

L'utilisation de la notion de coût marginal dans la recherche opérationnelle industrielle.

par J. LESOURNE (France)

Directeur Général de la S.E.M.A.

1. Nombreux sont ceux pour qui le terme de recherche opérationnelle évoque nécessairement des modèles mathématiques complexes à base de processus stochastiques et de programmes linéaires ou non. Particulièrement sensibles à l'essor des mathématiques qui a accompagné le développement de la recherche opérationnelle, ils mettent davantage l'accent sur les outils que sur l'esprit général de la méthode. Or, beaucoup de problèmes de recherche opérationnelle, abordés avec soin et dans une optique pas trop scolaire, peuvent être résolus, sur la base d'une étude poussée de données numériques, par des modèles simples relevant le plus souvent de l'analyse marginale. C'est ce que nous voudrions montrer dans cet article qui comprendra deux parties nettement distinctes, consacrées l'une à l'exposé des éléments essentiels de l'analyse marginale, l'autre à trois exemples de problèmes traités de cette façon.

Principes de l'analyse marginale.

2. Dans la plupart des études de recherche opérationnelle, le groupe de travail s'efforce de déterminer une politique optimum, c'est-à-dire les valeurs de certains paramètres de décision $x, y, z, \dots$ qui, compte-tenu des liaisons, rendent maximum une certaine expression $R(x, y, z)$ que l'on peut toujours interpréter comme le revenu que l'unité considérée tire de l'opération.

Si x, y, z sont des variables indépendantes qui peuvent prendre toutes les valeurs d'un certain domaine fermé et continu D et si la fonction $R(x, y, z)$ est définie et doublement dérivable sur tout le domaine D , il est bien connu qu'une condition nécessaire et suffisante pour que l'ensemble de valeurs x_0, y_0, z_0 , toutes distinctes des bornes de variation, constitue un maximum *relatif* pour $R(x, y, z)$ est que :

$$\frac{\partial R}{\partial x}(x_0, y_0, z_0) = 0$$

$$\frac{\partial R}{\partial y}(x_0, y_0, z_0) = 0$$

$$\frac{\partial R}{\partial z}(x_0, y_0, z_0) = 0$$

$$\begin{aligned} & \frac{\delta^2 R}{\delta x^2} dx^2 + \frac{\delta^2 R}{\delta y^2} dy^2 + \frac{\delta^2 R}{\delta z^2} dz^2 \\ & + 2 \frac{\delta^2 R}{\delta x \delta y} dx dy + 2 \frac{\delta^2 R}{\delta x \delta z} dx dz + 2 \frac{\delta^2 R}{\delta y \delta z} dy dz \leq 0 \\ & \text{pour } x = x_0, y = y_0, z = z_0 \end{aligned}$$

quels que soient dx , dy , dz . D'où l'idée enfantine de chercher la politique optimum en déterminant d'abord les politiques au voisinage desquels la variation de revenu est du second ordre. Il reste ensuite à distinguer les minimums relatifs des maximums relatifs et, parmi ces derniers, par calcul direct, le maximum absolu.

Ces considérations extrêmement simples s'étendent, à de légères modifications près, aux maximums relatifs situés sur les frontières du domaine. Ainsi, si $x_1 \leq x \leq x_2$, la condition nécessaire et suffisante pour que x_1 soit un maximum relatif de $R(x)$ est que la première dérivée non nulle en x soit négative (1). Généralement, ce sera évidemment la première.

L'analyse marginale peut aussi s'étendre au cas où les variables de décision ne sont pas indépendantes mais doivent satisfaire à une contrainte de la forme :

$$f(x, y, z) = 0 \quad \text{ou} \quad f(x, y, z) \leq 0$$

On sait que l'on peut alors introduire un multiplicateur de Lagrange λ et raisonner comme précédemment (2) sur la fonction :

$$R(x, y, z) + \lambda f(x, y, z)$$

où λ peut s'interpréter comme le coût de la contrainte liant les paramètres de décision. Dans le cas du second type de contraintes λ est nul si $R(x, y, z)$ est maximum pour des valeurs de x, y, z telles que $f(x, y, z) < 0$.

La recherche du maximum du revenu est particulièrement simple lorsque la fonction de revenu et, éventuellement, les fonctions caractérisant les contraintes possèdent des propriétés, de convexité par exemple, qui assurent l'unicité du maximum.

Ces propriétés très simples sont d'ailleurs au cœur de toute la théorie mathématique des programmes et de la dualité. Elles ont depuis longtemps

(1) Naturellement, on suppose que la fonction $R(x)$ possède des dérivées de tous ordres, sur tout le domaine D .

