Méthodes mises à l'échelle pour la reconstruction tomographique en coordonnées cylindriques

Guillaume Mestdagh

Département de génie électrique Polytechnique Montréal

Direction : Yves Goussard et Dominique Orban

29 août 2019

Tomographie à rayons X

Tomographe médical

Objectif

Déterminer la disposition intérieure d'un patient à partir de mesures de transmission

Axes de recherche actuels

- Diminution des doses de rayons X
- Réduction du bruit sur les images
- Perfectionnement des méthodes statistiques

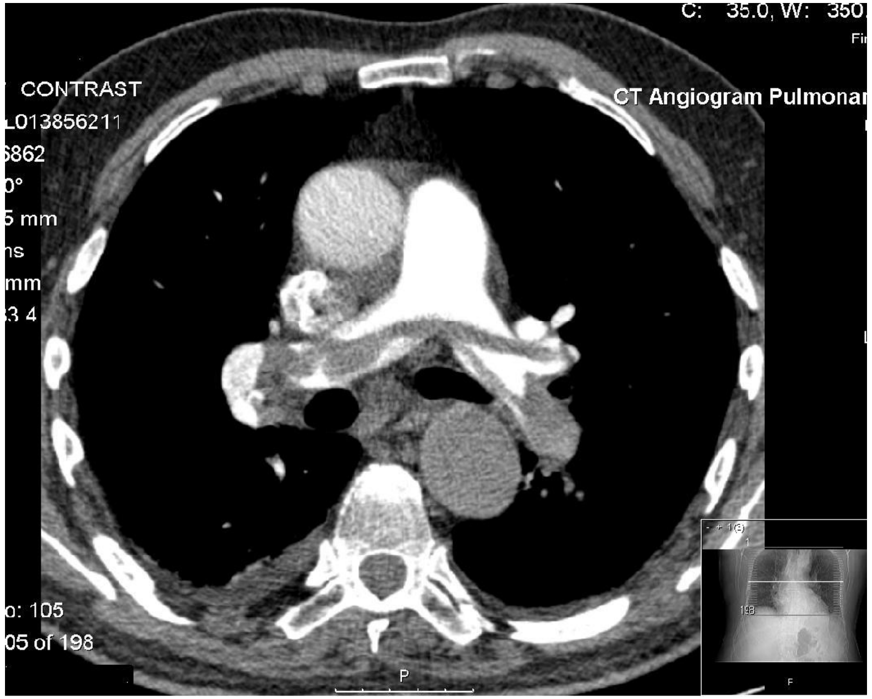


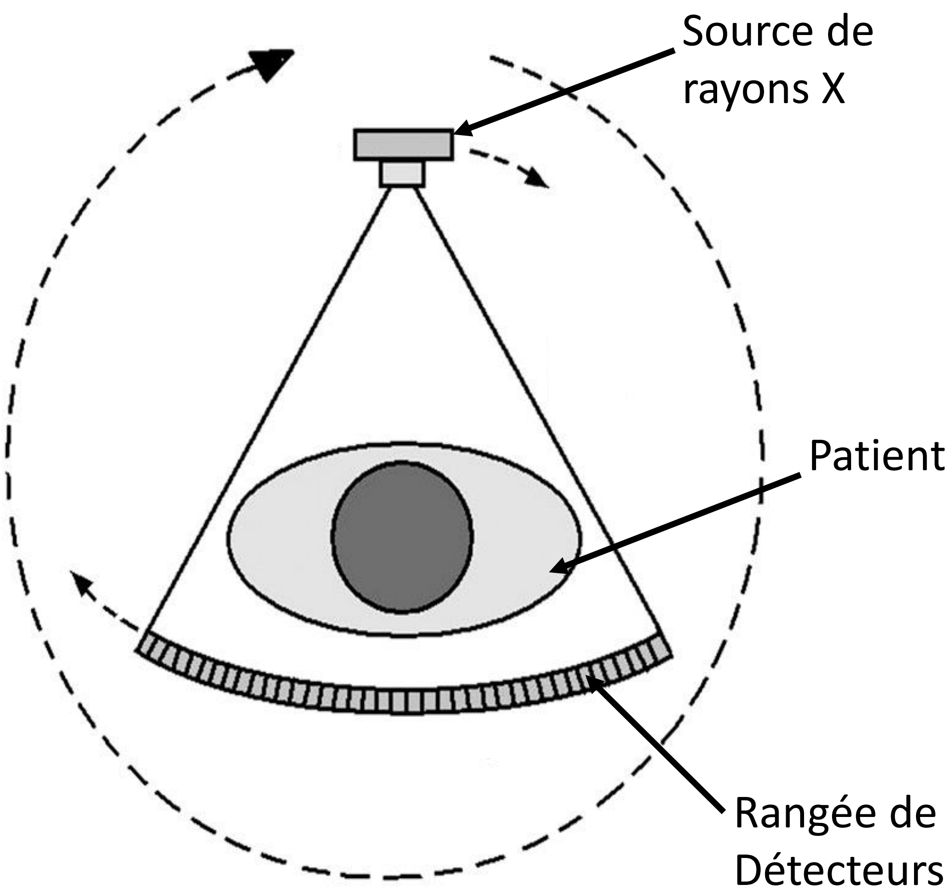
Image tomographique (https://commons.wikimedia.org/wiki/File:SADDLE_PE.JPG)

Introduction

Sommaire

- Problème de reconstruction d’image
- Algorithmes d’optimisation mis à l’échelle
- Résultats numériques

Reconstruction tomographique



Géométrie du tomographe (Goldman, 2007)

But

Reconstruire une image à partir de mesures sous différents angles

Données disponibles

Atténuation des rayons X pour chaque angle et pour chaque détecteur

Image reconstruite

Coefficient d’atténuation en chaque point du domaine.

