A MAJORIZE-MINIMIZE LINE SEARCH ALGORITHM FOR BARRIER FUNCTION OPTIMIZATION

Emilie Chouzenoux, Saïd Moussaoui, and Jérôme Idier

IRCCYN, CNRS UMR 6597, 1, rue de la Noë, BP 92101, F-44321 Nantes Cedex 03, France
phone: + 33 2 40 37 69 24, fax: + 33 2 40 37 69 30,
{emilie.chouzenoux,said.moussaoui,jerome.idier}@irccyn.ec-nantes.fr

ABSTRACT

Many signal and image estimation problems such as maximum entropy reconstruction and positron emission tomography, require the minimization of a criterion containing a *barrier* function i.e., an unbounded function at the boundary of the feasible solution domain. This function has to be carefully handled in the optimization algorithm. When an iterative descent method is used for the minimization, a search along the line supported by the descent direction is usually performed at each iteration. However, standard line search strategies tend to be inefficient in this context. In this paper, we propose an original line search algorithm based on the majorize-minimize principle. A tangent majorant function is built to approximate a scalar criterion containing a barrier function. This leads to a simple line search ensuring the convergence of several classical descent optimization strategies, including the most classical variants of nonlinear conjugate gradient. The practical efficiency of the proposal scheme is illustrated by means of two examples of signal and image reconstruction.

1. INTRODUCTION

The solution of several problems in signal and image estimation involves the minimization of a criterion F in the strictly feasible domain $\mathcal{C}$ defined by some concave inequalities $c_i(\boldsymbol{x}) > 0$, $i = 1, \ldots, N$. These constraints are implicitly taken into account when F contains a *barrier* function, which makes the criterion unbounded at the boundary of $\mathcal{C}$ so that its minimizers belong to $\mathcal{C}$. For example, the minimizer of $F(\boldsymbol{x}) = F_0(\boldsymbol{x}) - \sum_i \log x_i$ is strictly positive because of the unboundness of the logarithmic function at the neighborhood of zero.

This property is used by interior point methods [14] to minimize $F_0(\boldsymbol{x})$ subject to $c_i(\boldsymbol{x}) \geqslant 0$, a barrier function B being artificially introduced to keep the solution inside the feasible domain. The augmented criterion can be expressed as $F_\mu(\boldsymbol{x}) = F_0(\boldsymbol{x}) + \mu B(\boldsymbol{x})$, where $\mu \geqslant 0$ is the barrier parameter and B is the barrier function associated to the constraints $c_i(\boldsymbol{x}) > 0$. For instance, $B(\boldsymbol{x}) = -\sum_i \log(c_i(\boldsymbol{x}))$. The minimization of F_μ must be performed for a sequence of parameter values μ that decreases to 0. This method can be applied to sparse signal reconstruction and compressed sensing [5] as illustrated in section 4.

The barrier term can also be part of the criterion itself such as in maximum entropy terms and Poissonian log-likelihoods. For instance, let us consider the inverse problem of recovering a signal or an image $\boldsymbol{x}$ from a set of noisy observations $\boldsymbol{y}$, where the measurement process is represented by the linear model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\epsilon}$ with $\boldsymbol{H}$ is a known matrix and $\boldsymbol{\epsilon}$ a noise term. An estimate of $\boldsymbol{x}$ can be obtained as the minimizer of a cost function F depending on the noise statistics and on the desired solution properties.

In the case of a Gaussian likelihood, maximum entropy reconstruction [7, 13] consists in minimizing $F(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 + \lambda S(\boldsymbol{x})$ where λ is the regularization parameter and $S(\boldsymbol{x})$ is an entropy term such as

$$S(\boldsymbol{x}) = \begin{cases} \sum_i x_i \log x_i & \text{Shannon entropy} \\ -\sum_i \log x_i & \text{Burg entropy} \end{cases}$$

In both cases, S acts not only as a regularization function but also as a barrier function for positivity constraints. In the case of Shannon entropy, although S remains bounded in the nonnegative orthant, positivity is enforced by the unboundness of the norm of $\nabla S(\boldsymbol{x})$ for small positive values of $\boldsymbol{x}$.

