Cours 11 : Gradient conjugué

Christophe Gonzales

LIP6 – Université Paris 6, France

Point faible de la méthode du gradient (1/2)

$\varphi'(\widetilde{\lambda}) = -\vec{\nabla} f(\widetilde{x})^T . \vec{\nabla} f(x) = 0$

$\Longrightarrow$ les gradients en x et $\widetilde{x}$ sont orthogonaux

$\Longrightarrow$ à chaque pas on prend une direction orthogonale à la direction précédente

cheminement de la méthode du gradient : «en zigzag»

Point faible de la méthode du gradient (2/2)

$\Longrightarrow$ pour éviter les zigzags et accélérer la convergence, on peut avoir recours à l'un des procédés suivants :

- diminuer le pas : ne pas aller jusqu'en $\widetilde{x}$.

  Polyak (66) : effectuer des pas prédéterminés en imposant une suite (λ^k) telle que $\lambda^k \downarrow 0$ et $\sum_k \lambda^k = +\infty$
  $\Longrightarrow (x^k)$ tend vers $\widehat{x}$

- utiliser d'autres directions que l'anti-gradient :

  Forsythe (1968), Luenberger (1973) :
  toutes les m itérations, au lieu de partir dans la direction de l'anti-gradient en x^k, «couper» en partant dans la direction $\delta = x^k - x^{k-m}$

- utiliser des directions «conjuguées»

Matrices définies positives

Définition d'une matrice définie positive

- Matrice M carrée $n \times n$
- M est symétrique : $M = M^T$, i.e., $M_{jk} = M_{kj}\ \forall j, k$
- $x^T M x > 0\ \forall x \neq 0$

Matrices définies positives

Définition d'une matrice définie positive

- Matrice M carrée $n \times n$
- M est symétrique : $M = M^T$, i.e., $M_{jk} = M_{kj}\ \forall j, k$
- $x^T M x > 0\ \forall x \neq 0$

$\Longrightarrow$ M correspond à une matrice de changement de base dans un espace vectoriel de dimension n :

Propriété

- si M = matrice définie positive
- alors $\exists Q$ matrice carrée
- $\exists P$ matrice diagonale à coeffs > 0 telles que :
  $M = Q^T P^T P Q$
  $\Longrightarrow x^T M x = x^T Q^T P^T P Q x = (PQx)^T (PQx)$

Forme quadratique définie positive (QDP)

Définition d'une forme quadratique définie positive

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ forme quadratique définie positive
- $f(x) = \frac{1}{2}x^T Cx + p^T x \quad \forall x \in \mathbb{R}^n$
- C : matrice définie positive ; p : vecteur de taille n

Forme quadratique définie positive (QDP)

Définition d'une forme quadratique définie positive

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ forme quadratique définie positive
- $f(x) = \frac{1}{2}x^T Cx + p^T x \quad \forall x \in \mathbb{R}^n$
- C : matrice définie positive ; p : vecteur de taille n

$\Longrightarrow$ forme développée de f :

$$f(x_1, \ldots, x_n) = \frac{1}{2}\left[\sum_{j=1}^{n} x_j \left[\sum_{k=1}^{n} c_{jk} x_k\right]\right] + \sum_{j=1}^{n} p_j x_j$$

$$= \frac{1}{2}\sum_{j=1}^{n} c_{jj} x_j^2 + \sum_{j=1}^{n}\sum_{k=1}^{j-1} c_{jk} x_j x_k + \sum_{j=1}^{n} p_j x_j$$

Propriétés d'une forme QDP (1/3)

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0,1[$

$$f(\lambda x + (1-\lambda)y)$$
$$= \frac{1}{2}(\lambda x + (1-\lambda)y)^T C(\lambda x + (1-\lambda)y) + p^T(\lambda x + (1-\lambda)y)$$

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0,1[$

$$
\begin{aligned}
& f(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}(\lambda x + (1-\lambda)y)^T C(\lambda x + (1-\lambda)y) + p^T(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}\left[\lambda^2 x^T Cx + (1-\lambda)^2 y^T Cy + \lambda(1-\lambda)\left[x^T Cy + y^T Cx\right]\right] \\
& \qquad + \lambda p^T x + (1-\lambda) p^T y
\end{aligned}
$$

