l'on préfère souvent le réduire à un système plus simple.

Si l'on examine les valeurs propres de la matrice jacobienne près de l'équilibre, on trouve, selon l'analyse de FitzHugh, deux valeurs propres réelles strictement négatives, et deux valeurs propres complexes de partie réelle légèrement positive ; en "trichant un peu" pour appliquer le théorème de la variété centrale, on se ramène asymptotiquement à l'étude d'une surface correspondant à ces deux dernières valeurs propres. La dynamique résultante est alors très proche de celle de FitzHugh–Nagumo, ce qui justifie le remplacement de Hodgkin–Huxley par ce dernier système. Il est aussi possible d'effectuer cette réduction par des arguments heuristiques, en analysant qualitativement les comportements attendus pour les différentes variables et en les groupant deux par deux.

• D'autres variantes encore plus sophistiquées font intervenir des termes supplémentaires avec des dérivées spatiales, du style d^2V/dx^2, issus des équations de Maxwell ; cela transforme l'EDO en EDP. D'autres modèles utilisent aussi des termes stochastiques qui en fonct une équation différentielle stochastique (EDS). Enfin on peut incorporer des modélisations plus fines de certains phénomènes biologiques et thermodynamiques.

Les modèles que nous avons passés en revue ont été développés et utilisés pour des fins diverses, que l'on retrouvera dans les articles de Wikipedia et Scholarpedia aux rubriques Van der Pol, FitzHugh–Nagumo et Hodgkin–Huxley ; pour l'historique et les développements du sujet on pourra aussi consulter les sources mentionnées dans le cours déjà cité de Knill [**21**]. Ces équations sont avant tout utilisées pour faire des calculs numériques et des études théoriques sur toutes sortes de questions importantes mettant en jeu des propagations électriques (processus cérébraux, battements du cœur, etc.) Sous un angle purement qualitatif, le fait que ces équations donnent naissance à des cycles d'oscillations bien déterminées nous aide à comprendre comment des phénomènes réguliers tels que le battement du cœur peuvent se mettre en place spontanément.

On peut aussi utiliser ces modèles pour étudier la réponse à des stimuli variables, aléatoires ; la théorie du chaos a été appliquée avec succès à cette problématique.

Un autre sujet capital est l'exploration des transitions de phase entre différents états d'un système complexe à seuil (excité ou non excité). Ainsi, on trouvera sur

`http://en.wikipedia.org/wiki/Hodgkin%E2%80%93Huxley_model`

une petite animation sur la réponse du modèle de Hodgkin–Huxley à un courant qui augmente peu à peu : pour un courant bas, le système est en équilibre et le potentiel V varie peu au cours du temps ; quand le courant augmente, on observe un pic suivi d'un retour à l'équilibre ; mais quand le courant atteint une certaine valeur, la réponse à l'impulsion est une série de pics correspondant à un cycle limite. L'analyse de ce phénomène entraîne naturellement vers une branche délicate et importante de la théorie des EDO : la **bifurcation**, c'est à dire la capacité d'un système à changer brusquement de propriétés qualitatives en fonction des variations d'un paramètre.

En guise de dernière remarque, on notera que la simulation de toutes ces équations "à cycle" du type de celle de Van der Pol n'est pas si simple ! On verra des exemples en exercices.

CHAPITRE 6

Initiation aux systèmes hamiltoniens

6.1. Définition et exemples

Les équations du second ordre surgissent très fréquemment dans l'étude des phénomènes naturels conservatifs, et mènent à des espaces de phases de type "position/vitesse" comme $\mathbb{R}^n \times \mathbb{R}^n$, ou plus généralement l'espace tangent à une variété. En fait, ces situations sont si courantes qu'il est naturel de s'intéresser à des classes de systèmes conservatifs dans lesquels l'inconnue est par construction une double variable. Les systèmes hamiltoniens appartiennent dans cette catégorie.

DÉFINITION 94. Soit $H = H(x,p)$ une fonction de classe C^1 définie sur $\mathbb{R}^n \times \mathbb{R}^n$. On appelle système hamiltonien associé à H l'équation différentielle du premier ordre

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p} \\ \dot{p} = -\dfrac{\partial H}{\partial x}. \end{cases} \tag{98}$$

La fonction H est appelée le hamiltonien (ou l'hamiltonien) du système.

Ici $\partial H/\partial p = \nabla_p H$ est un vecteur à n composantes, de sorte que l'on peut écrire plus explicitement

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial x_i}, \qquad 1 \leq i \leq n.$$

On dit que les variables x_i et p_i sont *conjuguées*. Le champ de vecteurs apparaissant au membre de droite de (98) est le *champ de vecteurs hamiltonien* associé à H.

Nous verrons plus tard comment définir des systèmes hamiltoniens dans un cadre géométrique bien plus général, mais pour le moment, nous nous contenterons de (98). Deux choses sautent aux yeux : la première est la ressemblance avec les flots gradients, qui sont aussi des équations du premier ordre dans lesquelles le second membre est défini par la dérivée d'une fonction. La seconde est la "symétrie subtilement brisée" dans le rôle des variables x et p, avec le changement de signe entre les deux équations de (98). (Pour insister sur cette dualité entre

variables, il est extrêmement courant d'utiliser la notation q à la place du traditionnel x.)

Le formalisme hamiltonien a été introduit vers 1820–1830 par Rowan Hamilton, le plus célèbre des mathématiciens irlandais (également connu comme le père des quaternions, et comme un savant aux capacités intellectuelles extraordinaires). Le génie de Hamilton, en l'occurrence, était d'avoir reconnu une structure extrêmement générale qui allait pouvoir mener à une étude unifiée de très nombreuses équations. Il faut noter que Cauchy travaillait également sur ce formalisme en même temps que Hamilton ; et que cette théorie prenait sa source dans les travaux antérieurs de Lagrange et Poisson en astronomie.

Aujourd'hui les "systèmes hamiltoniens", convenablement généralisés, représentent l'une des branches les plus respectables de la théorie des équations d'évolution ; il s'agit surtout d'EDO, mais certaines EDP sont également des systèmes hamiltoniens. Le sujet est empli de techniques variées et puissantes, développées durant bientôt deux siècles...

On peut penser à la fonction H dans (98) comme à une sorte d'énergie. En fait, la première propriété fondamentale d'un système hamiltonien est la conservation de cette fonction.

PROPOSITION 95. *Le hamiltonien H est préservé au cours de l'évolution hamiltonienne correspondante.*

DÉMONSTRATION. Il suffit de calculer :

$$\begin{aligned}\frac{d}{dt}H\big(x(t),p(t)\big) &= \frac{\partial H}{\partial x}\cdot\dot{x} + \frac{\partial H}{\partial p}\cdot\dot{p}\\ &= \frac{\partial H}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial H}{\partial p}\cdot\frac{\partial H}{\partial x} = 0.\end{aligned}$$

□

COROLLAIRE 96. *Si H est de classe C^2 et que les composantes connexes de ses lignes de niveau de H sont compactes, alors le système hamiltonien est défini globalement en temps.*

Ce corollaire résulte immédiatement du théorème de Cauchy–Lipschitz (la régularité C^2 de H entraînant la régularité C^1 du champ de vecteurs) et du critère de compacité.

La preuve de la Proposition 95 nous a montré, au passage, l'expression de la dérivée temporelle d'une observable quelconque :

$$\frac{d}{dt}f\big(x(t),p(t)\big) = \frac{\partial f}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial f}{\partial p}\cdot\frac{\partial H}{\partial x}.$$

Cette expression est si commode qu'on lui a donné un nom :

DÉFINITION 97 (crochet de Poisson). Soient f et g deux fonctions différentiables sur $\mathbb{R}^n \times \mathbb{R}^n$, à valeurs réelles. On appelle *crochet de Poisson* de f et g la fonction

$$\begin{aligned}\{f,g\} &= \frac{\partial f}{\partial x}\cdot\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\cdot\frac{\partial g}{\partial x} \\ &= \sum_i \left(\frac{\partial f}{\partial x_i}\,\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\,\frac{\partial g}{\partial x_i}\right).\end{aligned}$$

On remarque que $\{g,f\} = -\{f,g\}$ et donc $\{f,f\} = 0$; que $\{x_i,x_j\} = 0$, $\{p_i,p_j\} = 0$, $\{x_i,p_j\} = \delta_{ij}$ (c'est à dire 1 si $i = j$ et 0 si $i \neq j$). On a alors, en variante simple de la preuve de la Proposition 95 :

PROPOSITION 98. *Soient f et H deux fonctions différentiables $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, et soit $(x(t),p(t))$ une solution des équations hamiltoniennes de hamiltonien H. Alors*

$$\left.\frac{d}{dt}\right|_{t=0} f\big(x(t),p(t)\big) = \{f,H\}.$$

Cette proposition donne une recette algébrique simple et efficace pour calculer les dérivées temporelles des observables le long d'un flot hamiltonien. La conservation de l'énergie se réécrit ainsi $\{H,H\} = 0$.

Après l'évolution des observables, on s'intéresse ensuite aux équilibres.

PROPOSITION 99. *Les équilibres d'un système hamiltonien sont les points où la différentielle du hamiltonien s'annule : $dH = 0$. En particulier, tout maximum local ou tout minimum local de H est un équilibre.*

DÉMONSTRATION. Il suffit de revenir à (98) : il y a équilibre si et seulement si $\partial H/\partial x = 0$ et $\partial H/\partial p = 0$, c'est à dire si et seulement si la différentielle tout entière s'annule. □

Le réflexe suivant est de considérer la matrice jacobienne du champ de vecteurs, et d'en calculer la divergence. Soit donc

$$\xi = \xi_H = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial x}\right)$$

le champ de vecteurs associé à H. En dérivant ξ on obtient la matrice jacobienne $D\xi = (\partial_j \xi_H^i)$; c'est une matrice carrée $(2n)\times(2n)$, avec n indices pour les variables x_i et n indices pour les variables p_i ; en notation matricielle par blocs,

$$D\xi = \begin{pmatrix} \dfrac{\partial^2 H}{\partial p\,\partial x} & \dfrac{\partial^2 H}{\partial p\,\partial p} \\ -\dfrac{\partial^2 H}{\partial x\,\partial x} & -\dfrac{\partial^2 H}{\partial x\,\partial p} \end{pmatrix}.$$

On en déduit en particulier que la trace est nulle, c'est à dire que *la divergence d'un champ de vecteurs hamiltonien est identiquement nulle.* On en déduit deux conséquences majeures :

(i) *le volume est préservé par le flot* dans l'espace des phases : si complexe que soit le flot, on sait que le volume des états initiaux est le même que le volume des états correspondant à un temps ultérieur. Il est important de noter que cette propriété a lieu seulement dans l'espace des phases.

(ii) *il n'y a jamais stabilité asymptotique* : en effet, la somme des valeurs propres du système linéarisé étant nulle, les parties réelles ne peuvent être toutes strictement négatives. Soit toutes les valeurs propres sont imaginaires pures (le système étant à coefficients réels, les valeurs propres non réelles sont conjuguées deux à deux) ; soit il y a au moins une valeur propre de partie réelle strictement positive.

Même en l'absence de stabilité asymptotique, on peut étudier la stabilité orbitale via les propriétés de H : si par exemple on a un maximum local strict de la fonction H, alors les lignes de niveau confinent le système au voisinage de cet équilibre pour tous les temps.

Pour conclure cette introduction, voici quelques **exemples** de systèmes hamiltoniens sur $\mathbb{R}^n \times \mathbb{R}^n$. Il suffit de les présenter par leur hamiltonien, l'équation en découle. On rappelle que la dimension est toujours paire, les systèmes hamiltoniens les plus simples sont donc ceux de dimension 2, puis viennent les systèmes de dimension 4, etc. Attention à la terminologie : quand on dit "en dimension 2", cela correspond donc à $n = 1$.

Exemple 100. En dimension 2, posons

$$H(x,p) = V(x) + \frac{p^2}{2m}.$$

Alors l'équation associée est

$$\begin{cases} \dot{x} = \dfrac{p}{m} \\ \dot{p} = -\nabla V(x) \end{cases}$$

D'où l'équation $m\ddot{x} = -\nabla V(x)$, ce qui est l'équation de Newton dans un champ de potentiel V. On reconnaît dans $p = m\dot{x}$ la quantité de mouvement, et dans H l'énergie totale du système. Au passage, cet exemple simple permet de se souvenir de l'emplacement du signe négatif dans les équations (98).

Exemple 101. Considérons le système de N planètes en interaction autour du Soleil, où l'on néglige les mouvements du Soleil (supposé

à l'origine), les dimensions des planètes, et les interactions planète-planète. On note m_0 la masse du Soleil, et $m_1, \dots, m_N$ les masses des planètes ; $x_1, \dots, x_N \in \mathbb{R}^3$ leurs positions, et $p_1, \dots p_N \in \mathbb{R}^3$ leurs quantités de mouvement. C'est donc un système de dimension $6N$, dont le hamiltonien est

$$H(x,p) = \sum_{1\leq i\leq N} \left(\frac{|p_i|^2}{2m_i} - \frac{\mathcal{G}\, m_i\, m_0}{|x_i|} \right) - \sum_{1\leq i<j\leq n} \frac{\mathcal{G}\, m_i\, m_j}{|x_i - x_j|}.$$

On retrouve ainsi les équations approchées des planètes autour du Soleil, menant aux lois de Kepler.

Exemple 102. Si l'on veut raffiner l'exemple précédent en tenant compte des mouvements du Soleil, les formules deviennent un peu plus compliquées et il faut corriger quelques coefficients. Pour apprécier l'influence des masses relatives du Soleil et des planètes, notons

$$\varepsilon = \frac{\sum_{i\neq 0} m_i}{m_0}.$$

En supposant que $\sum p_i = 0$, le nouveau hamiltonien s'écrit alors

(99)
$$H(X,Y) = \sum_{1\leq i\leq N} \left(\frac{|Y_i|^2}{2\mu_i} - \frac{\mathcal{G}\, \mu_i\, m_0^i}{|X_i|} \right) + \varepsilon \sum_{1\leq i<j\leq n} \left(\frac{Y_i \cdot Y_j}{m_0} - \frac{\mathcal{G}\, M_i\, M_j}{|X_i - X_j|} \right).$$

où $M_i = m_i/\varepsilon$, $m_0^i = m_0 + m_i$,

Exemple 103. Le système de Hénon–Heiles est un modèle simple de hamiltonien non linéaire en dimension 4 :

$$H(x,p) = \frac{|p|^2}{2} + \frac{|x|^2}{2} + \lambda \left(x_1^2 x_2 - \frac{x_2^3}{3} \right).$$

La partie quadratique de H correspond à des termes linéaires de l'EDO, alors que la correction cubique implique des termes non linéaires, quadratiques, dans l'EDO. Ce système est utilisé dans l'étude du chaos déterministe.

Exemple 104. La toupie, ou toupie d'Euler, est le modèle mathématique d'un solide X soumis à la pesanteur, que l'on fait tourner autour d'un point. Petit "rappel" de mécanique du solide [**23**] : un solide en rotation autour d'un point a une "matrice d'inertie" I, qui donne des indications sur sa "corpulence" vue depuis le point de rotation. Plus précisément I représente l'application linéaire $\omega \longmapsto \int_X x \wedge (\omega \wedge x)\, \rho(dx)$, où $\wedge$ désigne le produit vectoriel et $\rho(dx)$ la masse volumique de l'élément dx (penser à ω comme à un vecteur rotation). Les directions propres

de la matrice d'inertie sont appelés axes principaux d'inertie, et les valeurs propres sont les moments principaux d'inertie. Appelons I_1, I_2, I_3 ces valeurs propres, et e_1, e_2, e_3 les directions propres. On se place alors dans un repère mobile dont les axes sont ces directions propres, et on note n_1, n_2, n_3 les composantes du vecteur unitaire vertical n dans ce repère mobile. Le hamiltonien s'écrit alors

$$H(x,p) = \sum \frac{\ell_k^2}{I_k} + mgx \cdot n,$$

où m est la masse, x la position du centre de gravité (qui n'est pas forcément le centre de rotation) et g l'intensité du champ gravitationnel. Ce système a fait l'objet d'études à partir de Leonhard Euler, et Sofia Kowalevskaya s'y est distinguée spectaculairement.

EXEMPLE 105. "Tous" les systèmes lagrangiens "strictement convexes" sont aussi des systèmes hamiltoniens. Avant d'expliquer ce que recouvre cette assertion, nous allons prendre du temps pour étudier les systèmes lagrangiens plus en profondeur.

6.2. Du lagrangien à l'EDO

Dans cette section nous allons voir comment écrire l'équation différentielle associée à un lagrangien : il s'agit de transformer un problème variationnel (la minimisation de l'action) en une EDO, que l'on appelle **équation d'Euler–Lagrange** de la fonctionnelle d'action.

Soit $L = L(x,v)$ un lagrangien défini sur $\mathbb{R}^n \times \mathbb{R}^n$. On pourrait aussi considérer un lagrangien défini sur un espace plus général de dimension $2n$, comme $\mathbb{T}^n \times \mathbb{R}^n$ ou TM (le fibré tangent à une variété différentiable M de dimension n) ; ou encore ajouter une dépendance en temps et considérer un lagrangien plus général $L(x,v,t)$; mais au-delà de la complexité accrue du formalisme, cela ne changerait rien de fondamental, ni aux calculs, ni aux conclusions. Nous allons donc nous contenter ici d'un lagrangien autonome $L = L(x,v)$ défini sur $\mathbb{R}^n \times \mathbb{R}^n$.