In a penalized Poissonian likelihood case, the criterion to minimize reads $F(\boldsymbol{x}) = L(\boldsymbol{y}, \boldsymbol{H}\boldsymbol{x}) + \lambda R(\boldsymbol{x})$, where R is the penalization function and L is the fidelity to data term defined as

$$L(\boldsymbol{y}, \boldsymbol{H}\boldsymbol{x}) = \sum_m [\boldsymbol{H}\boldsymbol{x}]_m - y_m \log([\boldsymbol{H}\boldsymbol{x}]_m). \qquad (1)$$

The logarithmic term in L plays the role of a barrier function ensuring the positivity of $[\boldsymbol{H}\boldsymbol{x}]_m$. In section 4, we will consider an emission tomography problem [12] that calls for the minimization of this criterion form.

The aim of this paper is to address optimization problems that read

$$\min_{\boldsymbol{x}} \left(F(\boldsymbol{x}) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x}) \right), \quad \mu \geqslant 0 \qquad (2)$$

where P is a differentiable function and B is a differentiable barrier function ensuring the fulfillment of some linear constraints $[\boldsymbol{A}\boldsymbol{x}]_i + \rho_i > 0$.

Many optimization algorithms are based on iteratively decreasing the criterion by moving the current solution $\boldsymbol{x}_k$ along a direction $\boldsymbol{d}_k$,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \qquad (3)$$

where $\alpha_k > 0$ is the *stepsize* and $\boldsymbol{d}_k$ is a *descent direction* i.e., satisfying $\nabla F(\boldsymbol{x}_k)^T \boldsymbol{d}_k < 0$. In practice, such iterative descent direction methods consist in alternating the following steps

THIS WORK IS SUPPORTED BY THE FRENCH ANR PROJECT OPTIMED

1. Construction of $\boldsymbol{d}_k$: the direction depends on the gradient of the criterion at the current value $\boldsymbol{x}_k$. More elaborate methods also involve the Hessian matrix (e.g., Newton, Quasi-Newton) or also depend on the previous descent directions (e.g., conjugate gradient methods, L-BFGS).
2. Determination of α_k (*line search*): the value of α_k is obtained by minimizing the scalar function $f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$.

However, the barrier function causes the inefficiency of standard line search strategies [9]. In the next section we propose an original line search procedure based on the majorize-minimize (MM) principle [3] by deriving an adequate form of a tangent majorant function well suited to approximate a criterion containing a barrier function.

2. LINE SEARCH STRATEGIES

2.1 Problem statement

According to Wolfe conditions, α_k is acceptable if there exists $(c_1, c_2) \in (0;1)$ such that

$$F(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k) + c_1 \alpha_k \boldsymbol{g}_k^T \boldsymbol{d}_k \quad (4)$$

$$|\nabla F(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^T \boldsymbol{d}_k| \leqslant c_2 |\boldsymbol{g}_k^T \boldsymbol{d}_k| \quad (5)$$

where $\boldsymbol{g}_k \triangleq \nabla F(\boldsymbol{x}_k)$. There exist several strategies [11] for finding such an acceptable stepsize: exact minimization of $f(\alpha)$, backtracking or more generally dichotomy, approximation of $f(\alpha)$ using a cubic interpolating function [8, 11] or approximation of $f(\alpha)$ by a quadratic function [6]. However, the barrier term $B(\boldsymbol{x})$ implies that $f(\alpha)$ tends to infinity when α is equal to the smallest positive step $\bar{\alpha}$ cancelling some constraint at $\boldsymbol{x} + \bar{\alpha}\boldsymbol{d}$. Consequently, we must ensure that during the line search, the step values remain in the interval $[0; \bar{\alpha})$ since the function f is undefined for $\alpha \geqslant \bar{\alpha}$. Moreover, due to the vertical asymptote at $\bar{\alpha}$, methods using cubic interpolations or quadratic approximations are not suited [9].

Some line search strategies adapted to barrier function optimization have been proposed in [9]. They make use of specific interpolating functions accounting for the barrier term in $f(\alpha)$. Unfortunately, the resulting algorithms are not often used in practice, probably because the proposed interpolating functions are difficult to compute. In contrast, our approach is not based on interpolation, but rather on majorization, with a view to devise a simple line search strategy with strong convergence properties.