(2) Mais en utilisant le théorème de Kuhn et Tucker dans le cas $f(x, y, z) \leq 0$.

donné naissance, sous la plume des économistes, à quelques notions fondamentales d'usage courant.

3. Nous en citerons trois, le coût marginal, la recette marginale, le taux de rentabilité marginal.

Tel qu'il est défini par les économistes, le coût marginal est le coût de production, d'une unité supplémentaire ou l'économie de dépense qui est la conséquence de la baisse de la production d'une unité. Généralement, il y a donc autant de coûts marginaux que de moyens permettant techniquement de faire varier la production. Néanmoins, l'usage a montré que l'on pouvait grouper les coûts marginaux en deux classes assez distinctes, les coûts marginaux à court terme dans le cadre d'une capacité de production donnée et les coûts marginaux à long terme qui supposent que la variation de production est obtenue par une variation de capacité. Les coûts marginaux à court terme sont généralement faibles au dessous de la capacité de production mais croissent extrêmement vite au voisinage de cette capacité (cette croissance ne fait d'ailleurs que traduire économiquement la notion physique de capacité). Les coûts marginaux à long terme, au contraire, ont une allure qui dépend beaucoup de la branche d'industrie considérée. On a assisté, dans les dernières années, à une utilisation de plus en plus large de la notion de coût marginal.

La recette marginale est la variation de recette entraînée par la vente d'une unité supplémentaire. Si le prix de vente est peu sensible à la quantité vendue, la recette marginale se confond naturellement avec le prix, mais il n'en est plus de même si le prix de vente baisse lorsque la quantité vendue augmente. La recette marginale est alors inférieure au prix de vente. La notion de recette marginale peut d'ailleurs se généraliser et s'appliquer aux différents paramètres qui caractérisent une politique de vente. On peut parler de recette marginale procurée par un représentant supplémentaire, de recette marginale liée à l'accroissement d'un budget de publicité, etc.

La notion de *taux de rentabilité marginal* est plus délicate à introduire. On sait qu'à chaque investissement caractérisé par sa taille x , on peut associer un revenu actualisé, fonction de x et du taux d'intérêt i , $R(x, i)$. $R(x, i)$ décroît lorsque le taux d'intérêt augmente et s'annule pour une valeur de i qui a reçu le nom de taux de rentabilité moyenne. Il y a par contre pour chaque x une valeur $i_m(x)$ du taux d'intérêt telle que, lorsque $i = i_m(x)$, le revenu $R(x, i_m)$, fonction de x seul, soit maximum. C'est cette valeur i_m qui a été définie comme le taux de rentabilité marginal; i_m est solution de l'équation en i :

$$\frac{\delta R}{\delta x}(x, i) = 0$$

La taille x du projet qui rend le revenu actualisé maximum est telle que :

$$i_m(x) = i_0$$

où i_0 est le taux d'intérêt du marché. Le taux de rentabilité marginale joue donc un rôle très important dans le choix des investissements.

4. *En résumé*, l'analyse marginale, qui introduit des notions telles que coûts marginaux, recettes marginales, taux de rentabilité marginale, s'efforce de déterminer les politiques optimums en recherchant les politiques au voisinage desquelles les variations de revenus sont négatives pour tous les déplacements infinitésimaux compatibles avec les liaisons. L'avantage de l'analyse marginale est dans sa facilité d'emploi et sa clarté d'interprétation. En mettant en évidence les aspects concrets du problème, elle conduit naturellement les chercheurs à attacher toute l'importance souhaitable à l'évaluation correcte des données numériques. Elle les guide dans ce travail en les éloignant des faux problèmes de dépenses passées ou de répartition arbitraire de dépenses. Par contre, de par sa structure, elle suppose le plus souvent que les politiques possibles peuvent varier d'une façon continue. Elle ne détecte que des extrema relatifs et se prête mal au calcul lorsque les variables doivent satisfaire à un grand nombre de contraintes. Son champ d'application n'en reste pas moins très important ainsi que nous allons chercher à le montrer dans les exemples qui suivent.