On rappelle que par définition du formalisme lagrangien, les trajectoires sont, sur de petits intervalles de temps, des minimiseurs (ou tout au moins des points critiques) de la fonctionnelle d'action

$$I(x) = \int_{t_1}^{t_2} L\big(x(t), \dot{x}(t)\big)\, dt$$

parmi les courbes de classe C^1 qui coïncident avec x aux temps t_1 et t_2.

Si $x(t)$ est un minimiseur de I, alors toute variation $x(t)+z(t)$ telle que $z(t_1)=z(t_2)=0$ sera moins performante :

$$\int_{t_1}^{t_2} L\big(x(t)+z(t),\dot{x}(t)+\dot{z}(t)\big)\,dt \geq \int_{t_1}^{t_2} L\big(x(t),\dot{x}(t)\big)\,dt.$$

Ici z est quelconque, on peut donc le multiplier par une quantité arbitraire, disons $\varepsilon>0$. On obtient

$$\int_{t_1}^{t_2} L\big(x(t)+\varepsilon z(t),\dot{x}(t)+\varepsilon\dot{z}(t)\big)\,dt \geq \int_{t_1}^{t_2} L\big(x(t),\dot{x}(t)\big)\,dt.$$

Par hypothèse le lagrangien L est différentiable. Donc quand $\varepsilon\to 0$ le membre de gauche dans (6.2) devient

$$\int_{t_1}^{t_2} L\big(x(t),\dot{x}(t)\big)\,dt + \varepsilon\left[\int_{t_1}^{t_2}\frac{\partial L}{\partial x}\big(x(t),\dot{x}(t)\big)\cdot z(t)\,dt + \int_{t_1}^{t_2}\frac{\partial L}{\partial \dot{x}}\big(x(t),\dot{x}(t)\big)\cdot \dot{z}dt\right] + o(\varepsilon). \tag{100}$$

On reporte dans (6.2) et on simplifie par le terme d'ordre 0 ; on divise ensuite tout par ε (ce qui ne change pas l'inégalité car $\varepsilon>0$), et on fait tendre ε vers 0. On trouve

$$\int_{t_1}^{t_2}\frac{\partial L}{\partial x}\big(x(t),\dot{x}(t)\big)\cdot z(t)\,dt + \int_{t_1}^{t_2}\frac{\partial L}{\partial \dot{x}}\big(x(t),\dot{x}(t)\big)\cdot \dot{z}(t)\,dt \geq 0. \tag{101}$$

Cela est vrai pour tout z vérifiant les conditions mentionnées plus haut ; on peut donc remplacer z par $-z$, et la même inégalité reste vraie, alors que le membre de gauche est changé en son opposé ; cela prouve que finalement l'expression dans le membre de gauche de (6.2) est égale à 0.

À ce stade on a démontré que

$$\int_{t_1}^{t_2}\frac{\partial L}{\partial x}\big(x(t),\dot{x}(t)\big)\cdot z(t)\,dt + \int_{t_1}^{t_2}\frac{\partial L}{\partial \dot{x}}\big(x(t),\dot{x}(t)\big)\cdot \dot{z}(t)\,dt = 0, \tag{102}$$

où z est une fonction arbitraire dans $C^1([t_1,t_2];\mathbb{R}^n)$ qui s'annule aux bornes de l'intégrale. Malgré cette dernière restriction, on s'attend à ce que la condition d'annulation entraîne des contraintes très particulières sur la fonction L ; pour le voir on va "mettre z en facteur" en réalisant une intégration par parties dans la seconde intégrale de (102). Du fait de l'annulation de z aux bornes, il n'y a pas de terme de bord :

$$\int_{t_1}^{t_2}\frac{\partial L}{\partial \dot{x}}\big(x(t),\dot{x}(t)\big)\cdot \dot{z}(t)\,dt = -\int_{t_1}^{t_2}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}(x(t),\dot{x}(t))\right)\cdot z(t)\,dt. \tag{103}$$

L'équation (102) devient alors

$$\int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right)\right](x(t), \dot{x}(t)) \cdot z(t)\, dt = 0. \tag{104}$$

Posons $w = \partial_x L - (d/dt)\partial_{\dot{x}} L$, évalué le long du mouvement. On sait que $\int w(t) \cdot z(t)\, dt = 0$ pour tout z dans une large classe de fonctions. Il suffirait de choisir $z = w$ pour conclure que $w = 0$; mais cela n'est pas autorisé car w ne s'annule pas forcément au bord, et n'est pas forcément de classe C^1. Admettons cependant cette propriété, sous la forme d'un lemme :

LEMME 106. *Soit $w = w(t)$ une fonction continue, définie sur $[t_1, t_2]$, à valeurs dans $\mathbb{R}^n$. Si $\int w(t) \cdot z(t)\, dt = 0$ pour toute fonction $z \in C^1([t_1, t_2]; \mathbb{R}^n)$ s'annulant en t_1 et t_2, alors w est identiquement nulle.*

Avec ce lemme, on conclut que $w = 0$, ce qui est l'équation recherchée : on l'appelle l'*équation d'Euler–Lagrange* du problème variationnel de moindre action.

Énonçons la conclusion sous la forme d'un théorème précis :

THEORÈME 107 (Équation d'Euler–Lagrange). *Soit $L = L(x, v)$ un lagrangien de classe C^1 sur $\mathbb{R}^n \times \mathbb{R}^n$, et $x = x(t)$ une courbe de moindre action de classe C^1 sur $[t_1, t_2]$, à valeurs dans $\mathbb{R}^n$. Alors, le long de la trajectoire $(x(t), \dot{x}(t))$ dans l'espace des phases,*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x}. \tag{105}$$

REMARQUE 108. Il est assez facile de retenir l'équation (105) quand on a remarqué que "les dérivées temporelles se simplifient". En effet, si l'on écrit formellement $\partial L/\partial \dot{x}$ comme $dL/d(dx/dt)$, le membre de gauche de (105) s'écrit

$$\left(\frac{d}{dt}\right) \frac{dL}{d\left(\frac{d}{dt}\right) x},$$

et en "barrant" les (d/dt) on trouve dL/dx, ce qui se trouve être le bon résultat, exprimé au membre de droite de (105) !

Le Théorème 107 se généralise à plusieurs niveaux :

- On peut considérer un lagrangien L défini sur $\mathbb{T}^n \times \mathbb{R}^n$ plutôt que $\mathbb{R}^n \times \mathbb{R}^n$; les démonstrations sont alors inchangées.
- Cela ne pose pas de problème non plus d'ajouter une dépendance en temps du lagrangien (qui devient alors une fonction de x, v et t).

- On peut aussi élargir, dans une certaine mesure, l'espace des courbes de moindre action : l'énoncé s'applique à des courbes qui sont seulement supposées continues et C^1 par morceaux.
- On peut encore généraliser l'énoncé à peu de frais au cas d'une géométrie courbe de l'espace des états : le lagrangien est alors défini sur TM (ou $TM \times [t_1, t_2]$), où TM désigne le fibré tangent à une variété M. Les calculs en géométrie courbée demandent cependant une certaine familiarité avec la géométrie différentielle ; nous aurons l'occasion d'en reparler.

Pour conclure cette section, il reste juste à démontrer le Lemme 106.

Preuve du Lemme 106. En considérant des vecteurs $z(t)$ dont seule une composante est non nulle, on se ramène à démontrer l'énoncé séparément pour chaque composante de z ; il suffit donc de traiter le cas $n = 1$.

Par le théorème d'approximation de Weierstrass, on peut approcher uniformément w, sur l'intervalle $[t_1, t_2]$, par une suite de fonctions polynomiales, a fortiori C^∞. Pour tout $\eta > 0$ il existe donc $z_\eta : [t_1, t_2] \to \mathbb{R}$ telle que $\sup |w - z_\eta| \leq \eta$.

Soit maintenant une fonction plateau : une fonction χ_δ de classe C^∞ (disons), qui prend toutes ses valeurs entre 0 et 1, identiquement égale à 0 dans un voisinage de t_1 et de t_2, et identiquement égale à 1 dans le sous-intervalle $[t_1 + \delta, t_2 + \delta]$, où δ est assez petit. On a ainsi $|z_\eta - z_\eta \chi_\delta| \leq \|z_\eta\| \, 1_{|t-t_i| \leq \delta}$. Finalement

$$\begin{aligned}\left| \int |w(t)|^2 \, dt - \int w(t) z_\eta(t) \, dt \right| &\leq |t_1 - t_2| \, \|w - z_\eta\| \, (\|w\| + \eta) \\ &\leq |t_1 - t_2| \eta (\|w\| + \eta);\end{aligned}$$

$$\left| \int w(t) z_\eta(t) \, dt - \int w(t) z_\eta(t) \chi_\delta(t) \right| \leq 2 \, \delta \|w\| \, (\|w\| + \eta).$$

Or par hypothèse $\int w(t) z_\eta(t) \chi_\delta(t) \, dt = 0$, puisque la fonction $z_\eta \chi_\delta$ vérifie les propriétés attendues. En notant $M = \|w\|_\infty = \sup |w|$, on a donc

$$\left| \int |w(t)|^2 \, dt \right| \leq |t_1 - t_2| \, M\eta + 2M(M + \eta)\delta.$$

On fait alors tendre η et δ vers 0 pour obtenir $\int |w|^2 = 0$, ce qui prouve bien que $w = 0$. □

6.3. Du lagrangien au hamiltonien

Dans cette section, nous allons étudier la transformation d'une équation d'Euler–Lagrange (105) en équation hamiltonienne, dès lors que le lagrangien a une structure particulière. On commence avec deux définitions.

DÉFINITION 109 (impulsion). Soit $L = L(x, v)$ un lagrangien de classe C^1 sur $\mathbb{R}^n \times \mathbb{R}^n$. On appelle

$$p = p(x, v) = \frac{\partial L}{\partial v}(x, v)$$

l'impulsion associée au lagrangien L et à x, v.

On pense à p comme à la différentielle de L dans la variable v ; on peut aussi y penser comme à un vecteur $\nabla_v L(x, v)$.

DÉFINITION 110 (stricte convexité). On dit que le lagrangien $L = L(x, v)$ est strictement convexe en v si

(i) $v \longmapsto L(x, v)$ est strictement convexe, pour tout x, c'est à dire que

$$L(x, (1-t)v + tw) \leq (1-t)L(x, v) + tL(x, w),$$

avec égalité stricte si $0 < t < 1$ et $v \neq w$;

(ii) $L(x, v)$ est superlinéaire en v, uniformément en x, c'est à dire

$$\lim_{|v| \to \infty} \frac{\inf_x L(x, v)}{|v|} = +\infty.$$

On préfère souvent qualifier la propriété (ii) de "surlinéarité" ou "superlinéarité", mais cela simplifiera la terminologie que de l'imposer dans la définition de la stricte convexité. On montre que si une fonction φ est C^1 et strictement convexe, alors *son gradient $\nabla\varphi$ est une bijection continue* $\mathbb{R}^n \to \mathbb{R}^n$.

THEORÈME 111 (Équivalence lagrangien-hamiltonien). *Soit $L = L(x, v)$ un lagrangien de classe $C^2(\mathbb{R}^n \times \mathbb{R}^n; \mathbb{R})$. Si L est strictement convexe en v, le changement de variables $(x, v) \longmapsto (x, p)$ transforme l'équation d'Euler–Lagrange* (105) *en l'EDO équivalente hamiltonienne associée au hamiltonien*

$$H(x, p) = L^*(x, p) = \sup_{v \in \mathbb{R}^n} \Big(\langle p, v \rangle - L(x, v) \Big).$$

REMARQUES 112. 1. L'opération $\varphi \longmapsto \varphi^*$, où

$$\varphi^*(p) := \sup_v \Big(\langle p, v \rangle - \varphi(v) \Big),$$

est appelée **transformée de Legendre**.

2. En supposant que L est de classe C^2 on garantit que le changement de variables est de classe C^1 ; mais l'équivalence reste vraie sous des hypothèses de régularité moins fortes. En revanche, l'hypothèse de stricte convexité est essentielle.

3. Le même théorème reste vrai, sans changement, si L est défini sur $\mathbb{T}^n \times \mathbb{R}^n$ plutôt que $\mathbb{R}^n \times \mathbb{R}^n$.

Avant de prouver le Théorème 111, on rappelle les principales propriétés de la transformée de Legendre, opération fondamentale en théorie des fonctions convexes [**32**] où elle joue le même rôle que la transformée de Fourier en théorie de l'intégration.

D'abord la définition : pour toute fonction $\varphi(v)$ à valeurs réelles, la fonction $\varphi^*(p) = \sup(\langle p, v\rangle - \varphi(v))$ est convexe, en tant que supremum de fonctions affines. Elle peut, en général, prendre la valeur $+\infty$; mais si φ est strictement convexe alors $\varphi^*(p)$ est toujours un nombre réel bien défini. De plus, la fonction $p \cdot v - \varphi(v)$ est alors strictement concave et tend vers $-\infty$ quand $|v| \to +\infty$; elle a donc exactement un maximum, qui correspond aussi au point d'annulation de la différentielle ; comme $d(p \cdot v - \varphi(v)) = p - d\varphi(v)$, ce maximum est obtenu pour

$$p = d\varphi(v).$$

On peut voir p ici comme une (forme) différentielle, c'est à dire l'application qui à z associe $d\varphi(v) \cdot z$; mais on peut aussi le voir comme un vecteur, qui est le gradient de φ en v, soit le vecteur des dérivées partielles. Tant que l'on travaille sur l'espace euclidien $\mathbb{R}^n$, cette identification n'a aucune importance ! Le crochet $\langle p, v\rangle$ pourra ainsi être interprété soit comme un crochet de dualité, soit comme un simple produit scalaire.

Voici l'interprétation géométrique de la transformée de Legendre : si l'on fixe la pente p, on cherche v tel que $d\varphi(v) = p$; alors $\varphi^*(p) = \langle p, v\rangle - \varphi(v)$ est la différence entre la fonction affine de pente p passant par l'origine et la fonction φ.

L'inégalité de Young,

$$\langle p, v\rangle \leq \varphi(v) + \varphi^*(p)$$

découle immédiatement de la définition ; en outre, il y a égalité si et seulement si $p = d\varphi(v)$. On dit que v et p sont alors conjugués.

En partant à nouveau de l'inégalité de Young, on trouve que $\varphi(v) \geq \langle p, v\rangle - \varphi^*(p)$, pour tout p, et donc $\varphi(v) \geq \sup_p[\langle p, v\rangle - \varphi^*(p)]$. On trouve l'égalité pour $p = d\varphi(v)$, ce qui prouve finalement

$$\varphi(v) = \sup_{p \in \mathbb{R}^n} \Big(\langle p, v\rangle - \varphi^*(p) \Big). \tag{106}$$

En d'autres termes,

$$\varphi^{**} = \varphi.$$

En outre, l'identité $p = d\varphi(v)$ montre que les fonctions $\nabla\varphi$ et $\nabla\varphi^*$ sont inverses l'une de l'autre ; en particulier, $\nabla\varphi^*$ est une bijection continue $\mathbb{R}^n \to \mathbb{R}^n$. On dit que la transformée de Legendre réalise une *dualité* sur l'ensemble des fonctions C^1 strictement convexes.

En fait cette relation de dualité s'étend dans un cadre nettement plus général : elle s'applique à des fonctions convexes quelconques (strictement convexes ou non, différentiables ou non), pourvu que l'on autorise ces fonctions à prendre la valeur $+\infty$, et que l'on impose la semi-continuité inférieure. Mais dans ce cours nous nous limiterons au cadre de la stricte convexité C^1.

Ces rappels étant faits, comment exploiter le formalisme de la transformée de Legendre dans le problème qui nous intéresse ? On peut, bien sûr, appliquer la transformée de Legendre (dans la variable v) à la fonction L, pour tout x ; et ainsi définir $L^*(x,p) = L(x,\cdot)^*(p)$. On peut alors définir $p = \partial L/\partial v$, de sorte que l'équation d'Euler–Lagrange (105) devient

$$\frac{dp}{dt} = \frac{\partial L}{\partial x}(x,v), \tag{107}$$

et il faut exprimer le membre de droite en fonction de L^*.