2.2 MM algorithms

In MM algorithms [3, 4], the minimization of a function f is obtained by performing successive minimizations of *tangent majorant* functions for f. Function $h(\boldsymbol{u}, \boldsymbol{v})$ is said tangent majorant for $f(\boldsymbol{u})$ at $\boldsymbol{v}$ if $h(\boldsymbol{u}, \boldsymbol{v}) \geqslant f(\boldsymbol{u})$ and $h(\boldsymbol{v}, \boldsymbol{v}) = f(\boldsymbol{v})$. The initial optimization problem is then replaced by a sequence of easier subproblems, corresponding to the MM update rule

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

Recently, a line search procedure based on an MM algorithm has been introduced [6]. In this strategy, the stepsize value α_k results from J successive minimizations of quadratic tangent majorant functions for the scalar function $f(\alpha)$. The convergence of a family of non-linear conjugate gradient methods associated to this line search strategy is proved in [6] whatever the value of J. However, since the function $f(\alpha)$ resulting from problem (2) is unbounded, there is no quadratic that majorizes $f(\alpha)$ in the whole definition domain of α. Actually, it would be sufficient to majorize $f(\alpha)$ within the level set $\mathcal{L}_k = \{\alpha, F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k)\}$ but this set is difficult to determine or even to approximate.

2.3 A new tangent majorant for MM line search

Instead of a quadratic, we propose the following form of tangent majorant function:

$$h(\alpha) = p_0 + p_1\alpha + p_2\alpha^2 - p_3 \log(\bar{\alpha} - \alpha), \quad (6)$$

which is reminiscent of interpolation functions proposed in [9, 10]. According to MM theory, the stepsize α_k is defined by

$$\begin{aligned} \alpha_k^0 &= 0, \\ \alpha_k^{j+1} &= \arg\min_{\alpha} h_k^j(\alpha, \alpha_k^j), \quad j = 0, \ldots, J-1, \\ \alpha_k &= \alpha_k^J, \end{aligned} \quad (7)$$

where $h_k^j(\alpha, \alpha_k^j)$ is the tangent majorant function

$$\begin{aligned} h_k^j(\alpha,\alpha_k^j) &= f(\alpha_k^j) + (\alpha - \alpha_k^j)\dot{f}(\alpha_k^j) + \frac{1}{2} m_k^j (\alpha - \alpha_k^j)^2 \\ &+ \gamma_k^j \Big[(\bar{\alpha}_k - \alpha_k^j) \log\Big(\frac{\bar{\alpha}_k - \alpha_k^j}{\bar{\alpha}_k - \alpha} \Big) - \alpha + \alpha_k^j \Big] \end{aligned} \quad (8)$$

which depends on two parameters m_k^j, γ_k^j. It is easy to check that $h_k^j(\alpha, \alpha) = f(\alpha)$ for all α. There remains to find values of m_k^j, γ_k^j such that $h_k^j(\alpha, \alpha_k^j) \geqslant f(\alpha)$ holds for all $\alpha \in [0; \bar{\alpha}_k)$.

In [1], the case of a logarithmic barrier associated with linear inequality constraints

$$B(\boldsymbol{x}) = -\sum_i t_i \log([\boldsymbol{A}\boldsymbol{x}]_i + \rho_i), \quad \mu,\, t_i > 0,$$

is dealt with (Figure 1 illustrates a example of scalar criterion and the obtained majorant). Assuming that $p(\alpha) = P(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$ is majorized by the quadratic function

$$p(\alpha_k^j) + (\alpha - \alpha_k^j)\, \dot{p}(\alpha_k^j) + \frac{1}{2} m_p (\alpha - \alpha_k^j)^2, \quad (9)$$

the majorization of f is given by the following property.

Property 1. *Let $\boldsymbol{a} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{\rho}$ and $\boldsymbol{\delta} = \boldsymbol{A}\boldsymbol{d}_k$, so that f has a barrier located at*

$$\bar{\alpha} = \min_{i|\delta_i < 0} -a_i/\delta_i. \quad (10)$$

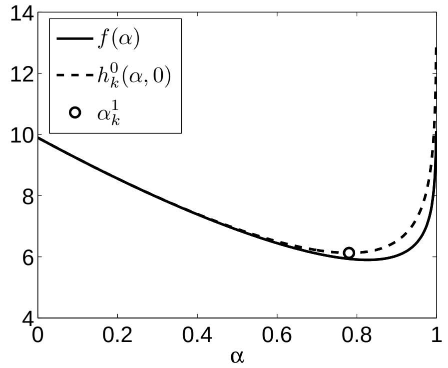

Figure 1: Example of a tangent majorant function $h_k^0(\alpha, 0)$ for $f(\alpha) = (\alpha - 5)^2 - \sum_{i=1}^{10} \log(i - \alpha)$. $h_k^0(\alpha, 0)$ is defined by (8) with $m_k^0 = 2$, $\gamma_k^0 = 1.55$ and $\bar{\alpha} = 1$

