Majorize-Minimize linesearch for inversion methods involving barrier function optimization

E. Chouzenoux‡, S. Moussaoui and J. Idier

IRCCyN, CNRS UMR 6597, Ecole Centrale de Nantes,
1, rue de la Noë, BP 92101, F-44321 Nantes Cedex 03, France.
Phone: + 33 2 40 37 69 19, fax: + 33 2 40 37 69 30

E-mail: `said.moussaoui@irccyn.ec-nantes.fr`

Abstract. This paper focuses on the issue of stepsize determination (linesearch) in iterative descent algorithms applied to the minimization of a criterion containing a barrier function associated to linear constraints. Such an issue arises in inversion methods involving the minimization of a penalized criterion where the barrier function comes either from the data fidelity term or from the regularizing functional. In order to circumvent the inefficiency of general-purpose linesearch strategies in the case of barrier functions, we propose to adopt a majorization-minimization scheme by deriving a new form of a majorant function well suited to approximate a criterion containing barrier terms. We also establish the convergence of classical descent algorithms when this linesearch strategy is employed. Its efficiency is illustrated by means of numerical examples of signal and image restoration.

Submitted to: *Inverse problems*

‡ E. Chouzenoux is with the Université Paris-Est, LIGM, CNRS-UMR 8049

1. Introduction

A common inverse problem arising in many application domains is to estimate an object from a set of observations depending on this object through a measurement process. In this paper, we consider the frequent situation where the dependence of the observations $\boldsymbol{y} \in \mathbb{R}^M$ on the unknown discretized object $\boldsymbol{x}^o \in \mathbb{R}^N$ is represented by a linear model

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^o + \boldsymbol{\epsilon}, \tag{1}$$

with $\boldsymbol{K}$ a known ill-conditioned matrix and $\boldsymbol{\epsilon}$ an additive noise term representing measurement errors and model uncertainties. This simple formalism covers many real situations such as deblurring, denoising, and inverse-Radon transform in tomography [1]. It can also be used as a first order approximation of a non-linear observation model [2]. To handle the ill-posedness of such problems, several efficient inversion methods are based on the minimization of a composite criterion (See for instance [3, 4] and references therein):

$$F(\boldsymbol{x}) = S(\boldsymbol{x}) + \lambda R(\boldsymbol{x}), \quad \lambda \geqslant 0. \tag{2}$$

The first term $S(\boldsymbol{x})$ aims at enforcing some fidelity of the solution to the data. It typically corresponds to a neg-log-likelihood, which derives from the statistics of the noise $\boldsymbol{\epsilon}$. The second term $R(\boldsymbol{x})$, whose weight is set by the parameter λ, is a regularization term that allows to account for additional information not carried out by the data alone. Its design is linked to some *a priori* assumptions one can have concerning the sought object. Both terms will be assumed differentiable in the sequel.

The effective resolution of the inverse problem is then expressed as that of finding the minimizer of the composite criterion (2). However, in several cases, the solution cannot be given explicitly or cannot be computed directly since it requires the inversion of large scale matrices. Instead, iterative descent algorithms are employed. Starting from an initial guess $\boldsymbol{x}_0$, these algorithms generate a sequence of iterates $\{\boldsymbol{x}_k\}$ until the fulfillment of a stopping condition. In practice, from the current value $\boldsymbol{x}_k$, the update $\boldsymbol{x}_{k+1}$ is obtained according to

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

where $\alpha_k > 0$ is the *stepsize* and $\boldsymbol{d}_k$ is a *descent direction i.e.*, a vector such that $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$, where $\boldsymbol{g}_k = \nabla F(\boldsymbol{x}_k)$ denotes the gradient of F at $\boldsymbol{x}_k$. The determination of α_k is called the *linesearch*. Linesearch strategies perform an inexact minimization of $f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$ to find a stepsize value that ensures the convergence of the whole descent algorithm [5, 6].

The strategies used for computing the direction and the stepsize strongly depend on the mathematical properties of the criterion. In this paper, we focus on penalized criteria that contain a *barrier* function associated to some constraints $\boldsymbol{x} \in \mathcal{C}$. A fundamental property of barrier functions is to ensure that any minimizer of F belongs to the interior of a feasible domain $\mathcal{C}$ by making the gradient of F unbounded at the boundary of $\mathcal{C}$. This property is used by interior point algorithms [7] to solve inequality-constrained optimization problems, a barrier function being artificially introduced to the objective function. Interior-point methods have been applied for instance to sparse signal reconstruction [8] and to image reconstruction under positivity constraints [9].

Table 1 reports several examples of barrier criteria that can be encountered in the context of signal or image reconstruction. In the first two examples, the barrier results

Name	Function	Constraints
Log-likelihoods		
Poisson [12]	$\sum_{m=1}^{M} [\boldsymbol{K}\boldsymbol{x}]_m - y_m + y_m \log \frac{y_m}{[\boldsymbol{K}\boldsymbol{x}]_m}$	$[\boldsymbol{K}\boldsymbol{x}]_m > 0$
Gamma [13, 14]	$\sum_{m=1}^{M} -\log \frac{y_m}{[\boldsymbol{K}\boldsymbol{x}]_m} + \frac{y_m}{[\boldsymbol{K}\boldsymbol{x}]_m}$	$[\boldsymbol{K}\boldsymbol{x}]_m > 0$
Prior log-densities		
Gamma [15, 16]	$\sum_{n=1}^{N} (1-\alpha_n)\log x_n + \frac{\alpha_n}{\beta_n} x_n$	$x_n > 0$
Beta [16]	$\sum_{n=1}^{N} (1-\alpha_n)\log(x_n - a_n) + (1-\beta_n)\log(b_n - x_n)$	$x_n \in (a_n, b_n)$
Rayleigh [17]	$\sum_{n=1}^{N} -\log(x_n) + \alpha_n x_n^2$	$x_n > 0$
Entropies		
Shannon [18]	$\sum_{n=1}^{N} x_n \log x_n$	$x_n > 0$
Burg [11]	$-\sum_{n=1}^{N} \log x_n$	$x_n > 0$
Hyperbolic [19]	$-\sum_{n=1}^{N} \sqrt{x_n}$	$x_n > 0$
Cross entropy [20]	$\sum_{n=1}^{N} x_n \log \frac{x_n}{r_n} + x_n - r_n$	$x_n > 0$
Generalized Fermi-Dirac [21]	$\sum_{n=1}^{N} (x_n - a_n)\log(x_n - a_n) + (b_n - x_n)\log(b_n - x_n)$	$x_n \in (a_n, b_n)$
Roughness penalties		
Kullback-Leibler [22]	$\sum_{n=1}^{N} x_n \log \frac{x_n}{x_{n-1}} + x_n - x_{n-1}$	$x_n > 0$
Itakura-Saito [22]	$\sum_{n=1}^{N} -\log \frac{x_n}{x_{n-1}} + \frac{x_n}{x_{n-1}}$	$x_n > 0$