Repartons de l'identité de Young,

$$L(x,v) + L^*(x,p) = \langle p, v\rangle. \tag{108}$$

Pour trouver une relation entre les dérivées en x, on dérive cette identité. Attention : il faut se souvenir que p dans cette identité dépend de x et v. Plus explicitement,

$$L(x,v) + L^*(x,p(x,v)) = \langle p(x,v), v\rangle.$$

D'où

$$\frac{\partial L}{\partial x}(x,v) + \frac{\partial L^*}{\partial x}(x,p) + \frac{\partial L^*}{\partial p}(x,p)\cdot\frac{\partial p}{\partial x} = \left\langle \frac{\partial p}{\partial x}, v\right\rangle.$$

Dans le dernier terme du membre de gauche, nous avons $\partial_p L^*(x,p(x,v)) = \partial_p L^*(x,\partial_v L(x,v))$, qui vaut v puisque $\partial_p L^*$ et $\partial_v L$ sont inverses l'une de l'autre. Ce dernier terme à gauche se simplifie donc avec le terme de droite, et l'on trouve

$$\frac{\partial L}{\partial x}(x,v) + \frac{\partial L^*}{\partial x}(x,p) = 0 \tag{109}$$

(exactement comme si l'on avait dérivé dans (108) en oubliant la dépendance de p en x !). L'équation (107) est donc équivalente à

$$\frac{dp}{dt} = -\frac{\partial L^*}{\partial x}(x,p). \tag{110}$$

Il reste à exprimer l'équation $dx/dt = v$ en fonction de p. Cela s'obtient en disant à nouveau que $\partial_v L$ et $\partial_p L^*$ sont inverses l'une de l'autre : $v = (\partial_v L)^{-1}(p) = (\partial_p L^*)(p)$. En d'autres termes,

$$\frac{dx}{dt} = \frac{\partial L^*}{\partial p}(x,p). \tag{111}$$

La combinaison de (111) et (110) montre bien que le système est hamiltonien dans les variables (x,p) avec hamiltonien L^*.

Exemple 113. Posons $L(x,v) = m|v|^2/2$ (énergie cinétique). Alors la transformée de Legendre vaut $L^*(x,p) = |p|^2/(2m)$ (la transformée de Legendre d'une fonction quadratique est une fonction quadratique), et l'impulsion $p = \nabla_v L = mv$ coïncide avec la "quantité de mouvement".

Exemple 114. Soit $L(x,v) = |v|^4/4$; alors la transformée de Legendre vaut $H = |p|^{4/3}/(4/3)$. On note que H n'est pas très régulier : il est de classe C^1, comme il se doit, mais pas deux fois différentiable, alors que L est de classe C^∞, et même analytique. Ce manque de différentiabilité de H est lié au fait que L "manque de stricte convexité" : son graphe est en effet très plat près de $v = 0$. De manière générale, la transformée de Legendre établit une dualité entre les propriétés de convexité et les propriétés de régularité.

Exemple 115. Soit $L(x,v) = m|v|^2/2 - V(x)$; alors $H(x,p) = |p|^2/(2m) + V(x)$, qui n'est autre que l'énergie totale associée à l'équation de Newton $m\ddot{x} = -\nabla V(x)$.

6.4. Transformations symplectiques

Que dire maintenant des changements de variables ?

Définition 116 (Transformation canonique). On appelle transformation symplectique, ou transformation canonique, un changement de variable $(x,p) \to (X,P)$ qui préserve la structure de flot hamiltonien.

Explicitement, une transformation canonique est un changement de variables tel qu'un système hamiltonien défini dans les anciennes variables est aussi un système hamiltonien défini dans les nouvelles

variables. Autrement dit, on souhaite que le système

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p} \\ \dot{p} = -\dfrac{\partial H}{\partial x} \end{cases}$$

soit équivalent au système

$$\begin{cases} \dot{X} = \dfrac{\partial H}{\partial P} \\ \dot{P} = -\dfrac{\partial H}{\partial X}, \end{cases}$$

si le nouveau hamiltonien $H(X, P)$ est obtenu à partir de l'ancien $H(x, p)$ simplement en effectuant le changement de variable dans la fonction :

$$H(X(x, p), P(x, p)) = H(x, p).$$

(Ici on utilise par abus de notation le même symbole H pour le nouveau et l'ancien hamiltonien.)

Reste à savoir quels sont ces changements de variables préservant la structure hamiltonienne. On cherche une condition qui s'applique pour tout hamiltonien H !

Commençons par effectuer quelques calculs en dimension 2, dans $\mathbb{R} \times \mathbb{R}$. On effectue un changement de variables $X = X(x, p)$, $P = P(x, p)$, supposé bijectif et de classe C^1.

Le hamiltonien dans les variables (x, p) est obtenu en composant le hamiltonien dans les variables (X, P) par le changement de variables : c'est $H(X(x, p), P(x, p))$. Par dérivation des fonctions composées ("chain-rule"),

$$\dot{X} = \frac{\partial X}{\partial x}\,\dot{x} + \frac{\partial X}{\partial p}\,\dot{p}, \qquad \dot{P} = \frac{\partial P}{\partial x}\,\dot{x} + \frac{\partial P}{\partial p}\,\dot{p}. \tag{112}$$

On suppose le système hamiltonien dans les variables (x, p), de sorte que

$$\dot{x} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{\partial H}{\partial x}. \tag{113}$$

Enfin, par chain-rule à nouveau,

$$\frac{\partial H}{\partial x} = \frac{\partial H}{\partial X}\frac{\partial X}{\partial x} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial x}, \qquad \frac{\partial H}{\partial p} = \frac{\partial H}{\partial X}\frac{\partial X}{\partial p} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial p}. \tag{114}$$

On combine (112), (113) et (114) pour trouver

$$\dot{X} = \frac{\partial X}{\partial x}\left(\frac{\partial H}{\partial X}\frac{\partial X}{\partial p} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial p}\right) - \frac{\partial X}{\partial p}\left(\frac{\partial H}{\partial X}\frac{\partial X}{\partial x} + \frac{\partial H}{\partial P}\frac{\partial P}{\partial x}\right). \tag{115}$$

Deux termes ont le bon goût de se simplifier, et il reste

$$\dot{X} = \frac{\partial H}{\partial P}\left(\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x}\right);$$

on en déduit que pour garantir $\dot{X} = \partial H/\partial P$ il faut imposer

$$\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x} = 1. \tag{116}$$

De manière symétrique, on trouve

$$\dot{P} = -\frac{\partial H}{\partial X}\left(\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x}\right);$$

et pour garantir $\dot{P} = -\partial H/\partial X$, on retrouve la même condition (116). On peut la réécrire au moyen d'un crochet de Poisson : $\{X, P\} = 1$.

Finalement, nous avons obtenu la

Proposition 117. *Un changement de variables $(x, p) \to (X, P)$ sur $\mathbb{R}^2$, de classe C^1, est canonique si et seulement si $\{X, P\} = 1$, c'est à dire*

$$\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x} = 1.$$

Remarque 118. On aurait pu également exprimer (x, p) en fonction de (X, P), supposer le système hamiltonien dans les variables (X, P) et demander à ce qu'il le soit aussi dans les variables (x, p) ; on aurait trouvé la même condition. En fait la propriété de transformation canonique est invariante par inversion (si un changement de variable est canonique, son inverse l'est aussi) ; cela sera clair par la Proposition 121 par exemple. En conséquence, on pourrait aussi remplacer la propriété ci-dessus par la propriété "réciproque"

$$\frac{\partial x}{\partial X}\frac{\partial p}{\partial P} - \frac{\partial x}{\partial P}\frac{\partial p}{\partial X} = 1,$$

où l'on note $(x(X, P), p(X, P))$ l'inverse de $(X(x, p), P(x, p))$.

Passons maintenant à la dimension $2n$. On peut refaire tous les calculs "avec des indices" : on trouve alors, à la place de (115),

$$\dot{X}^i = \sum_{jk} \frac{\partial X^i}{\partial x^j}\left(\frac{\partial H}{\partial X^k}\frac{\partial X^k}{\partial p_j} + \frac{\partial H}{\partial P_k}\frac{\partial P_k}{\partial p_j}\right) - \frac{\partial X^i}{\partial p_j}\left(\frac{\partial H}{\partial X^k}\frac{\partial X^k}{\partial x^j} + \frac{\partial H}{\partial P_k}\frac{\partial P_k}{\partial x^j}\right) \tag{117}$$

... et cette fois la simplification ne se produit plus ! On obtient

• des termes en H_{X^k} (on note la variable de dérivation en indice), qui valent

$$\sum_j \left(\frac{\partial X^i}{\partial x^j} \frac{\partial X^k}{\partial p_j} - \frac{\partial X^i}{\partial p_j} \frac{\partial X^k}{\partial x^j} \right),$$

et devraient tous être nuls (c'était automatique pour $n = 1$, mais cela ne l'est plus) ;

• des termes en H_{P_k}, qui valent

$$\sum_j \left(\frac{\partial X^i}{\partial x^j} \frac{\partial P^k}{\partial p_j} - \frac{\partial X^i}{\partial p_j} \frac{\partial X^k}{\partial x^j} \right),$$

et devraient donc être tous nuls, sauf quand $k = i$. Alors on aura effectivement $\dot{X}^i = -H_{p_i}$.

On effectue un calcul similaire pour l'équation sur $\dot{P}_i$, et on obtient les mêmes conclusions, quitte à permuter X et P. Finalement, on a prouvé la

PROPOSITION 119. *Un changement de variables $(x, p) \to (X, P)$ sur $\mathbb{R}^n$, de classe C^1, est canonique si et seulement si, pour tous indices i et k dans $\{1, \dots, n\}$,*

$$\{X^i, X^k\} = 0, \qquad \{P_i, P_k\} = 0, \qquad \{X^i, P_k\} = \delta^i_k, \tag{118}$$

où δ^i_k vaut 1 si $i = k$, et 0 sinon. Plus explicitement,

$$\begin{cases} \displaystyle\sum_j \left(\frac{\partial X^i}{\partial x^j} \frac{\partial X^k}{\partial p_j} - \frac{\partial X^i}{\partial p_j} \frac{\partial X^k}{\partial x^j} \right) = 0 \\ \displaystyle\sum_j \left(\frac{\partial P_i}{\partial x^j} \frac{\partial P_k}{\partial p_j} - \frac{\partial P_i}{\partial p_j} \frac{\partial P_k}{\partial x^j} \right) = 0 \\ \displaystyle\sum_j \left(\frac{\partial X^i}{\partial x^j} \frac{\partial P^k}{\partial p_j} - \frac{\partial X^i}{\partial p_j} \frac{\partial X^k}{\partial x^j} \right) = \delta^i_k. \end{cases}$$

REMARQUE 120. Le lecteur se demande peut-être pourquoi on a toujours mis les indices en haut pour les variables x et X, et en bas pour les variables p et P. C'est une habitude calculatoire bien commode de géométrie différentielle, consistant à noter en haut les indices des vecteurs tangents, et en bas ceux des vecteurs cotangents (formes différentielles). Quand on s'y prend bien, les sommations sur des indices répétés se font toujours avec l'appariement d'un indice en bas et d'un indice en haut : $\sum p_i v^i$ représente la dualité $\langle p, v \rangle$ entre la forme linéaire p et le vecteur tangent v. (Souvent d'ailleurs, on omet la sommation et on considère qu'elle est implicite sur les indices répétés : c'est la "convention d'Einstein" ; ainsi $p_i v^i$ est une abréviation de $\sum p_i v^i$.)

Si l'on ne fait pas d'erreur de calcul, les indices répétés apparaissent alors toujours une fois en haut et une fois en bas (un indice en bas au dénominateur compte pour un indice en haut, et vice versa) ; cela clarifie le sens des expressions et permet de repérer plus facilement des erreurs de calcul. On note que, pour garder la cohérence de la convention, on écrit indifféremment $\delta_{ij} = \delta_i^j = \delta^{ij}$.

Nous allons maintenant réinterpréter la condition canonique d'une manière plus géométrique, en considérant l'action du changement de variable sur la forme quadratique ω, dite *forme symplectique*, définie par

$$\omega\big((\xi,\pi),(\xi',\pi')\big) = \sum_{j=1}^{n}\big(\xi_j\,\pi'_j - \xi'_j\,\pi_j\big). \tag{119}$$

Proposition 121. *Un changement de variables $z = (x,p) \to Z = (X,P)$, de classe C^1, défini sur $\mathbb{R}^n \times \mathbb{R}^n$, est une transformation canonique si et seulement si il préserve la forme symplectique : pour tous vecteurs ζ, ζ' dans $\mathbb{R}^n \times \mathbb{R}^n$, $\omega(dZ\zeta, dZ\zeta') = \omega(\zeta,\zeta')$.*

Remarques 122. 1. Ainsi les transformations canoniques sont-elles les analogues des isométries, modulo le remplacement du produit scalaire (géométrie euclidienne) par la forme ω (géométrie symplectique).

2. On peut considérer ζ et ζ' comme des vecteurs tangents à $\mathbb{R}^n \times \mathbb{R}^n$; on pourra alors penser à l'équation ci-dessus comme à $\omega_Z(dZ\zeta, dZ\zeta') = \omega_z(\zeta,\zeta')$. Dans $\mathbb{R}^n \times \mathbb{R}^n$ cela n'a guère d'importance ; quand on travaille dans une géométrie plus générale, ce n'est pas négligeable.

Démonstration. Effectuons la démonstration en dimension 2 pour simplifier. On notera $\zeta = (\xi,\pi)$. Repartons de

$$\omega((\xi,\pi),(\xi',\pi')) = \xi\pi' - \pi\xi'$$

Le changement de variable $z \longmapsto Z$ transforme les vecteurs (tangents) ξ, π, de sorte que $\omega(dZ\zeta, dZ\zeta')$ vaut

$$\left(\frac{\partial X}{\partial x}\,\xi + \frac{\partial X}{\partial p}\,\pi\right)\left(\frac{\partial P}{\partial x}\,\xi' + \frac{\partial P}{\partial p}\,\pi'\right) - \left(\frac{\partial P}{\partial x}\,\xi + \frac{\partial P}{\partial p}\,\pi\right)\left(\frac{\partial X}{\partial x}\,\xi' + \frac{\partial X}{\partial p}\,\pi'\right)$$
$$= (\xi\pi' - \xi'\pi)\left(\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial P}{\partial x}\frac{\partial X}{\partial p}\right),$$

et l'on retrouve la condition (116). □

Remarque 123. En fait, si l'on exprime les flots hamiltoniens à l'aide de la forme symplectique, on se convainc facilement que les propriétés "préserver les flots hamiltoniens" et "préserver la forme symplectique" sont équivalentes.

Nous avons maintenant bien compris les conditions que doit vérifier une transformation canonique, mais nous ne sommes pas encore convaincus que cette notion soit vraiment utile : les conditions (118), qui sont au nombre de $3n^2 - n$, sont peut-être si contraignantes que l'on aura bien du mal à les satisfaire toutes en même temps. Pourtant, il est assez facile de construire des transformations canoniques avec des recettes appropriées. La proposition suivante permet d'en construire des exemples à volonté : elle affirme que tout flot hamiltonien est une transformation canonique !

THEORÈME 124. *Soient $H \in C^2(\mathbb{R}^n \times \mathbb{R}^n; \mathbb{R})$ un hamiltonien engendrant un flot (Φ_t) sur un intervalle de temps $[t_1, t_2]$ contenant 0. Alors, pour tout $t \in [t_1, t_2]$, Φ_t est une transformation canonique.*

DÉMONSTRATION. Nous allons vérifier que

$$\frac{d}{dt}\Big[\omega\big(d\Phi_t(\sigma), d\Phi_t(\sigma')\big)\Big] = 0,$$

autrement dit que la variation de ω le long du flot est nulle. Par propriété de semigroupe, il suffira de le vérifier en $t = 0$. Pour faire le calcul, on va donc se contenter de développements limités à l'ordre 1 en temps.

On note $x = x_0$ et $p = p_0$ les valeurs initiales de (x, p), et X, P les valeurs au temps t ; c'est à dire que $\Phi_t(x, p) = (X, P)$. Alors, par les équations (98),

$$X = x + t\nabla_p H(x, p) + o(t), \qquad P = p - t\nabla_X H(x, p) + O(t^2),$$

ou le terme $O(t^2)$ l'est au sens de la topologie C^1.

Faisons le calcul en dimension 2 ; on laisse le cas général en exercice :

$$\begin{aligned}\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x} &\\ = \left(1 + t\,\frac{\partial^2 H}{\partial x \partial p}\right)\left(1 - t\,\frac{\partial^2 H}{\partial p \partial x}\right) - \left(t\,\frac{\partial^2 H}{\partial x \partial x}\right)\left(t\,\frac{\partial^2 H}{\partial p \partial p}\right) + O(t^2)& \\ = 1 + O(t^2).&\end{aligned}$$

On a donc

$$\left.\frac{d}{dt}\right|_{t=0}\left(\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x}\right) = 0,$$

ce qu'il fallait démontrer. □

Une première conséquence du Théorème 124 est la possibilité de construire des transformations canoniques à volonté : il suffit de définir n'importe quel hamiltonien et de considérer le flot correspondant (à condition qu'il n'y ait pas d'explosion prématurée).

Une seconde conséquence est que tout flot hamiltonien admet des invariants géométriques intéressants : il ne préserve pas que le volume (théorème de Liouville), mais aussi la forme symplectique ω.

En fait le théorème de Liouville découle de la préservation de la forme symplectique, car le déterminant (forme infinitésimale du volume), vu comme une forme n-linéaire alternée, se déduit de la forme symplectique :

$$\mathrm{vol} = \omega^{\wedge n} = \omega \wedge \ldots \wedge \omega,$$

où $\wedge$ est une opération appelée *produit extérieur*, qui à deux formes multilinéaires alternées en associe une troisième. (On rappelle qu'une forme multilinéaire est dite alternée si la permutation de ses arguments entraîne une multiplication du résultat par la signature de la permutation ; c'est le cas, bien sûr, du déterminant qui est une n-forme linéaire.) Explicitement, si α est une m-forme alternée et β une n-forme alternée

$$\begin{aligned}(120)\quad &\alpha \wedge \beta(v_1, \ldots, v_{m+n}) \\ &\qquad = \sum_\sigma \varepsilon(\sigma)\, \alpha\big(v_{\sigma(1)}, \ldots, v_{\sigma(m)}\big)\, \beta\big(v_{\sigma(m+1)}, \ldots, v_{\sigma(m+n)}\big)\end{aligned}$$

où σ varie sur l'ensemble des permutations de $m+n$ éléments, et $\varepsilon(\sigma)$ est la signature de σ.