If $\alpha_k^j = 0$, *let also* $m_k^j = m_p + \mu m_b$ *and* $\gamma_k^j = \mu \gamma_b$ *with* $m_b = \ddot{b}_1(0)$ *and* $\gamma_b = \dot{b}_2(0)\bar{\alpha}$. *Otherwise, let*

$$\begin{aligned} m_b &= \frac{b_1(0) - b_1(\alpha_k^j) + \alpha_k^j \dot{b}_1(\alpha_k^j)}{(\alpha_k^j)^2/2} \\ \gamma_b &= \frac{b_2(0) - b_2(\alpha_k^j) + \alpha_k^j \dot{b}_2(\alpha_k^j)}{(\bar{\alpha} - \alpha_k^j)\log(1 - \alpha_k^j/\bar{\alpha}) + \alpha_k^j} \end{aligned} \tag{11}$$

where $b_1(\alpha) = \sum_{i|\delta_i>0} -t_i \log(a_i + \alpha\delta_i)$ *and* $b_2(\alpha) = \sum_{i|\delta_i<0} -t_i \log(a_i + \alpha\delta_i)$. *Then, function* $h_k^j(\cdot, \alpha_k^j)$ *is a tangent majorant of* $f(\cdot)$

Moreover, (11) implies $m_k^j, \gamma_k^j > 0$, so $h_k^j(\cdot, \alpha_k^j)$ is strictly convex. Hence, it has a unique minimizer, which takes an explicit form:

$$\alpha_k^{j+1} = \alpha_k^j + \frac{-A_2 + \sqrt{A_2^2 - 4A_1A_3}}{2A_1}, \tag{12}$$

with $A_1 = -m_k^j$, $A_2 = \gamma_k^j - \dot{f}(\alpha_k^j) + m_k^j(\bar{\alpha} - \alpha_k^j)$ and $A_3 = (\bar{\alpha} - \alpha_k^j)\dot{f}(\alpha_k^j)$. In such conditions, (7) produces monotonically decreasing values $\{f(\alpha_k^j)\}$ and the series $\{\alpha_k^j\}$ converges to a stationnary point of $f(\alpha)$ [3].

3. CONVERGENCE ANALYSIS RESULTS

This section focuses on the convergence of the iterative scheme (3) where α_k is chosen according to our MM strategy. Only the main results are presented here, while a detailed analysis can be found in [1]. In the whole section, $\boldsymbol{d}_k$ is assumed to be a descent direction, so that $\dot{f}(0) = \boldsymbol{d}_k^T \boldsymbol{g}_k < 0$.

3.1 Lower and upper bounds for the stepsize

Property 2. *There exists* $\nu > 0$ *such that*

$$\alpha_k^1 \geqslant -\nu \boldsymbol{g}_k^T \boldsymbol{d}_k / \|\boldsymbol{d}_k\|^2. \tag{13}$$

Moreover, $\forall j > 1$, *there exist* $(c^{\min}, c_j^{\max})$ *such that*

$$c^{\min}\alpha_k^1 \leqslant \alpha_k^j \leqslant c_j^{\max}\alpha_k^1. \tag{14}$$

3.2 Wolfe conditions

Given a current solution $\boldsymbol{x}_k$ and a current descent direction $\boldsymbol{d}_k$, the chosen stepsize α_k must induce a sufficient decrease of F. The first Wolfe condition (4) measures this decrease. It is equivalent to

$$f(\alpha) - f(0) \leqslant c_1 \alpha_k \dot{f}(0). \tag{15}$$

The following result holds.

Property 3. *The stepsize* (7) *fulfills* (15) *with*

$$c_1 = (2c_J^{\max})^{-1} \in (0; 1) \tag{16}$$

On the other hand, it turned out difficult or even impossible to fulfill the second Wolfe condition (5) for any value of J. Fortunately, it is easier to show that the so-called Zoutendijk condition holds, the latter being a weaker condition that is nonetheless sufficient to lead us to convergence results.