Table 1. Examples of barrier functions encountered in penalized signal or image reconstruction. The first two functions are data fidelity functions $S(\boldsymbol{x})$ while the others are penalty functions $R(\boldsymbol{x})$. We emphasize that Gamma log-likelihood and the two roughness penalties do not fall within the scope of this study.

from the presence of singular terms in the data fidelity term. For example, when a Poisson noise distribution is assumed, $S(\boldsymbol{x})$ corresponds to the Kullback-Leibler divergence of $\boldsymbol{K}\boldsymbol{x}$ from $\boldsymbol{y}$, which plays the role of a barrier function associated to the constraints $[\boldsymbol{K}\boldsymbol{x}]_m > 0$, $m = 1, \ldots, M$. In section 4.1, we will consider a positron emission tomography problem [10] which involves this form of likelihood. In the other examples, the barrier function is part of the regularization term. For instance, Shannon and Burg entropic penalty terms, used in the maximum entropy strategy for image reconstruction [11], act as barrier functions for the positivity orthant. The maximum entropy approach will be applied in section 4.2 to the reconstruction of one-dimensional nuclear magnetic resonance spectra.

As discussed in [23], general-purpose linesearch techniques tend to be inefficient in the case of criteria containing a barrier function. In this paper, we propose a majorization-minimization approach by constructing a tangent majorant function suitable for a wide set of barriers. As it will be shown hereafter, the main advantage of this approach is to yield a simple scheme for stepsize determination that ensures the convergence of many descent algorithms whatever the number of linesearch iterations.

The rest of this paper is organized as follows. In Section 2 we recall the main properties of the barrier functions and discuss why specific linesearch strategies are called for when dealing with barrier functions. The proposed linesearch procedure is introduced and its properties are studied in Section 3. Section 4 illustrates the efficiency of the proposed approach through numerical examples in the field of signal and image processing.

2. Linesearch strategies for barrier functions

2.1. Formulation of the criterion involving barrier functions

In this paper, we focus on the cases when the composite criterion (2) can be rewritten as

$$F(\boldsymbol{x}) = P(\boldsymbol{x}) + B(\boldsymbol{x}), \tag{4}$$

where B is a barrier function associated to $\boldsymbol{x} \in \mathcal{C}$, with $\mathcal{C}$ defined by linear inequalities:

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^N \,|\, C_i(\boldsymbol{x}) = \boldsymbol{c}_i^T \boldsymbol{x} + \rho_i > \boldsymbol{0},\, \forall i \in \{1, \dots, I\}\}, \tag{5}$$

and P is a differentiable criterion over $\mathcal{C}$. A barrier function associated to $\boldsymbol{x} \in \mathcal{C}$ is built as

$$B(\boldsymbol{x}) = \sum_{i=1}^{I} \psi_i(C_i(\boldsymbol{x})), \tag{6}$$

where for all $i \in \{1, \dots, I\}$, $\psi_i(u)$ are scalar barrier functions associated to $u > 0$, i.e.:

$$\begin{gathered} \psi_i \text{ is continuous and strictly convex on } [0, +\infty[\\ \psi_i(u) \text{ is differentiable on } (0, +\infty[\\ \lim_{u \to 0} \dot{\psi}_i(u) = +\infty \end{gathered}$$

Barrier name	Logarithmic	Inverse	Entropic	Hyperbolic
Function $\psi(u)$	$-\log u$	u^{-1}	$u \log u$	$-u^r, \quad r \in (0, 1)$

Table 2. Examples of scalar barrier functions associated to $u > 0$. The first two are strict barrier functions since they grow to infinity as $u \to 0$. Note that (7) does not hold for the inverse barrier function.

When $\lim_{u \to 0} \psi_i(u) = +\infty$, the scalar barrier ψ_i is said *strict*. In the particular case where ψ_i is a strict scalar barrier function for all $i \in \{1, \dots, I\}$, $B(\boldsymbol{x})$ is called a strict barrier function. We restrict ourselves to barrier functions (6) formed of scalar barriers ψ_i that fulfill

$$-\frac{2}{u}\ddot{\psi}_i(u) \leqslant \dddot{\psi}_i(u) \leqslant 0, \quad \forall u > 0, \quad \forall i \in \{1, \dots, I\}. \tag{7}$$

This assumption allows us to consider, for the ψ_i in (6), logarithmic, entropic and hyperbolic scalar barrier functions presented in Table 2. Therefore, all barrier functions from Table 1 fall within the scope of (6)-(7) except the Gamma log-likelihood and the two roughness penalties.

2.2. Determination of the stepsize

Let $\boldsymbol{x}_k \in \mathcal{C}$ and $\boldsymbol{d}_k$ a descent direction for F at $\boldsymbol{x}_k$. In order to compute the new iterate $\boldsymbol{x}_{k+1}$, one has to perform a linesearch that identifies a step length α_k achieving sufficient reductions in $f(\alpha) = F(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k)$ [6, Chap.3]. The presence of scalar barrier functions ψ_i imply that the derivative of the scalar function

$$f(\alpha) = P(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) + \sum_{i=1}^{I} \psi_i(C_i(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k)) \tag{8}$$

is unbounded when α is such that $C_i(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) = 0$ for some i. Since functions C_i are assumed to be linear, this limits the stepsize value α_k to an interval (α_-, α_+) where

$$\begin{cases} \alpha_- = \max_{i\in\mathcal{I}_-} -\frac{\theta_i}{\delta_i} \\ \alpha_+ = \min_{i\in\mathcal{I}_+} -\frac{\theta_i}{\delta_i} \end{cases} \text{with} \quad \begin{cases} \mathcal{I}_- = \{i \in \{1,\dots,I\} \,|\, \delta_i > 0\} \\ \mathcal{I}_+ = \{i \in \{1,\dots,I\} \,|\, \delta_i < 0\}\,, \end{cases} \tag{9}$$

where, for all $i \in \{1,\dots,I\}$, $\theta_i = \boldsymbol{c}_i^T\boldsymbol{x}_k + \rho_i$, $\delta_i = \boldsymbol{c}_i^T\boldsymbol{d}_k$, and it is understood that $\alpha_- = -\infty$ (respectively, $\alpha_+ = +\infty$) if $\mathcal{I}_-$ (resp., $\mathcal{I}_+$) is empty. Moreover, the stepsize should fulfill some sufficient convergence conditions. The most popular are the strong Wolfe conditions that state that a stepsize series $\{\alpha_k\}$ is acceptable if there exist σ_1, $\sigma_2 \in (0,1)$ such that for all k and for all $\boldsymbol{x}_k$,

$$F(\boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k) + \sigma_1\alpha_k\boldsymbol{g}_k^T\boldsymbol{d}_k, \tag{10}$$

$$|\nabla F(\boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k)^T\boldsymbol{d}_k| \leqslant \sigma_2|\boldsymbol{g}_k^T\boldsymbol{d}_k|. \tag{11}$$