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0, 1[$

$$
\begin{aligned}
& f(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}(\lambda x + (1-\lambda)y)^T C(\lambda x + (1-\lambda)y) + p^T(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}\left[\lambda^2 x^T Cx + (1-\lambda)^2 y^T Cy + \lambda(1-\lambda)\left[x^T Cy + y^T Cx\right]\right] \\
& \qquad + \lambda p^T x + (1-\lambda)p^T y \\
& \quad = \lambda f(x) + (1-\lambda) f(y) \\
& \qquad + \tfrac{1}{2}\lambda(1-\lambda)\left[-x^T Cx + x^T Cy + y^T Cx - y^T Cy\right]
\end{aligned}
$$

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0, 1[$

$$
\begin{aligned}
& f(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}(\lambda x + (1-\lambda)y)^T C(\lambda x + (1-\lambda)y) + p^T(\lambda x + (1-\lambda)y) \\
& \quad = \tfrac{1}{2}\left[\lambda^2 x^T Cx + (1-\lambda)^2 y^T Cy + \lambda(1-\lambda)\left[x^T Cy + y^T Cx\right]\right] \\
& \qquad + \lambda p^T x + (1-\lambda)p^T y \\
& \quad = \lambda f(x) + (1-\lambda)f(y) \\
& \qquad + \tfrac{1}{2}\lambda(1-\lambda)\left[-x^T Cx + x^T Cy + y^T Cx - y^T Cy\right] \\
& \quad = \lambda f(x) + (1-\lambda)f(y) - \tfrac{1}{2}\lambda(1-\lambda)(y-x)^T C(y-x)
\end{aligned}
$$

Propriétés d'une forme QDP (1/3)

Une forme QDP est strictement convexe

Démonstration : Soit $x, y \in \mathbb{R}^n$, $y \neq x$, et $\lambda \in]0, 1[$

$$
\begin{aligned}
f(\lambda x + (1-\lambda) y) & \\
&= \tfrac{1}{2}(\lambda x + (1-\lambda)y)^T C (\lambda x + (1-\lambda)y) + p^T(\lambda x + (1-\lambda)y) \\
&= \tfrac{1}{2}\left[\lambda^2 x^T C x + (1-\lambda)^2 y^T C y + \lambda(1-\lambda)\left[x^T C y + y^T C x\right]\right] \\
&\quad + \lambda p^T x + (1-\lambda) p^T y \\
&= \lambda f(x) + (1-\lambda) f(y) \\
&\quad + \tfrac{1}{2}\lambda(1-\lambda)\left[-x^T C x + x^T C y + y^T C x - y^T C y\right] \\
&= \lambda f(x) + (1-\lambda) f(y) - \tfrac{1}{2}\lambda(1-\lambda)(y-x)^T C (y-x)
\end{aligned}
$$

Or C définie positive $\Longrightarrow (y-x)^T C (y-x) > 0$

$\Longrightarrow f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$

Propriétés d'une forme QDP (2/3)

toute forme QDP est strictement convexe $\Longrightarrow$ cours de la semaine dernière :

Une forme QDP a un unique point stationnaire
= minimum global

Propriétés d'une forme QDP (2/3)

toute forme QDP est strictement convexe $\Longrightarrow$ cours de la semaine dernière :

> Une forme QDP a un unique point stationnaire
> = minimum global

point stationnaire $\Longrightarrow \vec{\nabla} f(x) = 0$

Propriétés d'une forme QDP (2/3)

toute forme QDP est strictement convexe $\Longrightarrow$ cours de la semaine dernière :

> Une forme QDP a un unique point stationnaire
> = minimum global

point stationnaire $\Longrightarrow \vec{\nabla} f(x) = 0$

$$\vec{\nabla} f(x) = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dots \\ \dfrac{\partial f(x)}{\partial x_j} \\ \dots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} c_{11} x_1 + \sum\limits_{k \neq 1} c_{1k} x_k + p_1 \\ \dots \\ c_{jj} x_j + \sum\limits_{k \neq j} c_{jk} x_k + p_j \\ \dots \\ c_{nn} x_n + \sum\limits_{k \neq n} c_{nk} x_k + p_n \end{bmatrix} = \begin{bmatrix} C_1 x + p_1 \\ \dots \\ C_j x + p_j \\ \dots \\ C_n x + p_n \end{bmatrix}$$

$$\vec{\nabla} f(x) = Cx + p$$

$\Longrightarrow$ f atteint son minimum en $\widehat{x}$:

unique solution du système linéaire $Cx + p = 0$:

$$f \text{ minimal en } \widehat{x} = -C^{-1}p$$

Propriétés d'une forme QDP (3/3)

$$\vec{\nabla} f(x) = Cx + p$$

$\Longrightarrow$ f atteint son minimum en $\widehat{x}$:

unique solution du système linéaire $Cx + p = 0$:

