Master MIMSE - Année 2

Optimisation 2

Optimisation sans contraintes - (R)appels

Fonctions à une variable

- On veut optimiser une fonction $f : \Re \longrightarrow \Re$ sur un espace $X \subset \Re$.
- X est un intervalle ou une collection d'intervalles.
- On suppose que f est dérivable sur l'intérieur de X.
- L'optimum est soit sur une borne d'un des intervalles,
- soit en un point $\hat{x}$ de l'intérieur de X vérifiant :

$$f'(\hat{x}) = 0.$$

- Si de plus f est deux fois dérivable,
  - $f''(\hat{x}) > 0$ implique un minimum local,
  - $f''(\hat{x}) < 0$ implique un maximum local.
- S'il y a des points de X où f n'est pas dérivable, ils peuvent être optimaux.

Fonctions à une variable - Méthode de Newton

- Ne donne qu'un optimum local : information locale : résultat local.
- Pour avoir un optimum global, il faut des informations globales : convexité.
- Résoudre $f'(x) = 0$ n'est généralement pas facile, même pour des polynômes.
- Pas d'écriture analytique.
- On peut calculer de façon approchée une solution de $f'(x) = 0$ par la méthode de Newton :

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)},$$

- Partir d'un point initial x_0 pas trop loin de la solution !

Fonctions à plusieurs variables

- Conditions analogues au cas à une variable, mais qui font appel aux dérivées partielles et aux gradient.
- Optimums toujours éventuellement présents à la frontière de X.
- Soit $x = (x_1, \ldots, x_n)$.
- Si $\frac{\partial f}{\partial x_i}(x)$ existe pour tout x et tout i,
- Si f a un extremum local en $\hat{x}$ dans l'intérieur de X,
- Alors $\frac{\partial f}{\partial x_i}(\hat{x}) = 0 \ \forall i$
- Ou encore $\nabla f(\hat{x}) = 0$.
- Si f est deux fois différentiable, soit $\nabla^2 f$ sa matrice hessienne (matrice des $\frac{\partial^2 f}{\partial x_i \partial x_j}$).
- Si $\nabla^2 f(\hat{x}) \succ 0$ (toutes ses valeurs propres > 0) alors minimum local,
- Si $-\nabla^2 f(\hat{x}) \succ 0$ alors maximum local.
- En général, ni l'un ni l'autre : **point-selle !**

Fonctions à plusieurs variables - Méthode de Newton-Raphson

- Les mêmes remarques que précédemment s'appliquent :
  - Résultats seulement locaux,
  - Difficultés à résoudre $\nabla f = 0$
  - En plus, système de n équations à n inconnues.
- On peut toujours calculer de façon approchée une solution de $\nabla f(x) = 0$ par la méthode de Newton-Raphson :

$$x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n),$$

- Attention ! Il faut pouvoir calculer l'inverse de $\nabla^2 f$ en tous les points !
- Attention (bis) ! Le conditionnement de $\nabla^2 f$ peut causer des erreurs de calcul qui se répercutent.

Fonctions à plusieurs variables - Convexité

- On dit qu'un espace S est convexe si pour toute paire de points x, y de S et tout $0 \leq \alpha \leq 1$, le point $\alpha x + (1 - \alpha)y \in S$.
- Le segment $[x, y] \subset S$.
- Une fonction f définie sur S convexe est convexe si pour toute paire de points x, y de S et tout $0 \leq \alpha \leq 1$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

- De plus,

$$f(x) - f(y) \geq \nabla f(x)^T (x - y).$$

- La courbe de f est en-dessous de ses cordes et au-dessus de ses tangentes.
- Si f est convexe sur X convexe et si $\hat{x}$ est un minimum local de f alors $\hat{x}$ est un minimum global de f.
- Si f est concave sur X convexe et si $\hat{x}$ est un maximum local de f alors $\hat{x}$ est un maximum global de f.
- Si $\forall x\ \nabla^2 f(x) \succeq 0$ alors f est convexe.

Cas d'une fonction quadratique

- On considère le cas d'une fonction quadratique $f(x) = \sum a_i x_i^2 + \sum \sum a_{ij} x_i x_j + \sum b_i x_i + c$.
- On veut minimiser f sur $\Re^n$.
- On peut écrire f sous la forme

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

  avec $A_{ii} = 2a_i$ et $A_{ij} = A_{ji} = a_{ij}$.
- $\nabla f = Ax + b$.
- $\nabla^2 f = A$.
- Si $A \succ 0$ alors f admet un minimum unique au point $x^* = -A^{-1} b$.
- Sinon, pas de minimum (peut partir à $-\infty$) en général.