Formulation du problème

Modèle direct

Pixels de l'image (coefficients d'atténuation)

$$b = Ax + \mathrm{bruit}$$

Données mesurées

Matrice du système

Problème inverse (Bouman & Sauer, 1993)

Terme de moindres carrés

$$\min_{x \geq 0} \frac{1}{2} \| Ax - b \|_V^2 + R(x)$$

Contraintes de bornes

Matrice de pondération diagonale

Pénalisation sur les différences entre pixels voisins

Caractéristiques du problème

- Grande quantité de données et d'inconnues
- Empreinte mémoire importante
- Contraintes de bornes

Principaux algorithmes

Algorithmes couramment utilisés en imagerie

Méthodes de gradient projeté

- Gradient projeté accéléré
  (Nesterov, 1983)
- Gradient projeté spectral
  (Birgin et Martinez, 2002)

Méthodes avec séparation du problème

- Lagrangien augmenté
  (Hestenes, 1969)
- Algorithme primal-dual
  (Chambolle et Pock, 2010)

Méthodes spécifiques

- Sous-ensembles ordonnés
  (Erdoğan et Fessler, 1999)
- Descente par coordonnée
  (Sauer et Bouman, 1993)

Algorithmes peu utilisés en imagerie

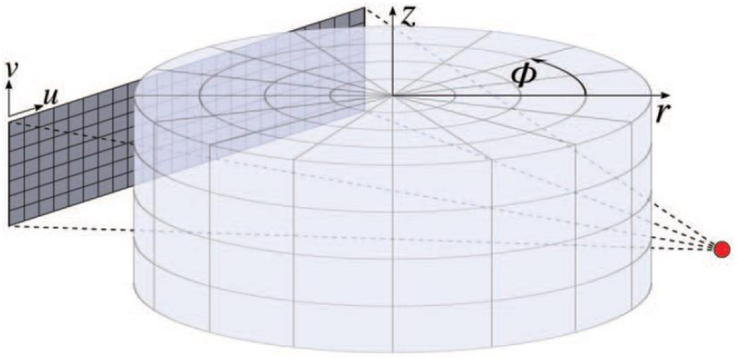
Méthodes de directions projetées

- Méthode de Newton projetée avec recherche linéaire
  (Bertsekas, 1982)
- Méthode de Newton avec région de confiance
  (Lin et Moré, 1999)
- Méthode de L-BFGS projetée
  (Kim, Sra et Dhillon, 2010)
- L-BFGS-B
  (Byrd, Lu, Nocedal et Zhu, 1995)

Méthodes de points intérieurs

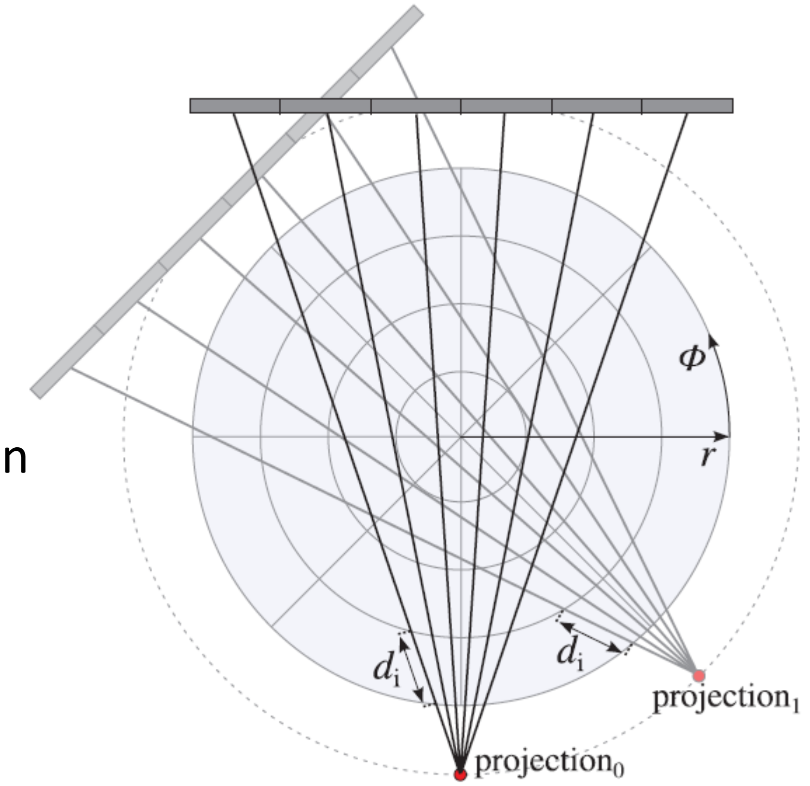
- Ipopt
  (Wächter et Biegler, 2005)
- PDCO
  (Chen, Donoho et Saunders, 2001)

Coordonnées cylindriques



Discrétisation en coordonnées cylindriques
(Thibaudeau, Leroux, Fontaine, Lecomte, 2013)

Discrétisation invariante par rotation d’un angle $\delta\phi$



Avantage

Invariance par rotation
→ Matrice du système structurée

Inconvénient

Pixels de tailles différentes
→ Problème mal conditionné

Mise à l'échelle du problème (Golkar, 2013)

Matrice du système circulante par blocs

$$A^{\mathrm{T}}A = \underbrace{F^*}_{\text{Transformée de Fourier discrète par blocs}} \begin{pmatrix} \hat{A}_1 & & \\ & \ddots & \\ & & \underbrace{\hat{A}_{n_b}}_{\text{Bloc dense}} \end{pmatrix} F$$

Matrice de mise à l'échelle circulante par blocs

$$C = F^* \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & \underbrace{D_{n_b}}_{\text{Bloc diagonal}} \end{pmatrix} F$$

Tel que $C^{\mathrm{T}}\left(A^{\mathrm{T}}VA + \nabla^2 R(x)\right)C$ est bien conditionnée

Taille de la matrice du système (Thibaudeau, 2013)

Pour une tranche de 512 × 512 pixels

Coordonnées cartésiennes	2765 MB
Coordonnées cylindriques	5.5 MB

Un produit par C coûte $\mathrm{O}(n \log n)$ opérations
(n est la taille de x)

Objectifs

- Exploiter la mise à l'échelle pour accélérer la résolution du problème
- Adopter une stratégie simple qui s'adapte à plusieurs algorithmes
- Limiter les coûts supplémentaires liés aux produits par C

Précédemment dans le projet (McLaughlin, 2017)

Changement de variable

$$x = Cu$$

Variable dans l'espace d'origine

Variable dans l'espace transformé

Problème dans l'espace transformé

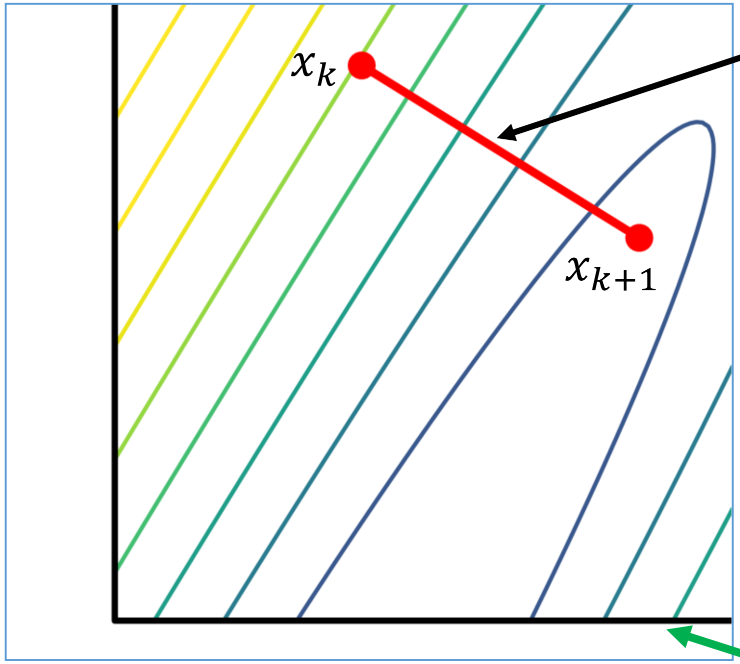
$$\min_{Cu \geq 0} \frac{1}{2} \| ACu - b \|^2 + R(Cu)$$