Ainsi, à partir de la préservation de ω on déduit que le flot hamiltonien préserve aussi $\omega \wedge \omega$, $\omega \wedge \omega \wedge \omega$, etc. Parmi cette suite de formes multilinéaires, la dernière non nulle est $\omega^{\wedge n}$ qui coïncide avec le déterminant. (Une forme m-linéaire alternée en dimension n est identiquement nulle si $m > n$, il est donc normal que l'on s'arrête à n.) Par suite, le flot hamiltonien préserve la version intégrale du déterminant, qui est le volume.

6.5. Méthode de la fonction génératrice

On a vu dans le Théorème 124 que tout hamiltonien (c'est à dire toute fonction H) induit une transformation symplectique. Comme on l'a utilisé dans la preuve, au premier ordre en t cette transformation est définie par des dérivées partielles : $X \simeq x + t\,\partial_p H$, $P \simeq p - t\,\partial_x H$. Nous allons maintenant voir comment développer cette idée, avec la méthode dite de la *fonction génératrice*, qui engendre automatiquement des transformations symplectiques via des dérivées partielles.

La méthode permet de passer d'un ancien jeu de variables (x, p) a un nouveau jeu (X, P). Pour cela, on utilise une fonction qui dépend de deux de ces quatre variables. Lesquelles ? On a le choix, et les formules s'adaptent à ce choix. Le formalisme le plus courant, peut-être, consiste

à choisir la première des variables de départ, et la seconde des variables d'arrivée : la fonction dépendra donc de (x, P).

PROPOSITION 125 (Méthode de la fonction génératrice). *On se donne une fonction $S = S(x, P)$ quelconque, telle que*

$$(a) \quad S \in C^2(\mathbb{R}^n \times \mathbb{R}^n; \mathbb{R}); \qquad (b) \quad \det\left(\frac{\partial^2 S}{\partial x\, \partial P}\right) \neq 0;$$

$$(c) \quad \forall x, \quad P \longmapsto \frac{\partial S}{\partial x}(x, P) \text{ est une bijection.}$$

On pose alors

$$p = \frac{\partial S}{\partial x}, \qquad X = \frac{\partial S}{\partial P}. \tag{121}$$

La première équation définit implicitement $P = P(x, p)$; la seconde définit explicitement $X = X(x, p)$.

Écrivons les choses un peu plus en détail pour éviter toute ambiguïté : la première équation $\partial_x S(x, P) = p$ est à voir comme une équation dans la variable P ; les conditions (a), (b) et (c) permettent de résoudre cette équation par application du théorème classique des fonctions implicites. (Les conditions (a) et (b) suffisent localement.) Il suffit ensuite de poser $X(x, p) = \partial_p S(x, P(x, p))$.

Vérifions que la méthode fonctionne ! On peut le faire de manière "abstraite" en vérifiant la propriété de préservation de la forme symplectique, mais on peut aussi le voir directement. Cette deuxième approche est conceptuellement plus simple, mais attention, c'est un exercice plutôt traître en calcul différentiel. On commence encore par $n = 1$. Dérivons les deux formules (121) par rapport à x et p : on trouve les quatre identités

$$1 = \frac{\partial^2 S}{\partial x\, \partial P} \frac{\partial P}{\partial p} \tag{122}$$

$$0 = \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial x\, \partial P} \frac{\partial P}{\partial x} \tag{123}$$

$$\frac{\partial X}{\partial x} = \frac{\partial^2 S}{\partial x\, \partial P} + \frac{\partial^2 S}{\partial P^2} \frac{\partial P}{\partial x} \tag{124}$$

$$\frac{\partial X}{\partial p} = \frac{\partial^2 S}{\partial P^2} \frac{\partial P}{\partial p}. \tag{125}$$

De (124) et (125) on déduit

$$\frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x} = \left(\frac{\partial^2 S}{\partial x\,\partial P} + \frac{\partial^2 S}{\partial P^2}\frac{\partial P}{\partial x}\right)\frac{\partial P}{\partial p} - \frac{\partial^2 S}{\partial P^2}\frac{\partial P}{\partial p}\frac{\partial P}{\partial x}$$
$$= \frac{\partial^2 S}{\partial x\,\partial P}\frac{\partial P}{\partial p},$$

et par (122) cela vaut 1.

La dimension générale est un peu plus compliquée ! On commence par écrire

$$p_i = \frac{\partial S}{\partial x^i}, \qquad X^i = \frac{\partial S}{\partial P_i},$$

d'où, en dérivant par rapport à x^j et p_j,

$$\frac{\partial X^i}{\partial x^j} = \frac{\partial^2 S}{\partial x^j\,\partial P_i} + \frac{\partial^2 S}{\partial P_i\,\partial P_\ell}\frac{\partial P_\ell}{\partial x^j} \tag{126}$$

(sommation implicite sur l'indice ℓ)

$$\frac{\partial X^i}{\partial p_j} = \frac{\partial^2 S}{\partial P_i\,\partial P_\ell}\frac{\partial P_\ell}{\partial p_j} \tag{127}$$

$$\delta_i^j = \frac{\partial^2 S}{\partial x^i\,\partial P_\ell}\frac{\partial P_\ell}{\partial p_j} \tag{128}$$

$$0 = \frac{\partial^2 S}{\partial x^i\,\partial P_\ell}\frac{\partial P_\ell}{\partial x^j} + \frac{\partial^2 S}{\partial x^i\,\partial x^j}. \tag{129}$$

Lançons-nous dans la vérification de $\{X^i, P_k\} = 0$. On calcule comme précédemment, en utilisant cette fois (126) et (127),

$$\sum_j \left(\frac{\partial X^i}{\partial x^j}\frac{\partial P_j}{\partial p_j} - \frac{\partial X^i}{\partial p_j}\frac{\partial P_k}{\partial x^j}\right)$$
$$= \sum_j \left(\frac{\partial^2 S}{\partial x^j\,\partial P_i} + \sum_\ell \frac{\partial^2 S}{\partial P_i\,\partial P_\ell}\frac{\partial P_\ell}{\partial x^j}\right)\frac{\partial P_k}{\partial p_j} - \sum_{j\ell}\frac{\partial^2 S}{\partial P_i\,\partial P_\ell}\frac{\partial P_\ell}{\partial p_j}\frac{\partial P_k}{\partial x^j}$$
$$= \sum_j \frac{\partial^2 S}{\partial x^j\,\partial P_i}\frac{\partial P_k}{\partial p_j} + \sum_\ell \frac{\partial^2 S}{\partial P_i\,\partial P_\ell}\sum_j\left(\frac{\partial P_\ell}{\partial x^j}\frac{\partial P_k}{\partial p_j} - \frac{\partial P_\ell}{\partial p_j}\frac{\partial P_k}{\partial x^j}\right).$$

Le premier terme de cette dernière expression est le produit de la matrice $\partial^2_{xP}S$ par la matrice $\partial_p P$; la relation (128) dit que ces deux matrices sont inverses l'une de l'autre (leur produit dans (128) apparaît dans l'ordre inverse ; mais l'identité $AB = I$ implique $BA = I$!), le résultat est donc δ^i_k. Quant au reste de l'expression, il sera nul si $\{P_\ell, P_k\} = 0$ pour tous ℓ, k. Modulo la preuve de $\{P_\ell, P_k\} = 0$ (qu'il

faut de toute façon établir pour montrer que (X, P) est une transformation canonique), nous avons donc démontré que $\{X^i, P_k\} = \delta^i_k$.

Passons maintenant à la vérification de $\{P_\ell, P_k\} = 0$. Notre problème est que nous n'avons aucune information directe sur les dérivées de P : elles apparaissent toujours multipliées par une matrice telle que S_{PP} ou S_{xP}. L'idée est de profiter de l'inversibilité de S_{xP} : on va multiplier la matrice $(\{P_\ell, P_k\})$ à gauche et à droite par S_{xP} ou sa transposée, et prouver que le résultat est nul ; la conclusion en découlera.

On calcule donc, avec l'aide de (128) et (129), et en sommant implicitement sur (i, j, k),

$$\begin{aligned}\frac{\partial^2 S}{\partial x^\ell\,\partial P_i}\left(\frac{\partial P_i}{\partial x^j}\frac{\partial P_k}{\partial p_j} - \frac{\partial P_i}{\partial p_j}\frac{\partial P_k}{\partial x^j}\right)\frac{\partial^2 S}{\partial x^m\,\partial P_k}& \\ = -\frac{\partial^2 S}{\partial x^\ell\,x^j}\,\delta^j_m + \frac{\partial^2 S}{\partial x^j\,x^m}\,\delta^j_\ell& \\ = -\frac{\partial^2 S}{\partial x^\ell\,x^m} + \frac{\partial^2 S}{\partial x^\ell\,x^m} = 0.&\end{aligned}$$

Il reste enfin à prouver que $\{X^i, X^k\} = \delta^{ik}$; cette identité, un peu plus simple, est laissée en exercice.

6.6. Complément : Formalisme hamiltonien généralisé

Jusqu'à présent c'est dans $\mathbb{R}^n \times \mathbb{R}^n$ que nous avons pris contact avec le formalisme hamiltonien et la dualité Lagrangien/Hamiltonien. Reprenons maintenant la discussion dans un cadre plus général ! Toute cette section peut être vue comme un complément, reposant sur des concepts géométriques plus élaborés ; mais il est bon d'en avoir conscience, ne serait-ce que superficiellement.

L'espace des phases d'un système lagrangien est en général l'ensemble des couples (x, v), où la variable x, que nous appellerons "position" par convention, vit dans un espace géométrique qui peut en général être une variété différentiable M, et où la variable v est un vecteur tangent à M en x. On dit que v appartient à l'espace tangent T_xM. L'espace T_xM est aussi appelé la *fibre tangente* en x, et x est le point base de la fibre. En résumé, le lagrangien est une fonction définie sur le fibré tangent TM, à valeurs réelles.

Pour chaque x on applique la transformée de Legendre dans la variable v : cela donne une nouvelle variable p, que l'on peut voir comme une forme linéaire sur l'espace tangent. On écrit alors

$$L^*(x, p) = \sup_{v \in T_xM} \Big(\langle p, v \rangle - L(x, v) \Big),$$

où $\langle p, x\rangle = p(x)$ est le crochet de dualité entre la forme linéaire p et le vecteur (tangent) v. L'ensemble des paires (x, p), où p est une forme linéaire sur T_xM, est appelé le *fibré cotangent*, et on le note T^*M. Ainsi le hamiltonien H est une fonction définie sur le fibré cotangent, à valeurs réelles.

Pourquoi sommes-nous, soudainement, si soucieux de bien faire la distinction entre forme linéaire et vecteur tangent, alors que par exemple, dans les rappels sur la transformée de Legendre, nous passions allégrement de la forme linéaire au gradient ? La raison est simple et subtile à la fois : jusqu'à présent nous n'avions qu'un espace $\mathbb{R}^n$ à gérer, l'espace euclidien, avec son produit scalaire fixé une fois pour toutes ; mais maintenant que l'on considère le fibré tangent, avec un espace tangent qui varie en fonction de x, on peut être amené à faire des produits scalaires qui varient d'un point à l'autre. C'est le principe de la géométrie riemannienne, où chaque espace tangent vient avec son propre produit scalaire ! Comme l'identification entre les vecteurs et les formes linéaires dépend du choix du produit scalaire, on n'a plus de correspondance a priori bien définie entre les uns et les autres. On préfère alors revenir aux formes linéaires qui, elles, sont intrinsèquement définies indépendamment du choix des produits scalaires.

Cependant, il est fréquent que l'on travaille sur une variété M qui est non seulement une variété différentiable, mais aussi une variété riemannienne, c'est à dire munie d'un produit scalaire défini sur chaque espace tangent et variant de manière lisse en fonction du point base x. Alors, on peut réaliser sans ambiguïté l'identification entre forme linéaire et vecteur (via $\langle p, v\rangle = \langle p, v\rangle_x$, où l'on note avec un indice x le produit scalaire dans T_xM), et définir un hamiltonien comme une fonction $TM \to \mathbb{R}$:

$$H(x,p) = \sup_{v \in T_xM} \Big(\langle p, v\rangle_x - L(x,v) \Big).$$

Revenons à la fonction H définie sur le fibré cotangent. Que se passe-t-il quand on la dérive ? On peut toujours définir dH comme une forme différentielle sur T^*M (donc un élément de $T^*(T^*M)$!). Dans un cadre riemannien, on peut aussi voir H comme une fonction définie sur TM, et remplacer dH par un vecteur tangent ∇H (un élément de $T(TM)$...) Comme H dépend de x et p, on peut distinguer deux composantes : $\partial_x H$ et $\partial_p H$. Il semble ainsi que ce soit juste une question de calculer les dérivées composante par composante.

En fait les choses ne sont pas si simples : car s'il est facile de "dériver en p après avoir gelé x" (opération qui se passe dans un espace vectoriel fixé, T_xM), il est en revanche délicat de "dériver en x après avoir gelé

p", car il faut pour cela établir une correspondance entre les différents espaces tangents (une "connection"). On utilise d'ordinaire la connection déduite des géodésiques : allant d'un point x à un point proche y selon une courbe de plus courte distance, on identifie la vitesse en y et la vitesse en x. Les géomètres distinguent alors la dérivation "verticale" ∂/∂_p qui s'effectue dans une fibre, en fixant x et en faisant varier p, et la dérivation "horizontale" ∂/∂_x qui utilise la correspondance entre fibres infinitésimalement proches... Un casse-tête ? En pratique il y a des recettes simples pour effectuer les calculs, où l'on retrouve d'ailleurs les symboles de Christoffel de (10).

On peut aussi reproduire toute la Section 6.2 dans ce cadre géométrique, pourvu que ∂_v soit la dérivation verticale dans la fibre tangente, ∂_x la dérivation horizontale, et d/dt (appliqué à un champ de vecteurs tangent) la "dérivation covariante".

La situation la plus simple est celle où le fibré est tout simplement un produit ; au-delà de $\mathbb{R}^n$, cela se produit quand l'espace géométrique est le tore $\mathbb{T}^n$: alors le fibré tangent et le fibré cotangent s'identifient tout deux à $\mathbb{T}^n \times \mathbb{R}^n$, et il devient évident de distinguer la composante verticale de la composante horizontale !

La dernière étape est le lien entre la dérivée de H et le flot hamiltonien. Pour obtenir le champ hamiltonien, il faut encore appliquer à ∇H un opérateur linéaire, J, que nous pouvons identifier à sa matrice dans l'espace tangent à T^*M :

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Que dire de l'opérateur J ? D'abord, il vérifie $J^2 = -I$, et évoque donc le nombre complexe i. Ensuite, on peut le considérer comme une forme bilinéaire antisymétrique : si on applique J à des vecteurs tangents (ξ, π) et (ξ', π') en (x, p), alors on peut écrire

$$J\big((\xi, \pi), (\xi', \pi')\big) = \sum_j \big(\xi_j \, \pi'_j - \xi'_j \, \pi_j\big). \tag{130}$$

On appelle *forme symplectique* une telle forme bilinéaire antisymétrique – plus précisément, une forme bilinéaire, définie sur un fibré tangent $T\Sigma$, variant de manière lisse avec le point base, qui dans un jeu de coordonnées approprié peut s'écrire sous la forme (130). Il est traditionnel d'utiliser la notation ω pour une telle forme. Le mot "symplectique", issu du grec ancien, signifie "complexe", et renvoie à l'identité $J^2 = -I$.

Les géomètres emploient une définition équivalente mais un peu différente, plus intrinsèque dans l'esprit : ils disent qu'une forme symplectique est une forme bilinéaire ω définie sur un fibré tangent $T\Sigma$ (il

faut penser à Σ comme T^*M), variant de manière lisse, antisymétrique, non dégénérée et fermée. La condition de non-dégénérescence signifie que le noyau de ω est nul, c'est à dire que pour tout $\sigma \neq 0$ il existe z tel que $\omega(\sigma, z) \neq 0$. Quant à la condition de fermeture, elle veut dire que $d\omega = 0$, où d est l'opérateur de différentielle extérieure, qui est une façon intrinsèque de dériver une k-forme en une $(k+1)$-forme :

$$d\alpha(v_0, \dots, v_k) = \sum_i d[\alpha(v_1, \dots, \widehat{v_i}, \dots, v_k)] \cdot v_i \\ + \sum_{i \neq j} (-1)^{i+j} \alpha\big([v_i, v_j], \dots, \widehat{v_i}, \dots, \widehat{v_j}, \dots, v_n\big),$$

où la notation $\widehat{v_j}$ signifie que l'on supprime cette composante des arguments. La condition de fermeture est satisfaite si et seulement si ω est localement la différentielle extérieure d'une 1-forme. Cette formulation ne fait pas référence à un jeu de coordonnées particulier, et ne nécessite pas non plus de savoir que Σ a la forme d'un fibré cotangent.