The stepsize is then determined with an iterative procedure that generates candidate values, until (10)-(11) are satisfied. One iteration usually consists in a bracketing phase that finds an interval containing acceptable stepsizes, followed by a polynomial cubic interpolation phase that computes a particular stepsize within this interval [5, 6]. However, cubic interpolation is not suited to interpolate function (8) since its derivative $\dot{f}(\alpha)$ tends to $-\infty$ when α tends to α_- or α_+. Therefore, new interpolating functions have been proposed in [23, 24] to account for the barrier singularity.

2.3. Interpolation based linesearch for barrier functions optimization

The particular case $\psi_i(u) = -\mu\log(u)$, $\mu > 0$, for all $i \in \{1,\dots,I\}$, is considered in [23, 25, 24]. In order to account for the logarithmic barrier term, Murray *et al* propose a log-quadratic interpolating function of the form

$$f_0 + f_1\alpha + f_2\alpha^2 - \mu\log(f_3 - \alpha) \tag{12}$$

where the coefficients f_i are chosen to fit f and its derivative at two or three trial points. More precisely, four interpolating schemes are considered where in each case, f_0, f_1 and f_2 have an analytical expression while the computation of f_3 requires to solve a scalar equation. In order to guarantee the uniqueness of f_3, some inequality has to be fulfilled. If this is the case, f_3 is computed from an iterative Newton procedure. Otherwise, f_3 is undefined and a cubic interpolation is rather used. The linesearch strategy consists in repeating this interpolation process over intervals $[a, b]$ until the fulfillment of Wolfe conditions [24] or Armijo condition [25]. Let us remark that the resulting algorithms are not often used in practice, possibly because the interpolating function is difficult to compute.

3. Majorize-Minimize linesearch

Recently, a linesearch procedure based on the Majoration-Minimization (MM) principle has been introduced [26]. In this strategy, the stepsize α_k results from successive minimizations of quadratic tangent *majorant functions* for $f(.)$. Function $h(., \alpha')$ is said tangent majorant for $f(.)$ at α' if for all α,

$$\begin{cases} h(\alpha, \alpha') \geqslant f(\alpha), \\ h(\alpha', \alpha') = f(\alpha'). \end{cases} \tag{13}$$

The convergence of conjugate-gradient [27, 28] and truncated Newton algorithms [29] associated to quadratic MM linesearch strategy has been established. A major advantage of quadratic majorization is that it gives an analytical formulation of the stepsize value. However, its application is not possible in the case of strict barrier function since there exists no quadratic function that majorizes f in the set (α_-, α_+). Actually, it would be sufficient to majorize f within the level set $\{\alpha \in \mathbb{R} | \, f(\alpha) \leqslant f(0)\}$, but this set reveals difficult to determine or even to approximate. In the case of non strict barriers, f is bounded at the boundary of the set (α_-, α_+). However, the curvature of f is unbounded and one can expect suboptimal results by majorizing the scalar function with a parabola. In particular, very high curvature will be obtained for stepsize values close to the singularity. In this section, we propose a new form of a tangent majorant function that is well suited to approximate a criterion containing a barrier function.

3.1. A new tangent majorant for MM linesearch

Let $\alpha' \in (\alpha_-, \alpha_+)$ a current stepsize value. Instead of a quadratic, we propose the following form of tangent majorant function of f at α':

$$h(\alpha, \alpha') = h_0 + h_1\alpha + h_2\alpha^2 - h_3 \log(h_4 - \alpha). \tag{14}$$

The majorizing function (14) takes a similar form than (12) but, here, parameters h_i are chosen to ensure the majorization properties (13) for all α and α' in (α_-, α_+). According to the MM principle, the stepsize is defined by $\alpha_k = \alpha^J$, with

$$\begin{aligned} \alpha^0 &= 0 \\ \alpha^{j+1} &= \arg\min_{\alpha} h(\alpha, \alpha^j), \quad j = 0, \ldots, J-1 \end{aligned} \tag{15}$$

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

$$h(\alpha, \alpha^j) = \\ f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2} m^j (\alpha - \alpha^j)^2 + \gamma^j \left[(\bar{\alpha}^j - \alpha^j) \log \frac{\bar{\alpha}^j - \alpha^j}{\bar{\alpha}^j - \alpha} - \alpha + \alpha^j \right], \tag{16}$$

which depends on three parameters m_j, γ_j and $\bar{\alpha}^j$. It is easy to check that $h(\alpha, \alpha) = f(\alpha)$ for all α. There remains to find values of $m^j, \gamma^j, \bar{\alpha}^j$ such that $h(\alpha, \alpha^j) \geqslant f(\alpha)$ holds for all $\alpha \in (\alpha_-, \alpha_+)$.

3.2. Construction of the majorant function

Let us introduce the following assumption on P.

Assumption 1. *For all $\boldsymbol{x}'$, there exists a symmetric matrix $\boldsymbol{A}(\boldsymbol{x}')$ such that*

$$Q(\boldsymbol{x},\boldsymbol{x}') = P(\boldsymbol{x}') + (\boldsymbol{x}-\boldsymbol{x}')^T\nabla P(\boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^T\boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x}-\boldsymbol{x}') \geqslant P(\boldsymbol{x}) \tag{17}$$

for all $\boldsymbol{x}$. Moreover, for any bounded set $\mathcal{V}$ included in the definition domain of $P(.)$, the set $\{\boldsymbol{A}(\boldsymbol{x})|\boldsymbol{x}\in\mathcal{V}\}$ has a positive bounded spectrum with bounds $(\nu^{\mathcal{A}}_{\min},\nu^{\mathcal{A}}_{\max})$, i.e. for all $\boldsymbol{x}\in\mathcal{V}$,

$$0 < \nu^{\mathcal{A}}_{\min} \leqslant \frac{\boldsymbol{v}^T\boldsymbol{A}(\boldsymbol{x})\boldsymbol{v}}{\|\boldsymbol{v}\|^2} \leqslant \nu^{\mathcal{A}}_{\max},\ \forall \boldsymbol{v}\in\mathbb{R}^N\setminus\{\boldsymbol{0}\} \tag{18}$$

As emphasized in [28, Lem.2.1], Assumption 1 holds if $P(.)$ is gradient Lipschitz with constant L_p by setting $\boldsymbol{A}(\boldsymbol{x}) = L_p\,\boldsymbol{I}_N$ for all $\boldsymbol{x}$, where $\boldsymbol{I}_N$ states for the identity matrix with size $N\times N$. Useful methods for constructing $\boldsymbol{A}(\boldsymbol{x})$ without requiring the knowledge of L_p are developped in [30, 31].

