OPTIMISATION ET ORDINATEUR

Une tentative d'optimisation de l'optimisation vers l'optimiseur relatif absolu:

la Méthode des Moindres Quelconques (MMQ)

Gérard A. Langlet

Proposition d'article, compréhensible par tout lecteur sensé et de bonne foi, pour les brèves du DRECAM, par exemple.

Introduction

La méthode des moindres carrés reste, et de loin, la plus connue et, probablement la plus utilisée, de toutes les méthodes d'optimisation. Simple à exposer, elle a conquis tous les domaines: on la retrouve utilisée par les politologues aussi bien que par les physiciens, par les analystes financiers aussi bien que les biologistes et des médecins.

Son principe initial ne fait pas appel à des mathématiques de haut niveau : il peut s'enseigner dans des classes élémentaires, dès que le représentation cartésienne des fonctions a été assimilée.

Notre but n'est pas d'exposer cette méthode de la manière classique, mais plutôt d'une façon telle qu'elle va se retrouver elle-même optimisée, si l'utilisateur désire la pratiquer sur une grande quantité de données, à l'aide d'un ordinateur et non plus à la main. Elle deviendra alors, sans effort, la plus efficace de toutes les méthodes connues, car de nouvelles propriétés vont apparaître, en raison du mode de fonctionnement de l'ordinateur, au niveau où il gère de façon optimale, donc sans aucune erreur, l'information qu'on lui soumet, celui de l'algèbre binaire.

La Base

Soit une fonction Y(t)=F(Xl, X2, ... ,Xn, t) exprimant, en fonction de différents paramètres ajustables Xi ... Xn, l'évolution d'un phénomène quelconque en fonction de t, le temps.

En réalité, les différentes valeurs de Y peuvent se représenter sur un graphe cartésien en fonction de la variation de n'importe quel paramètre au choix, t ne représentant alors qu'un choix particulier (très courant toutefois, aussi bien en physique qu'en biologie). L'examen de Y=F(t) suffit alors pour exposer la base de la méthode, sur ce que l'on appelle une série chronologique.

Supposons que F exprime l'état actuel d'une théorie, c'est-à-dire la meilleure fonction connue capable d'exprimer, le plus correctement possible, et pour des intervalles raisonnamment larges de variation de tous les autres paramètres X1 ... Xn, la variation de F en fonction de t.

Pour des valeurs connues de tous les paramètres, l'utilisateur va alors tracer, sur une feuille de papier ou un écran graphique, le graphe de Y=F(t), après avoir calculé un certain nombre, toujours fini, de valeurs de Y, pour le même nombre de valeurs de t. Appelons Yt cette série de nombres.

Parallèlement, l'utilisateur a effectué des mesures avec le plus grand soin, pour ces mêmes valeurs de t, en fixant les valeurs des autres paramètres aux mêmes valeurs que celles utilisées pour les calculs. Appelons Ym cette série de nombres.

Remarque. Il n'est pas nécessaire, pour la suite de l'exposé, que l'on ait fixé des valeurs constantes pour l'ensemble des autres paramètres lorsque Y varie en fonction de t; il est seulement indispensable que, pour chaque calcul, les paramètres intervenant dans F aient tous la même valeur que ceux utilisés au même point pour effectuer la mesure. Yt et Ym représentent alors deux trajectoires, l'une théorique, l'autre expérimentale, car mesurée, dans l'hyperespace Y=F(Xl, ... Xn) en fonction de t. Le raisonnement ci-après va alors acquérir un caractère très général.

Moindres carrés et autres moindres.

On peut représenter sur le même papier ou écran les deux graphes de Yt et Ym en fonction de t en utilisant des coordonnées cartésiennes, d'une manière très classique; c'est d'ailleurs ce que font la plupart des utilisateurs et les auteurs de publications ou d'ouvrages didactiques, quel que soit le domaine d'application considéré : un graphique bien présenté parle toujours plus que de longs et austères tableaux de nombres. L'information se transmet alors de manière visuelle.

En général, aucune théorie n'est complète... Il sera extrêmement rare que tous les points d'ordonnées Ym coïncident parfaitement avec les points d'ordonnées Yt pour chaque valeur de t. Même si la coïncidence dans le cas étudié est suffisamment bonne partout pour valider la théorie, il va apparaître des écarts si on essaie a) de prévoir le devenir du système étudié (tel est le but de toutes les séries chronologiques), au delà des valeurs de t mesurables et effectivement mesurées, b) de changer de trajectoire en modifiant les valeurs des autres paramètres du système, c) d'affiner les mesures en contrôlant mieux la précision de tous les paramètres.