La forme symplectique habituelle avec laquelle nous avons travaillé jusqu'ici, ω dans (119), est la différentielle extérieure de la "forme tautologique" Θ :

$$\Theta_{(x,p)}(\xi, \pi) = \langle p, \xi \rangle_x. \tag{131}$$

(Θ, souvent notée $p\,dq$, est la forme différentielle la plus simple que l'on puisse imaginer sur T^*M : la composante horizontale du vecteur tangent est un vecteur tangent à M, auquel on peut appliquer la forme linéaire p).

L'intérêt d'avoir une forme bilinéaire est qu'elle identifie les vecteurs tangents et les formes ; comme le fait une métrique riemannienne. On peut donc à ω associer un opérateur de "gradient symplectique", que nous noterons ∇^S, via l'identité

$$\omega(\nabla^S H, \sigma) = \langle dH, \sigma \rangle, \tag{132}$$

le crochet désignant encore l'opération de dualité. (On rencontre aussi la convention avec le signe opposé – cela n'a pas d'importance !) Calculons $\nabla^S H$ dans les coordonnées (130) : en notant $(g_{x_i})_{1 \leq i \leq n}$ et $(g_{p_i})_{1 \leq i \leq n}$ les composantes "verticales" et "horizontales", respectivement, de $\nabla^S H$, et de même $(\sigma_{x_i})_{1 \leq i \leq n}$ et $(\sigma_{p_i})_{1 \leq i \leq n}$ les composantes de σ, on réécrit (132) en

$$\sum_i \big(g_{x_i} \sigma_{p_i} - g_{p_i} \sigma_{x_i}\big) = \sum_i \big(H_{x_i} \sigma_{x_i} + H_{p_i} \sigma_{p_i}\big),$$

d'où l'on conclut que $g_{x_i} = H_{p_i}$ (dérivée de H dans la variable p_i) et $g_{p_i} = -H_{x_i}$ (dérivée de H dans la variable x_i). Finalement, $\nabla^S H$ s'identifie à $J \nabla H$, ce qui est le champ de vecteurs associé au hamiltonien

H ! Et le flot hamiltonien devient ainsi

$$\dot{z} = \nabla^S H(z), \qquad z \in \Sigma. \tag{133}$$

Pour parachever la généralisation, on note que le crochet de Poisson se réexprime simplement en fonction de ω et du gradient symplectique :

$$\{f, g\} = \omega(\nabla^S f, \nabla^S g).$$

Résumons : étant donné une variété Σ, une forme (bilinéaire) symplectique ω sur les espaces tangents à Σ, et une fonction $H : \Sigma \to \mathbb{R}$, on peut définir le gradient symplectique de H, et une équation associée (133), que l'on appellera flot hamiltonien généralisé. Et dans le cas où $\Sigma = T^*M$ et où ω prend la forme (130), on retrouve ainsi le flot hamiltonien "habituel" associé à H.

Finalement le formalisme hamiltonien partage de nombreux points communs avec le formalisme des flots gradients. L'un comme l'autre sont définis par une structure géométrique bilinéaire non dégénérée d'une part, et une fonction d'"énergie" d'autre part ; la différence fondamentale est dans la structure géométrique, qui est symétrique pour les flots gradients, et antisymétrique pour les flots hamiltoniens. On note que l'hypothèse d'antisymétrie, alliée à celle de non-dégénérescence, implique que la dimension de l'espace est forcément paire (une matrice antisymétrique en dimension impaire ne peut être injective !), alors que la dimension de l'espace des phases d'un flot gradient est quelconque. On peut dresser un tableau comparatif plus précis :

flot gradient	flot hamiltonien
espace des phases : variété riemannienne dimension quelconque m	espace des phases : variété symplectique dimension paire $N = 2n$
structure géométrique : forme bilinéaire symétrique g, non dégénérée (produit scalaire)	structure géométrique : forme bilinéaire antisymétrique ω, non dégénérée fermée (forme symplectique)
opérateur gradient ∇ : $g(\nabla f, \xi) = \langle df, \xi \rangle$ (pour tout vecteur tangent ξ)	opérateur gradient symplectique ∇^S : $\omega(\nabla^S f, \sigma) = \langle df, \sigma \rangle$ (pour tout vecteur tangent σ)
isométries	transformations canoniques
énergie E (potentiel) décroissante en temps	énergie H (hamiltonien) constante en temps
équation $\dot{x} = -\nabla E(x)$	équation $\dot{z} = \nabla^S H(z)$
dérivée de f le long du flot : $-\nabla f \cdot \nabla E = -g(\nabla f, \nabla E)$	dérivée de f le long du flot : $\{f, H\} = \omega(\nabla^S f, \nabla^S H)$ (crochet de Poisson avec H)
divergence : ΔE (non-préservation du volume) (non-préservation de g)	divergence nulle préservation du volume préservation de ω

Une remarque supplémentaire, qui empruntera à un formalisme un peu plus élaboré : le parallèle entre les structures géométriques inclut aussi une notion de *volume* qui découle naturellement de la géométrie. Pour la colonne de gauche, c'est le volume riemannien, que l'on peut définir comme $\sqrt{|\det g|}\, dx^1 \wedge \dots dx^m$, où $|\det g|$ est le déterminant de la matrice qui représente la forme bilinéaire g dans le système de coordonnées $(x^1, \dots, x^m)$; de manière plus intrinsèque mais plus savante, le volume riemannien (non orienté) est la mesure de Hausdorff de dimension n associée à la distance géodésique sur (M, g). Pour la colonne de droite, on peut définir à partir de ω une notion naturelle de volume, en considérant $\omega^{\wedge n}$, la n-ème puissance extérieure de ω.

Une dernière remarque : le crochet de Poisson vérifie l'identité de Jacobi :

$$\{f,\{g,h\}\}+\{g,\{h,f\}\}+\{h,\{f,g\}\}=0$$

Cette relation algébrique joue un rôle important en "mécanique symplectique", et on l'impose souvent dans les généralisations (encore plus) abstraites du crochet de Poisson.

6.7. Systèmes hamiltoniens intégrables

En général on ne peut "résoudre" les équations hamiltoniennes ; mais si le hamiltonien H a de nombreuses symétries, les lois de conservation contraindront le mouvement. En utilisant des transformations canoniques adéquates, on peut espérer choisir les invariants comme fonctions coordonnées. Si parmi les $2n$ fonctions coordonnées, il y a une paire de variables conjuguées invariantes (disons x^1 et p_1), on peut alors restreindre le système au sous-espace défini par les valeurs de ces deux variables, et l'on se retrouve avec un système hamiltonien défini dans un espace de dimension $2(n-1)$. En supposant que l'on a tenu compte de cela pour réduire la dimensionalité au maximum, la situation la plus favorable est celle dans laquelle n variables (disons $p_1,\dots,p_n$) sont conservées, mais pas leurs variables conjuguées. Comme alors $\partial_{x^i}H=-\dot{p}_i=0$, cela veut dire que H ne dépend pas des variables $x^1,\dots,x^n$. Cette situation, rare mais pas exceptionnelle, a donné lieu à de nombreux travaux ; on parle alors de système intégrable.

Définition 126. Soit $H=H(x,p)$ un hamiltonien appartenant à $C^1(\mathbb{R}^n\times\mathbb{R}^n;\mathbb{R})$. On dit que le système associé est *complètement intégrable*, ou simplement *intégrable*, s'il existe une transformation canonique $(x,p)\to(X,P)$ et une fonction $H_0:\mathbb{R}^n\to\mathbb{R}$ telle que $H(x,p)=H_0(P)$. Pas abus de langage, on dit aussi que le hamiltonien lui-même est intégrable.

Cette définition s'étend sans problème à $\mathbb{T}^n\times\mathbb{R}^n$. En revanche, sur un fibré tangent quelconque TM, l'espace tangent varie d'un point à l'autre, de sorte que cela n'a pas vraiment de sens de parler d'une fonction dépendant "seulement de p".

Dans le système de coordonnées (X,P) de la Définition 126, les équations du mouvement sont

$$\begin{cases}\dfrac{dP}{dt}=0\\[2ex]\dfrac{dX}{dt}=dH_0(P).\end{cases}\tag{134}$$

On peut alors résoudre les équations "explicitement" :

$$\begin{cases} P_i(t) = P_i(0) \\ X_i(t) = X_i(0) + \omega_i\, t; \qquad \omega_i = \dfrac{\partial H_0}{\partial P_i}(P(0)). \end{cases} \tag{135}$$

Exemple 127. On a reconnu dans les équations précédentes la solution du mouvement libre ; en effet, le hamiltonien d'une particule libre ponctuelle de masse m est $H = |p|^2/(2m)$, qui ne dépend que de p.

Exemple 128. L'exemple le plus fondamental de système intégrable confiné est le mouvement libre sur le tore. Soit $H(p) = |p|^2/2$ sur $\mathbb{T}^n \times \mathbb{R}^n$. Alors l'équation hamiltonienne correspondante est

$$p_i(t) = p_i(0) \qquad x_i(t) = x_i(0) + \omega_i\, t \text{ (modulo 1)}, \qquad \omega_i = p_i.$$

La trajectoire dessinée par le système dans l'espace des positions dépend des fréquences ω_i. Considérons le cas important où $n = 2$ (l'espace des phases est alors de dimension 4 mais le mouvement se fait dans un sous-espace de dimension 2). Pour éviter les situations triviales, on suppose que l'une des fréquences, disons ω_2, est non nulle. Alors

- si ω_1/ω_2 est rationnel, la trajectoire est périodique et dessine une *courbe de Lissajoux*. On dit alors que le vecteur (ω_1, ω_2) est en résonance.
- si ω_1/ω_2 est irrationnel (ce qui se produit avec probabilité 1 quand on choisit ω_1 et ω_2 "au hasard"), alors la trajectoire est dense dans $\mathbb{T}^2$. On dit que le vecteur de fréquences est non résonant, et que le mouvement est *quasipériodique*.

On note que les vecteurs résonants, bien que rares, forment un ensemble dense. Cet exemple montre que même pour un système intégrable, la topologie de la trajectoire dépend de détails très fins des conditions initiales : cette trajectoire peut être dense nulle part (périodicité) ou partout dense (quasipériodicité), selon qu'il y a résonance ou pas ; et l'on passe de l'un à l'autre de manière totalement discontinue.

En dimension n quelconque, la même analyse reste globalement valable ; l'image de la trajectoire peut prendre toutes les dimensions entre 1 et n, en fonction des relations algébriques liant les différentes fréquences ω_j. Dans le cas non résonant où il n'y a aucune relation linéaire à coefficients entiers entre les ω_j, la trajectoire remplit densément le tore.

Un système intégrable donne accès à des invariants, qui dans les notations précédentes sont toutes les variables P_i. Cependant, en pratique c'est souvent l'inverse qui se produit : on cherche à dénicher des

invariants pour en faire des variables commodes. D'où la question : étant donnés des invariants, quand pourra-t-on les choisir pour nouvelles variables ?

Les propriétés suivantes sont invariantes, et seront donc vraies dans tout jeu de coordonnées obtenu par transformation canonique :

- les différentielles des P_i sont linéairement indépendantes ;
- les crochets de Poisson des fonctions P_i entre elles sont tous nuls.

Le théorème de Liouville–Arnold [**34**], que nous énoncerons sans preuve, établit une forme de réciproque.

THEORÈME 129. *Soit H un hamiltonien sur une variété symplectique Σ, par exemple un fibré cotangent T^*M. On suppose qu'il existe n fonctions K_i, de différentielles linéairement indépendantes, telles que $\{K_i, H\} = 0$, $\{K_i, K_j\} = 0$. Alors*

- on peut construire des fonctions $\varphi_1, \dots, \varphi_n$ jouant le rôle de variables conjuguées aux $K_1, \dots, K_n$:

$$\{\varphi_i, \varphi_j\} = 0, \qquad \{K_i, \varphi_j\} = \delta_{ij}.$$

La fonction (φ, K) est alors une transformation canonique et dans ce nouveau jeu de variables l'équation hamiltonienne se réécrit

$$\dot{\varphi}_i = \frac{\partial H}{\partial K_i}, \qquad K_i = \text{const.};$$

- chaque fonction φ_i est uniquement déterminée, à addition près d'un terme $\partial S(K)/\partial K_i$;

- si en outre la sous-variété $\{K_i = \alpha_i\}$ est compacte connexe, alors cet ensemble est difféomorphe au tore $\mathbb{T}^n$ et les fonctions φ_i peuvent faire office de coordonnées, globalement définies sur cette sous-variété.

La première partie de ce théorème nous suggère une nouvelle définition, plus intrinsèque et a priori plus générale, d'un système intégrable : un système hamiltonien qui admet n lois de conservation "suffisamment indépendantes". La dernière partie de ce théorème nous indique alors que, en quelque sorte, "un système intégrable confiné se ramène toujours à $\mathbb{T}^n \times \mathbb{R}^n$".

Plus précisément, supposons que l'on considère un système intégrable, avec une condition initiale qui varie dans un petit ouvert (de Σ). Le vecteur $(K_1, \dots, K_n)$ prend alors ses valeurs dans un petit ouvert U de $\mathbb{R}^n$; et les trajectoires du système intégrable sont dessinées dans des composantes connexes des ensembles $\{(K_1, \dots, K_n) = \alpha\}$, $\alpha \in U$. Ajoutant l'hypothèse de compacité, on peut appliquer le théorème, et l'espace exploré est alors difféomorphe à $\mathbb{T}^n \times U$. Ce paramétrage ne vaut que localement en K, mais il est global dans la variable de temps.

Face à un système intégrable, y a-t-il un jeu de coordonnées meilleur qu'un autre ? Des coordonnées populaires et assez intrinsèques sont les variables dites d'*action–angle*. Soit un système intégrable, défini par un hamiltonien $H(x,p)$; on suppose que l'espace des phases est difféomorphe à $\mathbb{T}^n \times U$, avec U un ouvert de $\mathbb{R}^n$. Fixons ce difféomorphisme, disons Ψ, défini sur $\mathbb{T}^n \times U$. Pour $\alpha \in U$ et pour tout $j \in \{1, \dots, n\}$ on définit C_j comme "une courbe qui enserre le $j$ème facteur de $\mathbb{T}^n$", autrement dit l'image par Ψ du lacet $x \longmapsto (0, \dots, 0, x, 0, \dots, 0, \alpha_1, \dots, \alpha_n)$, avec le x situé en position j. Les C_j constituent une "base de lacets" du tore. On définit alors

$$I_j = \frac{1}{2\pi} \int_{C_j} p \cdot dx = \frac{1}{2\pi} \int_{C_j} \Theta,$$

où Θ est la "forme tautologique" définie par (131). (En dimension $n = 1$, I_1 correspond à l'aire enserrée par la courbe C_1.) On appelle $I_1, \dots, I_n$ les *variables d'action* : elles remplacent les variables α et permettent d'indexer les différents tores.

On définit ensuite $S(x,p) = \int p \cdot dx$, et $\varphi_j = \partial S / \partial I_j$. Les variables $(\varphi_1, \dots, \varphi_n)$ sont appelées *variables d'angle* : chaque φ_j varie dans $\mathbb{R}/\mathbb{Z}$ et fournit une coordonnée sur le tore.

Une fois le formalisme bien mis en place, la recherche de systèmes intégrables devint un sujet important en mécanique classique. Voici quelques exemples célèbres :

- les équations du système solaire képlerien, c'est à dire sans interactions planète–planète. La variable ϕ_i est alors une sorte de "position angulaire" sur l'orbite numéro i... Beaucoup des théorèmes classiques et célèbres d'astronomie ont été réinterprétés en tirant partie de cette propriété d'intégrabilité. Le mouvement est alors, soit périodique, soit quasipériodique, selon que les fréquences du système sont toutes multiples d'une fréquence commune, ou pas.
- À la fin du 19ème siècle, Sofia Kowalevskaya stupéfia les spécialistes de physique mathématique en mettant en évidence un nouveau cas d'intégrabilité dans le système de la toupie (corps en rotation autour d'un point) que l'on croyait bien connaître. Il s'agit de la situation dans laquelle les trois moments principaux d'inertie du solide sont en proportion $(2,2,1)$, et où le centre d'inertie appartient au plan défini par les deux premiers axes principaux. Pour prouver l'intégrabilité, elle exhiba un nouvel invariant, et fit le lien avec certaines fonctions transcendantes étudiées par Abel et Jacobi.
- Le système de Toda est défini en dimension 6 par le hamiltonien

$$H = \left(p_1^2 + p_2^2 + p_3^2\right) + e^{x_1 - x_2} + e^{x_2 - x_3} + e^{x_3 - x_1},$$

ce qui correspond à trois particules sur la droite réelle avec des potentiels d'interaction (asymétriques) exponentiels. On montre en effet que, en sus de H, on a conservation de

$$P = p_1 + p_2 + p_3,$$

$$K = (p_1+p_2-2p_3)(p_2+p_3-2p_1)(p_3+p_1-2p_2) - (p_1+p_2-2p_3)e^{x_1-x_2} \\ - (p_2+p_3-2p_1)e^{x_2-x_3} - (p_3+p_1-2p_2)e^{x_3-x_1},$$

et dH, dP, dK sont linéairement indépendants sur un sous-ensemble ouvert ; en outre $\{K, H\} = \{K, P\} = \{H, P\} = 0$, ce qui permet d'appliquer le Théorème 129.