Under Assumption 1, the majorization of (8) is given by the following theorem.

Theorem 1. *Let $F = P + B$ where P fulfills Assumption 1 and B takes the form (6) where ψ_i fulfills (7) for all $i\in\{1,\dots,I\}$. For all $\boldsymbol{x}_k\in\mathcal{C}$ and $\boldsymbol{d}_k\in\mathbb{R}^N$, the log-quadratic function (16) is tangent majorant for (8) at α^j for the following parameters:*

$$\begin{cases}\bar{\alpha}^j = \alpha_- \\ m^j = m^j_p + Z_2(\alpha^j) & \forall\alpha\in(\alpha_-,\alpha^j] \\ \gamma^j = (\alpha_- - \alpha^j)Z_1(\alpha^j)\end{cases} \tag{19}$$

$$\begin{cases}\bar{\alpha}^j = \alpha_+ \\ m^j = m^j_p + Z_1(\alpha^j) & \forall\alpha\in[\alpha^j,\alpha_+) \\ \gamma^j = (\alpha_+ - \alpha^j)Z_2(\alpha^j)\end{cases} \tag{20}$$

where

$$\begin{cases} m^j_p = \boldsymbol{d}_k^T\boldsymbol{A}(\boldsymbol{x}_k+\alpha^j\boldsymbol{d}_k)\boldsymbol{d}_k, \\ Z_1(\alpha^j) = \sum_{i\in\mathcal{I}_-}\zeta_i(\alpha^j), \\ Z_2(\alpha^j) = \sum_{i\in\mathcal{I}_+}\zeta_i(\alpha^j), \end{cases}$$

with $\zeta_i(\alpha) = \delta_i^2\ddot{\psi}_i(\theta_i+\alpha\delta_i)$ for all $i = 1,\dots,I$.

Proof. See Appendix A. □

Remark 1. *If the set $\mathcal{I}_-$ is empty (i.e., $\alpha_- = -\infty$), it is understood that $Z_1(\alpha^j)$ equals zero. Thus, for all $\alpha\in(-\infty,\alpha^j]$, $\gamma^j = 0$ and the tangent majorant function has the following expression*

$$h(\alpha,\alpha^j) = f(\alpha^j) + (\alpha-\alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}(m^j_p + Z_2(\alpha^j))(\alpha-\alpha^j)^2. \tag{21}$$

Respectively, if $\mathcal{I}_+$ is empty (i.e., $\alpha_+ = +\infty$), $Z_2(\alpha^j) = 0$ so that for all $\alpha\in[\alpha^j,+\infty)$,

$$h(\alpha,\alpha^j) = f(\alpha^j) + (\alpha-\alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}(m^j_p + Z_1(\alpha^j))(\alpha-\alpha^j)^2. \tag{22}$$

Although Theorem 1 separately defines $h(\alpha, \alpha^j)$ whether α is in $(\alpha_-, \alpha^j]$ or $[\alpha^j, \alpha+)$ (see Fig. 1 for an illustration), the resulting function is twice differentiable and convex according to the following lemma.

Lemma 1. *Under Assumption 1, $h(., \alpha^j)$ is C^2 and strictly convex in (α_-, α_+).*

Proof. See Appendix B. □

3.3. Minimization of the tangent majorant

According to the MM theory, the stepsize α_k is defined by (15) where $h(\alpha, \alpha^j)$ is the tangent majorant function (16). The MM recurrence (15) involves the computation of the minimizer of $h(\alpha, \alpha^j)$ for $j \in \{0, \ldots, J-1\}$. Thanks to Lemma 1, the tangent majorant $h(., \alpha^j)$ has a unique minimizer, which can be expressed as an explicit function of $\dot{f}(\alpha^j)$ as follows:

$$\alpha^{j+1} = \begin{cases} \alpha^j - \dfrac{2q_3}{q_2 + \sqrt{q_2^2 - 4q_1q_3}} & \text{if} \quad |\bar{\alpha}^j| < \infty \quad \text{and} \quad \dot{f}(\alpha^j) \leqslant 0 \\ \alpha^j - \dfrac{2q_3}{q_2 - \sqrt{q_2^2 - 4q_1q_3}} & \text{if} \quad |\bar{\alpha}^j| < \infty \quad \text{and} \quad \dot{f}(\alpha^j) > 0 \\ \alpha^j - \dfrac{\dot{f}(\alpha^j)}{m^j} & \text{if} \quad |\bar{\alpha}^j| = \infty \end{cases} \tag{23}$$

with

$$\begin{cases} q_1 = & -m^j \\ q_2 = & \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha}^j - \alpha^j) \\ q_3 = & (\bar{\alpha}^j - \alpha^j)\dot{f}(\alpha^j) \end{cases} \tag{24}$$

Finally, (15) produces monotonically decreasing values $\{f(\alpha^j)\}$ and the series $\{\alpha^j\}$ converges to a stationary point of $f(\alpha)$ [32].

3.4. Convergence analysis

This section studies the convergence of the iterative descent algorithm

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad k \geqslant 0, \tag{25}$$

when $\boldsymbol{d}_k$ satisfies $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$ and the stepsize value α_k results from (15). The proposed analysis requires the following assumption on F:

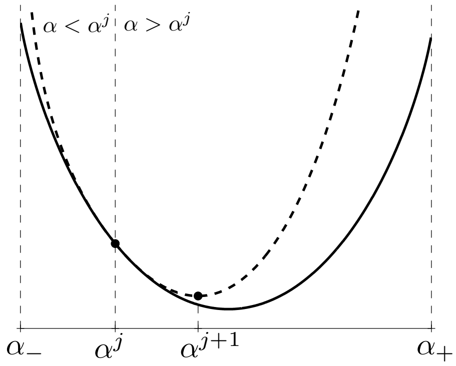
Assumption 2. *For some $\boldsymbol{x}_0 \in \mathcal{C}$, there exists a neighborhood $\mathcal{V}_0$ of the level set $\mathcal{L}_0 = \{\boldsymbol{x} \in \mathbb{R}^N | F(\boldsymbol{x}) \leqslant F(\boldsymbol{x}_0)\}$ such that:*

