A Majorize-Minimize line search algorithm for barrier function optimization

Emilie Chouzenoux
emilie.chouzenoux@irccyn.ec-nantes.fr

Institut de Recherche en Communication et Cybernétique de Nantes (FRANCE)

EUSIPCO, Glasgow, August 27th 2009

Barrier functions

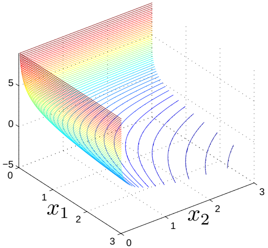
Definition

B is said *barrier function* associated to the constraint $\boldsymbol{x} \in \mathcal{C}$ if $B(\boldsymbol{x})$ is unbounded at the boundary of $\mathcal{C}$

E.g.: Logarithmic barrier function for positivity contraints

$$B(\boldsymbol{x}) = -\sum_n \log(x_n)$$

$$\mathcal{C} = \{\boldsymbol{x} > 0\}$$



$$B(x) = -\log(x_1) - \log(x_2)$$

$\Rightarrow$ If a criterion contains a barrier function, its minimizers belong to $\mathcal{C}$.

Criteria involving barrier functions

① *Interior point methods:*

$$\min_{c_i(\boldsymbol{x})>0} P(\boldsymbol{x}) \Leftrightarrow \min P(\boldsymbol{x}) - \mu \sum_i \log(c_i(\boldsymbol{x})), \{\mu\} \to 0$$

② *Emission tomography:*

$$\hat{\boldsymbol{x}} = \arg\min \sum_i [\boldsymbol{H}\boldsymbol{x}]_i - y_i \log[\boldsymbol{H}\boldsymbol{x}]_i + \lambda R(\boldsymbol{x}) \Rightarrow \boldsymbol{H}\hat{\boldsymbol{x}} > 0$$

③ *Maximum entropy:*

$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda \sum_n x_n \log x_n \Rightarrow \hat{\boldsymbol{x}} > 0$$

General formulation

Objective function

$$\text{minimize } \left(F(\boldsymbol{x}) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x})\right), \quad \mu > 0 \tag{1}$$

- $B(\boldsymbol{x}) = \sum_{i=1}^{I} b_i(\boldsymbol{a}_i^T \boldsymbol{x} + \rho_i)$: barrier function,
  e.g., $b_i(u) = -\log u$ or $u \log u$

Algorithmic scheme

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad \text{for} \quad k = 1, \dots, K$$

- $\boldsymbol{d}_k$ *descent direction* i.e., $\boldsymbol{d}_k^T \nabla F(\boldsymbol{x}_k) < 0$: (P)CG, L-BFGS, ...
- $\alpha_k > 0$ *stepsize* given by a **1D nonquadratic MM algorithm**

Outline

Line search strategies

Goal: Given $\boldsymbol{x}_k$ and $\boldsymbol{d}_k$, find α_k that ensures the convergence of the algorithm

- Classical strategies: Dichotomy, backtracking, cubic interpolation, quadratic approximation...
- Iterative minimization of $F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$
- Identifying α_k that fulfills some convergence conditions

e.g.: Wolfe conditions

Problem statement

$$f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) = P(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) + \mu \underbrace{B(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)}_{\text{barrier term } b(\alpha)}$$

b undefined for $\alpha > \bar{\alpha}$ if there exists i such that $c_i(\boldsymbol{x}_k + \bar{\alpha}\boldsymbol{d}_k) = 0$.

- Line search constrained to $\alpha \in [0, \bar{\alpha})$
- Vertical asymptote at $\bar{\alpha}$
- Classical methods not suited

The Majorize-Minimize (MM) principle *[Hunter04]*

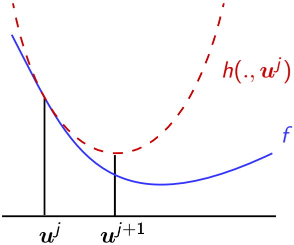
Goal: find $\boldsymbol{u}$ that minimizes f over E

For all $\boldsymbol{v} \in E$, let $h(., \boldsymbol{v})$ a *tangent majorant* for f at $\boldsymbol{v}$ i.e.,

$$h(\boldsymbol{u}, \boldsymbol{v}) \geqslant f(\boldsymbol{u}), \quad \forall \boldsymbol{u} \in E,$$
$$h(\boldsymbol{v}, \boldsymbol{v}) = f(\boldsymbol{v})$$

MM algorithm:

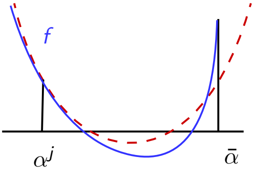
$$\boldsymbol{u}^{j+1} = \arg\min_{\boldsymbol{u} \in E} h(\boldsymbol{u}, \boldsymbol{u}^j)$$



Proposed 1D nonquadratic MM stepsize strategy

Curvature of the barrier term unbounded at $\bar{\alpha}$

$\Rightarrow$ No quadratic majorizing approximation for f.



