Les méthodes de gradient réduit

J. Frédéric Bonnans

INRIA-Saclay & CMAP, Ecole Polytechnique, France

Journées en l'honneur de Pierre Huard
JFRO, 24-25 novembre 2008, Carré des Sciences, Paris

Programmation linéaire

Format standard :

$$\operatorname*{Min}_{x \in \mathbb{R}^n} c \cdot x; \quad Ax = b \quad x \geq 0. \tag{LP}$$

Base B : (B, N) partition de $\{1, \ldots, n\}$ et A_B inversible.

Solution de base : $x_N = 0$ et $x_B = A_B^{-1} b$ (admissible si $x_B \geq 0$).

Si la valeur de (LP) est finie, il existe une solution basique.

Algorithme du simplexe

Dantzig 1947 :
Point courant x basique, $x_B > 0$ (pour chaque composante)

Estimations du multiplicateur et du gradient réduit

$$\lambda := -A_B^{-\top} c_B; \quad g := c_N + A_n^\top \lambda$$

Si $g \geq 0$, x est optimal : arrêt

Soit $i \in N$ tel que $g_i < 0$; $d_N = e_i$; $d_B = -A_B^{-1} A_i$

Si $d_B \geq 0$, $\mathrm{val}(P) = -\infty$; arrêt

Déplacement maximum dans la direction d pour annuler une variable de base k

Pivotage : Rentrée dans la base de i, sortie de k.

Méthodes de descente de descente

Cauchy 1847 : globalisation de la méthode de Newton pour un système “carré” d'équations non linéaires
Méthode de gradient sur le critère de moindres carrés associé
Recherche linéaire : pas optimal et heuristique

Curry 1944 : plus petit pas assurant la stationarité
Preuve de convergence

Hestenes et Stiefel 1952 : gradient conjugué (cas quadratique)

Davidon 1959, Fletcher et Powell 1963 : métrique variable (DFP)

Rosen 1960 : gradient avec projection

Ce qui n'est pas encore connu !

Armijo 1966 : Règle de recherche linéaire

Fletcher-Reeves 1964, Polak-Ribière 1969 : Extensions du gradient conjugué

Fletcher, Powell 1969 : Lagrangien augmenté

Wolfe 1972 Conditions de recherche linéaire

Wolfe 1974, Lemaréchal 1974 : faisceaux
(minimisation de fonctions convexes non différentiables)

Han 1976, Powell 1977 : programmation quadratique successive
pénalisation exacte, combinaison avec métrique variable

Algorithmes polynomiaux (ellipsoïdes, points intérieurs)

Algorithme initial du gradient réduit : P. Wolfe, 1962

$$\underset{x\in\mathbb{R}^n}{\mathrm{Min}}\, f(x); \quad Ax = b \quad x \geq 0. \qquad (NLP)$$

Point courant x basique, $x_B > 0$, $x_N \geq 0$.

Estimations du multiplicateur et du gradient réduit

$$\lambda := -A_B^{-\top}\nabla_B f(x); \quad g := \nabla_N f(x) + A_N^\top \lambda$$

Gradient "projeté" g_P : $g_{Pi} = g_i$ si $x \neq 0$, $g_{Pi} = (g_i)_-$ sinon.

Si $g_P = 0$, x est optimal : arrêt

$d_N = -g_P$, $d_B = -A_B^{-1} d_N$

Calcul du déplacement maximum ρ_{max} dans la direction d pour annuler une variable (de base ou non)

Choix du pas optimal $\rho \in]0, \rho_{max}]$.

Pivotage si variable de base annulée

Limitations de l'approche de P. Wolfe

Limité au gradient projeté et au pas optimal

Pas d'analyse de convergence

Pas de test numérique

Article de P. Faure et P. Huard, 1965

Premier article donnant une analyse théorique et des résultats théoriques sur la méthode du gradient réduit

Principales caractéristiques :

Analyse de convergence

Evitement du "cyclage immédiat"

Recherche linéaire "1/4, 1/2, 3/4"

Sous relaxation, diagonalisation des directions

Conjugaison des directions

Implémentation en Fortran IV (1962) sur IBM 7094

Problèmes d'investissement : 209 variables, 61 contraintes

Analyse de convergence de P. Faure et P. Huard, 1965

Hypothèses

Critère convexe, minoré, "C^2 borné".

Pivots uniforméments non nuls

Taille des pas de recherche linéaire uniforméments bornés

Résultat : alternatives suivantes

1) Arrêt en une solution

2) Sous suite convergeant vers la solution, de gradient réduit projeté tendant vers 0

3) Gradient réduit projeté ne tendant pas vers 0 : convergence vers un point $\bar{x}$, et $\rho_k \to 0$.