- *$\mathcal{V}_0$ is bounded;*
- *F is differentiable on $\mathcal{V}_0$ and $\nabla F(\boldsymbol{x})$ is Lipschitz continuous on $\mathcal{V}_0$ with the Lipschitz constant $L > 0$, i.e.,*

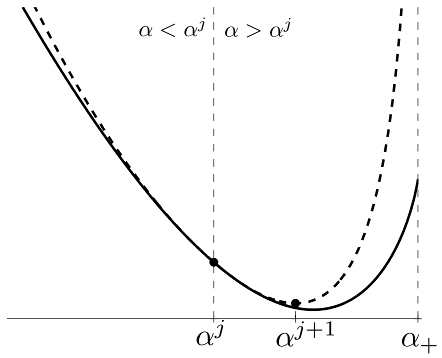
$$\|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\| \leqslant L\|\boldsymbol{x} - \boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}_0. \tag{26}$$

Let us emphasize that Assumption 1 holds for the particular case $\mathcal{V} = \mathcal{V}_0$. Moreover, the boundedness assumption on $\mathcal{V}_0$ holds if F is coercive, that is:

$$\lim_{\|\boldsymbol{x}\| \to +\infty} F(\boldsymbol{x}) = +\infty. \tag{27}$$



(a) Case α_- and α_+ finite



(b) Case $\alpha_- = -\infty$ and α_+ finite

Figure 1. Schematic principle of the MM linesearch procedure. The tangent majorant function $h(\alpha, \alpha^j)$ (dashed line) for $f(\alpha)$ (solid line) at α^j is piecewise defined on the sets $(\alpha_-, \alpha^j]$ and $[\alpha^j, \alpha_+)$. The new iterate α^{j+1} is taken as the minimizer of $h(., \alpha^j)$. Two cases are illustrated. The third and last case where α_- is finite and $\alpha_+ = +\infty$ is the mirror image of case (b).

3.4.1. Properties of the stepsize series. First, let us recall some essential properties of the MM recurrence.

Lemma 2 ([31, 32])**.** *Let $\boldsymbol{x}_k \in \mathcal{V}_0$ and $\boldsymbol{d}_k$ such that $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$. For all $j \geqslant 1$, the series $\{\alpha^j\}$ defined by* (15) *fulfills*

- $f(\alpha^j) \leqslant f(\alpha^{j-1})$
- $\text{sign}(\alpha^j - \alpha^{j-1}) = -\text{sign}(\dot{f}(\alpha^{j-1}))$
- $\alpha^j > 0$

and converges to a stationary point of f.

The first item of Lemma 2 implies that for all k,

$$F(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k),$$

which means that the iterates $\{\boldsymbol{x}_k\}$ remain in $\mathcal{V}_0$. However, this simple monotonicity condition does not imply that the descent algorithm (3) converges [6, Chap.3].

3.4.2. Sufficient decrease condition. In order to ensure that the descent algorithm makes reasonable progress, the stepsize value must yield a *sufficient decrease* in the objective function F, as measured by the first Wolfe condition (10).

Property 1. *Let $\boldsymbol{x}_k \in \mathcal{V}_0$ and $\boldsymbol{d}_k$ satisfying $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$. Under Assumptions 1 and 2, the iterates of* (15) *fulfill*

$$F(\boldsymbol{x}_k + \alpha^j \boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k) + \sigma_1^j \alpha^j \nabla F(\boldsymbol{x}_k)^T \boldsymbol{d}_k, \tag{28}$$

for all $j \geqslant 1$, with $\sigma_1^j = (2\sigma_{\max}^j)^{-1} \in (0,1)$ for some $\sigma_{\max}^j > 0$.

Proof. See Appendix C. □

Property 1 is a strong result since it means that the MM linesearch produces a sufficient decrease of the criterion, whatever the number of linesearch iterates J.

3.4.3. Stepsize minoration. The first Wolfe condition alone is not sufficient to ensure the convergence since it does not prevent arbitrarily small steps. A second condition is required, such as the second Wolfe condition (11). Here, the proposed convergence study rather relies on a direct minoration of the stepsize values.

Property 2. *Let $\boldsymbol{x}_k \in \mathcal{V}_0$ and $\boldsymbol{d}_k$ satisfying $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$. Under Assumptions 1 and 2, for all $j \geqslant 1$, the iterates of* (15) *fulfill*

$$\alpha^j \geqslant \sigma_{\min} \alpha^1 \tag{29}$$

and

$$\alpha^j \geqslant \sigma_{\min} \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\nu \|\boldsymbol{d}_k\|^2} \tag{30}$$

for some $\sigma_{\min}$, $\nu > 0$.

Proof. See Appendix D. □

3.4.4. Zoutendijk condition. Obviously, the global convergence of a descent direction method is not only ensured by a good stepsize strategy, but also by well-chosen search directions $\boldsymbol{d}_k$. Convergence proofs often rely on the fulfillment of Zoutendijk condition

$$\sum_{k=0}^{\infty} \|\boldsymbol{g}_k\|^2 \cos^2 \theta_k < \infty, \tag{31}$$

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

$$\cos \theta_k = \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{g}_k\| \, \|\boldsymbol{d}_k\|}. \tag{32}$$

In the case of the proposed linesearch, Properties 1 and 2 lead to the following result.

Property 3. *Let α_k be defined for all k by* (15) *with $J \geqslant 1$. Under Assumptions 1 and 2, $(\boldsymbol{g}_k, \boldsymbol{d}_k)_{k \geqslant 0}$ fulfills Zoutendijk condition* (31).

Proof. See Appendix E. □

3.4.5. Gradient related directions. Finally, a general convergence result can be established from Property 3 by using the concept of *gradient related* direction [33].

Definition 1. *A direction sequence* $\{\boldsymbol{d}_k\}$ *is said gradient related to* $\{\boldsymbol{x}_k\}$ *if the following property holds: for any subsequence* $\{\boldsymbol{x}_k\}_{k\in\mathcal{K}}$ *that converges to a nonstationary point, the corresponding subsequence* $\{\boldsymbol{d}_k\}_{k\in\mathcal{K}}$ *is bounded and satisfies*

$$\limsup_{k\to\infty, k\in\mathcal{K}} \boldsymbol{g}_k^T \boldsymbol{d}_k < 0. \tag{33}$$

Property 4. *[33] If there exist* $\nu_1 > 0$, $\nu_2 > 0$ *such that for all* k, $\boldsymbol{d}_k$ *fulfills*

$$\nu_1 \|\boldsymbol{g}_k\|^2 \leqslant -\boldsymbol{g}_k^t \boldsymbol{d}_k, \qquad \|\boldsymbol{d}_k\|^2 \leqslant \nu_2 \|\boldsymbol{g}_k\|^2, \tag{34}$$

then $\{\boldsymbol{d}_k\}$ *is gradient related to* $\{\boldsymbol{x}_k\}$.