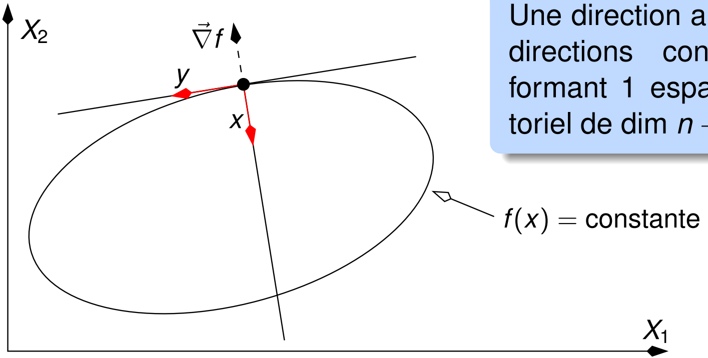
$$f \text{ minimal en } \widehat{x} = -C^{-1}p$$

méthode des directions conjuguées

appliquée à une fonction quadratique, c'est une méthode itérative de calcul de la solution $\widehat{x}$ du système $Cx + p = 0$

Directions conjuguées

- 2 vecteurs $x, y \in \mathbb{R}^n \setminus \{0\}$
- f : forme quadratique (C matrice définie positive)
- x et y définissent des directions conjuguées si $x^T C y = 0$



Une direction a 1 ∞ de directions conjuguées formant 1 espace vectoriel de dim $n-1$

Directions conjuguées (2/2)

Cas particulier de direction conjuguée :

direction $\vec{\nabla} f(x) = Cx + p$

minimum de f atteint en $\widehat{x}$ tel que $p = -C\widehat{x}$

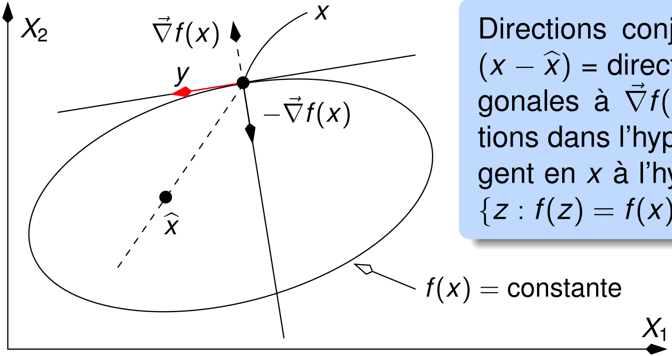
$$\Longrightarrow y^T \vec{\nabla} f(x) = y^T(Cx + p) = y^T Cx + y^T p = y^T C(x - \widehat{x})$$

Cas particulier de direction conjuguée :

direction $\vec{\nabla} f(x) = Cx + p$

⚠ minimum de f atteint en $\widehat{x}$ tel que $p = -C\widehat{x}$

$\Longrightarrow y^T \vec{\nabla} f(x) = y^T(Cx + p) = y^T Cx + y^T p = y^T C(x - \widehat{x})$



Directions conjuguées de $(x - \widehat{x})$ = directions orthogonales à $\vec{\nabla} f(x)$ = directions dans l'hyperplan tangent en x à l'hypersurface $\{z : f(z) = f(x)\}$

Changement de repère pour gradient conjugué

Lemme 1

- C une matrice définie positive
- $y^1, \ldots, y^k, \ldots, y^n$ = n vecteurs linéairement indépendants

Alors, on peut construire n vecteurs $\delta^1, ..., \delta^k, ..., \delta^n$ tels que :

- $\delta^1, \ldots, \delta^n$ linéairement indépendants
- les δ^i conjugués deux à deux
- $\delta^1 = y^1$

  $\delta^2 = y^2 - \left(\frac{{y^2}^T C \delta^1}{{\delta^1}^T C \delta^1}\right) \delta^1$

  $\ldots$

  $\delta^k = y^k - \sum_{l=1}^{k-1} \left(\frac{{y^k}^T C \delta^l}{{\delta^l}^T C \delta^l}\right) \delta^l$