Détails techniques : petits pas

$$x_{k+1} = x_k + \rho_k d_k$$

$$\|d_k\| = 1; \quad g^k \cdot d_N^k \leq -c\|g_P^k\|;$$

$$\begin{aligned} f(x_k + \rho d_k) &= f(x_k) + \rho g^k \cdot d_N^k + o(\rho) \\ &\leq f(x_k) - c\rho\|g_P^k\| + o(\rho) \end{aligned}$$

On choisit ρ_k assez petit pour que

$$f(x_{k+1}) \leq f(x_k) - c'\rho_k\|g_P^k\|$$

Donc $\sum_k \rho_k\|g_P^k\| < \infty$: Le long d'une sous suite $\mathcal{K}$ telle que $\liminf_{k\in\mathcal{K}} \|g_P^k\| > 0$, on a

$$\sum_{k\in\mathcal{K}} \rho_k < \infty$$

Détails techniques : petits pas, mais pas trop petits

Prendre le pas assez grand : si $\rho_k < \rho_{max,k}$, alors (Wolfe 1972)

$$\nabla f(x_{k+1})d_k \geq \frac{1}{2}\nabla f(x_k)d_k$$

Au voisinage de $\hat{x}$ tel que $g(\hat{x})_- \neq 0$, implique

$$\rho_k > \rho(\hat{x}) \quad \text{ou } \rho_k = \rho_{max,k}$$

Points d'adhérence non stationnaires

Propriété des **paquets d'itérations locales** :

Si $\hat{x}$ point d'adhérence non stationnaire

Sous suite associée k_1, k_2, ... telle que $x_{k_i} \longrightarrow \hat{x}$

Alors pour tout $\ell > 0$: $x_{k_i+\ell} \longrightarrow \hat{x}$

On appelle **paquets d'itérations locales** des itérations successives dans un petit voisinage de $\hat{x}$.

Zigzagging possible seulement pour les composantes du gradient réduit nulles au point d'adhérence.

Cyclage des variables de base ?

Règle de Markowitz d'entrée en base (1957)

Choisir la variable hors base la plus positive sous contrainte de pivot (supérieur à une fraction du pivot maximum)

Hyp. de non dégérescence : Base non dégénérée associée à $\hat{x}$

Alors (sauf dégérescence au cours des itérations) le paquet d'itérations locales finit par trouver une base non dégénérée associée à $\hat{x}$

La base du paquet d'itérations change un nombre fini de fois.

Ceci résout la question du cyclage des variables de base

Cyclage des variables hors base

Contre exemple de P. Wolfe (1966-1972)

$$\operatorname*{Max}_{x,y,z\geq 0} f(x,y,z) = -\frac{4}{3}(x^2 - xy + y^2)^{3/4} + z$$

Zigzagging (Zoutendijk 1960)

Pour le pas optimal, x_k et y_k s'annulent à tour de rôle, empêchant la convergence.

Anti cyclage : P. Huard, 1982

Direction obtenue par projection du déplacement dans la direction du gradient réduit projeté.

Pour un certain $\beta > 0$, on projette le déplacement non basique :

$$d_N^k = -\frac{(x_N^k + \beta g_P^k)_+}{\beta} \quad d_B^k = -A_B^{-1} d_N^k$$

On a la condition de descente uniforme : pour un certain $\alpha > 0$

$$g^k \cdot d_N^k \leq -\alpha \|g_P^k\| \|d_N^k\|$$

$\rho_{max,k} \geq \beta$: **tout point d'adhérence est stationnaire**

De plus si $\rho_{max,k} = \beta$: on sature un paquet de contraintes simultanément : réduction du nombre d'itérations

Métrique variable I

Refs Bertsekas (1982), Bonnans (1983)

Réduction au cas de seules contraintes : $x \geq 0$

Séparation des variables grandes et petites; $\varepsilon > 0$:

$$I^k := \{1 \leq i \leq n;\ x_i \leq \varepsilon\}; \quad J^k := \{1, \ldots, n\} \setminus I^k.$$

Direction de descente telle que

$$d_i^k = -g_{P,i}^k,\ i \in I^k;$$

Recherche linéaire par projection sur $\mathbb{R}_+^n$
(pas maximum assurant la positivité de la base)
Possibilité de (dé)saturer simultanément plusieurs contraintes

Métrique variable II

Forme de métrique variable :

$$d^k = -(M^k)^{-1} g_P^k;$$

M^k bloc diagonale pour (I^k, J^k), bloc identité sur I^k
Bloc sur J^k : approximation du hessien $\nabla^2 f(x^k)$

Amélioration possible : $d^k_{J^k}$ solution de

$$\nabla^2_{J^k I^k} f(x^k) d^k_{I^k} + \nabla^2_{J^k J^k} f(x^k) d^k_{J^k} + g_{J^k} = 0$$

Anti cyclage par blocage des variables

Blanchon, Bonnans, Dodu (1991), Bonnans (2006)

Paquets d'itérations à l'intérieur desquelles on ne libère aucune variable

Fin du paquet si optimalité approchée pour les variables non bloquées

References

1. P. Faure et P. Huard, *Résolution de programmes mathématiques à fonction non linéaire par la méthode du gradient réduit*, Rev. Franç. de Rech. Opér., 36(1965).
2. P. Huard, *Convergence of the reduced gradient method*. In Nonlinear programming 2, Academic Press, 1974.
3. P. Huard *Un algorithme général de gradient réduit*. EDF Bull. Direction Études Rech. Sér. C Math. Inform, 1982.
4. J.F. Bonnans, *Optimisation Continue*, Dunod, 2006.