Theorem 2. *Let* $\{\boldsymbol{x}_k\}$ *a sequence generated by a descent method* $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$. *Assume that the sequence* $\{\boldsymbol{d}_k\}$ *fulfills* (34) *and* α_k *is defined by* (15). *Under Assumptions 1 and 2, the descent algorithm* (25) *converges in the sense* $\lim_{k\to\infty} \|\boldsymbol{g}_k\| = 0$.

Proof. According to Property 3, Zoutendijk condition (31) holds. Theorem 2 results from [34, Theo.5.1]. □

As emphasized in [35, Sec.6.2], Theorem 2 yields convergence of several classical descent optimization schemes such as the steepest descent method, truncated Newton method and the projected gradient method for constrained optimization. Let us remark that it does not cover nonlinear conjugate gradient algorithms (NLCG) defined by the following recurrence

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \\ \boldsymbol{d}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k. \end{aligned} \tag{35}$$

where β_k is the conjugacy parameter. However, a deeper analysis of the MM stepsize properties shows that specific convergence results hold for standard NLCG methods such as Fletcher-Reeves, Polak-Ribière-Polyak and the modified Polak-Ribière-Polyak [35, Sec.6.3].

4. Numerical results

In this section, two application examples are considered to illustrate the practical efficiency of the proposed linesearch procedure. In both cases, a reference descent optimization algorithm is considered, and the MM linesearch is tested against two Wolfe linesearch strategies taken from [5] and [23], respectively based on polynomial and log-quadratic interpolation.

4.1. Image reconstruction under Poisson noise

A simulated positron emission tomography (PET) ([10]) reconstruction problem is first considered. The measurements in PET are modeled as Poisson random variables:

$$\boldsymbol{y} \sim Poisson(\boldsymbol{K}\boldsymbol{x} + \boldsymbol{r}) \tag{36}$$

where the nth entry of $\boldsymbol{x}$ represents the radioisotope amount in pixel n and $\boldsymbol{K}$ is the projection matrix whose elements K_{mn} model the contribution of the nth pixel to the mth datapoint. The components of $\boldsymbol{y}$ are the counts measured by the detector pairs and $\boldsymbol{r}$ models the background events (scattered events and accidental coincidences). The aim is to reconstruct the image $\boldsymbol{x} \geqslant 0$ from the noisy measurements $\boldsymbol{y}$.

4.1.1. Objective function. According to the noise statistics, the neg-log-likelihood of the emission data is

$$S(\boldsymbol{x}) = \sum_{m=1}^{M} \left([\boldsymbol{K}\boldsymbol{x}]_m + r_m - y_m \log([\boldsymbol{K}\boldsymbol{x}]_m + r_m) \right). \tag{37}$$

A useful penalization favoring smoothness of the estimated image is given by

$$R(\boldsymbol{x}) = \sum_{\ell} \omega_\ell \phi([\boldsymbol{D}\boldsymbol{x}]_\ell),$$

where ϕ is the edge preserving potential function $\phi(u) = \sqrt{\delta^2 + u^2} - \delta$ and $\boldsymbol{D}\boldsymbol{x}$ is the vector of difference between neighboring pixel intensities [36]. The weights depend on the relative position of the neighbors: $\omega_\ell = 1$ for vertical and horizontal neighbors and $\omega_\ell = 2^{-\frac{1}{2}}$ for diagonal neighbors. The estimated image is the minimizer of the following objective function

$$F(\boldsymbol{x}) = S(\boldsymbol{x}) + \lambda R(\boldsymbol{x}), \tag{38}$$

over the positive orthant $\{\boldsymbol{x} \geqslant \boldsymbol{0}\}$.

An efficient approach for solving this constrained optimization problem is to use the split-gradient method (SGM) from [37] associated with a convergent linesearch strategy ([12, 38]). The first part of the criterion implies the presence of a logarithmic barrier in $S(.)$ associated to the domain $\boldsymbol{K}\boldsymbol{x} + \boldsymbol{r} > 0$. We propose to analyse the performance of the SGM algorithm when the stepsize is computed using the proposed MM linesearch.

4.1.2. Optimization strategy. The SGM is a descent algorithm aimed at minimizing a criterion under nonnegativity constraints. Assuming that the gradient can be splitted into positive and negative parts $\nabla F(\boldsymbol{x}) = V(\boldsymbol{x}) - U(\boldsymbol{x})$, $U(\boldsymbol{x}), V(\boldsymbol{x}) \geqslant 0$, for all $\boldsymbol{x} \geqslant \boldsymbol{0}$, the SGM iteration is defined as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + s_k \boldsymbol{d}_k, \quad \boldsymbol{d}_k = -\mathrm{Diag}\left(\frac{\boldsymbol{x}_k}{V(\boldsymbol{x}_k)} \right) \boldsymbol{g}_k, \tag{39}$$

where s_k is a stepsize ensuring the positivity of the iterates. When $s_k = 1$ for all k, iteration (39) takes a very simple multiplicative form

$$\boldsymbol{x}_{k+1} = \mathrm{Diag}\left(\frac{\boldsymbol{x}_k}{V(\boldsymbol{x}_k)} \right) U(\boldsymbol{x}_k). \tag{40}$$

However, the unit step size does not guarentee the convergence of the iterates and a linesearch along $\boldsymbol{d}_k$ has to be performed. More precisely, according to [39], the convergence is ensured if

$$s_k = \min(\tau s_{\max}, \alpha_k), \quad s_{\max} = \max\{s | \boldsymbol{x}_k + s\boldsymbol{d}_k \geqslant 0\}, \quad \tau \in (0, 1), \tag{41}$$

as soon as α_k results from a linesearch along $\boldsymbol{d}_k$ that satisfies both (10) and (31) conditions.

