

Du bon dosage de l'aléatoire dans nos recherches

Un monde trop complexe

Vous connaissez peut-être déjà la légende du sage Sissa, inventeur d'après une tradition indienne du jeu des échecs. À son roi qui lui demandait quelle récompense il désirait, il proposa que sur la première case d'un échiquier soit posé un grain de riz ; sur la seconde, 2 grains, puis 4, puis 8, etc ; et qu'on lui offre ce riz. En fait, le sage avait le sens de l'humour : pour remplir ainsi les 64 cases, il eût fallu, à la louche, 1000 ans de notre production mondiale actuelle. Voilà un phénomène typique des problèmes dits combinatoires, où les permutations, les combinaisons, les associations de quelques objets élémentaires amènent immédiatement à des ordres de grandeurs faramineux. Il y a ainsi bien plus de parties d'échecs possibles que d'atomes estimés dans l'univers. Comment, parmi toutes ces parties possibles, décider du meilleur coup à jouer ?

Cette complexité émergeant de briques de base pourtant simples se retrouve dans des domaines très variés : optimiser un emploi du temps, un itinéraire ou la position de composants électroniques sur une puce ; en statistiques, isoler parmi un énorme jeu de données (les omniprésentes *big data* – par exemple du code génétique, des relevés météorologiques, des mesures d'imagerie médicale ou des données tirées de réseau de production ou internet) de l'information pertinente et intelligible, qui ne fasse intervenir dans l'idéal que quelques paramètres ; en biochimie, prévoir par ordinateur (moins cher, plus rapide, moins douloureux pour les rats de laboratoire que des expériences *in vivo*) la configuration stable d'une protéine (*fig. 1*), celle qui minimise les tensions internes entre les acides aminés qui la composent et détermine sa fonction (par exemple l'hémoglobine a une forme de réceptacle à oxygène). Chacune de ces questions ouvre un abîme insondable de réponses envisageables, parmi lesquelles il faut dénicher la meilleure.



Figure 1 : Le repliement d'une protéine, un vrai casse-tête combinatoire.



Pour un mathématicien, tous ces problèmes n'en sont en réalité qu'un seul et même (après tout, « la mathématique est l'art de donner le même nom à des choses différentes », disait Poincaré). On se retrouve devant un espace de possibilités de très grande dimension (toutes les parties d'échecs possibles, toutes les configurations de la chaîne d'acides aminés, tous les jeux de paramètres explicatifs pour des données statistiques, etc.) et on cherche parmi elles celle qui optimise un paramètre donné (les chances de victoire aux échecs, l'énergie interne pour la protéine, l'adéquation aux données en statistiques...). D'un point de vue abstrait, on peut voir les choses ainsi : on se trouve, aveugle, dans un relief inconnu, on peut se déplacer (c'est-à-dire explorer les différentes configurations longitude/latitude possibles) et on cherche l'endroit de plus basse altitude.

Les algorithmes stochastiques

Pour résoudre notre (nos) problème(s), une première idée serait d'utiliser la force brute et de tester toutes les possibilités, d'explorer tout l'espace. Après tout, n'avons-nous pas de nos jours une incroyable puissance de calcul ? En fait, la complexité est rédhibitoire. Rappelons-nous de Sissa. Le plus puissant des ordinateurs actuels mettrait des siècles à parcourir toutes les parties d'échec possible.

En gardant à l'esprit l'image du paysage vallonné, une autre stratégie serait la suivante : puisque l'on cherche l'altitude la plus basse, il suffit de suivre la pente comme une bille. Mais si l'on démarre dans un cratère de volcan, on va aboutir avec cette méthode au fond de celui-ci, qui est beaucoup plus haut que la vallée juste à côté (fig. 2). On ferait mieux d'abord de remonter la pente, sortir du cratère, puis redescendre, mais on ne peut comprendre cela qu'avec une vision d'ensemble qui, on l'a vu plus haut, est hors de portée du fait de la complexité combinatoire. Si le paysage est simple, ce problème de minimum local n'apparaît pas, mais dans les exemples que j'ai donnés, on a plutôt affaire aux champs de bataille de la Marne envahi par un bataillon de taupe. Ceci illustre le défaut de la méthode, dans la recherche ou dans la vie, qui consiste à vouloir à tout prix améliorer des détails, qui amène à s'enfoncer toujours plus avant dans un cadre fixé alors qu'il faudrait changer du tout au tout, quitte à être contre-productif au début.