  $\ldots$

$\Longrightarrow$ tout vecteur = combinaison linéaire des directions conjuguées fonction de $\vec{\nabla} f$

Lemme 2

- f = forme quadratique définie positive
  $f(x) = \frac{1}{2}x^T Cx + p^T x$
- x^0 = un point quelconque
- n vecteurs $\delta^1, \ldots, \delta^n$ linéairement indépendants et conjugués deux à deux

Alors le minimum de f est atteint en $\widehat{x} = x^0 + \sum_{k=1}^{n} \widehat{\lambda}_k \delta^k$ où :

- $\forall k$, $\widehat{\lambda}_k = -\dfrac{\vec{\nabla} f(x^0)^T \delta^k}{\delta^{k^T} C \delta^k}$
- $\widehat{\lambda}_k$ minimise $\varphi_k(\lambda_k) = f(x^0 + \lambda_k \delta^k)$
- $\widehat{\lambda}_k$ minimise $\psi_k(\lambda_k) = f(x^{k-1} + \lambda_k \delta^k)$ où $x^{k-1} = x^0 + \sum_{l=1}^{k-1} \lambda_l \delta^l$
- en fait on a aussi $\widehat{\lambda}_k = -\dfrac{\vec{\nabla} f(x^{k-1})^T \delta^k}{\delta^{k^T} C \delta^k}$

Méthode du gradient conjugué

1. *Initialisation :*
   Choix de x^0 point initial et d'une direction $\delta^1 = -\vec{\nabla} f(x^0)$
2. $k^{\text{ème}}$ *étape :*
   1. Calcul de $\widehat{\lambda}_k$ minimisant $\varphi_k(\lambda_k) = f(x^{k-1} + \lambda_k \delta^k)$:
      $$\widehat{\lambda}_k = -\frac{\vec{\nabla} f(x^{k-1})^T \delta^k}{\delta^{k^T} C \delta^k} = \frac{\vec{\nabla} f(x^{k-1})^T \vec{\nabla} f(x^{k-1})}{\delta^{k^T} C \delta^k}$$
   2. $x^k = x^{k-1} + \widehat{\lambda}_k \delta^k$
   3. calcul de $\vec{\nabla} f(x^k)$
   4. si $\vec{\nabla} f(x^k) = 0$ alors x^k est optimal fin ; sinon :
      $\delta^{k+1} = -\vec{\nabla} f(x^k) + \beta^k \delta^k$, avec :
      $$\beta^k = \frac{\vec{\nabla} f(x^k)^T C \delta^k}{\delta^{k^T} C \delta^k} = \frac{\|\vec{\nabla} f(x^k)\|^2}{\|\vec{\nabla} f(x^{k-1})\|^2}$$
3. $k \longleftarrow k + 1$

Efficacité de la méthode du gradient conjugué

proposition

La méthode du gradient conjugué pour les fonctions quadratiques atteint le minimum en n itérations.

Extension aux fonctions quelconques

Fletcher et Reeves (1964)

Adaptation l'algorithme du gradient conjugué à une fonction différentiable quelconque

Méthode du gradient conjugué pour les fonctions quelconques

1. *Initialisation :*
   Choix de x^0 point initial et d'une direction $\delta^1 = -\vec{\nabla} f(x^0)$
2. $k^{\text{ème}}$ *étape :*
   1. Calcul de $\widehat{\lambda}_k$ minimisant $\varphi_k(\lambda_k) = f(x^{k-1} + \lambda_k \delta^k)$
   2. $x^k = x^{k-1} + \widehat{\lambda}_k \delta^k$
   3. calcul de $\vec{\nabla} f(x^k)$
   4. si $\|\vec{\nabla} f(x^k)\| < \varepsilon$ alors fin ; sinon :
      $$\delta^{k+1} = -\vec{\nabla} f(x^k) + \beta^k \delta^k \text{ avec } \beta^k = \frac{\|\vec{\nabla} f(x^k)\|^2}{\|\vec{\nabla} f(x^{k-1})\|^2}$$
3. $k \longleftarrow k+1$

Convergence de l'algo de Fletcher et Reeves

Convergence de l'algorithme de Fletcher et Reeves

La convergence de l'algorithme de Fletcher et Reeves n'est assurée que si on le rapplique à partir du point 1 périodiquement en prenant comme nouveau point de départ le dernier x^k calculé, e.g., toutes les n itérations.