Chaque point des graphes de Ym et de Yt doit, honnêtement, devenir le barycentre (centre de gravité) d'une fenêtre rectangulaire, dont la longueur et la largeur de part et d'autre du point central, exprimeront les erreurs absolues sur la détermination théorique et expérimentale de t et des valeurs de Y.

Dans le cas général, les erreurs sur t sont négligeables devant celles sur Y. Avec la précision dont on dispose pour le calcul théorique de Yt sur ordinateur, (mais pas toujours si F est une fonction très complexe exigeant des millions d'itérations par exemple), les erreurs sur Yt sont négligeables devant celles sur Ym, sauf si la formule est vraiment fausse...

Considérons seulement le cas où le rectangle sur chaque point de Yt se réduit effectivement à un point, donc, sur l'écran de visualisation, à un pixel élémentaire (de l'anglais "picture element", petit carré noir rempli par le point visible sur l'écran, et qui n'est jamais un point euclidien (théoriquement sans dimension). Le rectangle construit sur chaque point de Ym, puisque l'épaisseur (largeur) sur t vaut aussi 1 pixel, devient visuellement représentable par des flèches verticales opposées

l'espace blanc entre les flèches ayant alors aussi 1 pixel de largeur et 1 pixel d'épaisseur.

Si théorie et mesures coïncident exactement, c'est-à-dire au mieux pour la précision affichable sur l'écran, le pixel théorique (noir) vient boucher le trou entre les deux flèches, exactement. Si l'accord est acceptable, le pixel noir théorique devient invisible, car il se situe dans l'une des flèches, soit en haut, soit en bas. Il suffit alors, sur un écran polychrome, de lui affecter par exemple une couleur verte pour le voir néanmoins; sur un écran monochrome, on peut, soit remplacer le pixel noir théorique par un marqueur (croix, carré, losange, petit rond, etc...), soit, encore plus simplement, inverser sa couleur de sorte que l'une des deux flèches noires apparaisse affectée d'un trou blanc.

Remarque. Les flèches n'ont pas nécessairement toutes la même hauteur pour tous les points, car la précision expérimentale peut varier d'un point à l'autre.

Lorsque le point mesuré se situe hors des flèches, c'est-à-dire ne correspond pas à la théorie, il sera toujours visible, au dessus ou en dessous de sa paire de flèches. Sur un écran polychrome, on pourra alors le faire apparaître comme un pixel rouge; l'expérience prouve que cette représentation colorée transmet l'information de coïncidence, parfaite, acceptable ou inacceptable, visuellement, pour l'ensemble du graphe, au cerveau humain d'une manière très efficace donc

rapide.

Mais l'expérience reste ce qu'elle est : un absolu... relatif.

Absolu, parce que si les conditions expérimentales sont reproductibles et suffisamment fines, les flèches conserveront, sur tout graphe dessiné dans les mêmes conditions, la même position de leur trou central pour tous les points. Honnêtement, et dans tous les cas, seuls les points rouges sont à prendre en considération pour affiner la seule entité maintenant modifiable... la théorie.

Relatif, parce que, si, par un moyen quelconque on parvient à réduire la hauteur des flèches, les points verts vont, petit à petit mais toujours discrètement, devenir rouges et ce, pour l'ensemble des points possibles de toutes les trajectoires possibles. L'optimiseur idéal, car le travail de l'utilisateur, économiste, physicien ou biologiste, est toujours une optimisation permanente, doit être capable de corriger F pour toutes les trajectoires possibles dans l'hyperespace des paramètres, donc de supprimer un maximum de points rouges.

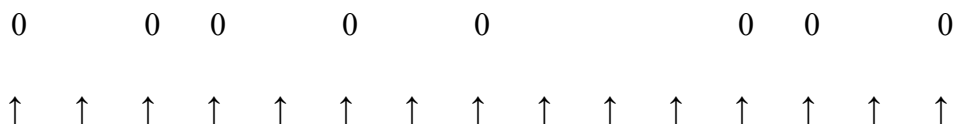
On assiste donc à un match contradictoire: Plus on corrige F, moins il reste de points rouges; plus on affine l'expérience, plus il apparaît de point rouges. A ce stade, il importe d'effectuer plusieurs remarques :

Si, sur ce graphe visuel, on décide d'appliquer à chaque fois une transformation sur les ordonnées ramenant les trous des couples de flèches sur une même ligne horizontale, on perd l'information concernant les ordonnées absolues, mais on conserve la seule information nécessaire pour pratiquer l'optimisation, indépendamment du choix de la méthode d'optimisation, les écarts. Ceux-ci sont alors donnés, pour chaque point rouge par un nombre entier nécessairement non nul, exprimant une altitude, positive ou négative, en pixels.

Les suites de valeurs Yt et Ym peuvent être avantageusement remplacées par une nouvelle suite que l'on va définir maintenant, avant de choisir la méthode d'optimisation.