Contraintes affines

Caractéristiques

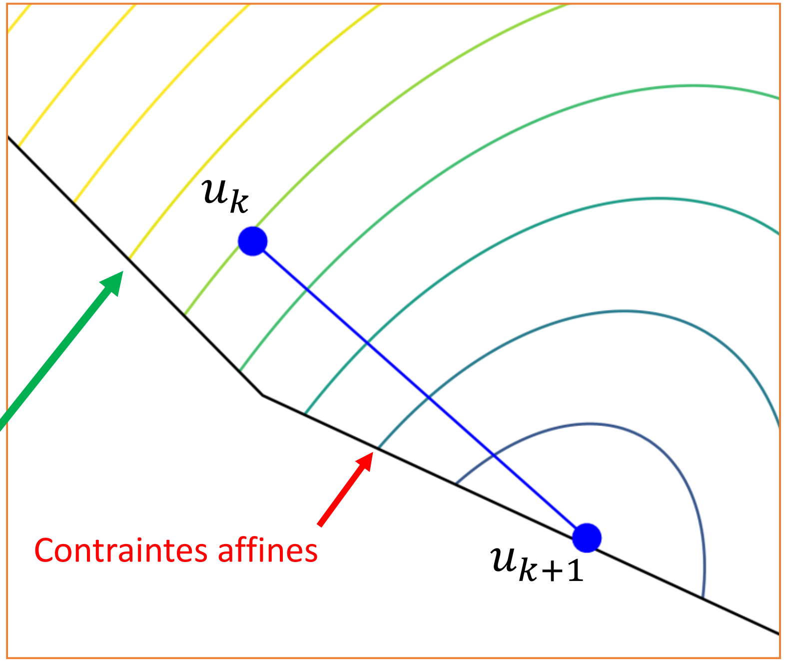
- Meilleur conditionnement
- Projections orthogonales calculées par une procédure itérative

Précédemment dans le projet (McLaughlin, 2017)



Pas dans la direction du gradient

$x = Cu$



Frontière de l’ensemble réalisable

$$\min_{x\geq 0} \frac{1}{2}\|Ax - b\|^2 + R(x)$$

$$\min_{\boxed{Cu}\geq 0} \frac{1}{2}\|ACu - b\|^2 + R(Cu)$$

Espace d’origine
- ☺ Projections peu coûteuses
- ☹ Mauvais conditionnement

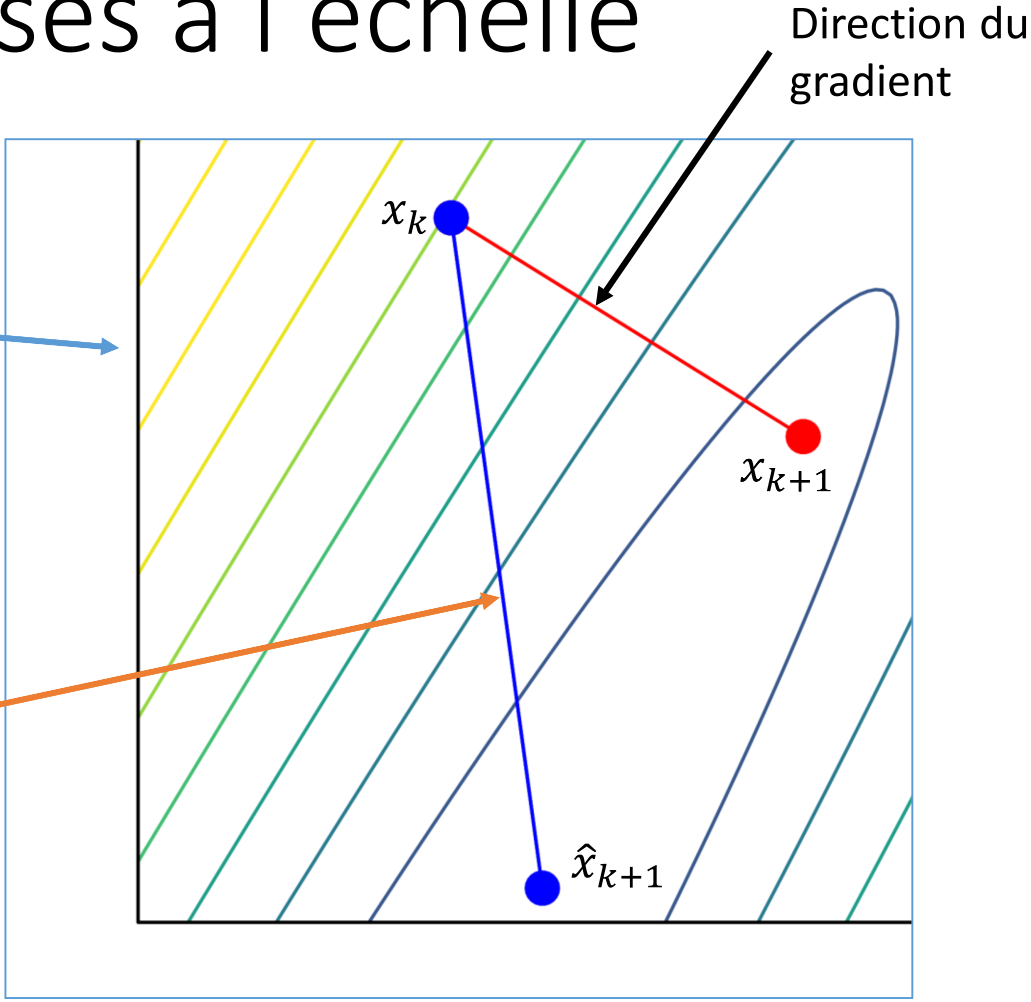
Espace transformé
- ☹ Projections coûteuses et inexactes
- ☺ Bon conditionnement