6.8. Théorie perturbative hamiltonienne

À ce stade nous avons planté le décor des systèmes hamiltoniens : définitions, propriétés algébriques, transformations, etc. Nous avons passé en revue beaucoup de structure, mais on peut légitimement se demander ce que l'on va en faire ! C'est en fait maintenant que commencent les applications : le formalisme hamiltonien est le point de départ de nombreuses questions, parfois théoriques et parfois pratiques, qui emplissent des ouvrages de référence comme ceux d'Arnold [**2**] et Thirring [**34**]. Citons

- la résolution de systèmes intégrables
- l'étude de l'asymptotique de trajectoires, en particulier la correspondance entre états asymptotiques en temps $-\infty$ et en temps $+\infty$ en présence d'une interaction localisée : c'est la théorie du *scattering*, importante pour l'étude de tous les phénomènes faisant intervenir des interactions brusques, collisions entre particules, ondes etc.
- l'étude de bifurcations de familles de systèmes dépendant d'un paramètre.
- l'étude de l'influence qualitative et quantitative de petites perturbations sur le temps long.

Dans cette section et la suivante, nous allons développer ce dernier sujet. Considérons un système mécanique, bien compris, que l'on perturbe par un effet "petit" en un certain sens. Comment quantifier l'effet de cette perturbation ?

Sur un temps fini, on peut appliquer le théorème de dépendance lisse par rapport à un paramètre ; mais quand le temps est grand, tout devient possible. Arnold aimait évoquer l'exemple du seau d'eau avec une petite fuite. Si la fuite est minuscule, en une journée le seau ne s'est presque pas vidé, et le seau d'eau avec petite fuite (en tant que système dynamique !) est très proche du seau d'eau sans fuite. Mais au

bout d'un temps très grand, le seau d'eau qui fuit s'est complètement vidé, et ce, quelle que soit la taille de la fuite ! D'où les questions : a-t-on stabilité en temps grand malgré la perturbation ? et sinon, sur quelle échelle de temps a-t-on stabilité ? Dans le cas du seau, il est clair que l'échelle de temps caractéristique est l'inverse du débit de la fuite (rapporté à la contenance du seau) : sur toute échelle de temps plus petite, on ne verra aucun effet significatif. Que dire alors de systèmes plus généraux ?

Considérons ce problème sous sa forme abstraite hamiltonienne. On se donne H_0 un hamiltonien (une fonction de (x,p)), et εH_1 un autre hamiltonien, représentant la perturbation ; le paramètre $\varepsilon > 0$ sera supposé très petit pour refléter la petitesse de la perturbation. On considère donc le hamiltonien complet $H = H_0 + \varepsilon H_1$, et l'on se pose la question de l'influence du terme supplémentaire en fonction de ε. On suivra dans cette section la présentation de Thirring [**34**].

Pour un premier contact avec le problème, considérons une observable quelconque, f, c'est à dire une fonction lisse sur l'espace des phases. Le problème physiquement pertinent est l'évolution de l'observable $f(x(t),p(t))$ le long du système hamiltonien. On peut effectuer ces calculs à coups de crochets de Poisson :

$$\left.\frac{d}{dt}\right|_{t=0} f \circ \Phi_t = \{f, H\},$$

$$\left.\frac{d^2}{dt^2}\right|_{t=0} f \circ \Phi_t = \{\{f, H\}\},$$

etc. de sorte que l'on peut toujours écrire une "série exponentielle"

$$f \circ \Phi_t = \sum_{k\geq 0} \{\{...\{f,H\},H\},...H\}\, \frac{t^k}{k!}.$$

Si tout est analytique, c'est une formule exacte ; sinon c'est une série formelle.

Les perturbations liées à H_1 peuvent être déduites de cette formule : ainsi,

$$\{f, H\} = \{f, H_0\} + \varepsilon\{f, H_1\},$$

$$\{\{f, H\}, H\} = \{f, H_0\} + \varepsilon\Big(\{\{f, H_0\}, H_1\} + \{\{f, H_1\}, H_0\}\Big) + \varepsilon^2\{\{f, H_1\}, H_1\},$$

etc. L'influence de ε peut ainsi se calculer formellement à tous les ordres.

En général, cette approche n'apporte rien pour ce qui est des ordres de grandeur en temps : le terme ε se retrouve multiplié par des puissances arbitrairement grandes du temps, de sorte que l'on est même incapable d'identifier une échelle de stabilité.

Une façon de voir les choses, a priori plus prometteuse, est de comparer l'évolution liée à H avec celle qui est liée à H_0 ; c'est aussi la philosophie qui inspire la théorie du scattering, où l'on compare l'évolution à une évolution asymptotique "simple". Nous allons donc distinguer le flot Φ_t engendré par H et le flot Φ_t^0 engendré par H_0, et étudier $\Phi_{-t}^0 \circ \Phi_t$.

Faisons un petit aparté calculatoire. Appelons ξ le générateur du flot, c'est à dire l'opérateur de dérivation le long du flot :

$$\xi f = \left.\frac{d}{dt}\right|_{t=0} f \circ \Phi_t = \{f, H\};$$

bien sûr ξf est aussi égal à $df \cdot \xi$, où l'on utilise la même notation ξ pour le champ de vecteurs hamiltonien. L'opérateur ξ commute avec le flot :

$$\xi(f \circ \Phi_s) = (\xi f) \circ \Phi_s.$$

(C'est une conséquence de la formule $(d/ds)\Phi_s = \xi \circ \Phi_s$, et on peut voir cela comme une variante de la formule habituelle $A\, e^{sA} = e^{sA} A$.) Soient maintenant deux flots Φ et Φ', associés à des opérateurs ξ et ξ', que peut-on dire de $f \circ (\Phi_t \circ \Phi'_t)$? L'argument t apparaît à deux reprises, et quand on dérive par rapport au temps il y a donc deux termes, obtenus en gelant respectivement la première et la seconde occurrence de cette variable :

$$\frac{d}{dt} f \circ (\Phi_t \circ \Phi'_t) = \big((\xi f) \circ \Phi_t\big) \circ \Phi'_t + \big(\xi'(f \circ \Phi_t) \circ \Phi'_t\big).$$

(Il est facile de formaliser cela à l'aide du théorème de dérivation des fonctions composées, même s'il est non moins facile de se noyer dans le formalisme.) Attention, dans le premier terme on peut écrire indifféremment $(\xi f) \circ \Phi_t$ ou $\xi(f \circ \Phi_t)$), mais dans le second on ne peut pas a priori écrire $(\xi'(f \circ \Phi_t) = (\xi' f) \circ \Phi_t$! En tout cas nous obtenons finalement une expression raisonnablement compacte pour la dérivée :

$$\frac{d}{dt} f \circ (\Phi_t \circ \Phi'_t) = \Big((\xi + \xi')(f \circ \Phi_t)\Big) \circ \Phi'_t.$$

En particulier, si Φ_t et Φ'_t sont engendrés par des hamiltoniens H et H' respectivement, on trouve

$$\frac{d}{dt} f \circ (\Phi_t \circ \Phi'_t) = \big\{(f \circ \Phi_t), H + H'\big\} \circ \Phi'_t. \tag{136}$$

On applique maintenant cette formule à $H = H_0 + \varepsilon H_1$ et $H' = -H_0$, de façon à isoler le rôle de la perturbation. Le flot associé au hamiltonien $-H_0$ n'est autre que (Φ^0_{-t}), de sorte que finalement

$$\frac{d}{dt} f \circ (\Phi_t \circ \Phi^0_{-t}) = \varepsilon \big\{ (f \circ \Phi_t), H_1 \big\} \circ \Phi^0_{-t}. \tag{137}$$

On peut ensuite intégrer cela et en déduire une série itérative qui donne $f \circ \Phi_t \circ \Phi^0_{-t}$ en puissances de ε. Si g est une fonction quelconque, on notera, pour alléger les notations, $g(t) = g \circ \Phi^0_t$ (c'est l'évolution non perturbée) : alors

$$f \circ \Phi_t = f(t)$$
$$+ \sum_{k \geq 1} \varepsilon^k \int_0^t dt_1 \dots \int_{t_{k-1}}^t dt_k \, \{\{\{\dots\{f(t), H_1(t_k)\}, \dots, H_1(t_2)\}, H_1(t_1)\}.$$

Nous avons ainsi exprimé l'évolution comme une série perturbative en le petit paramètre ε. Si l'évolution est confinée et tout infiniment dérivable, il est raisonnable d'imaginer que l'intégrande du terme d'ordre k ci-dessus est de taille M^k ; supposons que ce soit le cas. Comme la région d'intégration est de taille $t^k/k!$, on aura alors

$$f \circ \Phi_t = f(t) + O\left(\sum_{k \geq 1} \frac{\varepsilon^k M^k t^k}{k!} \right) = f \circ \Phi^0_t + O\big(e^{\varepsilon M t} - 1\big).$$

Cela montre que la perturbation se fait sentir seulement pour des échelles de temps d'ordre au moins $O(1/\varepsilon)$, et permet en outre d'avoir des estimations de la correction.

Jusqu'ici il n'y a rien de bien surprenant : il est tout à fait naturel qu'une perturbation de taille ε entraîne une perturbation importante pour $t = O(\varepsilon^{-1})$. Par ailleurs, jusqu'ici le formalisme hamiltonien n'a entraîné que des simplifications mineures. Nous allons maintenant voir que si le système est *intégrable*, alors la situation change et l'on s'attend à une certaine forme de stabilité sur des échelles de temps bien plus longues ! En fait, nous allons voir que le système reste, en un certain sens, (localement) intégrable à l'ordre ε, de sorte que les corrections au comportement intégrable ne sont attendues qu'à l'ordre $O(\varepsilon^{-2})$. Cet énoncé demande à être nuancé, car il y a des conditions non triviales à respecter ; nous allons expliquer la méthode, qui elle tire très efficacement parti du formalisme hamiltonien.

Considérons donc un hamiltonien intégrable sur $\mathbb{T}^n \times \mathbb{R}^n$, exprimé dans les variables d'action–angle : $H_0 = H_0(I) = H_0(I_1, \dots, I_n)$. La forme des trajectoires dépend des relations algébriques entre les $\omega_j = \partial H_0 / \partial I_j$; dans le cas général non résonant, ces trajectoires sont denses

et dans le tore $\mathbb{T}^n$. Ajoutons ensuite une petite perturbation, a priori non intégrable, prenant la forme d'un hamiltonien $\varepsilon H_1(\varphi, I)$:

$$H = H_0(I) + \varepsilon H_1(\varphi, I),$$

où $\varepsilon > 0$ est un petit paramètre. On se place au voisinage d'un certain vecteur d'actions, disons $I = I_0$, que l'on voit comme l'état "non perturbé" ; on notera

$$\omega^0 = \frac{\partial H}{\partial I}(I_0),$$

et l'on supposera $\omega^0 = (\omega_1^0, \dots, \omega_n^0)$ non résonant. On cherche alors une transformation canonique

$$(\varphi, I) \longmapsto (\varphi', I')$$

qui préserve le caractère intégrable jusqu'à l'ordre 1 inclus.

On cherche cette transformation canonique par la méthode de la fonction génératrice ; par exemple avec une fonction $S = S(\varphi, I')$. Comme la transformation doit être proche de l'identité, on cherche S sous la forme

$$S_\varepsilon(\varphi, I') = \varphi \cdot I' + \varepsilon\, S_1(\varphi, I'),$$

où S_1 est une fonction $\mathbb{Z}^n$-périodique en φ, de classe au moins C^2, à déterminer ultérieurement. (Quand on écrit $\varphi \cdot I'$, on considère φ comme un élément de $\mathbb{R}^n$ et non $\mathbb{R}^n/\mathbb{Z}^n$; mais le φ' que l'on en déduit est bien un élément de $\mathbb{R}^n/\mathbb{Z}^n$.) On écrit alors les identités liant les anciennes et nouvelles variables :

$$\varphi' = \frac{\partial S_\varepsilon}{\partial I'} = \varphi + \varepsilon\, \frac{\partial S_1}{\partial I'}, \qquad I = \frac{\partial S_\varepsilon}{\partial \varphi} = I' + \varepsilon\, \frac{\partial S_1}{\partial \varphi}. \tag{138}$$

On note que pour ε suffisament petit, les conditions de non-degénérescence de S (Conditions (a), (b), (c) de la Proposition 125) sont bien satisfaites.

Le nouveau hamiltonien, lu dans les variables φ', I', coïncide avec $H(\varphi, I)$ (car la transformation est canonique), et vaut donc

$$\begin{aligned} H(\varphi, I) &= H_0(I) + \varepsilon H_1(\varphi, I) \\ &= H_0\left(I' + \varepsilon\, \frac{\partial S_1}{\partial \varphi}(\varphi, I')\right) + \varepsilon\, H_1(\varphi, I) \\ &= H_0(I') + \varepsilon\, \frac{\partial H_0}{\partial I}(I') \cdot \frac{\partial S_1}{\partial \varphi}(\varphi, I') + \varepsilon\, H_1(\varphi, I) + O(\varepsilon^2). \end{aligned}$$

Il est trop gourmand de demander l'intégrabilité à $O(\varepsilon^2)$ près ; mais nous allons demander cela seulement au voisinage de $I = I_0$, en travaillant à $O(\varepsilon |I - I_0|)$ près. En tenant compte de ce que $I = I' + O(\varepsilon)$,

$\varphi = \varphi' + O(\varepsilon)$, on trouve

$$H(\varphi', I') = H_0(I') + \varepsilon \frac{\partial H_0}{\partial I}(I_0) \cdot \frac{\partial S_1}{\partial \varphi}(\varphi', I_0) + \varepsilon H_1(\varphi', I_0) + O(\varepsilon^2) + O(\varepsilon |I - I_0|). \tag{139}$$

(Comme auparavant, on a utilisé un abus de notation en écrivant $H(\varphi', I')$ comme l'expression du hamiltonien dans les nouvelles variables ; noter que, comme $I' = I + O(\varepsilon)$, c'est la même chose d'écrire $O(\varepsilon^2) + O(\varepsilon|I - I_0|)$ ou $O(\varepsilon^2 + O(\varepsilon|I' - I_0|)$.) Le terme d'ordre ε dans (139) vaut

$$\varepsilon \left[\omega^0 \cdot \frac{\partial S_1}{\partial \varphi}(\varphi', I_0) + H_1(\varphi', I_0) \right]. \tag{140}$$

Peut-on maintenant choisir S_1 de telle sorte que ce terme d'ordre ε s'annule ? Ce serait trop demander, car si l'on intègre (140) sur $\mathbb{T}^n$ on trouve

$$\varepsilon \left(\omega^0 \cdot \left\langle \frac{\partial S_1}{\partial \varphi}(\,\cdot\,, I_0) \right\rangle_{\mathbb{T}^n} + \langle H_1(\,\cdot\,, I_0) \rangle_{\mathbb{T}^n} \right) = \varepsilon \left\langle H_1(\,\cdot\,, I_0) \right\rangle_{\mathbb{T}^n},$$

où l'on note avec des crochets l'opération d'intégration (moyenne).

Cependant, ce terme de moyenne en soi ne contrarie pas nos plans, car on peut le réécrire comme $\varepsilon \langle H_1(\,\cdot\,, I') \rangle_{\mathbb{T}^n} + O(\varepsilon |I - I_0|)$, et c'est à dire un terme intégrable, avec une petite erreur. Soustrayons donc la moyenne, et cherchons seulement à résoudre

$$\omega^0 \cdot \frac{\partial S_1}{\partial \varphi}(\,\cdot\,, I_0) + H_1(\,\cdot\,, I_0) - \langle H_1(\,\cdot\,, I_0) \rangle = 0. \tag{141}$$

C'est une équation aux dérivées partielles du premier ordre sur la fonction S_1, appartenant à la classe des équations dites de Hamilton–Jacobi. Il n'y a que peu de méthodes connues pour résoudre (141) dans le cas présent : on va passer aux séries de Fourier multidimensionnelles. On rappelle que l'on a un aller-retour entre les fonctions sur $\mathbb{T}^n$ et les suites indexées par $\mathbb{Z}^n$, via les formules

$$\widehat{f}(k) = \int_{\mathbb{T}^n} e^{-2i\pi k \cdot \varphi} f(\varphi)\, d\varphi, \qquad f(\varphi) = \sum_{k \in \mathbb{Z}^n} \widehat{f}(k)\, e^{2i\pi k \cdot \varphi}.$$

Pour abréger on notera $S_1(\,\cdot\,, I_0) = S_1$, $H_1(\,\cdot\,, I_0) = H_1$. La transformation de Fourier a la propriété bien connue de transformer la dérivation en multiplication par $2i\pi k$: ainsi (141) devient

$$2i\pi (k \cdot \omega^0) \widehat{S}_1(k) + \widehat{H}_1(k) = 0, \qquad \forall k \in \mathbb{Z}^n \setminus \{0\}. \tag{142}$$

C'est alors que l'hypothèse de non-résonance joue un rôle clé ! Par non-résonance du vecteur ω^0, le facteur $k\cdot\omega^0$ ne s'annule jamais pour $k \neq 0$; on peut donc écrire

$$\widehat{S}_1(k) = -\frac{\widehat{H}_1(k)}{2i\pi k\cdot\omega^0}. \tag{143}$$

On obtient S_1 en appliquant la formule de reconstruction de Fourier :

$$S_1(\varphi', I_0) = -\sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{\widehat{H}_1(k)}{2i\pi(k\cdot\omega^0)}\, e^{2i\pi k\cdot\varphi'}. \tag{144}$$

Comme H_1 est à valeurs réelles, $\widehat{H}_1(k)$ et $\widehat{H}_1(-k)$ sont conjugués ; on peut donv grouper les termes de la série deux à deux, selon les paires $(k,-k)$, et la série ainsi "regroupée" est à valeurs réelles.

