Hors-série :

Techniques d'optimisation convexe

Introduction

Problème

Trouver le minimiseur $\widehat{\boldsymbol{x}}$ de $J(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{H}\boldsymbol{x}\|^2 + \alpha \sum_{m=1}^{M} \phi(x_m)$

ou plus généralement de $\sum_{n=1}^{N} \psi(z_n - \boldsymbol{h}_n^{\mathrm{t}}\boldsymbol{x}) + \alpha \sum_{c=1}^{C} \phi(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c)$

sur $\mathbb{R}^M$, $\mathbb{R}_+^M$ ou un autre ensemble convexe $F \subset \mathbb{R}^M$, ϕ, ψ étant convexes

Convexité (au sens large)

- Pour un ensemble $F \subset \mathbb{R}^M$:

$$\forall \boldsymbol{x}_1,\, \boldsymbol{x}_2 \in F, \quad F \text{ contient le segment } [\boldsymbol{x}_1,\, \boldsymbol{x}_2]$$

- Pour une fonction f :

$$\forall \theta,\, \bar{\theta} = 1 - \theta, \quad f(\theta \boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) \leqslant \theta f(\boldsymbol{x}_1) + \bar{\theta} f(\boldsymbol{x}_2)$$

Propriétés...

... de conservation

① f, g convexes $\Longrightarrow$ $f+g$ convexe

② f convexe, $\alpha > 0$ $\Longrightarrow$ αf convexe

③ f convexe, g linéaire $\Longrightarrow$ $f \circ g$ convexe

... pour la minimisation

④ F, f convexes $\Longrightarrow$ l'ensemble des minimiseurs de f dans F est cvx

⑤ F, f convexes $\Longrightarrow$ pas de minima locaux dans F

⑥ $\begin{cases} F \text{ convexe,} \\ f \text{ strictement cvx} \end{cases}$ $\Longrightarrow$ au plus un minimiseur dans F

... géométriques [Luenberger 1969]

⑦ F convexe fermé $\Longrightarrow$ $\forall \boldsymbol{x} \in \mathbb{R}^M$, $\exists! \, \mathcal{P}_F(\boldsymbol{x})$ projection de $\boldsymbol{x}$ sur F telle que $\mathcal{P}_F(\boldsymbol{x}) = \arg\min_{\boldsymbol{t} \in F} \|\boldsymbol{t} - \boldsymbol{x}\|$

Méthodes de descente

Principe

Générer une suite $(\boldsymbol{x}^k)$ tel que $(J(\boldsymbol{x}^k))$ est décroissante, sous la forme

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k, \quad \alpha_k > 0 : \text{pas}, \quad \boldsymbol{d}_k : \text{direction de descente}$$

Propriété

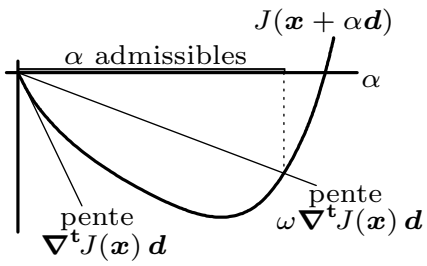
$$\begin{aligned} &\boldsymbol{d} \in \mathbb{R}^M \text{ est une direction de descente de } J \text{ en } \boldsymbol{x} \in \mathbb{R}^M \\ \Longleftrightarrow \quad & J(\boldsymbol{x} + \alpha \boldsymbol{d}) < J(\boldsymbol{x}), \quad \forall \alpha > 0 \text{ petit} \\ \Longleftrightarrow \quad & \boldsymbol{d}^{\mathrm{t}} \nabla J(\boldsymbol{x}) < 0 \end{aligned}$$

Schéma algorithmique

1. Choisir $\boldsymbol{x}^0 \in \mathbb{R}^M$;
2. Test d'arrêt : si $\|\nabla J(\boldsymbol{x}^k)\| \simeq 0$, stop ;
3. Choix de $\boldsymbol{d}^k$;
4. *Recherche linéaire* : choisir $\alpha^k > 0$ pour que J ■décroisse suffisamment■ ;
5. $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$, $k = k+1$, retour en 2.

Choix du pas

- Pas constant $\alpha^k = \alpha$: ne garantit ni la descente, ni la convergence
- Recherche linéaire exacte : $\alpha^k = \arg\min_\alpha \; J(\boldsymbol{x}^k + \alpha\, \boldsymbol{d}^k)$
  convergeant mais non réaliste, sauf cas particulier
- Pas décroissant
  condition typique de convergence : $\begin{cases} \lim\limits_{k\to\infty} \sum\limits_k \alpha^k = \infty, \\ \lim\limits_{k\to\infty} \sum\limits_k (\alpha^k)^2 < \infty \end{cases}$
- Pas adaptatif



(A) Condition d'Armijo : α^k *admissible* si
$J(\boldsymbol{x}^k + \alpha\, \boldsymbol{d}^k) \leqslant J(\boldsymbol{x}^k) + \omega\alpha^k\, \nabla^{\mathrm{t}} J(\boldsymbol{x}^k)\, \boldsymbol{d}_k, \quad \omega \in]0,1[$

(B) Recherche de α^k par *backtracking* :
adopter $\alpha^k = s\tau^p$ ($s > 0$, $\tau \in]0,1[$) pour le plus petit entier p tel que (A) est vrai.