Ladite suite est nécessairement un masque binaire, de même longueur que les suites Yt ou Ym : appelons-la R comme rouge. Elle contient 0 si le point n'est pas rouge, donc n'intervient pas dans la future optimisation, et 1 dans le cas contraire. Numériquement, ces mêmes valeurs 0 et 1 sont aussi des coefficients de pondération réduits à leur plus simple expression : une opposition logique : soit "inutile de considérer", soit "à considérer", l'une des propositions excluant l'autre ipso facto.

Par une autre transformation conceptuelle, ramenons toutes les flèches à la même longueur (elles sont déjà à la même altitude) en affichant un point au dessus s'il est rouge sur l'écran polychrome et rien du tout s'il n'est pas rouge (peu importe que dans le graphe original le point rouge se situe au-dessus ou en dessous des flèches). Supprimant alors la flèche du bas, nous obtenons un dessin ou plutôt un diagramme tel que



Maintenant, le choix d'une méthode d'optimisation plutôt qu'une autre, sur un critère plutôt qu'un autre, donc d'une fonction de coût quelconque, constitue toujours un arbitraire.

Seul le but final a de l'importance faire disparaître les points rouges, ramener leur ensemble à un ensemble vide.

Indépendamment du domaine d'application, de la complexité de F (sa non-linéarité), du nombre de mesures mises en jeu, des valeurs numériques mesurées ou calculées, de la trajectoire choisie pour ce faire, on a parfaitement le droit d'essayer d'optimiser la théorie sur le seul critère d'existence ou non d'un écart entre l'expérience et cette dernière, et ce, pour simplifier, en faisant appel à la méthode effectivement la plus simple et la plus connue, celle des moindres carrés.

Toutefois, avant de tenter de l'appliquer, il convient de rapporter, le plus objectivement possible, certaines critiques couramment émises à son sujet.

On a souvent reproché à la méthode des moindres carrés de ne pas considérer les écarts entre les points situés n'importe où sur les graphes de la même façon c'est-à-dire avec la même importance : la méthode optimise alors mieux les points pour lesquels les écarts sont plus grands en valeur absolue.

Déjà, on peut s'apercevoir qu'une optimisation sur la seule existence (1) ou la non-existence (0) doit être indépendante de l'ordre des points sur le graphe. On sait en effet qu'il est possible d'effectuer sur t un changement de variable tel qu'il en résultera, sur le diagramme ci-dessus, une certaine permutation de l'ordre des points qui n'aura aucune incidence sur le raisonnement suivi jusqu'à maintenant l'objection s'écroule d'elle-même, sans qu'il soit nécessaire d'écrire un seule expression mathématique à ce sujet. La simple logique binaire suffit, encore une fois.

Généralisons d'ores et déjà aussi la méthode des moindres carrés en critiquant cette fois la signification du mot "carrés".

Si Z est un vecteur, suite de valeurs toutes nulles exprimant, par définition un idéal, l'absence totale d'écart (ou d'erreur décelable) pour un ensemble d'échantillons quelconques relevés à la meilleure précision possible sur un phénomène quelconque, Z représente l'absolu.

Soit V un autre vecteur de même longueur (ou cardinal) que Z, exprimant soit par 0, pour le même ensemble, l'absence d'écart perceptible par rapport à Z, soit par 1 qui correspond à une non-absence (donc une existence) d'écart.

La stricte application des moindres carrés va consister à minimiser, par un procédé adéquat, la somme des différences V-Z élevée au carré.

Mais Z étant partout nul, ce procédé revient à minimiser V seul élevé au carré.

Mais V, étant sur toute son étendue, égal, soit à 0 soit à 1 par définition, V au carré est toujours égal à V lui-même.

On raisonnera maintenant sur V, élevé, terme à terme, à une puissance quelconque mais non nulle, sachant que 0 élevé à ladite puissance vaut toujours 0 , et que 1 élevé à la même puissance vaut toujours 1.

On se rend compte que ce raisonnement exprime dans toute sa généralité un théorème essentiel de l'optimisation idéale :

Toute optimisation réalisée sur un système ou ensemble quelconque exprimé par des parités d'existence ou de non-existence d'écarts observables, est, a priori, optimale, car elle ne dépend ni de l'ordre choisi pour exprimer ces écarts, ni de la puissance à laquelle on va opérer.