Conclusion : nous avons montré que

$$H(\varphi', I') = \Big[H_0(I') + \varepsilon\,\big\langle H_1(\,\cdot\,, I')\big\rangle_{\mathbb{T}^n}\Big] + O(\varepsilon^2) + O(\varepsilon|I - I_0|). \tag{145}$$

Tant que l'on considère des valeurs de l'action I qui restent à distance $O(\varepsilon)$ de I_0, on a donc préservation de l'intégrabilité à un terme $O(\varepsilon^2)$, et non $O(\varepsilon)$, près ; en outre, à l'ordre ε, on peut remplacer le hamiltonien H par sa version moyennée en φ.

Cela ne veut pas dire que les effets de la perturbation ne se font pas sentir sur une échelle de temps $O(1/\varepsilon)$: ils se manifesteront dans la variation des fréquences caractéristiques $\partial H/\partial I$: le vecteur ω^0 doit ainsi être remplacé par

$$\omega^0 + \varepsilon \int_{\mathbb{T}^n} \frac{\partial H_1}{\partial I}(\phi, I_0)\, d\varphi.$$

Cependant, ce n'est qu'à l'ordre 2 que l'intégrabilité disparaîtra.

Ce raisonnement est à la base des résultats de Lagrange et Laplace sur la stabilité du système solaire en l'absence de résonance, sur de grandes échelles de temps. On trouvera dans le cours de Laskar [**24**] des explications détaillées sur l'application de cette méthode aux problèmes d'astronomie ; comme le note l'auteur, le formalisme hamiltonien, bien utilisé, permet de retrouver en quelques lignes, ou presque, l'ensemble des résultats établis dans des centaines de pages de mémoires écrits au 19ème et au 20ème siècle.

Rétrospectivement, la condition de non-résonance ne doit pas surprendre. Les résonances sont ennemis de la stabilité, cela peut se comprendre intuitivement comme suit. Supposons que des astres dans le système solaire soient en résonance, ils se retrouveront donc alignés (quasiment) à l'identique périodiquement. Ce qui veut dire que toute

variation minime sur une période pourra être répétée sur la période suivante, et il n'y aura finalement aucun espoir de compensation. Dans le cas non-résonant au contraire, l'effet d'une perturbation se fera sentir tantôt dans un sens, tantôt dans l'autre, et globalement les perturbations se compenseront presque sur le long terme.

Dans le système solaire cela n'est d'ailleurs "pas loin" de se produire : l'année de Saturne et celle de Jupiter sont presque en rapport 5/2 (en fait environ 2,48) : ainsi, toutes les 10 années jupitériennes, c'est presque la même configuration qui se répète ; c'est pourquoi nous voyons Jupiter se rapprocher du Soleil depuis plusieurs siècles... Mais bien sûr, ce n'est pas une vraie résonance en 5/2, et en fait sur des échelles de temps plus longues, on s'attend bien à la stabilité.

Il reste cependant un obstacle technique que nous avons esquivé : la convergence de la série (144). En effet, même si les coefficients de la série sont tous bien définis, ils pourraient ne pas tendre vers 0 assez vite, voire ne pas tendre vers 0 du tout, quand $|k| \to \infty$. Cela est dû à ce que les coefficients $\omega^0 \cdot k$, bien que non nuls, peuvent prendre des valeurs arbitrairement petites. Il s'agit du *problème des petits diviseurs*, qui a fait faire des cauchemars aux astronomes du 18ème et du 19ème siècle... Nous aurons l'occasion d'en reparler dans la section suivante.

6.9. Complément : Théorème KAM

Le traitement des effets perturbatifs en temps grand a laissé quelques questions en plan. Il y a eu celle des petits diviseurs et de la convergence de la série (144) ; il y a aussi le problème des temps extrêmement longs : que se passe-t-il sur des échelles de temps encore bien plus grandes que ε^{-2} ?

Henri Poincaré s'est attaqué à ce problème à la fin du 19ème siècle, dans un célèbre mémoire qui a fait date. Il y réalisait de nombreuses avancées fondamentales sur les systèmes hamiltoniens.

L'une d'entre elles était ce que l'on appelle le *théorème de récurrence de Poincaré* : si l'on considère un système hamiltonien borné dans l'espace des phases, alors pour presque toute condition initiale il revient arbitrairement près de l'état initial, une infinité de fois. Ce résultat semble incroyable par sa généralité : par exemple, si l'on mélange de l'eau et du jus de bissap, et que l'on considère cela comme un système hamiltonien classique fait d'une foule de molécules, alors on a évidemment une borne sur l'énergie totale, et l'on devrait voir le système revenir au bout d'un certain temps à un état où les deux liquides sont non mélangés ! La faille dans le raisonnement est double : d'une part, on

ne peut faire l'économie des fluctuations quantiques, qui se retrouveront rapidement amplifiées par les chocs entre molécules ; d'autre part, même dans un modèle purement classique, dans ce système de très grande dimension, le temps de récurrence de Poincaré est tellement monstrueusement grand (bien plus que des milliards de milliards de fois l'âge supposé de l'Univers) qu'il n'a aucune signification physique.

Dans son mémoire, Poincaré croyait aussi montrer la stabilité ; mais il y avait une erreur béante dans son manuscrit, et finalement il conclut à la possibilité d'instabilités. On aurait oublié cette erreur (nous en faisons tant...) si elle n'était pas survenue dans un contexte dramatique : le manuscrit de Poincaré était en lice pour un concours du roi Oscar II de Suède, et avait valu au mathématicien français le grand prix, une forte somme d'argent, et une notoriété mondiale... quel émoi quand on découvrit qu'il était faux ! Mais le secret fut bien gardé, les exemplaires de la revue où il avait été publié furent (presque tous) détruits, et Poincaré parvint à corriger son mémoire ! L'article résultant [**31**] est considéré comme l'acte fondateur de la théorie moderne des systèmes dynamiques. Poincaré montra en particulier que les séries perturbatives sont génériquement divergentes (divergentes pour un ensemble de conditions initiales dense). L'une de ses principales conclusions, du point de vue de la physique, est que le comportement en temps grand d'un système dépend de paramètres si fins qu'ils sont en pratique incalculables... un grain de poussière mal placé peut changer la stabilité en instabilité ! C'est l'un des principes fondateurs de la théorie du chaos, et cela amena les gens à penser qu'une petite perturbation d'un système intégrable, sur des temps extrêmement longs, est en général complètement imprédictible.

La conclusion de Poincaré était également en phase avec le paradigme de l'*ergodicité*, qui prenait sa source dans les travaux de Ludwig Boltzmann, et avait émergé avec les résultats de George David Birkhoff et Janos Von Neumann. Un système est dit ergodique quand, au cours du temps, il explore densément l'ensemble de l'espace des phases a priori disponible, c'est à dire la partie qui n'est pas interdite par des lois de conservation. La théorie ergodique permet de remplacer alors le comportement moyen des observables le long des trajectoires, par des moyennes sur l'espace des phases, a priori bien plus faciles à manipuler.

Cependant, en 1954 (exactement 100 ans après la naissance de Poincaré !) le mathématicien russe Andrei Kolmogorov annonça un théorème extraordinaire [**19**], en clôture du Congrès International des Mathématiciens : si l'on perturbe un système hamiltonien intégrable par une petite perturbation hamiltonienne, l'effet de la perturbation reste, le plus souvent, minime *jusqu'à la fin des temps* ! La légende

dit que Kolmogorov avait été si heureux d'apprendre le décès de Joseph Staline qu'il avait voulu célébrer l'événement avec un théorème incroyable... il pouvait difficilement mieux faire, car son résultat a inspiré des développements passionnants jusqu'à nos jours, à travers la fameuse théorie KAM, ou Kolmogorov–Arnold–Moser. Pour les mathématiciens comme pour les physiciens, le théorème de Kolmogorov a transformé le regard que l'on pouvait avoir sur la stabilité ; il a aussi détruit l'espoir d'une théorie générale basée sur l'ergodicité. Des résultats de recherche récents sur la stabilité de certaines équations aux dérivées partielles s'inspirent de cette théorie [**27, 38**].

Voici une version du Théorème de Kolmogorov :

THEORÈME 130. *Pour (x,p) dans $\mathbb{T}^n \times \mathbb{R}^n$, soit $H_0 = H_0(p)$ un hamiltonien intégrable, et soit $H_1 = H_1(x,p)$ une perturbation (a priori non intégrable). On pose*

$$H_\varepsilon(x,p) = H_0(p) + \varepsilon\, H_1(x,p), \qquad \varepsilon > 0.$$

On suppose que H_0 et H_1 sont analytiques, et on suppose également que $\nabla^2 H_0(p)$ est inversible pour tout p dans un ouvert borné $\mathcal{V}$ de $\mathbb{R}^n$. Alors, pour ε assez petit, il existe,

- un ensemble mesurable $\mathcal{V}_\varepsilon \subset \mathcal{V}$, de grande mesure, c'est à dire que $|\mathcal{V} \setminus \mathcal{V}_\varepsilon| \longrightarrow 0$ quand $\varepsilon \to 0$;

- un changement de variables $(x,p) \longmapsto (x',p')$ défini sur $\mathbb{T}^n \times \mathcal{V}_\varepsilon$, qui transforme de manière lisse les trajectoires hamiltoniennes de H_ε en trajectoires de H_0.

En d'autres termes, localement en p, la plupart des trajectoires du hamiltonien H_ε sont très régulières et stables, puisqu'elles sont des déformations de trajectoires bien choisies pour le système intégrable de hamiltonien H_0.

REMARQUES 131. 1.Le changement de variables $(x,p) \to (x',p')$ n'est pas une transformation canonique ! Dans la démonstration, pour chaque p_0 dans $\mathcal{V}_\varepsilon$, on construit un changement de variables canonique (analytique) qui sera adapté à l'étude de la trajectoire correspondant à p_0. C'est comme dans la section précédente, où l'on avait fixé I_0 pour construire un changement de variables ad hoc.

2. L'hypothèse d'inversibilité de la matrice seconde $\nabla^2 H(p)$ peut être remplacée par une condition bien plus faible, au prix de difficultés techniques non négligeables ; cela a été accompli par les efforts successifs d'Arnold, Herman et Féjoz [**12**]. Cette extension est importante au plan conceptuel, car l'hypothèse d'inversibilité n'est pas vérifiée par le système solaire, qui est dégénéré en un certain sens.

Le théorème de Kolmogorov a fait des vagues considérables. C'était un coup de tonnerre aussi bien dans la structure mathématique de son énoncé que dans sa signification physique et dans sa preuve même.

D'abord, dans son énoncé, on note l'hypothèse de *régularité analytique* qui est inhabituelle pour nous. Jusqu'à présent, dans ce cours, nous avons surtout travaillé avec la régularité C^1, voire C^r quand on avait besoin de dérivées d'ordre supérieur. Mais le théorème de Kolmogorov consomme beaucoup de régularité, sans qu'on ait rien demandé de tel dans la conclusion ! Peu après Kolmogorov, Jürgen Moser montrait que l'on pouvait se contenter d'une régularité C^∞, ou plutôt C^r pour r assez grand.

Pour l'anecdote, Moser devait préparer une recension de la note succinte [**19**] dans laquelle Kolmogorov annonçait ses résultats ; ne parvenant à se convaincre des arguments avancés par Kolmogorov, il se mit en tête de le redémontrer. Cependant il ne parvint pas à résoudre le problème en régularité analytique, avec des changements de variables analytiques aussi. À la place, il obtint des résultats en régularité C^r, avec des transformations C^r également ; ce qu'il considérait comme un échec fut perçu par la communauté mathématique comme un succès, car les hypothèses de régularité étaient bien plus générales. (Moser avait indiqué qu'il avait besoin de 333 dérivées, mais c'était une plaisanterie, un nombre bien inférieur suffisait.)

Arnold publia alors de son côté une démonstration complète de la preuve en régularité analytique, de sorte que l'on baptisa finalement cette théorie du nom de Kolmogorov–Arnold–Moser. Il faut aussi noter que bien plus tard, Luigi Chierchia [**5**] proposa une reconstitution de la preuve de Kolmogorov, bien plus simple que celle d'Arnold. Nous avons là un exemple, parmi bien d'autres, des complications qui surgissent quand on veut démêler la paternité d'un théorème ou d'une théorie...

Toujours au niveau de l'énoncé, le deuxième élément de surprise majeur est le rôle de la théorie de la mesure. Jusqu'ici nous avons rencontré des énoncés qui étaient vrais pour toutes les conditions initiales, ou pour toutes les conditions initiales dans un ouvert... parfois, dans les compléments, nous avons vu des énoncés qui étaient vrais pour presque toute condition initiale. Mais ici nous découvrons un énoncé qui est vrai pour "la plupart" des conditions initiales, disons 95%. Et il est très délicat de démêler quelles sont ces conditions initiales ! Cela dépendra de détails très fins, de conditions de résonance ou de non-résonance...

Enfin, du côté de la physique, ce théorème détruit l'espoir d'un comportement ergodique "universel". À la place de cela, le système perturbé se comporte bien, la plupart du temps, même s'il n'y est

pas contraint par les lois de conservation ! C'est un peu comme si un "gendarme invisible" le confinait...

Mais les sources d'étonnement et d'inspiration sont présentes aussi dans la preuve. Comme dans la section précédente, l'idée principale est, pour chaque valeur p_0 dans $\mathcal{V}_\varepsilon$, de rechercher une fonction génératrice qui transforme le tore invariant d'équation $(p = p_0)$ pour le flot de hamiltonien H_0 en un autre tore, déformé, invariant pour le flot de hamiltonien H. Si le vecteur de fréquences pour le tore $p = p_0$ était $\omega^0 = \partial_p H_0(p_0)$, celui du tore déformé sera

$$\omega^0_\varepsilon = \omega^0 + \varepsilon \langle \partial_p H_1(\cdot, p_0) \rangle.$$

L'application $\omega^0 \longmapsto \omega^0_\varepsilon$ est un difféomorphisme, ce qui garantit que nous trouverons ainsi de très nombreux tores invariants distincts pour la dynamique de H_ε.

Le raisonnement de la section précédente est la brique élémentaire de toute la construction : on a vu au préalable comment obtenir l'intégrabilité à $O(\varepsilon^2)$ près, l'idée est de recommencer pour avoir un ordre arbitrairement grand en ε, et de passer à la limite pour finalement avoir vraiment intégrabilité sur le tore considéré. Comme dans la section précédente, on se retrouve donc à étudier des équations de la forme

$$\omega^0 \cdot \nabla S_1 = H_1,$$

et on fait cela en développant S_1 et H_1 en séries de Fourier. Une nouvelle fois, on étudie donc l'EDO en résolvant une EDP !

Cette EDP appartient à une classe d'équations problématiques : on est incapable de fixer a priori la classe de régularité dans laquelle chercher l'équation, et ∇u est peut-être beaucoup moins régulière que f... c'est le vecteur ω^0 qui permettra de le dire :

- si ω^0 est résonant, on ne peut pas résoudre l'équation en général ;
- si ω^0 est non-résonant, on peut au moins écrire le développement formel en séries de Fourier ;
- si ω^0 est en outre *diophantien*, et si f est analytique, alors la série converge et en outre u est aussi analytique.

Par définition, un vecteur ω est diophantien si le produit scalaire $k \cdot \omega$ est non seulement non nul, mais en outre "jamais trop petit". C'est ce que formalise la

Définition 132. Un vecteur $\omega \in \mathbb{R}^n$ est dit *diophantien*, ou plus rigoureusement *vérifiant une condition diophantienne de non-résonance*, s'il existe $c = c(\omega) > 0$ et $\gamma > 0$ tel que pour tout $k \in \mathbb{Z}^n \setminus \{0\}$,

$$|k \cdot \omega| \geq \frac{c}{|k|^\gamma}. \tag{146}$$

(Ici on choisit une norme arbitraire sur $\mathbb{Z}^n$, une fois pour toutes.) Un vecteur diophantien est donc un vecteur dont les composantes ne sont "pas du tout en résonance" ! En général on fixe γ dans (146), et l'on laisse c libre de varier.

Le Théorème KAM parvient à démontrer que les vecteurs diophantiens sont préservés : pour eux, l'on peut boucler la procédure itérative et trouver une transformation canonique adéquate. En fait l'ensemble $\mathcal{V}_\varepsilon$ est fait de petits îlots autour des vecteurs diophantiens, la taille de chaque îlot dépendant de ε et c.

Les vecteurs diophantiens sont très nombreux : on peut montrer par un argument de théorie de la mesure qu'en fait, dès que $\gamma > n-1$, alors presque tout vecteur est diophantien ! Donc l'ensemble des vecteurs diophantiens est de mesure pleine... mais d'un autre côté, l'ensemble complémentaire est dense, car il contient tous les vecteurs à composantes rationnelles ! Nous sommes donc ici immergés dans les paradoxes subtils de théorie de la mesure, qui font douter de l'existence des nombres "réels"... d'un autre côté, c'est très frappant de trouver ces paradoxes dans des problèmes d'énoncé aussi concret que l'étude de la stabilité du système solaire ! Il s'agissait d'une vraie révolution conceptuelle, non moins importante que celle de Poincaré.