5. *Le premier exemple* concerne la recherche de la localisation optimum d'un dépôt central d'approvisionnement destiné à alimenter n centres secondaires. Le dépôt reçoit de fournisseurs i des quantités q_i et distribue aux centres j des quantités q'_j ($\sum q_i = \sum q'_j$). Les voies de communication (voies ferrées et routes) sont supposées données. Les transports du dépôt aux centres se font par la route. Il est facile d'étudier pour chaque localisation du dépôt, les frais de transport totaux. Si une localisation du dépôt située le long d'une route, mais distincte d'un nœud de communication, est optimum, tout léger déplacement du dépôt le long de cette rue ne doit entraîner qu'une variation du second ordre des dépenses. Si une localisation du dépôt en un nœud de communication est optimum, tout déplacement du dépôt le long d'une route quelconque partant du nœud de communication doit se traduire par une augmentation des frais de transport. Souvent, le réseau de transport n'est pas trop complexe; il est donc très facile de déterminer rapidement, par approxi-

mation, les minimums relatifs de dépenses envisagées. Le calcul des frais de transport en chacun d'eux permet ensuite d'obtenir la meilleure solution. Dans les problèmes de ce genre, l'emploi de l'analyse marginale est d'autant plus intéressant que les économies à réaliser sont souvent modérées : leur ordre de grandeur ne justifierait pas, s'il n'existait cette méthode, des recherches poussées susceptibles d'absorber des sommes un peu importantes. ✓

6. Nous emprunterons *le second exemple* aux problèmes posés par la fixation du niveau de production d'un gisement de gaz naturel. Pour exploiter ce gisement dont la réserve est supposée connue, il faut construire des puits et des unités d'épuration et de traitement qui vivent p années. Des études de marché ont permis de se faire une idée, en fonction du niveau de production, du prix de vente du mètre cube de gaz à la sortie des unités de traitement. Le prix est évidemment une fonction décroissante du niveau; et, si l'on admet qu'il y a réversibilité, il baisse en fonction de la production dans la période de mise en exploitation du champ, mais croît en fin d'exploitation, lorsque la production diminue. Dans la pratique, cette éventualité apparaît douteuse, mais nous nous y limiterons ici, puisqu'il ne s'agit que d'illustrer un mécanisme de raisonnement. Toutes les unités de traitement sont supposées avoir une même capacité, déjà choisie.

Si l'on admet qu'un nombre quelconque d'unités peut être construit simultanément, à la période initiale, le diagramme de production en fonction du temps résulte de l'empilement d'un certain nombre de rectangles dont la hauteur correspond à la production d'une unité et la base à p années ($0, p$; $p, 2p$; $2p, 3p$; etc.). Les réserves étant données, le nombre total de rectangles l'est également. En fin d'exploitation, le débit des puits baisse, et il faut construire un plus grand nombre de puits pour avoir un niveau de production donné.

Mais un calcul montre que cet événement se produit suffisamment tard pour pouvoir être, du fait de l'actualisation, pratiquement négligé. Il est donc facile de calculer la valeur, actualisée au début de la marche d'une unité de traitement, des dépenses relatives à cette unité et au puits correspondant. Il est également aisé de déterminer le supplément de revenu ou *revenu marginal* que l'on tire de la construction, à partir d'un nombre d'unités donné, d'une unité supplémentaire pendant la période $k_p, (k+1)p$. On peut donc associer un revenu marginal à chaque portion d'un rectangle sur le graphique de production, ce revenu marginal décroît de la gauche vers la droite et du bas vers le haut. On choisira donc les rectangles dans l'ordre décroissant des revenus marginaux jusqu'à ce que l'on ait pris un nombre de rectangles compatibles

avec les réserves. En fait, les premières unités de production ne peuvent être construites simultanément; il faudra donc ensuite étudier l'influence sur les différents revenus marginaux du décalage dans le temps de la construction des différentes unités de traitement de la première génération.

L'analyse marginale peut également s'appliquer à l'étude de la politique optimum en fin d'exploitation. Cette politique peut se rechercher parmi les politiques suivantes à partir de l'instant pris comme origine des temps, où la baisse de pression du gisement commence à se faire sentir sur le débit des puits :

- construction de puits supplémentaires pour maintenir la pression constante de O à T_R
- fixation du nombre de puits à partir de l'instant T_R .

A partir de T_R , le débit baisse régulièrement et on retire du service les unités de traitement les unes après les autres. Au temps T_F où le débit des puits atteint une valeur minimum d_m , on cesse l'exploitation. La valeur optimum de T_R se calcule en écrivant qu'une légère variation de T_R est sans influence sur le revenu total actualisé.

7. Le troisième exemple portera sur la détermination d'un contrat d'abonnement optimal à Electricité de France pour une entreprise ayant une production autonome d'énergie électrique.

Le nouveau tarif appliqué par Electricité de France, désigné sous le nom de Tarif Vert, est caractérisé par des coûts d'abonnement et des coûts du kWh variables suivant le type d'abonnement souscrit, la puissance souscrite et la période d'utilisation.