Finding majorizing approximations of f of the form

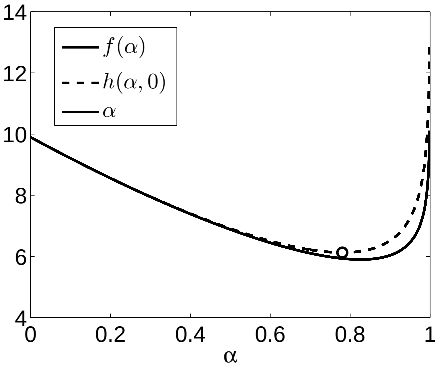
$$h(\alpha, \alpha^j) = p_0 + p_1\alpha + p_2\alpha^2 - p_3 \log(\bar{\alpha} - \alpha)$$

- Construction of h fairly easy for $-log$ and "entropic" barriers
- $\arg\min_\alpha h(\alpha, \alpha^j)$ is a root of degree 2 polynomial

Toy Example

$$f(\alpha) = (\alpha - 5)^2 - \sum_{i=1}^{10} \log(i - \alpha)$$

$\bar{\alpha} = 1$
$m^0 = 2$, $\gamma^0 = 1.55$
$\alpha^1 = 0.7805$
comparing with $\alpha^* = 0.8258$



Convergence analysis results *[Chouzenoux09]*

Goal: Discussing the convergence of the iterative descent algorithm

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad k = 1, \ldots, K$$

Stepsize properties

After any number of MM subiterates,

- Armijo and Zoutendijk conditions hold
- The stepsize bounded away from 0

$\Rightarrow$ Convergence of several optimization algorithms:

Steepest descent, CG, truncated Newton, (L)BFGS ...

Monodimensional Nuclear Magnetic Resonance

NMR model

$$y(\tau) = \int_0^{\infty} x(T) \exp^{-\frac{\tau}{T}} \, dT$$

T: relaxation time
τ: echo time
$y(\tau)$: measured echo
$x(T)$: spectrum to estimate
After discretization,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\epsilon}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{y} \in \mathbb{R}^m$$

Goal: Estimate $\boldsymbol{x}$ from $\boldsymbol{y}$ subject to $\boldsymbol{x} \geqslant 0$

Maximum entropy method

$$\min_{\boldsymbol{x} \geqslant 0} F(\boldsymbol{x}) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x})$$

- Fit to data: Least square $P(\boldsymbol{x}) = \|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\|^2$
- Regularization: Entropy measure $B(\boldsymbol{x}) = \sum_n x_n \ln x_n$

Optimization strategy

- NLCG algorithm with specific SVD preconditionner
- Comparison of a classical Wolfe line search *[Moré94]* with the proposed MM line search.

	c_1	c_2	K	$T(s)$
MT	10^{-3}	0.1	25	1.08
	10^{-3}	0.5	28	1.27
	10^{-3}	0.9	34	1.28
	10^{-3}	0.99	49	1.69
MM	J		K	$T(s)$
	1		24	0.86
	2		26	1.17
	5		28	1.73
	10		27	2.44

K: Iterates number
T: Time before convergence
(c_1, c_2): Wolfe parameters
J: Number of MM subiterates

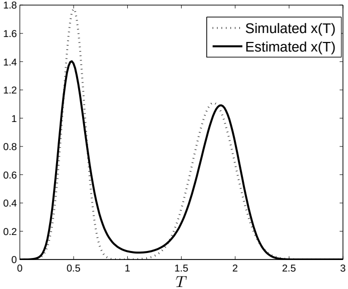


Figure: NMR reconstruction of synthetic data with $SNR = 40dB$

Interior point for quadratic programming

Convex quadratic programming problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F_0(\boldsymbol{x}) = \rho_0 + a_0^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}_0\boldsymbol{x} \tag{2}$$

$$\text{s.t. : } c_i(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}_i\boldsymbol{x} + a_i^T\boldsymbol{x} + \rho_i \geqslant 0, \quad 1 \leqslant i \leqslant m$$

Augmented criterion

$$F_\mu(\boldsymbol{x}) = F_0(\boldsymbol{x}) - \mu\sum_{i=1}^{m}\log(c_i(\boldsymbol{x}))$$

Interior point: Solve $\arg\min F_\mu$ for a series $\{\mu\} \rightarrow 0$.

Interior Point algorithm *[Boyd04]*

1) Set $\mu = 1$, tolerances ϵ, ξ and select a feasible point $\boldsymbol{x}$
2) **WHILE** $\frac{1}{2}(\boldsymbol{d}^T \nabla F_\mu)^2 > \xi$,

    Compute Newton direction $\boldsymbol{d}$ of F_μ
    Compute step size α
    Update $\boldsymbol{x} \leftarrow \boldsymbol{x} + \alpha \boldsymbol{d}$

3) **IF** $\mu < \epsilon$, **RETURN**
   ELSE Set $\mu \leftarrow \theta\mu$ and **GO TO** step 2.