Let $F = P + B$ with

$$B(\boldsymbol{x}) = \sum_{m=1}^{M} -y_m \log([\boldsymbol{K}\boldsymbol{x}]_m + r_m),$$
$$P(\boldsymbol{x}) = \sum_{m=1}^{M} [\boldsymbol{K}\boldsymbol{x}]_m + r_m + \lambda \sum_{\ell} \omega_\ell \phi([\boldsymbol{D}\boldsymbol{x}]_\ell).$$

The linear operators $\boldsymbol{K}$ and $\boldsymbol{D}$ are such that $\ker(\boldsymbol{K}^T\boldsymbol{K}) \cap \ker(\boldsymbol{D}^T\boldsymbol{D}) = \{\boldsymbol{0}\}$. Thus, it is straightforward that Assumption 2 holds for all $\boldsymbol{x}_0 > 0$. Moreover, according to [30], Assumption 1 holds for

$$\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{D}^T \mathrm{Diag}(\omega(\boldsymbol{D}\boldsymbol{x}))\boldsymbol{D}, \quad \omega(u) = \frac{1}{u}\dot{\phi}(u).$$

Therefore, according to Properties 1 and 3, the proposed MM linesearch ensures the convergence of the SGM algorithm. We propose to compare its performances with the Moré and Thuente's linesearch [5] (MT) based on the fulfillment of the strong Wolfe conditions (10)-(11). Two interpolation schemes will be considered for the MT linesearch, namely the cubic interpolation procedure (MTcubic) and the Murray and Wright's log-quadratic interpolation procedure (MTlog) [23].

The SGM iteration (39) is employed with the same splitting functionnals $U(.)$ and $V(.)$ as in [38, Eq.20-22]. The algorithm is initialized with a uniform positive object and the convergence is checked using the following stopping rule ([40])

$$\|\nabla_{\mathcal{P}} F(\boldsymbol{x}_k)\|_\infty < 10^{-3} \|\nabla_{\mathcal{P}} F(\boldsymbol{x}_0)\|_\infty, \tag{42}$$

where $\nabla_{\mathcal{P}} F(\boldsymbol{x})$ denotes the projected gradient of $F(.)$ at $\boldsymbol{x}$,

$$\nabla_{\mathcal{P}} F(\boldsymbol{x}) = \max(\boldsymbol{x} - \nabla F(\boldsymbol{x}), \boldsymbol{0}) - \boldsymbol{x}. \tag{43}$$

4.1.3. Results and discussion. We present a simulated example using data generated with J.A. Fessler's code available at `http://www.eecs.umich.edu/~fessler`. For this simulation, we consider an image $\boldsymbol{x}^o$ of size $N = 128 \times 128$ pixels and $M = 30720$ pairs of detectors. 10^5 counts are considered in the Poisson degradation model (36). The regularization parameters (λ, δ) are set to $\lambda = 10^{-1}$, $\delta = 50$ to get the best result in terms of similarity between the simulated and the estimated images (in the sense of quadratic error). The two images are illustrated in Fig. 2.

Tab. 3 summarizes the performance results in terms of iteration number and computation time in seconds on an Intel Core 2 CPU 6700, 3 GHz, 3 GB RAM. The same strategy as in [9, Sec.4.1.2] has been used for the implementation of the three linesearch methods, reducing the gradient computation counts to the descent algorithm outer iteration number. The design parameters are the number of sub-iterations J for the MM procedure and the Wolfe condition constants (σ_1, σ_2) for the MTcubic and MTlog methods. For the two latter methods, we give the mean number J of sub-iterations that are necessary to fulfill the two Wolfe conditions.

It can be noted that the split-gradient algorithm with MM linesearch (SGM-MM) requires about the same number of iterations than the MTcubic or MTlog

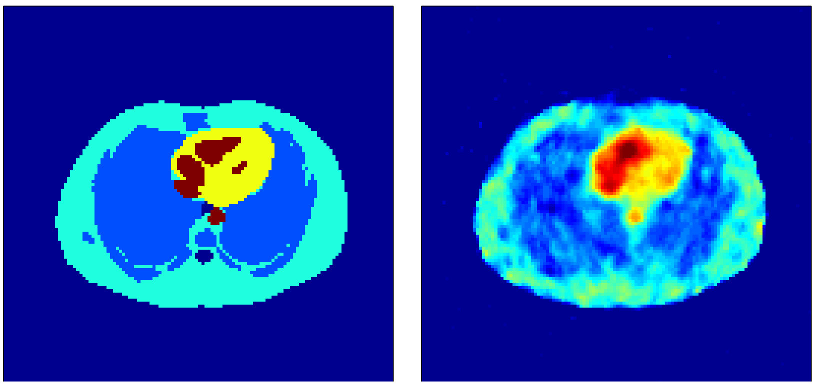
SGM-MTcubic					SGM-MTlog					SGM-MM		
σ_1	σ_2	J	Iter.	Time	σ_1	σ_2	J	Iter.	Time	J	Iter.	Time
10^{-4}	0.5	5.9	353	66	10^{-4}	0.5	5.4	349	60	1	353	**44**
10^{-4}	0.9	4.3	356	56	10^{-4}	0.9	4.9	350	58	2	349	48
10^{-4}	0.99	2.5	389	54	10^{-4}	0.99	3.3	350	54	3	350	54
10^{-3}	0.99	2.5	389	57	10^{-3}	0.99	3.3	350	54	4	350	60
10^{-2}	0.99	2.5	389	57	10^{-2}	0.99	3.3	350	53	5	349	66
10^{-1}	0.99	2.5	389	56	10^{-1}	0.99	3.3	350	53	10	349	94

Table 3. Comparison between MM, MTcubic and MTlog linesearch strategies for a PET reconstruction problem solved using the split-gradient algorithm, in terms of iteration number and time (in s.) before convergence, considered in the sense of (42). As a comparison, the multiplicative split-gradient (i.e., $s_k = 1$, $\forall k \geqslant 0$) requires 788 iterations and 58 s to fulfill the stopping criterion.

approaches (SGM-MTcubic and SGM-MTlog), provided that the parameters (σ_1, σ_2) are appropriately chosen.

The effect of the Wolfe parameters (σ_1, σ_2) differs according to the interpolation strategy. For the cubic linesearch, a decrease of the first Wolfe parameter σ_1 accelerates the convergence rate, but at a price of a larger cost per iteration. On the contrary, it appears that the number of iterations for SGM-MTlog remains stable towards the tuning (σ_1, σ_2), which shows that the use of Murray and Wright's log-quadratic interpolation enhances the performances of the MT linesearch.

In terms of time before convergence, the SGM algorithm performs better when the stepsize is obtained with the proposed MM search, because of smaller computation cost per sub-iteration. The MM strategy admits a unique tuning parameter, namely the sub-iteration number J, and it appears that the simplest choice $J = 1$ leads to the best results. This indicates that the best strategy corresponds to a rough minimization of $f(\alpha)$. Such a conclusion meets that of [28] in the context of quadratic MM linesearch.