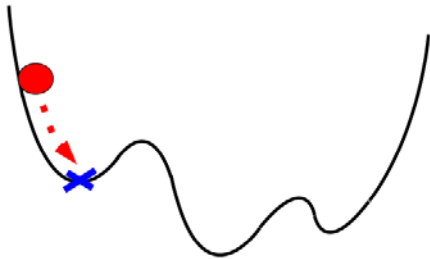


Figure 2 : le problème des minima locaux

Face à l'immensité des possibles et contre les pièges des minima locaux, une solution : explorer l'espace au hasard. Ce qui ne veut pas dire n'importe comment ! Par exemple, si aux échecs on choisit un pion uniformément et qu'on lui fait effectuer indifféremment n'importe quel mouvement autorisé, très vraisemblablement ce coup sera absurde et pas du tout représentatif d'une vraie partie. En revanche on peut associer des scores aux positions des pièces, basés sur l'analyse et l'expérience des bons joueurs, et décider du prochain coup en fonction de ce score. Prendre systématiquement LE coup qui donne le meilleur score, c'est suivre la bille qui roule jusqu'au fond du cratère : ça ne marche pas. Par contre on peut donner une plus ou moins grande probabilité de jouer un coup selon son score, de sorte que les coups stupides ne seront concrètement jamais proposés tandis que ceux qui semblent dans l'immédiat donner l'avantage seront mis en avant. Il faut un équilibre entre l'optimisation directe et un petit grain de folie.

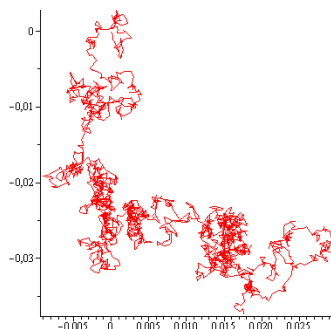


Figure 3 : Exploration aléatoire de l'espace

Au lieu d'une bille condamnée à dévaler la pente, on suit ainsi un ivrogne hésitant qui, tout en cherchant à descendre, va brouiller cette tendance par des mouvements erratiques qui, in fine, lui permettront de sortir de n'importe quel piège. On utilise donc un *algorithme stochastique*, synonyme technique pour *méthode aléatoire*.

Historiquement, le premier grand exemple en est l'algorithme de Monte-Carlo (en référence aux casinos) qui, comme toutes les grandes avancées scientifiques, a bien évidemment été conçu dans un but noble soutenu par des moyens financiers désintéressés, à Los Alamos, vers 1943, pour effectuer des calculs de physique nucléaire dans le cadre du projet Manhattan (ce qui prouve bien, au moins, qu'il fonctionne). Ces algorithmes sont maintenant partout ; la finance par exemple en est très friande.

Le recuit simulé

Dans ma thèse j'ai surtout étudié un algorithme particulier, le recuit simulé. Au départ bien sûr je n'y connaissais rien. L'idée est qu'au lieu de choisir une fois pour toutes le bon dosage entre optimisation et aléa, on le fait évoluer au cours du temps. Plus précisément on commence avec beaucoup d'aléa, c'est-à-dire un explorateur très ivre ou encore, si l'on imagine une particule de gaz soumise à une agitation thermique, une haute température. L'espace est rapidement exploré, mais pas dans le détail. Au fur et à mesure, on dégrise l'ivrogne (ou bien on refroidit la particule) ce qui va rendre plus difficile de s'échapper des pièges locaux (on a moins d'énergie pour remonter la pente) mais le fond des vallées sera plus précisément investigué. Ainsi travaille-t-on successivement à des échelles de plus en plus petites.