3.3 Zoutendijk condition

The global convergence of a descent direction method is non only ensured by a 'good choice' of the step but also by well-chosen search directions $\boldsymbol{d}_k$. Convergence proofs are often based on the fulfillment of Zoutendijk condition

$$\sum_{k=0}^{\infty} \|\boldsymbol{g}_k\|^2 \cos^2\theta_k < \infty, \tag{17}$$

where θ_k is the angle between $\boldsymbol{d}_k$ and the steepest descent direction $-\boldsymbol{g}_k$,

$$\cos\theta_k = -\boldsymbol{g}_k^T \boldsymbol{d}_k / (\|\boldsymbol{g}_k\| \, \|\boldsymbol{d}_k\|). \tag{18}$$

Inequality (17) implies that $\cos\theta_k \|\boldsymbol{g}_k\|$ vanishes for large values of k. Moreover, provided that $\boldsymbol{d}_k$ is not orthogonal to $-\boldsymbol{g}_k$ (i.e., $\cos\theta_k > 0$), condition (17) implies the convergence of the algorithm in the sense

$$\lim_{k\to\infty} \|\boldsymbol{g}_k\| = 0. \tag{19}$$

In the case of the proposed line search, the following result holds [1].

Property 4. *Let* α_k *be defined by* (7). *Then, according to Properties 2 and 3, Zoutendijk condition* (17) *holds.*

3.4 Convergence of Newton-like methods

The Newton-like methods are defined by the following recurrence

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \\ \boldsymbol{d}_k &= -\boldsymbol{B}_k^{-1} \boldsymbol{g}_k. \end{aligned} \tag{20}$$

Property 5. *Assume that matrices* $\boldsymbol{B}_k$ *are positive definite for all* k *and that there exists* $M > 0$ *such that*

$$\|\boldsymbol{B}_k\| \, \|\boldsymbol{B}_k^{-1}\| \leqslant M, \ \forall k. \tag{21}$$

Algorithm (20) *is convergent when* α_k *is defined by* (7) *in the sense*

$$\lim_{k\to\infty} \|\boldsymbol{g}_k\| = 0. \tag{22}$$

This result covers the following classical methods

- Steepest descent: $\boldsymbol{d}_k = -\boldsymbol{g}_k$,
- Newton: $\boldsymbol{d}_k = -\nabla^2 F(\boldsymbol{x}_k)^{-1}\boldsymbol{g}_k$ in the convex case,
- Quasi Newton (BFGS) in the convex case.

3.5 Convergence of conjugate gradient methods

The nonlinear conjugate gradient algorithm (NLCG) is defined by the following recurrence

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \\ \boldsymbol{d}_k &= -\boldsymbol{c}_k \operatorname{sign}(\boldsymbol{g}_k^T \boldsymbol{c}_k) \\ \boldsymbol{c}_k &= -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1} \end{aligned} \tag{23}$$

Property 6. *The NLCG algorithm is convergent when α_k is defined by (7) and β_k is chosen according to a classical conjugacy formula such as Polak Ribière Polyak (PRP), Fletcher Reeves (FR) or Hestenes Stiefel (HS), in the sense*

$$\lim_{k\to\infty} \inf \|\boldsymbol{g}_k\| = 0. \tag{24}$$

4. APPLICATIONS

We consider here two image/signal processing examples with the aim to analyse the performances of descent optimization algorithms when the step size is obtained by the proposed MM line search procedure.

4.1 Positron emission tomography

The measurements in positron emission tomography (PET) [12] are modeled as Poisson random variables

$$\boldsymbol{y} \sim Poisson(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{r}) \tag{25}$$

where the ith entry of $\boldsymbol{x}$ represents the radioisotope concentration in pixel i and $\boldsymbol{H}$ is the projection matrix whose elements H_{mi} model the contribution of the ith pixel to the mth datapoint. The components of $\boldsymbol{y}$ are the counts measured by the pairs detectors of and $\boldsymbol{r}$ models the background events. We consider a simulated example using data generated with J.A. Fessler's code available at `http://www.eecs.umich.edu/~fessler`. For this simulation, we take an object of size 128×128 pixels and assume $M = 24924$ pairs of detectors.