Directions de descente mises à l’échelle

Problème d’origine
avec contraintes de bornes
(pas de changement de variable)

Directions de descente
issues de l’espace transformé

$$d = -CC^T \nabla f(x)$$



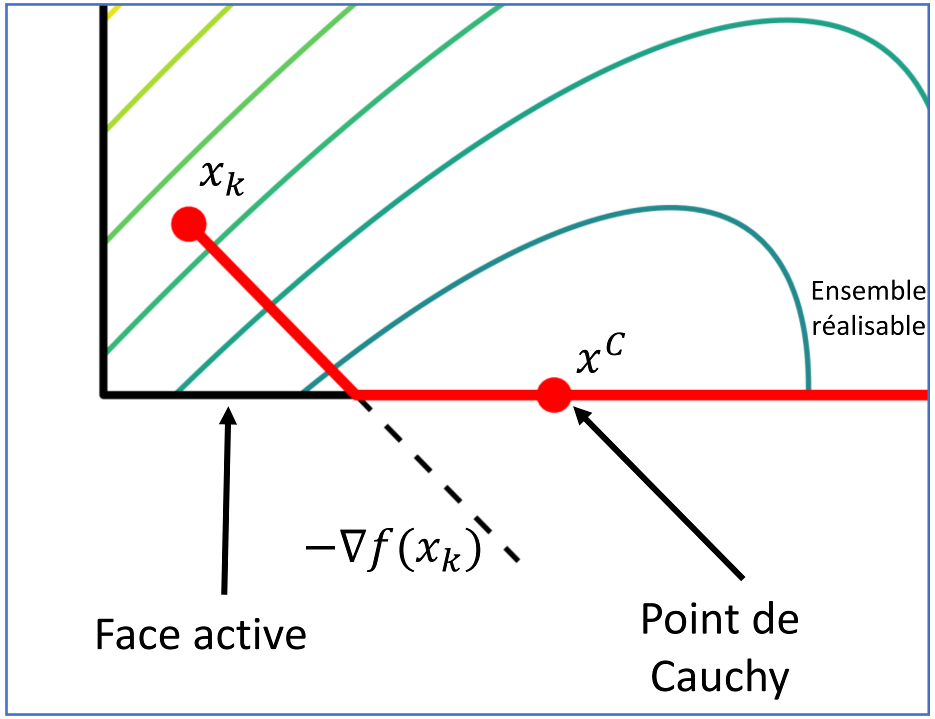
→ Quels éléments de l’espace transformé peut-on utiliser tout en restant compatible avec les contraintes de bornes ?

Méthodes de directions projetées

But
Deviner sur quelle face de l’ensemble réalisable se trouve la solution du problème.

Principales étapes

- Identifier une face active F
- Résoudre un sous-problème quadratique sans contraintes dans la face active F
- Identifier le nouvel itéré



Gradient projeté

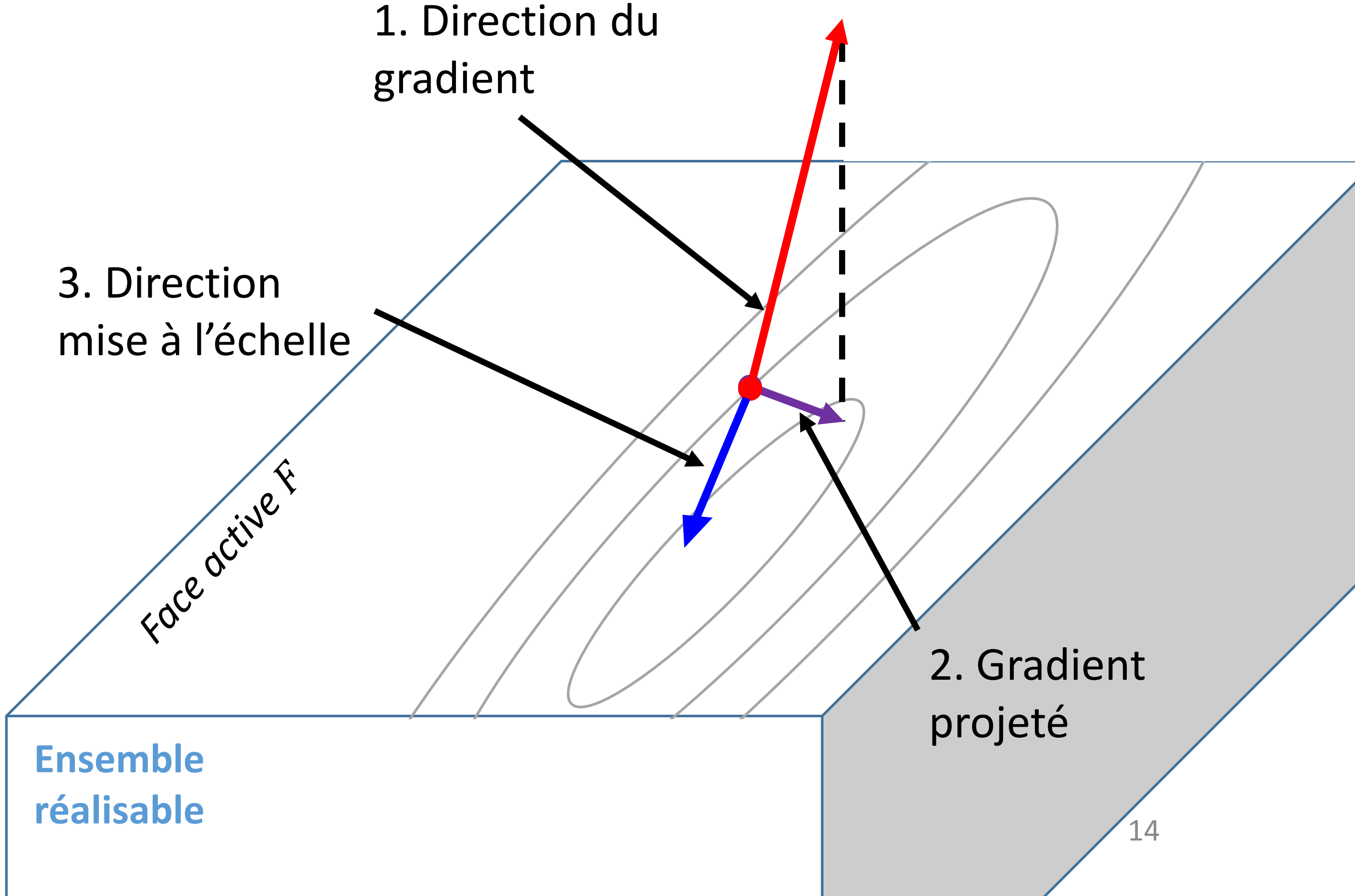
Définie positive

$$d_{\text{proj}} = \begin{pmatrix} d_F \\ d_{F^\perp} \end{pmatrix} = -\begin{pmatrix} [CC^{\mathrm{T}}]_{FF} & 0 \\ 0 & 0 \end{pmatrix} \nabla f(x)$$

Problème
Comment s’assurer que la direction mise à l’échelle reste une direction de descente quand on la projette ?