À l'origine le recuit est une méthode de forgeron. On chauffe un métal, ses molécules sont agitées et ont assez d'énergie pour changer de configuration, puis on le refroidit très lentement, de sorte que tout en gardant la possibilité de se déplacer, les molécules vont préférentiellement former une structure qui minimise son énergie interne, c'est-à-dire un cristal bien ordonné, qui donne un bon matériau. Refroidies abruptement, les molécules vont se retrouver d'un coup figées en plein milieu de leur mouvement, dans une structure désordonnée. C'est le même mécanisme en géologie : le même magma, selon qu'il refroidisse lentement en remontant progressivement dans la couche terrestre, ou subitement en émergeant au fond de l'océan, va donner du diamant ou du verre amorphe (*fig 4*).

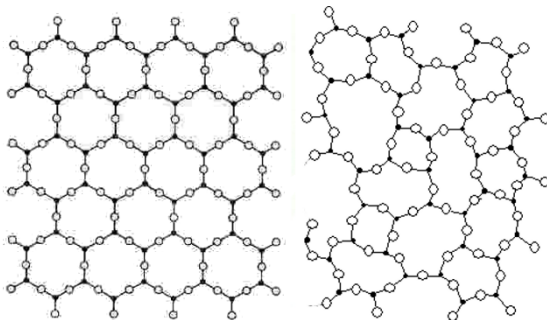


Figure 4 : refroidissement lent versus rapide

C'est l'utilisateur qui choisit la température pour le recuit simulé. S'il va trop vite, il risque de geler l'explorateur au beau milieu de sa recherche, dans un piège local, autrement dit une configuration inefficace. S'il va trop lentement l'algorithme va être très long. Pour calibrer en pratique un bon schéma de refroidissement, il faut comprendre en théorie à quelle vitesse la particule explore l'espace. Ça, c'était mon boulot.

Relaxation à l'équilibre

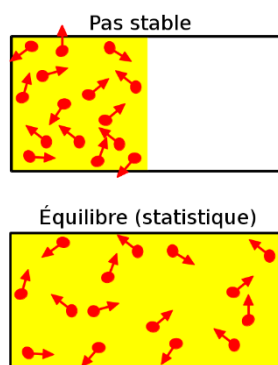
Maintenant que nous avons les idées plus claires sur le contexte avec des sages indiens saouls dans un paysage inconnu et froid qui fabriquent des bombes atomiques pour replier des protéines en jouant aux échecs, nous allons pouvoir parler de mathématiques. On ne fait hélas pas des articles scientifiques qu'avec des métaphores et des intuitions. Pourtant, ces images que j'ai utilisées, je les avais en tête dès le début de ma thèse. Si au départ d'ailleurs j'ai choisi le domaine des processus stochastiques (c'est-à-dire des choses qui évoluent au cours du temps avec de l'aléa, tel notre explorateur ivre, les cours de la bourse ou la position d'une particule de gaz), c'est que je les trouve extrêmement visuels.

Pour décrire mathématiquement ces dynamiques aléatoires, au lieu de considérer une seule trajectoire, on considère simultanément une infinité de mondes parallèle, l'un dans lequel la particule part à droite, l'un à gauche, devant, derrière, etc., de sorte qu'on regarde tout un nuage de particules. Le problème est le même qu'en physique statistique, où l'on considère effectivement un nuage de gaz. La densité du gaz mesure la probabilité que l'explorateur se trouve à cet endroit. Puisque notre (ou nos) particule(s) se déplace(nt), cette densité évolue. Ceci est codé par une équation, dite équation aux dérivées partielles (EDP) qui fait le bilan, dans chaque zone de l'espace, des entrées et des sorties. S'il y a plus de particules qui, sous l'action de l'agitation aléatoire ou de la tendance à descendre la pente (ou d'autres mécanismes de déplacement si l'on veut), entrent dans une zone qu'il n'en sort, la densité augmente, et inversement dans le cas contraire. Avec un peu d'expérience, en regardant l'EDP, suite de symboles mathématiques, on l'interprète tout de suite comme une façon de se déplacer, et réciproquement si on imagine une dynamique on peut immédiatement écrire son équation. Je trouve qu'il y a un plaisir semblable à l'assemblage de Lego à manipuler ces équations.