Choix de la direction de descente

Quelques méthodes classiques... [Bertsekas 1995]

1. méthode d'ordre zéro : relaxation
2. méthode d'ordre un : gradient (pseudo-conjugué)
3. méthode d'ordre deux : Newton, quasi-Newton
4. méthode de projection : méthode de Bregman [Bregman 1967]

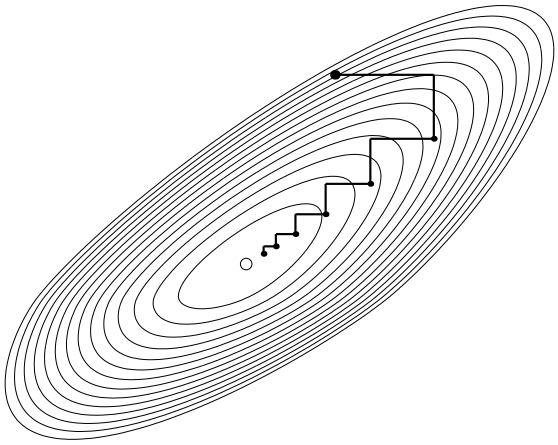
① Relaxation

Gauss-Seidel, descente coordonnées par coordonnées

Principe

Partant de $\boldsymbol{x}^0 = \{x_1^0, \ldots, x_M^0\}$, $\boldsymbol{x}^k$ connu :

x_m^{k+1} minimise $J(x_1^{k+1}, \ldots, x_{m-1}^{k+1}, x, \ldots, x_M^k)$ en x



Convergence

❶ $J : \mathbb{R}^M \to \mathbb{R}$ est de classe C^1

❷ J est strictement convexe

❸ J est coercif, *i.e.*, $\lim\limits_{\|\boldsymbol{x}\| \to \infty} J(\boldsymbol{x}) = +\infty \quad \Longrightarrow \quad \lim\limits_{k \to \infty} \boldsymbol{x}^k = \widehat{\boldsymbol{x}}$

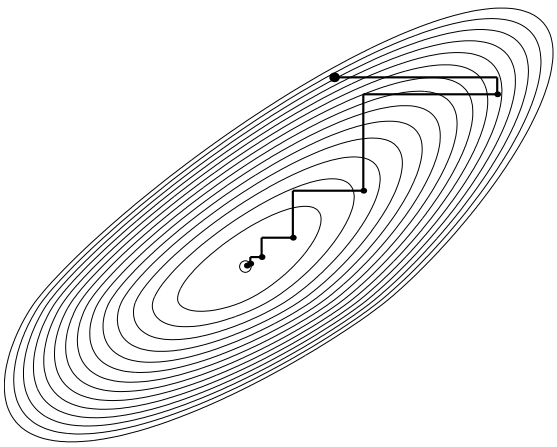
$\Big($ ou bien ❸' $F = [x_1^{\min}, x_1^{\max}] \times \cdots \times [x_M^{\min}, x_M^{\max}]$ $\Big)$

Variante : sur/sous-relaxation

$$x_m^{k+1} = x_m^k + \omega \alpha_m^k,$$

α_m^k : pas de la méthode standard

$$\begin{cases} \omega \in \,]0, 1[& \text{sous-relaxation} \\ \omega = 1 & \text{méthode standard} \\ \omega > 1 & \text{sur-relaxation} \end{cases}$$