4.1.1 Objective function

According to the noise statistics, the log-likelihood of the emission data is

$$J(\boldsymbol{x}) = \sum_m \big([\boldsymbol{H}\boldsymbol{x}]_m + r_m - y_m \log([\boldsymbol{H}\boldsymbol{x}]_m + r_m)\big).$$

A useful penalization aiming at favorizing smoothness of the estimated image is given by

$$R_{\text{Hub}}(\boldsymbol{x}) = \sum_{t\in\mathcal{T}} \omega_t \psi([\boldsymbol{D}\boldsymbol{x}]_t),$$

where ψ is the edge preserving potential fonction $\psi(u) = \sqrt{\delta^2 + u^2} - \delta$ and $[\boldsymbol{D}\boldsymbol{x}]_t$ is the vector of difference between neighboring pixel intensities. The weights depend on the relative position of the neighbors: $\omega_t = 1$ for vertical and horizontal neighbors and $\omega_t = 1/\sqrt{2}$ for diagonal neighbors. To ensure the positivity of the estimate, a logarithmic barrier term is added:

$$R_{\text{Pos}}(\boldsymbol{x}) = -\sum_{i=1}^{N} \log x_i. \tag{26}$$

Finally, the estimated image is the minimizer of the following objective function

$$F(\boldsymbol{x}) = J(\boldsymbol{x}) + \lambda_1 R_{\text{Hub}}(\boldsymbol{x}) + \lambda_2 R_{\text{Pos}}(\boldsymbol{x}). \tag{27}$$

4.1.2 Optimization strategy

The NLCG algorithm with PRP conjugacy is employed with or without preconditioning. The aim is to compare the performance of the proposed MM line search with Moré and Thuente's cubic interpolation procedure (MT) [8]. The algorithm is initialized with a uniform positive object and the convergence is checked using the following stopping rule [11]

$$\|\boldsymbol{g}_k\|_\infty < 10^{-5}(1 + |F(\boldsymbol{x}_k)|). \tag{28}$$

The regularization and barrier parameters are set to $\lambda_1 = 10$, $\delta = 0.01$ and $\lambda_2 = 0.1$. This choice leads to a fairly acceptable reconstructed image quality.

4.1.3 Results and discussion

Table 1 summarizes the performance results in terms of iteration number n_1 and computation time T on an Intel Pentium 4 3.2 GHz, 3 GB RAM. The design parameters are the Wolfe condition constants (c_1, c_2) for the MT method and the number of subiterations J for the MM procedure.

			NLCG		PNLCG	
	c_1	c_2	n_1	T(s)	n_1	T(s)
MT	10^{-4}	0.1	403	1043.3	18	78.66
MT	10^{-4}	0.5	207	454.6	17	73.64
MT	10^{-4}	0.9	196	418.5	19	82.81
MT	10^{-4}	0.999	170	362.7	19	83.27
	J		n_1	T(s)	n_1	T(s)
MM	1		232	287.45	18	58.76
MM	2		275	375.27	16	55.34
MM	5		304	512.28	22	83.02
MM	10		474	1017.7	24	100.45

Table 1: Comparison between MM and MT line search strategies for a PET reconstruction problem.

It can be noted that the NLCG algorithm with MM line search requires more iterations than the NLCG-MT approach provided that the parameters (c_1, c_2) are appropriately chosen. However, the NLCG-MM is faster because of a smaller computational cost per iteration. Moreover, the proposed MM procedure admits a unique tuning parameter, namely the subiteration number J, and $J = 1$ always seems a good choice. Furthermore, resorting to the diagonal Hessian preconditioner significantly speeds up the convergence in the sense of both n_1 and T.

4.2 Sparse spike deconvolution

The observation vector $\boldsymbol{y} \in \mathbb{R}^M$ results from the noisy convolution of a sparse spike train sequence $\boldsymbol{x} \in \mathbb{R}^N$ with a filter $\boldsymbol{h}$ of length L. The added noise is centered, white Gaussian. In this experiment, $M = 1020$, $L = 20$, $N = 1000$, the spike train sequence is simulated from a Bernoulli-Gaussian distribution with parameter $\beta = 0.06$, and the signal to noise ratio is 13dB.

4.2.1 Objective function

The ℓ_1 norm is a suited regularization function to account for the sparseness of $\boldsymbol{x}$, which leads to the following optimization problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{h} \star \boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 . \quad (29)$$

To tackle the non differentiability of the ℓ_1 norm, problem (29) can be classicaly reformulated as a quadratic programming problem [5]:

$$\min_{\boldsymbol{x},\boldsymbol{u}} F(\boldsymbol{x}, \boldsymbol{u}) = \|\boldsymbol{y} - \boldsymbol{h} \star \boldsymbol{x}\|_2^2 + \lambda \sum_i u_i$$
$$\text{subject to } -u_i \leqslant x_i \leqslant u_i, \quad i = 1, \ldots, N \quad (30)$$