Une remarque : nous sommes en train de coder le hasard. Il ne faut pas y voir une entreprise métaphysique. C'est simplement un langage, une modélisation, un cadre abstrait qui, par certains aspects, colle assez bien à des choses plus ou moins concrètes. De la même façon, on définit les triangles dans le plan, dont on démontre que la somme des angles fait 180 degré ; et cela peut bien être utile dans la vie réelle même si les triangles parfaits des mathématiciens n'existent pas.

Notre exploration aléatoire est donc devenue un nuage de gaz qui se répand dans un paysage vallonné.

Que fait un gaz, libéré dans une pièce hermétique vide ? Il se mélange. Mettons qu'il se trouve initialement dans une petite boîte que l'on ouvre. Au départ, il y a plus de particules qui quittent la boîte que d'autres n'y rentrent : la densité diminue dans celle-ci et augmente alentours. Au bout d'un moment, même la densité de gaz à l'autre bout de la pièce commence à augmenter. Après une heure, le gaz s'est complètement mélangé, sa densité est la même partout. Bien sûr au niveau microscopique les particules continuent de bouger. Mais au niveau macroscopique, dans chaque zone de l'espace, à chaque instant on voit autant de particules rentrer que sortir, donc l'état global est inchangé. On dit que le gaz a relaxé (ou convergé) vers son équilibre statistique, qui est ici la répartition uniforme dans toute la pièce.



On dispose même d'une façon de mesurer à quel point le nuage de gaz est loin de son équilibre : l'entropie. Une définition précise m'emmènerait trop loin d'autant que, fidèle à l'adage de Poincaré, ce mot désigne des choses différentes selon le domaine des maths, de la physique ou de l'informatique dans lequel on se trouve. Dans mon contexte, c'est un nombre qui est grand lorsque le nuage ne ressemble pas du tout à son équilibre (par exemple quand il est concentré dans une petite boîte), qui diminue à mesure qu'il se répand, et qui tend vers 0

quand le nuage se répartie uniformément dans la pièce.

Comme l'entropie est un simple nombre, elle est bien plus facile à manipuler que la densité du gaz qui est un objet compliqué. À partir de l'EDP satisfaite par la densité (dont il est illusoire de vouloir calculer une solution explicite) on peut déduire une équation beaucoup plus simple satisfaite par l'entropie (en réalité une inéquation, mais peu importe). Ceci permet de quantifier la vitesse à laquelle le gaz se mélange. Ou, dans notre cas, à quelle vitesse notre ivrogne se perd et donc explore l'espace : exactement l'information qui nous manque pour définir un schéma de refroidissement.

D'un point de vue technique, une bonne part de ma thèse a consisté à manipuler des objets semblables à l'entropie et à déduire des équations qu'ils satisfaisaient à partir de l'EDP du mouvement microscopique. Avec ça, j'ai rempli de gribouillis une dizaine de grands cahiers de recherche, ai eu droit à quelques ascenseurs émotionnels en alternant des démonstrations fausses, des corrections, des changements de stratégie complets, des retournements de situations, le tout à une fréquence de trois fois par jours pendant des semaines et une amplitude nerveuse que je ne me connaissais pas, et même quelques insomnies (moi l'homme le plus placide du monde qui ai toujours dormi comme une pierre). Étonnamment, quand on est dedans, cette agitation n'a pas l'air vaine.