Algorithmes de descente

- Pour trouver un minimum local de $\boldsymbol{f}$ quelconque, on a des algorithmes de descente.
- Principe :
  1. Point initial $\boldsymbol{x_0}$.
  2. Trouver une direction de descente $\boldsymbol{d}$: une direction telle que $\boldsymbol{d^T \nabla f(x_n) < 0}$.
  3. Eventuellement calculer un pas de descente $\boldsymbol{\rho}$.
  4. $\boldsymbol{x_{n+1} = x_n + \rho d}$.
  5. Si $\boldsymbol{\nabla f(x_{n+1}) \neq 0}$ réitérer.
- $\boldsymbol{\rho}$ peut être fixe, mais $\boldsymbol{\rho}$ petit : convergence longue, $\boldsymbol{\rho}$ grand : risque de diverger.
- $\boldsymbol{\rho}$ décroissant selon un schéma donné.
- $\boldsymbol{\rho}$ obtenu par recherche linéaire.

Recherche linéaire

- But : Minimiser selon ρ la fonction $f(x_n + \rho d)$.
- Approche 1 : exprimer $\phi(\rho) = f(x_n + \rho d)$ et optimiser ϕ. On doit tout recalculer quand on change x_n, risques d'erreur, difficultés à programmer.
- Approche 2 : il existe des méthodes pour optimiser $f(x_n + \rho d)$ sans avoir à l'exprimer explicitement, par réduction d'intervalles.
- On suppose ϕ **unimodale** : elle décroît jusqu'à ρ^* puis elle recroît.
- Méthode de Fibonacci :

$$s_i = a_i + (b_i - a_i)A_{n-i}/A_{n+2-i}$$

$$\bar{s}_i = a_i + (b_i - a_i)A_{n+1-i}/A_{n+2-i}$$

avec les A_i les éléments de la suite de Fibonacci 0,1,1,2,3,5,...

- On évalue aux points s_i et $\bar{s}_i$ et on réduit l'intervalle.
- Méthode de la section dorée : on approche les points précédents avec la formule :

$$s_i = a_i + (b_i - a_i)\frac{3 - \sqrt{5}}{2}$$

$$\bar{s}_i = a_i + (b_i - a_i)\frac{\sqrt{5} - 1}{2}$$

- Approche 3 : Théorème fondamental :

$$\nabla f(x_n + \rho^* d)^T d = 0.$$

Algorithme de plus forte pente

- On montre que la direction de descente qui fait décroître le plus vite $f(x_n + \rho d)$ est

$$d = -\nabla f(x_n).$$

- Algo très basique.
- En pratique, convergence pas toujours très bonne : voir exemple graphique.
- Si les x_n convergent vers x^* et que f est C^1 sur un voisinage de x^* alors x^* est un minimum local pour f.
- Algo du gradient accéléré : on fait plusieurs itérations du gradient, et le résultat nous donne la “vraie” direction de recherche.

Algorithmes de quasi-Newton

- Newton-Raphson : amélioration du gradient (tient compte de la géométrie). Optimise instantanément des fonctions quadratiques.
- En général, on l'améliore en introduisant un pas $\boldsymbol{\rho}$.
- Difficultés calculatoires : on approche la méthode (Quasi-Newton).
- Ex. Davidon-Fletcher-Powell
  1. $\boldsymbol{H_1 = I}$, $\boldsymbol{x_1}$ point initial.
  2. Itération $\boldsymbol{k}$:
     (a) $\boldsymbol{d_k = -H_k g_k}$ (on note $\boldsymbol{g_k = \nabla f(x_k)}$)
     (b) Trouver $\boldsymbol{\rho^*}$
     (c) $\boldsymbol{\sigma_k = \rho^* d_k}$
     (d) $\boldsymbol{x_{k+1} = x_k + \sigma_k}$
     (e) $\boldsymbol{\gamma_k = g_{k+1} - g_k}$
     (f) $\boldsymbol{H_{k+1} = H_k + A_k + B_k}$ avec $\boldsymbol{A_k = \sigma_k \sigma_k^T / \sigma_k^T \gamma_k}$ et $\boldsymbol{B_k = H_k \gamma_k \gamma_k^T H_k / \gamma_k^T H_k \gamma_k}$
     (g) Cas d'arrêt ? Sinon réitérer.