Solution
Appliquer la mise à l’échelle à l’intérieur de la face active (Bertsekas, 1982)

→ On obtient une direction de descente



Directions de gradient conjugué

Sous-problème quadratique
restreint à la face active F

$$\min_{x \in F} \ \frac{1}{2} x^{\mathrm{T}} B x + x^{\mathrm{T}} g$$

Gradient conjugué (Hestenes, 1952)
Génération de directions conjuguées par rapport à la matrice B

$$p_k^{\mathrm{T}} \, [B_{FF}] \, p_{k+1} = 0$$

Variables réduites de dimension $\dim F$

Sans mise à l'échelle

Résidu réduit

$$p_{k+1} = r_{k+1} + \beta_k p_k$$

Avec β_k tel que $p_k^{\mathrm{T}} [B_{FF}] p_{k+1} = 0$

Avec mise à l'échelle

$$p_{k+1} = [CC^{\mathrm{T}}]_{FF} r_{k+1} + \beta_k p_k$$

Avec β_k tel que $p_k^{\mathrm{T}} [B_{FF}] p_{k+1} = 0$

Équivalent à un préconditionnement à l'aide de $[CC^{\mathrm{T}}]_{FF}$

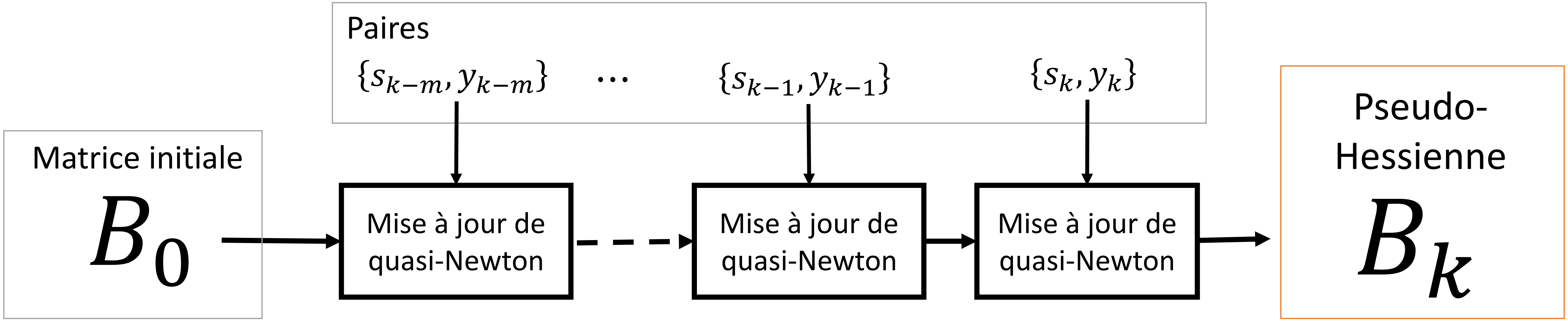
Matrice de quasi-Newton à mémoire limitée

Matrice de quasi-Newton

- On part d'une matrice initiale B_0
- On met à jour la matrice à l'aide des paires $\{s_k, y_k\}$ enregistrées en mémoire

Notations

$$\begin{cases} s_k &= x_{k+1} - x_k \\ y_k &= \nabla f(x_{k+1}) - \nabla f(x_k) \end{cases}$$



Matrice de quasi-Newton à mémoire limitée

Choix de la matrice initiale

Espace transformé

$$\bar{B}_0 = \bar{\theta} I$$

$$\bar{\theta} = \frac{\bar{y}_k{}^{\mathrm{T}} \bar{y}_k}{\bar{s}_k{}^{\mathrm{T}} \bar{y}_k}$$

Espace d’origine

$$B_0 = \theta \left(C C^{\mathrm{T}} \right)^{-1}$$

$$\theta = \frac{y_k^{\mathrm{T}} C C^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$$

Essais sur un problème de reconstruction 2D

Problème de reconstruction

$$\min_{x \geq 0} \quad \frac{1}{2} \|Ax - b\|_V^2 + \lambda \phi(x)$$

où $\phi(x) = \sum_i \sqrt{\delta^2 + [Kx]_i^2}$ et $\delta > 0$ (pénalisation $L_2 L_1$)

Matrice de différences finies

Taille du résidu
779 520 valeurs

Taille de l'inconnue
1 tranche de 262 160 pixels

Paramètres de régularisation
$\lambda = 10^{-4}, \delta = 10^{-1}$

Comparaison de 3 approches

- Pas de mise à l'échelle
- Changement de variable (problème avec contraintes affines)
- Directions mises à l'échelle (approche utilisée dans cette maîtrise)

3 algorithmes utilisés

- L-BFGS-B (Quasi-Newton)
- TRON (Newton à région de confiance)
- SPG (variante du gradient projeté)