L'ivrogne décidé

Quand j'ai commencé mon doctorat, tout ce que j'ai raconté plus haut était bien connu. Mon directeur de thèse Laurent Miclo et d'autres avaient déjà traité ces questions dans les années 90. Cependant dans ces études classiques (pour un mathématicien, *classique* signifie *avant moi*) on faisait l'hypothèse que l'explorateur n'avait pas de mémoire, qu'il décidait de ses prochains mouvements sans prendre en compte sa trajectoire passée. Et effectivement, c'est ainsi que l'algorithme était implémenté sur ordinateur. Mais, sans mémoire, la particule va sans cesse revenir sur ses pas. Pour explorer autre chose qu'un bocal de poisson rouge, c'est très inefficace.

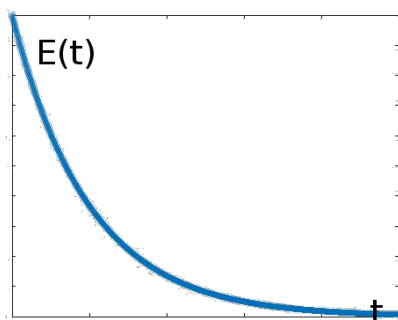


Figure 5 : convergence régulière au cours du temps

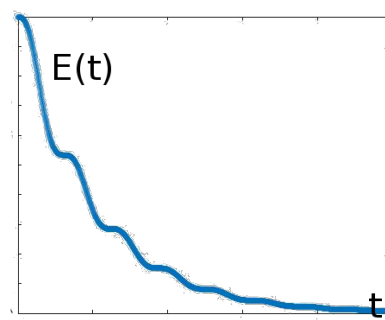


Figure 6 : convergence saccadée dans ma thèse

Pour des raisons pratiques on décide donc de garder des informations sur le passé, mais pas toute la trajectoire : dans un espace de grande dimension, repérer un point demande beaucoup de coordonnées, cela occuperait trop de mémoire numérique. Un compromis est de ne conserver qu'une mémoire instantanée : on se rappelle de l'endroit d'où l'on vient tout juste, et l'on évite d'y retourner immédiatement. Cela revient à avoir de l'inertie, comme une bille lâchée dans un toboggan peut remonter une rampe si elle a suffisamment d'énergie.

Donc, au lieu de suivre un ivrogne amnésique indécis, on confie l'exploration à un ivrogne décidé, plein d'élan. Ça n'a l'air de rien, mais en pratique c'est vraiment plus efficace, on quitte plus facilement les pièges locaux, et une amélioration sur cet algorithme a des retombés dans tous les domaines qui l'exploitent. En revanche, en théorie, tout s'effondre. Or,

rappelons-le, l'étude théorique est nécessaire pour le calibrage – en pratique – de la température.

D'où vient le problème ? Dans le cas amnésique on a un phénomène bien particulier : la vitesse à laquelle le nuage se rapproche de son équilibre (la variation de son entropie, donc) est proportionnelle à sa distance à ce même équilibre (son entropie, donc). Plus il en est loin, plus vite il s'en rapproche. Conséquence : il s'en rapproche à un rythme régulier (*fig 5*). Pour la dynamique à inertie, ça n'est plus le cas. Le nuage se comporte un peu comme une balançoire qui oscille plusieurs fois autour de sa position d'équilibre avant de s'y arrêter. En un sens, l'approche de l'équilibre se fait par à-coups, irrégulièrement (*fig 6*).