4.2.2 Optimization strategy

An interior-point method is proposed in [5] to solve (30). The augmented criterion has the form (2) where $P(\boldsymbol{x}) \equiv F(\boldsymbol{x}, \boldsymbol{u})$ and the barrier function is

$$B(\boldsymbol{x}, \boldsymbol{u}) = -\sum_i \log(u_i + x_i) - \sum_i \log(u_i - x_i). \quad (31)$$

For a decreasing sequence of μ, the augmented criterion $F_\mu(\boldsymbol{x}, \boldsymbol{u})$ is minimized using a truncated Newton method where the search direction is obtained by applying a preconditioned conjugate gradient (PCG) to the Newton equations. The stepsize α satisfying the first Wolfe condition (4) results from a backtracking line search and the barrier parameter μ is decreased when $\alpha \geqslant \alpha_{\min}$. The Matlab code of the algorithm is available at S. Boyd's homepage `http://www.stanford.edu/~boyd`. Here, we propose to compare the performances of this algorithm when the backtracking is replaced by our MM line search.

4.2.3 Results and discussion

Table 2 reports the computational results when Boyd's code is used with its default parameters $\alpha_{\min} = 0.5$ and $\lambda = 0.01$. n_1 is the number of μ updates, n_2 is the average number of iterations for minimizing $F_\mu(\boldsymbol{x}, \boldsymbol{u})$ and n_3 is the average number of PCG iterations. The use of our MM line search slightly enhances the performances of the deconvolution algorithm. On this particular problem, the best results have been obtained when J is larger or equal to 5.

		n_1	n_2	n_3	T
Backtracking		13	1	50.07	8.35
MM (J)	1	10	1.9	56.75	13.41
	2	10	1.7	58.67	13.81
	3	10	1.6	57.47	11.58
	4	11	1.27	51.67	9.32
	5	12	1	49.08	<u>7.77</u>
	6	12	1	49.08	7.85
	7	12	1	49.15	7.98

Table 2: Comparison between MM and backtracking line search for a spike train deconvolution problem.

5. CONCLUSION

In [6], a simple and efficient quadratic MM line search method has been proposed. However, it is restricted to gradient-Lipschitz criteria, which excludes the case of barrier functions. This case can be handled with the MM line search method presented in this paper. This method benefits from strong convergence results, it is still very easy to implement, and shows itself at least as efficient as classical techniques on practical problems.

REFERENCES

[1] E. Chouzenoux, S. Moussaoui, and J. Idier. A new line search method for barrier functions with strong convergence properties. Technical report, IRCCyN, 2009. `http://hal.archives-ouvertes.fr/IRCCYN-ADTSI`.

[2] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *Amer. Statist.*, 58(1):30–37, February 2004.

[3] M. Jacobson and J. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Trans. Image Processing*, 16(10):2411–2422, October 2007.

[4] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale L1-regularized least squares. *IEEE J. Selected Topics Sig. Proc.*, 1(4):606–617, December 2007.

[5] C. Labat and J. Idier. Convergence of conjugate gradient methods with a closed-form stepsize formula. *J. Optim. Theory Appl.*, 136(1):43–60, January 2008.

[6] A. Mohammad-Djafari and G. Demoment. Maximum entropy image reconstruction in X-ray and diffraction tomography. *IEEE Trans. Medical Imaging*, 7(4):345–354, 1988.

[7] J. J. Moré and D. J. Thuente. Line search algorithms with guaranteed sufficient decrease. *ACM Trans. on Math. Soft.*, 20(3):286–307, 1994.

[8] W. Murray and M. H. Wright. Line search procedures for the logarithmic barrier function. *SIAM J. Optimization*, 4(2):229–246, 1994.

[9] S. G. Nash and A. Sofer. A barrier method for large-scale constrained optimization. *ORSA J. Comput.*, 5(41), 1993.

[10] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer-Verlag, New York, NY, 1999.

[11] J. M. Ollinger and J. A. Fessler. Positron emission tomography. *IEEE Signal Processing Mag.*, 14(1):43–55, 1997.

[12] J. Skilling and R. K. Bryan. Maximum entropy image reconstruction: General algorithm. *Month. Not. Roy. Astr. Soc.*, 211:111–124, 1984.

[13] M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica 1992*, pages 341–407. Cambridge University Press, 1991.