Interior Point algorithm *[Boyd04]*

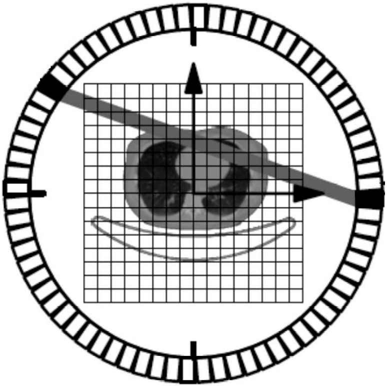
1) Set $\mu = 1$, tolerances ϵ, ξ and select a feasible point $\boldsymbol{x}$
2) **WHILE** $\frac{1}{2}(\boldsymbol{d}^T \nabla F_\mu)^2 > \xi$,
   Compute Newton direction $\boldsymbol{d}$ of F_μ
   Compute step size α
   Update $\boldsymbol{x} \leftarrow \boldsymbol{x} + \alpha \boldsymbol{d}$
3) **IF** $\mu < \epsilon$, **RETURN**
   ELSE Set $\mu \leftarrow \theta\mu$ and **GO TO** step 2.

Results

50 random problems of size $n = 400$, $m = 200$

Backtracking	273.24 iter	5636.87 s
Damping *[Nesterov94]*	135.30 iter	465.07 s
Proposed MM linesearch	64.40 iter	225.03 s

Emission tomography reconstruction



See article

Conclusion and future work

Main result

MM linesearch strategy for criteria involving barrier functions

- Simple stepsize scheme
- Strong convergence results
- Efficient in practice

Constrained optimization

Possible adaptation to

- Primal interior point *[Johnson2000]*
- L-BFGS-B *[Byrd1995]*
- Multiplicative algorithm *[Lanteri2001]*
- Kullback proximal algorithm *[Teboulle1997,Chrétien2000]*

Bibliography (1/3)

P. D. Bertsekas
Nonlinear Programming.
Athena Scientific, 1999.

S. Boyd and L. Vandenberghe
Convex Optimization.
Cambridge University Press, 2004.

Y. Nesterov and A. Nemirovskii
Interior point polynomial algorithms in convex programming.
Studies in Applied and Numerical Mathematics, 1994.

J. Nocedal and S. J. Wright.
Numerical Optimization.
Springer Series in Operations Research, 1999.

R. H. Byrd and P. Lu and J. Nocedal and C. Zhu.
A Limited Memory Algorithm for Bound Constrained Optimization.
SIAM Journal on Scientific Computing, 16:1190:1208,1995

E. Chouzenoux and S. Moussaoui and J. Idier
A new line search method for barrier functions with strong convergence properties
Technical Report - 27/02/09
http://hal.archives-ouvertes.fr/hal-00362304/fr/

S. Chrétien and A. O. Hero
Kullback proximal algorithms for maximum-likelihood estimation.
IEEE Transactions on Information Theory, 46(5):1800:1810,2000.

Bibliography (2/3)

H. Erdogan and J.A. Fessler.
Monotonic algorithms for transmission tomography
IEEE Medical Imaging, 18(9):801–814, 1999

J. Ch. Gilbert and J. Nocedal.
Global convergence properties of conjugate gradient methods for optimization.
Technical Report INRIA, n$^\circ$1268, 1990.

W. W. Hager and H. Zhang.
A survey of non linear conjugate gradient methods.
Pacific Journal of Optimization, 2 :35–58, 2006.

D. R Hunter and L. Kenneth.
A tutorial on MM algorithms
The American Statistician, 18(1):30–37, 2004

M.W. Jacobson and J.A. Fessler.
An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms
IEEE Image Processing, 16(10):2411–2422, 2007

C. A. Johnson and J. Seidel and A. Sofer
Interior-Point Methodology for 3-D PET Reconstruction.
IEEE Trans. on Medical Imaging, 19(4):271:285,2000.

Bibliography (3/3)

C. Labat and J. Idier.
Convergence of conjugate gradient methods with a closed-form stepsize formula.
Journal of Optimization Theory and Applications, 136(1):43–60, 2008.

H. Lanteri and M. Roche and O. Cuevas and C. Aime.
A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints.
Signal Processing, 81:945:974,2001.

F. Mariette and J. P. Guillement and C. Tellier and P. Marchal.
Continuous relaxation time distribution decomposition by MEM.
Signal Treatment and Signal Analysis in NMR, 10:218–234, 1996.

J. J. Moré and D. J. Thuente.
Line search algorithms with guaranteed sufficient decrease.
ACM Transactions on Mathematical Software, 20(3):286–307, 1994.

W. Murray and M. H. Wright.
Line search procedures for the logarithmic barrier function.
SIAM Journal on Optimization, 4(2):229–246, 1994.

M.Teboulle.
Convergence of proximal-like algorithms.
SIAM Journal on Optimization, 7(4):1069–1083, 1997.

A Majorize-Minimize line search algorithm for barrier function optimization

Emilie Chouzenoux
emilie.chouzenoux@irccyn.ec-nantes.fr

Institut de Recherche en Communication et Cybernétique de Nantes (FRANCE)

EUSIPCO, Glasgow, August 27th 2009