Le petit exercice qui suit, inspiré par Étienne Ghys [**16**], permet d'appréhender l'esprit du Théorème KAM sur un problème bien plus simple. Soit $f : \mathbb{T} \to \mathbb{R}$ une fonction lisse sur $\mathbb{R}/\mathbb{Z}$, telle que $\int f(t)\, dt = 0$. On considère maintenant une caricature de système dynamique en temps discret, définie par

$$\sigma_n(\alpha) = \varepsilon \Big[f(\alpha) + f(2\alpha) + \ldots + f(n\alpha). \Big]$$

Un célèbre théorème de Weyl (l'équipartition) dit que si α est irrationnel et f continue, alors $\sigma_n(\alpha)/n \longrightarrow \int f = 0$. Mais nous allons chercher ce qui se passe si l'on ne divise pas par n asymptotiquement : la question est de savoir si σ_n va diverger ou rester contrôlé. Si α est un peu quelconque, on a idée qu'il n'y aura pas de corrélation entre les différentes valeurs de $f(k\alpha)$, et dans la somme ci-dessus il devrait y avoir suffisamment de compensations pour que le tout reste borné.

Cela, c'est l'intuition ! Pour transformer cela en un raisonnement rigoureux, nous allons résoudre une équation auxiliaire :

$$v(x+\alpha) - v(x) = f(x), \qquad \forall x \in \mathbb{T}. \tag{147}$$

On passe en Fourier, écrivant

$$f(x) = \sum_{k \in \mathbb{Z}} f_k\, e^{2i\pi kx}, \qquad v(x) = \sum_{k \in \mathbb{Z}} v_k\, e^{2i\pi kx}.$$

(Les coefficients f_k sont donnés, les coefficients v_k sont les inconnues.) De (147) on déduit

$$\sum_{k\in\mathbb{Z}} v_k\big(e^{2i\pi k\alpha}-1\big)e^{2i\pi kx} = \sum_{k\in\mathbb{Z}} f_k\, e^{2i\pi kx}.$$

Pour $k=0$, nous trouvons $0=f_0$: c'est une condition nécessaire à la résolution, et c'est en fait la condition de moyenne nulle que nous avons imposée à f. On posera de manière arbitraire $v_0=0$.

Pour $k\neq 0$, nous trouvons

$$v_k(e^{2i\pi k\alpha}-1) = f_k.$$

Si α est irrationnel, il existe $k\in\mathbb{Z}\setminus\{0\}$ tel que $k\alpha\in\mathbb{Z}$, et on trouvera alors la condition nécessaire $0=f_k$: un coefficient, et en fait de nombreux coefficients, devront être nuls, ce qui conduirait à des hypothèses supplémentaires sur f. À la place, supposons que α est irrationnel : on peut alors écrire

$$v_k = \frac{f_k}{e^{2i\pi k\alpha}-1}. \tag{148}$$

Cela montre déjà que si la fonction v existe, ses coefficients de Fourier non nuls sont uniquement déterminés. En général cependant, cette somme n'a pas de raison de converger. Mais supposons maintenant que α est diophantien :

Définition 133. Un nombre réel α est dit diophantien s'il existe $c=c(\alpha)$ et $\gamma>0$ tel que pour tous entiers $p\in\mathbb{Z}$ et $q\in\mathbb{Z}$, $q>0$, on a

$$\left|\alpha-\frac{p}{q}\right| \geq \frac{c}{q^{1+\gamma}}. \tag{149}$$

Ce concept est bien sûr lié de très près à celui de vecteur diophantien ; on y reviendra. Si α est diophantien, on vérifie que $e^{2i\pi k\alpha}$ ne s'approche jamais trop près de 1 :

$$|e^{2i\pi k\alpha}-1| \geq \frac{c}{A\,|k|^{1+\gamma}},$$

où A est une constante numérique. Alors par (148),

$$|v_k| \leq \frac{A}{c}\,|k|^{1+\gamma}|f_k|, \tag{150}$$

où $c=c(\alpha)$ est une constante dépendant de α, mais pas de k, et A est une constante numérique. D'un autre côté, f étant de classe C^r, ses

coefficients de Fourier vérifient, par un argument classique d'intégration par parties répété,

$$|f_k| \leq \frac{\|f\|_{C^r}}{|k|^r}. \tag{151}$$

Combinant cela avec (150), on obtient l'estimation

$$|v_k| \leq \frac{A\,\|f\|_{C^r}}{c\,|k|^{r-1-\gamma}}. \tag{152}$$

Jusqu'ici les v_k n'étaient que des coefficients ; mais cette dernière estimation garantit la convergence de la série $\sum v_k\, e^{2i\pi kx}$, par convergence normale, dès que $r - \gamma > 2$. Nous avons donc montré : *dès que α est diophantien et f est de classe C^r avec $r > \gamma + 2$, on peut trouver une fonction continue v vérifiant* (147). En outre $\|v\|_\infty$ est borné par une constante qui dépend seulement des constantes c, γ et $\|f\|_{C^r}$.

Continuons le raisonnement. Pour tout $q > 1$, la mesure des $\alpha \in \mathbb{T}$ qui violent l'inégalité dans (149) est bornée par c/q^γ ; si $\gamma > 1$ la somme de toutes ces mesures est bornée en $O(c)$ car la série $\sum q^{-\gamma}$ converge. On en déduit que presque tout α est diophantien et que la mesure des α qui ne vérifient pas (149) tend vers 0 comme $O(c)$ pour $c \to 0$. Fixons maintenant $r = 4$, $\gamma = 3/2$, et nous voyons que

- pour presque tout α, on peut trouver une fonction v continue solution de (147) (unique à addition d'une constante près) ;
- l'ensemble des α tels que $\|v\|_\infty \leq 1/\delta$ est de mesure au moins $1 - C\delta$ où C est une constante qui dépend seulement de $\|f\|_{C^4}$.

La conclusion est facile : on a alors

$$\begin{aligned}\sigma_n(\alpha) &= \varepsilon\Big\{[v(2\alpha) - v(\alpha)] + [v(3\alpha) - v(2\alpha)] + \ldots + [v((n+1)\alpha) - v(n\alpha)]\Big\} \\ &= \varepsilon\Big(v((n+1)\alpha) - v(\alpha)\Big),\end{aligned}$$

d'où

$$\sup_{n\in\mathbb{N}} |\sigma_n(\alpha)| \leq 2\varepsilon\|v\|_\infty.$$

Et finalement, en choisissant δ égal à une fraction de ε nous trouvons

$$\sup_{n\in\mathbb{N}} |\sigma_n(\alpha)| \leq 1$$

pour tout α dans un ensemble de mesure $1 - O(\varepsilon)$.

Ce petit exercice capture beaucoup des sources d'étonnement liées à la théorie KAM : la régularité C^4 surgit sans qu'on l'ait vraiment prévu ; la mesure des α qui laissent σ_n uniformément bornée est grande quand $\varepsilon \to 0$; cet ensemble de conditions est obtenu via des conditions subtiles faisant intervenir l'approximation rationnelle.

Dans la théorie KAM, plutôt qu'une équation aux différences finies comme (147), c'est une équation aux dérivées partielles qu'il faut résoudre, du style $H_\varepsilon(x, \nabla\Phi + c) = h$. Cela est similaire à la théorie que nous avions mentionnée sous le nom de "KAM faible", sauf que cette fois-ci on veut prouver la régularité globale de la solution. C'est toute une affaire, un chapitre délicat de la théorie des EDP, faisant même intervenir des algorithmes d'approximation numérique ! On renvoie à Arnold [**2**], Thirring [**34**] et Chierchia [**5**] pour plus de détails.

Une question : à quoi ressemblent les vecteurs diophantiens ? Ce n'est pas simple de se le représenter. Considérons $n = 2$: le vecteur (ω_1, ω_2) est diophantien si et seulement si ω_1/ω_2 est diophantien au sens de la Définition 133 (à condition que $\omega_2 \neq 0$, cela va de soi). On est donc ramené à comprendre les nombres diophantiens. Ce sont en un sens des nombres "très irrationnels", et l'on pense peut-être à des nombres transcendants... mais au contraire, il y a des nombres transcendants qui peuvent être approchés de très très près par des nombres rationnels, comme les nombres de Liouville. En fait le nombre diophantien par excellence est solution d'une équation du second degré : c'est le célèbre nombre d'or,

$$\phi = \frac{\sqrt{5} - 1}{2},$$

solution de $\phi^2 + \phi - 1 = 0$. Le nombre ϕ est en effet, parmi tous les nombres de $[0, 1]$, le plus difficile à approcher par des rationnels, du fait de son développement en fraction continue : $\phi = 1/(1 + 1/(1 + 1/(1 + 1/.....$ La conséquence dans notre problème est remarquable : la théorie KAM prédit que dans un système intégrable en deux degrés de liberté, c'est le rapport de fréquences égal au nombre d'or qui est le plus stable aux perturbations ! C'est assez drôle de voir le nombre d'or surgir dans un problème aussi subtil, surtout associé à une notion de stabilité qui va bien avec les vertus d'harmonie qu'on lui prête... cependant, la raison fondamentale, aussi bien de son caractère diophantien que de ses propriétés légendaires, est le caractère "autoreproducteur" qui se lit dans l'équation $x = 1/(1 + x)$.

À l'autre extrême, les rapports les plus instables sont censés être les rapports rationnels. Mais que veut dire rationnel dans le monde physique, où l'on ne peut déterminer un nombre avec une précision infinie ? On peut, en pratique, s'accommoder d'une définition approximative, énoncée volontairement de manière floue : α est dit "pratiquement rationnel" s'il est très proche de p/q avec p et q pas trop grands.

Voici un exemple. Le système Soleil–Jupiter est complètement intégrable si l'on néglige les interactions avec les autres planètes. Ajoutons les

astéroïdes (qui sont ultra-légers face à ces deux astres), et considérons que leur stabilité est fonction des rapports entre leur vitesse angulaire et celle de Jupiter : la théorie KAM suggère que les trajectoires où ce rapport est pratiquement rationnel seront les plus touchées. Et de fait, dans la ceinture d'astéroïdes il y a effectivement des "trous" aux endroits correspondant à des rapports pratiquement rationnels : 2, 3, 5/3, 7/2, 7/3... La figure, reproduite dans l'ouvrage de Thirring, montre ainsi que la théorie KAM a été observée en pratique, dans son esprit sinon dans sa lettre, et sur certains sous-systèmes.

Pour le système solaire tout entier, en revanche cela ne marche pas vraiment : pendant de longues années on a cru, à la suite de Kolmogorov, que le système solaire était stable... jusqu'à ce que Laskar et Tremaine prouvent qu'il est finalement instable sur de longues périodes de temps. Mais c'est une autre histoire ! En fait, la théorie KAM fait intervenir de si petites valeurs de la perturbation que, en pratique, on ne peut jamais l'appliquer... Il n'empêche qu'elle a révolutionné notre façon de penser à la stabilité en temps grand, et que ses techniques se sont répandues dans d'autres domaines mathématiques.

Bibliographie

[1] V.I. Arnold, *Équations Différentielles Ordinaires*, Mir, 1974.

[2] V.I. Arnold, *Les Méthodes Mathématiques de la Mécanique Classique*, Mir, 1976.

[3] V.I. Arnold, *Chapitres supplémentaires de la théorie des Équations Différentielles Ordinaires*, Mir, 1984.

[4] S. Baigent, *Lotka–Volterra Dynamics – An introduction.* Preprint, University of College, London.

[5] L. Chierchia, A.N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. *Regul. Chaotic Dyn. 13*, No. 2 (2008), 130–139.

[6] A.J. Chorin, J.E. Marsden, *A Mathematical introduction to fluid mechanics*, Springer (2000), 3rd edition.

[7] M. Crouzeix, A.L. Mignot, *Analyse numérique des équations différentielles*, Masson 1984.

[8] S. Guerre-Delabrière, M. Postel, *Méthodes d'approximation, Équations différentielles, Applications Scilab.* Ellipses, 2004.

[9] C. De Lellis, Ordinary differential equations with rough coefficients and the renormalization theorem of Ambrosio (d'apr ?s Ambrosio, DiPerna, Lions). Bourbaki seminar 2007, Exp. No.972.

[10] P.M. Do Carmo, *Riemannian Geometry*, Birkhäuser, 1992.

[11] A. Fathi, *Weak Kam Theorem in Lagrangian Dynamics*, Cambridge Studies in Advanced Mathematics, 2014.

[12] J. Féjoz, Démonstration du "théorème d'Arnold" sur la stabilité du système planétaire (d'après Herman). *Ergod. Th. Dynam. Sys. 24*, No. 1, 1521-1582 (2004).

[13] *Handbook of Mathematical Fluid Dynamics*, North Holland, 2002-2007 (Quatre volumes), S. Friedlander & D. Serre, Eds.

[14] U. Frisch, *Turbulence, the legacy of A.N. Kolmogorov*, Cambridge University Press, 1996.

[15] S. Gallot, D. Hulin & Ph. Lafontaine, *Riemannian Geometry*, Universitext, Springer, 3e Éd., 2004.

[16] É. Ghys, Résonances et petits diviseurs. Consultable en ligne à `perso.ens-lyon.fr/ghys/articles/resonancesdiviseurs.pdf`

[17] J. Gleick, *Chaos : Making a new Science*, Vintage, 1987.

[18] I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products.* Academic Press, 1980.

[19] A.N. Kolmogorov, On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk. SSSR (N.S) 98* (1954), 527-530.

[20] M.W. Hirsch, S. Smale & R.L. Devaney, *Differential Equations*. Elsevier, Second Ed., 2013.

[21] O. Knill, *Lienhard Systems*. Notes de cours synthétiques sur les cycles limites stables. Consultable en ligne à `http://www.math.harvard.edu/archive/118r_spring_05/handouts/lienard.pdf`

[22] S. Kovalevskaya. Sur Le Probleme De La Rotation D'Un Corps Solide Autour D'Un Point Fixe, *Acta Mathematica 12* (1889), 177-232.

[23] L.D. Landau, E.M. Lifshitz, *Mécanique*, Mir, 1982.

[24] J. Laskar, Hamiltonien planétaire, Notes de cours, 2009.

[25] J. Laskar, Le système solaire est-il stable ? *Séminaire Poincaré 14* (2010), 221–246. Consultable en ligne à `www.bourbaphy.fr/laskar.pdf`

[26] J. Laskar, M. Gastineau, J.-B. Delisle, A. Farrès, A. Fienga, Strong chaos induced by close encounters with Ceres and Vesta. *Astronomy and Astrophysics, EDP Sciences (2011), 532*. Consultable en ligne à `hal.archives-ouvertes.fr/hal-00620783`

[27] C. Mouhot, C. Villani On Landau damping. *Acta Mathematica 207, 1 (2011) 29-201.*

[28] M.E. Nelson, *Hodgkin–Huxley Models*. Extrait de *Electrophysiological Models*, in *Databasing the Brain : From Data to Knowledge* (S. Koslow and S. Subramaniam, eds.) Wiley, New York, 2004. Consultable en ligne à `http://nelson.beckman.illinois.edu/courses/physl317/part1/Lec3_HHsection.pdf`

[29] A.F. Nikiforov, V. Ouvarov, *Fonctions spéciales de la physique math ?matique.* Mir, 1983.

[30] D. O'Shea, *Grigori Perelman face à la conjecture de Poincaré*, Quai des Sciences, Dunod (2007), traduit de l'anglais.

[31] H. Poincaré, Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica 13* (1890), 1–270.

[32] R.T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics.

[33] K. Siburg, *The Principle of Least Action in Geometry and Dynamics*. Springer-Verlag, Lecture Notes in Mathematics Vol. 1844, 2004.

[34] W. Thirring, *Classical Mathematical Physics : Dynamical Systems and Field Theories*. Springer-Verlag, 3ème édition, 2013. Traduit en anglais par E.M. Harrell.

[35] C. Villani, *A review of mathematical topics in collisional kinetic theory*, in *Handbook of Mathematical Fluid Dynamics*, North Holland, 2003 (Quatre volumes), S. Friedlander & D. Serre, Eds., Vol. 1. Consultable en ligne à `cedricvillani.org/wp-content/uploads/2012/08/preprint.pdf`

[36] C. Villani, *Optimal Transport, old and new*. Springer, Grundlehren der mathematischen Wissenschaften, Vol. 338, 2009.

[37] C. Villani, Paradoxe de Scheffer–Shnirelman revu sous l'angle de l'intégration convexe (d'après C. De Lellis et L. Székelyhidi). Séminaire Bourbaki, 61e année, 2008-2009, No.1001. Consultable en ligne à `www.bourbaki.ens.fr/TEXTES/1001.pdf`

[38] C. Villani, Particle systems and nonlinear Landau damping. *Phys. Plasmas 21*, 030901 (2014). Consultable en ligne à `http://cedricvillani.org/wp-content/uploads/2013/09/P21.Plasmas1.pdf`

[39] E.T. Whittaker, G.N. Watson *A course of modern analysis*. Cambridge University Press, 1996.