Ma méthode a consisté à ne pas simplement calculer l'entropie et sa vitesse (c'est-à-dire la variation de l'entropie) mais également l'accélération (c'est-à-dire la variation de la vitesse) et la variation de l'accélération elle-même, et de les comparer. C'est un peu comme si un politicien (et c'est là ma dernière métaphore), qui ne pourrait pas dire « le chômage est haut, certes, mais il diminue vite », se contenterait de : « le chômage est haut, et il augmente, et même il accélère, certes, mais au moins, son accélération, elle, ralentit, de sorte qu'un jour il se mettra à décélérer, de sorte qu'un jour il se mettra à diminuer, de sorte qu'un jour il sera bas ». C'est une méthode nouvelle d'exploiter un tel retard à l'allumage, qui apparaît dans de nombreux autres contextes : que ce soit en physique, en biologie ou en économie, rares sont les mécanismes qui ne dépendent pas du passé (les investissements dépendent des tendances boursières de la semaine dernière, le réchauffement climatique dépend des émissions de carbone des dernières décennies, etc.)

Cette idée était de Laurent, et c'était le point de départ de ma thèse car en l'état, telle quelle, elle ne fonctionnait pas. J'ai d'abord regardé ce qui se passait dans un cas simple (un ivrogne décidé sur un circuit en boucle sans relief) pour lequel des calculs explicites étaient possibles. Parallèlement je lisais des articles classiques et je discutais beaucoup avec Laurent, qui est un chercheur passionné et passionnant, ou d'autres collègues (comprendre : je buvais beaucoup de café ; c'est un peu l'équivalent d'une recherche de terrain pour un mathématicien). L'ambiance à l'institut de maths de Toulouse était excellente.

À force j'ai développé une certaine dextérité et un peu de recul sur mes calculs d'entropie et j'ai fini par trouver un moyen pour que tout s'emboîte bien, à l'aide d'une entropie auxiliaire nécessaire au début des calculs que l'on pouvait escamoter à la fin. Ceci dit ce premier résultat était un peu décevant, trop peu précis pour estimer la vitesse de relaxation quand la température de la particule est basse, ce qui est pourtant le régime nécessaire à comprendre pour le recuit simulé. C'était justement le défaut des méthodes qui existaient déjà et que nous voulions régler.

J'ai laissé ce problème reposer et ai cherché d'autres cobayes pour tester mes idées. Je suis parti enseigner deux semaines au Ghana dans un campus universitaire isolé où il n'y avait rien d'autre à faire entre les cours que travailler. J'y ai étudié des modèles plus simples, par exemple la taille d'une cellule vivante qui grossit et, à des moments aléatoires, se divise en deux. Le fait que l'aléa ne survenait qu'à certains moments précis posait problème et m'a poussé à écrire les choses d'une certaine façon (calculer, ça n'est jamais que réécrire). En revenant à mon problème de départ avec ce nouveau formalisme les calculs devenaient beaucoup plus clairs et j'ai pu isoler les paramètres qui étaient pertinents à basse température. Ma conclusion était que les méthodes existantes étaient suffisamment précises : simplement il fallait les écrire proprement pour s'en rendre compte.

... et au-delà

L'intérêt de travailler sur des dynamiques aléatoires, c'est qu'on peut souvent comparer notre propre processus de recherche au processus qu'on étudie (ça n'est pas Vincent Tejedor, lauréat 2013, qui me contredira). Je ne fais pas exception et, emporté par mon élan, je me suis maintenant éloigné de mon point initial, le recuit simulé, et poursuit, ivre, ma quête de nouveaux cobayes à qui infliger mes méthodes, pas tant pour obtenir des résultats sur ces problèmes que pour comprendre en retour les méthodes. Là aussi l'espace à explorer est incommensurable : rappelons-nous de Sissa.

Références :

- Jean-Luc Lutton et Ernesto Bonomi, *Le recuit simulé*, Pour la science, n° 129, juillet 1988, p. 68- 77
- Wikipedia, *recuit simulé* (ou en anglais : *simulated annealing*)
- Pierre Grihon, *Promenades aléatoires : vers les chaînes de Markov*. Bulletin de l'APMEP. Num. 501. p. 545-555.
- (plus technique) Michel Benaïm et Nicole El Karoui, *Promenade aléatoire. Chaînes de Markov et simulations ; martingales et stratégies*, les Éditions de l'École Polytechnique 2004