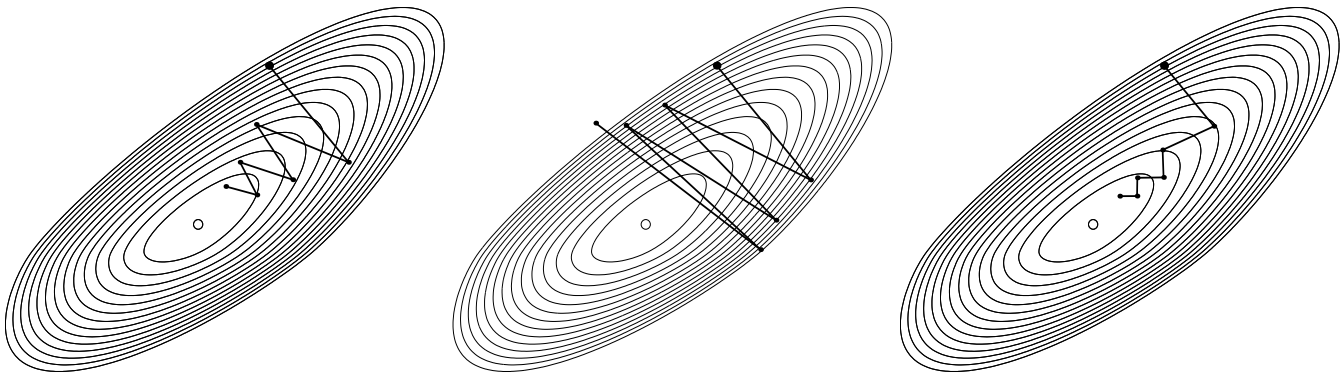


② Méthodes de gradient

Gradient standard

$$\boldsymbol{d}^k = -\nabla J(\boldsymbol{x}^k)$$

pas fixe, $\alpha = 0,55$ pas fixe, $\alpha = 0,65$ recherche ■exacte■

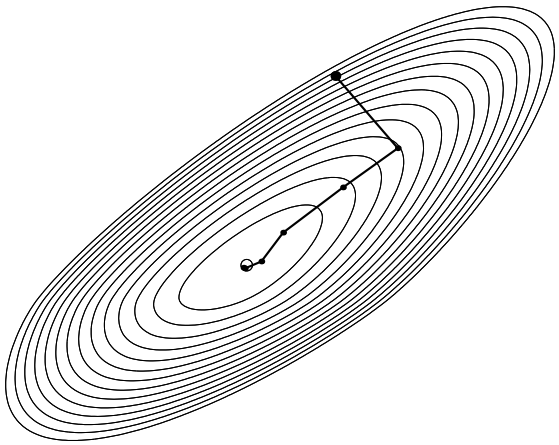


Convergence : même type de conditions que pour la relaxation

Gradient conjugué

$$\boldsymbol{d}^0 = -\nabla J(\boldsymbol{x}^0) \text{ puis } \boldsymbol{d}^k = -\nabla J(\boldsymbol{x}^k) + \beta_k \boldsymbol{d}^{k-1}$$

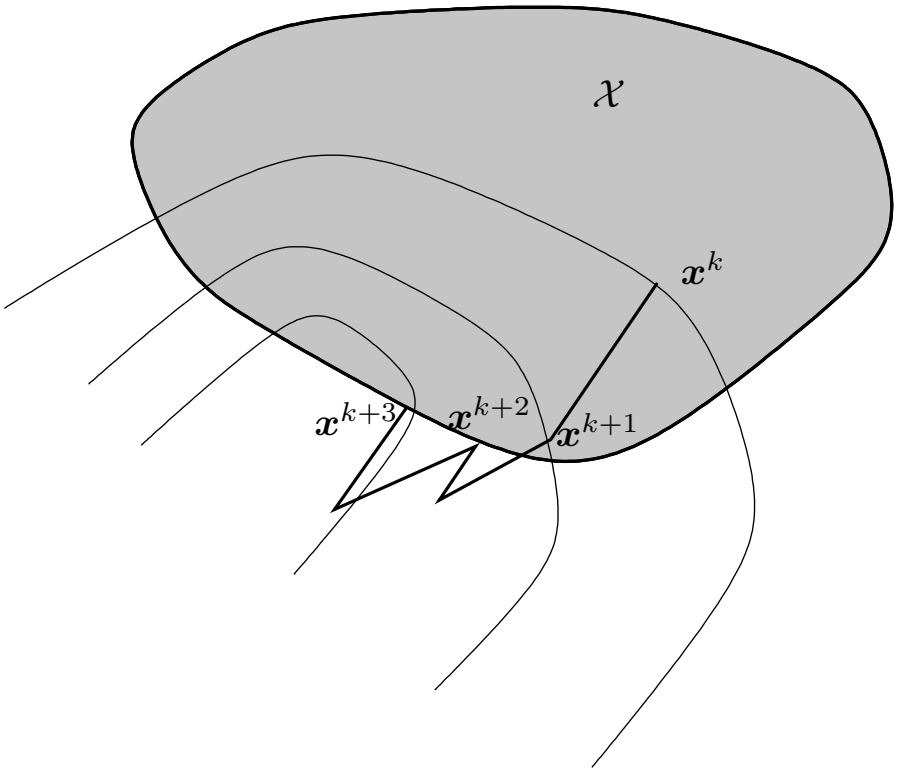
- Fletcher-Reeves : $\beta_k = \dfrac{\left\|\nabla J(\boldsymbol{x}^k)\right\|^2}{\left\|\nabla J(\boldsymbol{x}^{k-1})\right\|^2}$



- Polak-Ribiere : $\beta_k = \left\|\nabla J(\boldsymbol{x}^k)\right\|^2 - \dfrac{\left(\nabla J(\boldsymbol{x}^k)\right)^{\mathrm{t}} \nabla J(\boldsymbol{x}^{k-1})}{\left\|\nabla J(\boldsymbol{x}^{k-1})\right\|^2}$

Contraintes convexes

- Gradient projeté



[Bertsekas 1995] D. P. Bertsekas (1995), *Nonlinear programming*, Athena Scientific, Belmont, MA, USA.

[Bregman 1967] L. M. Bregman (1967), « The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming », *USSR Computational Mathematics and Mathematical Physics*, **7**, 200-217.

[Luenberger 1969] D. G. Luenberger (1969), *Optimization by Vector Space Methods*, John Wiley, New York, NY, USA, 1st ed.