L'algorithme L-BFGS-B

Caractéristiques

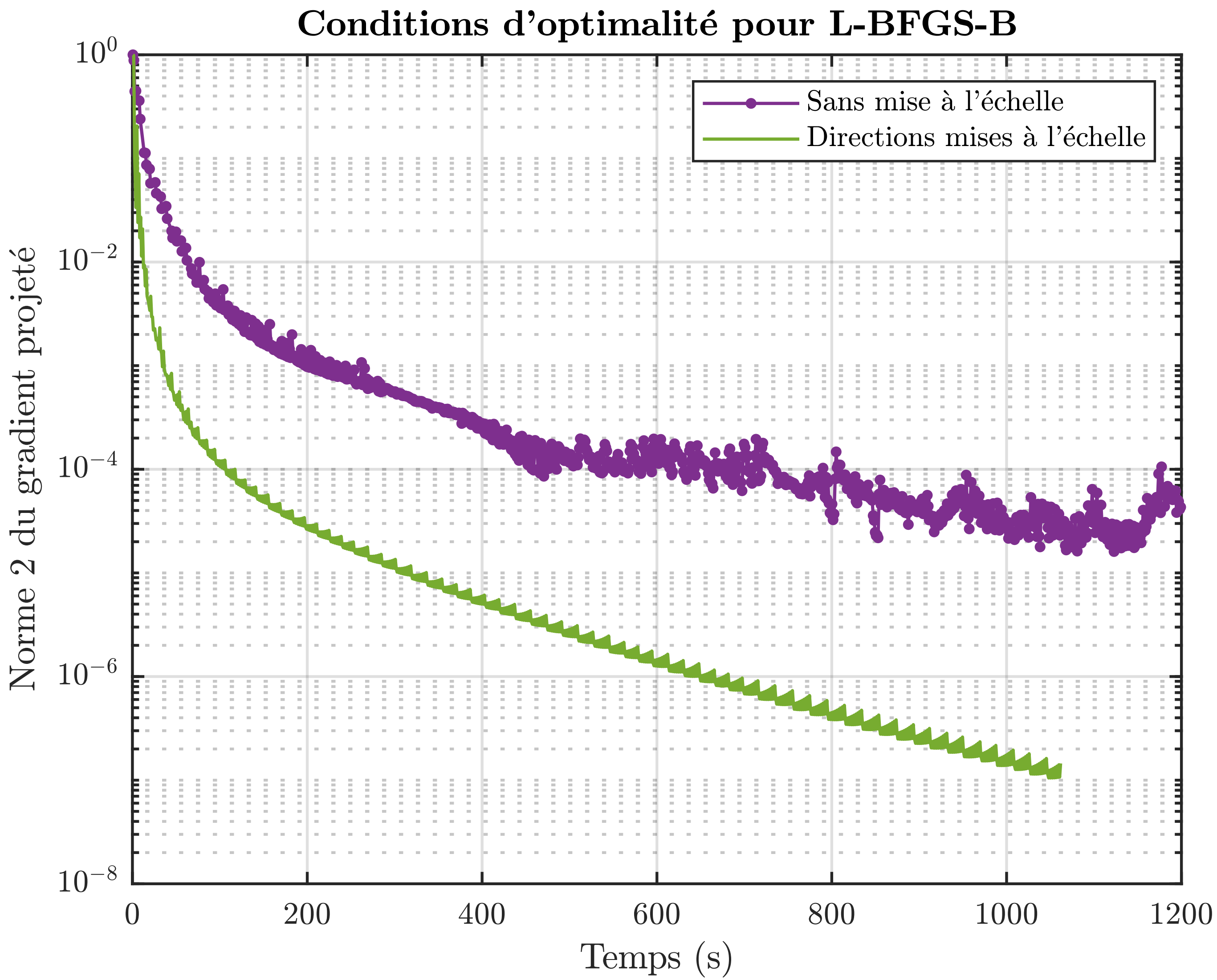
- Méthode de quasi-Newton
- Utilise la formule L-BFGS compacte
- Point de Cauchy inexact
- Résolution du sous-problème par gradient conjugué
- Écrit en Matlab (github.com:optimizers/NLPLab)

L-BFGS pour problèmes à contraintes de bornes
(Byrd, Lu, Nocedal et Zhu, 1995)

Pour chaque itération

- Calculer un point de Cauchy le long du chemin projeté
$$x(t) = \mathrm{Proj}\left(x_k - t \begin{pmatrix} [CC^{\mathrm{T}}]_{FF} & 0 \\ 0 & 0 \end{pmatrix} \nabla f(x) \right)$$
avec $t > 0$, et identifier la face active F
- Résoudre le sous-problème quadratique dans la face active F par gradient conjugué préconditionné
- Trouver l'itéré suivant x_{k+1} par une recherche linéaire de Wolfe
- Mettre à jour les données de l'opérateur L-BFGS

L'algorithme L-BFGS-B



Causes possibles des piètres performances

- Gestion des contraintes (recherche linéaire, etc.)
- Approximation de la Hessienne
- Opération L-BFGS difficiles à vectoriser

CPU: Intel® Xeon® CPU E5-2637 v4 @ 3.50GHz, 32GB RAM

L'algorithme TRON

Principales caractéristiques

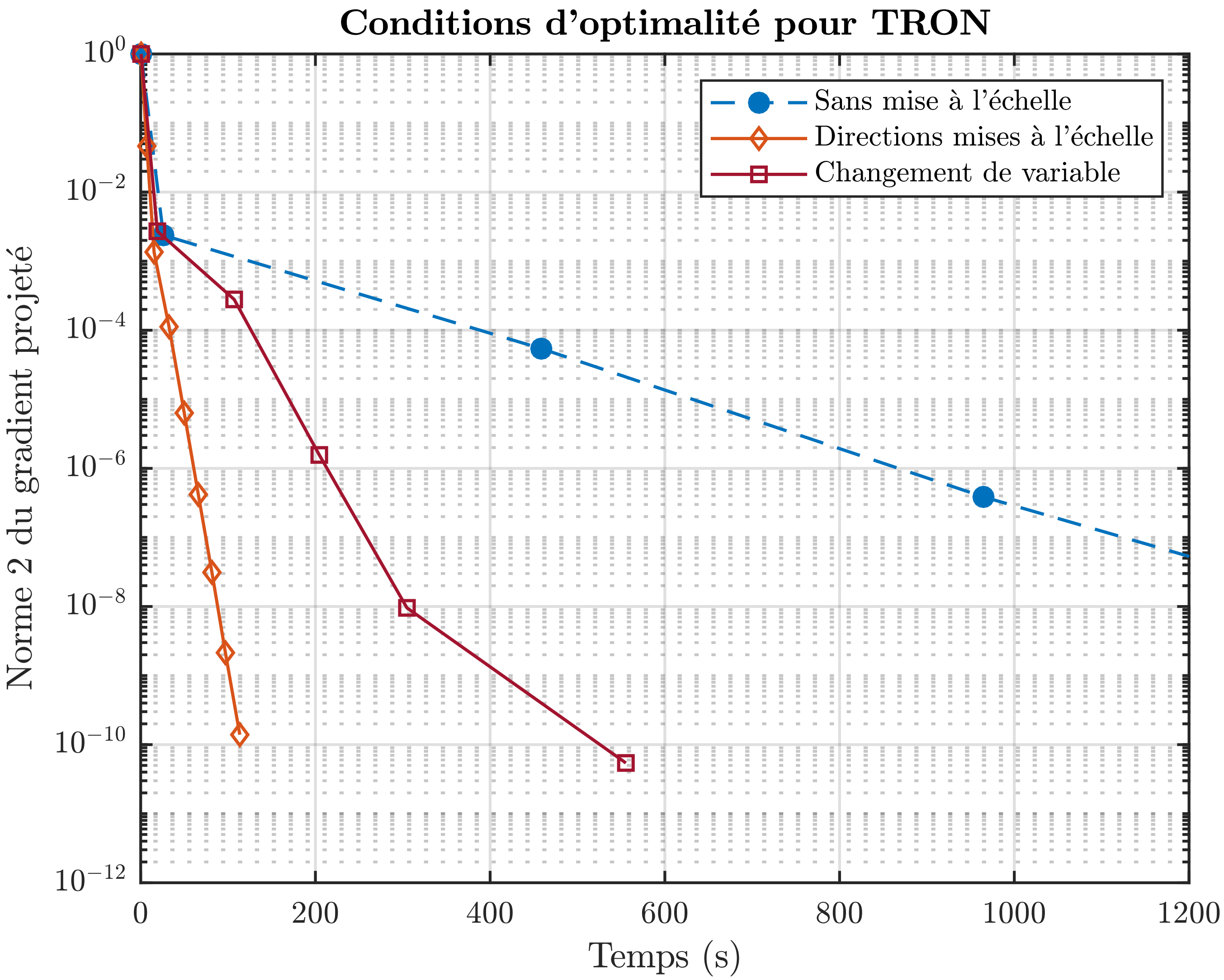
- Méthode de Newton
- Résolution du sous-problème par gradient conjugué
- Utilisation de recherches projetées
- Écrit en Matlab par Orban et Friedlander (github.com:optimizers/NLPLab)
- Variante pour contraintes affines écrite par McLaughlin (2017)