Réduit à ses principales dispositions, le Tarif Vert comporte trois variantes : le Tarif Général, le Tarif d'Appoint et le Tarif de Secours. Un contrat est caractérisé par l'abonnement pour une certaine puissance dans un des tarifs précédents ou dans une combinaison de ces trois tarifs. Si l'abonné possède des sources autonomes d'énergie électrique, il ne peut souscrire que des abonnements « d'Appoint » et de « Secours » : c'est le cas de l'entreprise étudiée.

Le montant facturé par E.d.F. se compose toujours de trois éléments : une prime fixe annuelle qui ne dépend que de la puissance souscrite, facturée à un taux dégressif, le prix de l'énergie effectivement consommée et, éventuellement, des pénalités pour dépassement de puissance.

Le coût de ces trois éléments est fonction des périodes tarifaires au nombre de cinq :

- les heures de pointe, de 6 heures à 8 heures et de 20 heures à 22 heu-

res pendant les mois de novembre, décembre, janvier et février (période 1);
 — les heures pleines (de 6 heures à 22 heures sauf les dimanches et les heures de pointe) d'hiver, du 1^{er} octobre au 31 mars et d'été, du 1^{er} avril au 30 septembre (période 2 et 3);

les heures creuses d'hiver et d'été (périodes 4 et 5).

Tout abonnement pour la période de rang i est acquis ipso facto pour les périodes de rang supérieur, mais il est possible de moduler l'abonnement, la puissance souscrite dans la période tarifaire ($i + 1$) devant alors être supérieure d'au moins 20% à celle qui est souscrite dans la période i (avec une différence minimum de 20 kW entre les deux périodes).

En Tarif d'Appoint, comme en Tarif de Secours, le taux de la prime fixe applicable au supplément de puissance souscrite pendant une période est égal au taux de la prime d'heures de pointe multiplié par un coefficient réducteur croissant avec le numéro d'ordre de la période. Les taux de primes dépendent aussi des puissances souscrites.

E.d.F. a enfin établi un système de pénalisation des dépassements de la puissance souscrite.

8. Pour résoudre analytiquement le problème il s'agit de déterminer deux groupes de variables :

- 1) les puissances P_i à souscrire dans chaque période tarifaire,
- 2) les tranches A_i et S_i à souscrire respectivement en « Appoint » et en « Secours » (telles que $A_i + S_i = P_i$).

Dans le cas de l'entreprise considérée, la demande aléatoire à E.d.F. avait la même distribution de probabilité dans toutes les périodes tarifaires et un disjoncteur était installé sur le branchement E.d.F. de façon à éviter les dépassements de puissance souscrite.

1) L'entreprise a estimé la pointe instantanée maximale M à laquelle il était possible de se limiter et la capacité minimale m de sa production autonome d'énergie électrique. $M - m$ est donc la valeur de P_1 et le disjoncteur doit donc être réglé à ce niveau de puissance. Les valeurs de P_i ($i \geq 2$) sont uniquement soumises, par suite des dispositions tarifaires, aux contraintes $P_{i+1} \geq P_i$.

2) Les valeurs des tranches A_i et S_i à souscrire en « Appoint » et en « Secours » pendant la période i sont également soumises aux conditions $A_{i+1} \geq A_i$ et $S_{i+1} \geq S_i$. La puissance souscrite $P_i = A_i + S_i$ est suffisante pour couvrir les besoins toute l'année, mais le kWh d'appoint coûtant moins cher que le kWh de secours, il peut y avoir intérêt à augmenter dans les autres périodes la puissance souscrite en appoint.

On aura donc : $S_i = S_{i-1}$ ($i \geq 2$).

La détermination des A_i nécessite la connaissance de la loi de probabilité de la demande instantanée. Appelons $F_i(A_i)$ la probabilité d'avoir une demande inférieure ou égale à A_i .

Dans le cadre de ce que nous avons dit dans la première partie, la valeur optimale de A_i , s'obtient en écrivant que la variation de la dépense totale est nulle pour une petite variation de A_i .

La prime fixe ayant, par suite des rabais, des valeurs décroissantes dans les tranches successives de puissance, il est nécessaire de faire une hypothèse sur la tranche dans laquelle se trouve A_i . Si le résultat est incompatible avec l'hypothèse, il faut reprendre le calcul en se plaçant dans une nouvelle tranche; exceptionnellement plusieurs hypothèses peuvent conduire à des optimums locaux auquel cas il est nécessaire de procéder à l'évaluation des coûts totaux pour déterminer le vrai minimum.