Ce théorème a plusieurs conséquences ou corollaires

a) Ni la théorie des Nombres ni celle des Fonctions n'interviennent plus dans le raisonnement, ni par un choix d'ordre (au sens de l'ordre dans un ensemble dit ordonné), ni par un choix d'ordre (au sens d'élévation à une puissance imposée ou arbitrairement décidée).

b) Le même raisonnement s'applique aussi si l'on désire optimiser non plus une fonction F quelconque, mais l'une quelconque de ses intégrale ou de ses dérivées, par rapport à n'importe quelle variable ou paramètre quelconque, par rapport à une combinaison quelconque des ces variables ou paramètres. Il s'appliquerait aussi dans le cas d'une intégration ou d'une dérivation d'ordre non entier, donc quelconque, couvrant ainsi, toujours a priori, l'ensemble des considérations relatives à la combinatoire ou la complexité. Le mot "ordre" devient le plus important du raisonnement, car il couvre maintenant, en outre, aussi l'ordre de dérivation et d'intégration.

Une optimisation basée sur des parités d'existence ou de nonexistence et strictement sur ces seules parités, peut donc s'effectuer dans le plus grand désordre apparent et rester pourtant optimale à tous les ordres à la fois, quelle que soit la signification donnée au mot "ordre" parmi les trois différentes considérées jusqu'ici. Il est d'ailleurs possible (mais non nécessaire) de démontrer, par un développement limité en série de Taylor ou de Mac Laurin, que les deux derniers sens du mot "ordre" (élévation à une puissance quelconque, et ordre d'intégration ou de dérivation) ne sont pas indépendants, indépendamment de la fonction F et de son domaine d'application.

L'axiomatique développée ici ne sera jamais soumise, même pour un nombre a priori infini de parités, aux conséquences du Théorème de Gödel (1931), lequel énonce que toute axiomatique aboutit un jour à tomber sur une proposition indécidable. Le théorème s'applique en effet à toute axiomatique basée sur la théorie des nombres. Or, il est lui-même démontré en admettant comme un axiome ladite théorie et devient indémontrable si l'on renonce à cet axiome.

L'axiomatique développée ici ne sera jamais soumise, même pour un nombre infini de parités, à l'axiome de continuité, toute prise de mesure par échantillonnage et tout procédé de calcul réalisable (algorithme) étant nécessairement une suite d'opérations discrètes. En outre, et par définition, il ne peut exister de continuité entre une existence et une non-existence et réciproquement : L'optimiseur idéal ne pourra donc jamais être décrit correctement par aucune fonction continue. Par contre, il pourrait être descriptible par un algorithme exprimé nécessairement dans la même algèbre que celle qui décrit déjà, encore une fois, nécessairement, les données à traiter.

Questions : A partir de ce raisonnement, et de ce raisonnement seul, peut-on prévoir :

a) si l'algorithme discret exprimant le fonctionnement de l'algorithme idéal d'optimisation sera unique, à égalité de propriétés;

b) s'il est possible d'en imaginer ou d'en trouver un encore supérieur donc plus général.

Il semble bien que la seule réponse possible à ces deux questions, à choix binaire, soit **NON**.

Déjà, et sans que l'on ait cherché à définir avec précision l'optimiseur idéal, le raisonnement ci-dessus, sauf s'il est démenti, soit par preuve du contraire (encore un choix binaire), soit par une preuve expérimentale, montre que toute tentative d'optimisation de quoi que ce soit, utilisant soit la théorie des fonctions continues soit la théorie des nombres, revêtira un caractère d'optimalité non satistaisant.

Il en serait de même de toute tentative d'infirmer le présent raisonnement, soit en utilisant une axiomatique basée sur la théorie du continu, soit une autre, même discrète, mais basée sur la théorie des nombres et sur la nécessité d'introduire la notion d'ensemble ordonné (attention à Gödel), soit, ce qui serait encore pire, sur une combinaison des deux axiomes inutiles contestés ici (cas de la théorie des fonctions continues). La seule infirmation possible ne peut provenir que d'un autre raisonnement exclusivement logique (ne comportant aucun nouveau postulat), ou, pire, que d'une illogique basée sur l'arbitraire, donc sur le hasard, ou sur la poursuite obstinée d'une croyance forte en certaines erreurs initiales.

(Toute théorie se comporte comme un système dynamique : quelle que soit la fonction ou l'algorithme qui la font évoluer, de préférence dans le bon sens, elle restera essentiellement soumise à ses conditions initiales, renfermées dans son axiomatique initiale, donc dans ses postulats.)

En outre, l'optimiseur idéal, basé sur la seule logique des parités (ou des contraires) deviendrait à la fois l'énoncé correct (et le seul possible) à la fois du Principe de Moindre Action sous sa forme primitive la plus générale, indépendante de tout postulat et de toute grandeur de nature macroscopique ("La Nature est économe dans toutes ses actions", Maupertuis), et aussi celui de l'interaction élémentaire, si elle existe, comme les physiciens le postulent et la recherchent à grand renfort de crédits.