Méthode de Newton à région de confiance
(Lin et Moré, 1999)

Pour chaque itération

- Calculer un point de Cauchy le long du chemin projeté
$$x(t) = \mathrm{Proj}\left(x_k - t \begin{pmatrix} [CC^{\mathrm{T}}]_{FF} & 0 \\ 0 & 0 \end{pmatrix} \nabla f(x) \right)$$
avec $t > 0$, et identifier la face active F
- Résoudre le sous-problème quadratique dans la face active F par gradient conjugué préconditionné
- Accepter ou refuser la solution du sous-problème comme itéré suivant
- Mettre à jour le rayon de la région de confiance

L’algorithme TRON



Part du temps de calcul associée aux produits hessienne-vecteur

Sans mise à l’échelle	98 %
Changement de variable	61 %
Directions mises à l’échelle	98 %

→ Gain de temps sur les projections

CPU: Intel® Xeon® CPU E5-2637 v4 @ 3.50GHz, 32GB RAM

L'algorithme SPG

Principales caractéristiques

- Méthode de premier ordre
- Longueur de pas issue du gradient spectral de Barzilaï et Borwein (1988)
- Recherche linéaire
- Écrit en Matlab (github.com:optimizers/NLPLab)
- Variante pour contraintes affines écrite par McLaughlin (2017)

Méthode du gradient spectral projeté
(Birgin et Martinez, 2002)

Pour chaque itération

- Calculer la longueur de pas

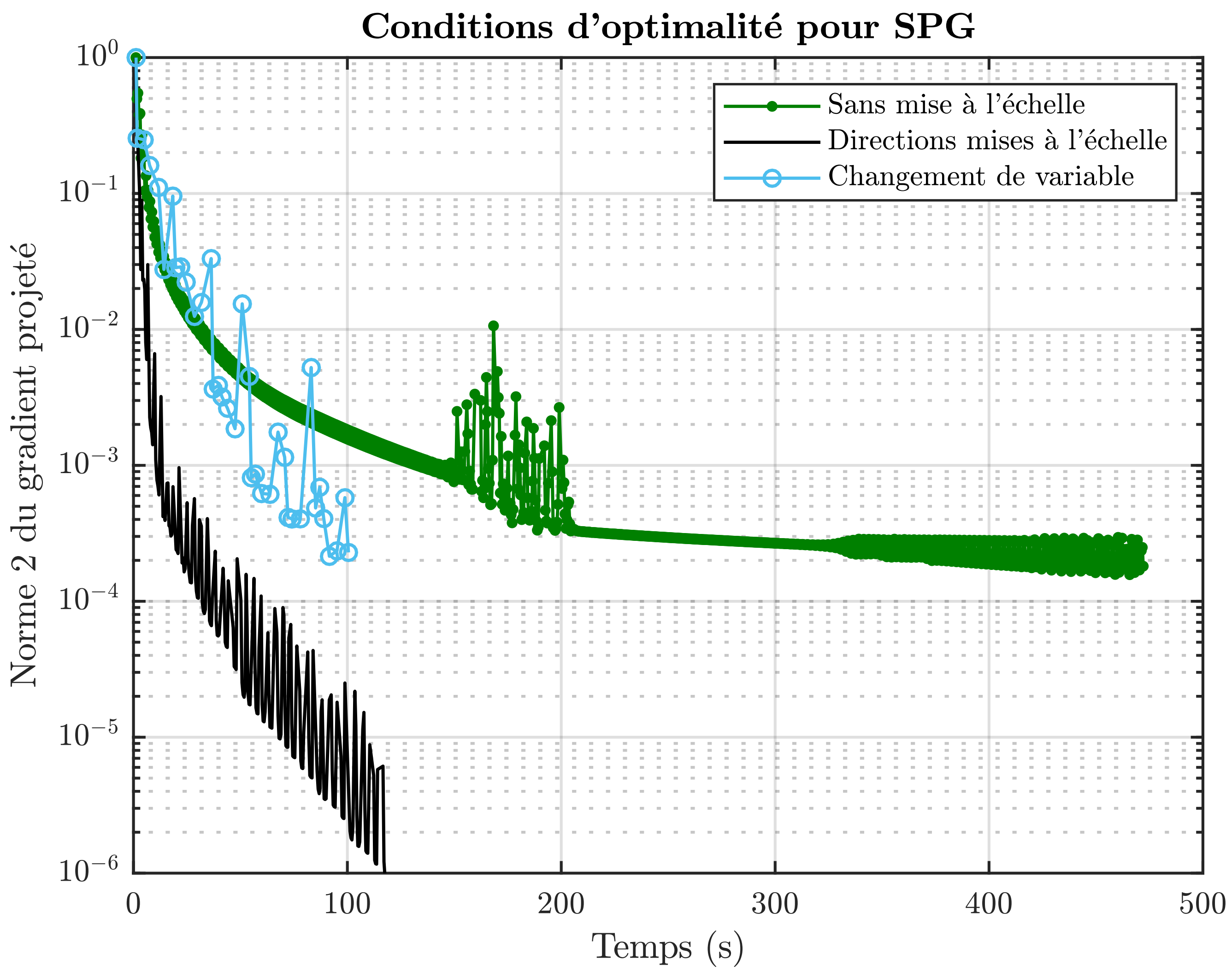
$$\theta_k = \frac{y_k^{\mathrm{T}} C C^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$$