Soient f_i^a et f_i^s les montants des primes fixes « d'Appoint » et de « Secours » dans la tranche considérée, e_i^a et e_i^s les coûts du kWh « d'Appoint » et de « Secours »; soit n_i le nombre d'heures de fonctionnement de l'usine pendant la période i ; la variation des dépenses de la période i , qu'entraîne un accroissement d'un kWh de l'abonnement « d'Appoint » est :

$$f_i^a - f_{i+1}^a + n_i (e_i^a - e_i^s) (1 - F_i) \quad (i \geq 2)$$

à condition que la période i soit indépendante de la période $(i+1)$, c'est-à-dire que la valeur de $F_i(A_i)$ obtenue en annulant cette expression donne une valeur de A_i satisfaisant à la relation $A_i \leq A_{i+1}$; en effet, dans ce cas, la souscription d'un kWh supplémentaire « d'Appoint » ne modifie pas la souscription de puissance de « Secours » (puisque $S_i = S_1$), et diminue d'une unité la puissance souscrite en « Appoint » dans la période suivante.

Pour la période $i = 1$, la variation de dépenses est légèrement différente, puisque la somme des abonnements « d'Appoint » et de « Secours » est constante.

$$f_1^a - f_2^a - f_1^s + n_1 (e_1^a - e_1^s) (1 - F_1) \quad (1)$$

Les cinq équations obtenues en annulant les expressions ci-dessus fournissent cinq valeurs de $[1 - F_i(A_i)]$, d'où cinq valeurs de A_i . Les données numériques du Tarif Vert sont telles que ces valeurs de $(1 - F_i)$ sont toujours

positives. Néanmoins, les valeurs ainsi obtenues isolément pour les A_i ne sont en général pas compatibles avec les liaisons.

Plusieurs cas sont alors possibles :

a) $1 - F_i > 1$. Il n'existe pas de valeur de A_i correspondant à $F_i < 0$;

la dépense croît avec F_i puisque $(e_i^a - e_i^s)$ est négatif et, par suite, la valeur optimale de $A_i = A_{i-1}$. On ne peut déterminer leur valeur commune qu'en étudiant simultanément les périodes i et $i - 1$.

b) $1 - F_i < 1$ et $F_i < F_{i+1}$, la différence $F_{i+1} - F_i$ étant suffisamment grande pour que :

$$A_{i+1} \geq A_i (1 + 20\%)$$

La valeur de $F_i(A_i)$ fournit une valeur acceptable de A_i qui est l'optimum dans la période i .

c) $1 - F_i < 1$ et $F_i > F_{i+1}$, ou $F_i < F_{i+1}$, mais avec une différence trop petite pour que A_{i+1} excède A_i d'au moins 20%. Les périodes i et $i + 1$ ne sont pas indépendantes comme dans le cas (a). Dans les deux cas où deux périodes consécutives ne sont pas indépendantes ((a) et (c)), il est nécessaire de recalculer l'optimum pour l'ensemble des deux périodes groupées. La valeur commune de F'_i pour les deux périodes i et $i + 1$ est alors donnée par l'équation :

$$f_i^a - f_{i+2}^a + [(e_i^a - e_i^s) n_i + (e_{i+1}^a - e_{i+1}^s) n_{i+1}] (1 - F'_i) = 0 \quad (2)$$

La valeur de F'_i détermine une valeur acceptable de A_i et A_{i+1} , ou exprime que les deux périodes i et $i + 1$ ne sont pas indépendantes de la période $i + 2$ ou de la période $i - 1$. Dans ce dernier cas, il faut reprendre le calcul pour l'ensemble des trois périodes groupées, et ainsi de suite jusqu'à ce que les optimums locaux soient compatibles avec toutes les liaisons.

Les calculs numériques peuvent être faits à la règle très simplement sur la base du tarif et de la loi de probabilité des consommations.

Remarque.

Si la demande à E.d.F. n'est pas la même dans les différentes périodes tarifaires, les niveaux P_i de l'abonnement total varient avec i . Les raisonnements marginaux permettant de déterminer P_i dans chaque période (quand il n'y a pas de disjoncteur) peuvent conduire à des valeurs inacceptables, c'est-à-dire qui ne soient pas croissantes avec i ; il est alors nécessaire dans ce cas de regrouper certaines périodes, comme il a été fait plus haut pour la détermination de l'abonnement « d'Appoint ».

Conclusion.

9. Ces quelques exemples donnent une idée des possibilités très vastes d'application de l'analyse marginale. Ils ne suffisent peut-être pas malheureusement à mettre en évidence son intérêt pédagogique. Il est certainement bien préférable d'initier un chercheur à la recherche opérationnelle en l'habituant à appliquer correctement les notions marginales qu'en le plongeant d'emblée dans des problèmes de programmation complexes où il perdra de vue la structure profonde des phénomènes et sera principalement attiré par les constructions purement mathématiques.