- Calculer un pas de gradient projeté

$$\tilde{x}_k = \mathrm{Proj}\left(x_k - \theta_k \begin{pmatrix} [CC^{\mathrm{T}}]_{FF} & 0 \\ 0 & 0 \end{pmatrix} \nabla f(x) \right)$$

- Calculer $x_{k+1} = x_k + \lambda_k (\tilde{x}_k - x_k)$ par recherche linéaire

L’algorithme SPG



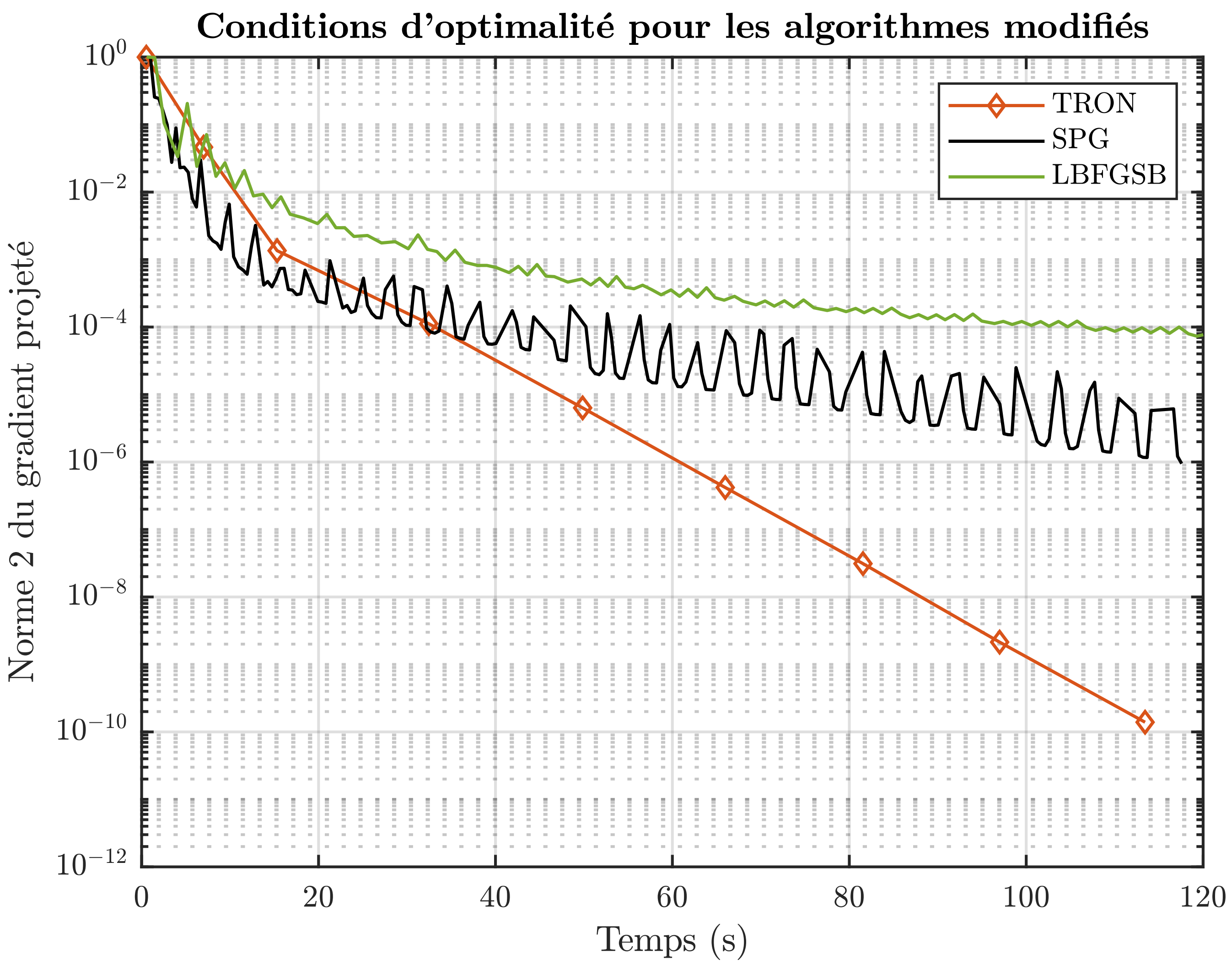
Part du temps de calcul associée aux produits hessienne-vecteur

Sans mise à l’échelle	99 %
Changement de variable	18 %
Directions mises à l’échelle	98 %

→ Gain de temps sur les projections

CPU: Intel® Xeon® CPU E5-2637 v4 @ 3.50GHz, 32GB RAM

Comparaison des trois algorithmes mis à l'échelle



Meilleurs méthodes	
Tolérance large	SPG
Tolérance stricte	TRON

CPU: Intel® Xeon® CPU E5-2637 v4 @ 3.50GHz, 32GB RAM

Conclusion

- Développement d'une stratégie de mise à l'échelle des algorithme de directions projetées.
- Mise en œuvre de la stratégie sur L-BFGS-B, TRON et SPG
(Production d'un L-BFGS-B avec gradient conjugué en matlab).
- Résultats numériques prouvant la supériorité de la stratégie par rapport au changement de variable.
- Faisabilité de la reconstruction en coordonnées cylindriques prouvée
- Stratégie adaptable à n'importe quel algorithme pour problème à contraintes de bornes.

Améliorations futures

Appliquer la mise à l'échelle pour résoudre des problèmes spécifiques en tomographie à rayons X.

Références

- Bouman, C., & Sauer, K. (1993). A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on image processing*, *2*(3), 296-310.
- Thibaudeau, C., Leroux, J. D., Fontaine, R., & Lecomte, R. (2013). Fully 3D iterative CT reconstruction using polar coordinates. *Medical physics*, *40*(11).
- Lin, C. J., & Moré, J. J. (1999). Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, *9*(4), 1100-1127.
- Birgin, E. G., & Martínez, J. M. (2002). Large-scale active-set box-constrained optimization method with spectral projected gradients. *Computational Optimization and Applications*, *23*(1), 101-125.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190-1208.
- Bertsekas, D. P. (1982). Projected Newton methods for optimization problems with simple constraints. *SIAM Journal on control and Optimization*, *20*(2), 221-246.
- McLaughlin, M. (2017). *Méthodes sans factorisation pour la tomographie à rayons-X en coordonnées cylindriques*(Master's thesis, Ecole Polytechnique, Montreal (Canada)).

Merci de votre attention !