(a) Simulated PET phantom (b) Reconstruction with similarity error 6%

Figure 2. Simulated PET reconstruction with split gradient method

4.2. Nuclear magnetic resonance reconstruction

We consider a mono-dimensional nuclear magnetic resonance (NMR) reconstruction problem. The NMR decay $y(t)$ associated with a continuous distribution of relaxation constants $x(T)$ is described in terms of a Fredholm integral of the first kind:

$$y(t) = \int_{T_{\min}}^{T_{\max}} x(T)\, k(t,T)\, dT. \tag{44}$$

with $k(t,T) = exp\left\{-\frac{t}{T}\right\}$. In practice, the measured signal $\boldsymbol{y}$ is a set of discrete experimental noisy data points $y_m = y(t_m)$ modeled as

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\epsilon} \tag{45}$$

where $\boldsymbol{K}$ and $\boldsymbol{x}$ are discretized versions of $k(t,T)$ and $x(T)$ with dimensions $M \times N$ and $N \times 1$, and $\boldsymbol{\epsilon}$ is an additive noise assumed centered white Gaussian. Given $\boldsymbol{y}$, the aim is to determine $\boldsymbol{x} \geqslant 0$. This problem is equivalent to a numerical inversion of the Fredholm integral (44) and is known as very ill-conditioned ([41]).

4.2.1. Objective function. In order to get a stabilized solution, an often used method minimizes the expression

$$F(\boldsymbol{x}) = S(\boldsymbol{x}) + \lambda R(\boldsymbol{x}) \tag{46}$$

under positivity constraints, where $S(.)$ is a fidelity to data term:

$$S(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{y}\|_2^2, \tag{47}$$

and $R(.)$ is an entropic regularization term, e.g., the Shannon entropy measure:

$$R(\boldsymbol{x}) = \sum_{n=1}^{N} x_n \log x_n \tag{48}$$

Moreover, the positivity constraint is implicitely handled because of the barrier property of the entropy function.

4.2.2. Optimization strategy. The truncated Newton (TN) algorithm is employed for minimizing (46). The direction $\boldsymbol{d}_k$ is computed by approximately solving the Newton system $\nabla^2 F(\boldsymbol{x}_k)\boldsymbol{d} = -\boldsymbol{g}_k$ using preconditioned conjugate gradient (PCG) iterations. We propose a preconditioning matrix $\boldsymbol{M}_k$ built as an approximation of the inverse Hessian of $F(.)$ at $\boldsymbol{x}_k$:

$$\boldsymbol{M}_k = \left(\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T + \lambda \mathrm{Diag}(\boldsymbol{x}_k)^{-1}\right)^{-1}, \tag{49}$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the truncated singular value decomposition of $\boldsymbol{K}$ with rank equal to 5, and $\boldsymbol{D} = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$. Direction $\boldsymbol{d}_k$ being gradient related, the convergence of the TN algorithm with the proposed linesearch is established in Theorem 2 under Assumptions 1 and 2, by defining $L \equiv P$ and $R \equiv B$. The verification of Assumption 1 is straightforward for $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{K}^T\boldsymbol{K}$. The fulfillment of Assumption 2 is more difficult to check since the level set $\mathcal{L}_0$ may contain an element $\boldsymbol{x}$ with zero components, contradicting the gradient Lipschitz assumption. In practice, we initialized the algorithm with $\boldsymbol{x}_0 > 0$ and we never noticed convergence issues in our practical tests. The extension of the convergence results under a weakened version of Assumption 2 remains an open issue in our convergence analysis.

The algorithm is initialized with a uniform positive object and, following [6], the convergence is checked using:

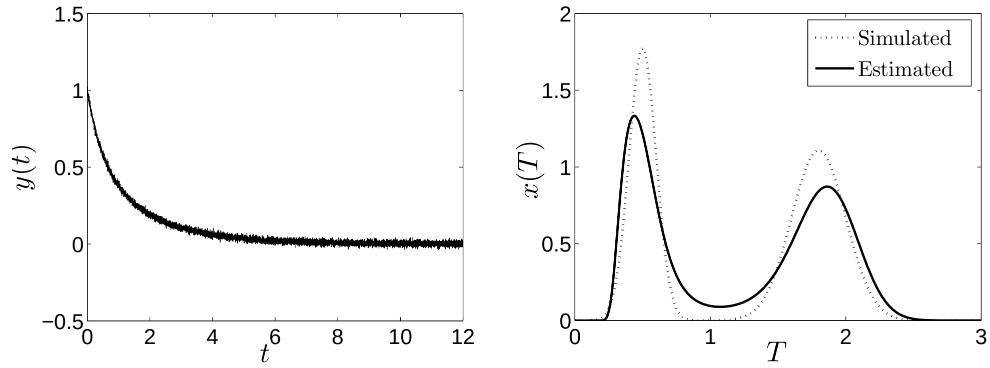
$$\|\nabla F(\boldsymbol{x}_k)\|_\infty \leqslant 10^{-9}(1 + |F(\boldsymbol{x}_k)|). \tag{50}$$

Following [42], the PCG iterations are stopped when:

$$\|\nabla F(\boldsymbol{x}_k) + \nabla^2 F(\boldsymbol{x}_k)\boldsymbol{d}_k\| \leqslant 10^{-5}\|F(\boldsymbol{x}_k)\|. \tag{51}$$

We propose to compare the performances of the MM linesearch with both MTcubic and MTlog strategies.

4.2.3. Results and discussion. Let $\boldsymbol{x}(T)$ a distribution to estimate. We consider the resolution of (45) when data $\boldsymbol{y}$ are simulated from $\boldsymbol{x}(T)$ via the NMR model (45) over $M = 10000$ sampled times t_m, with a SNR of 25 dB (Fig. 3). The regularization parameter λ is set to $\lambda = 7.2 \cdot 10^{-4}$ to get the best result in terms of similarity between the simulated and the estimated spectra (in the sense of quadratic error). Tab. 4 summarizes the performance results in terms of iteration number and computation time in seconds.



(a) Simulated NMR measurement (SNR: 25 dB)

(b) NMR reconstruction (similarity error: 8.5%)

Figure 3. Simulated NMR reconstruction with maximum entropy method

According to Table 4, the TN algorithm with the MM linesearch performs better than with Wolfe-based strategies with their best settings for σ_1 and σ_2. Concerning the choice of the sub-iteration number, it appears that $J = 1$ leads again to the best results in terms of computation time.

TN-MTcubic					TN-MTlog					TN-MM		
σ_1	σ_2	J	Iter.	Time	σ_1	σ_2	J	Iter.	Time	J	Iter.	Time
10^{-4}	0.5	1.9	35	8	10^{-4}	0.5	2.6	35	10	1	36	**6**
10^{-4}	0.9	1.4	41	10	10^{-4}	0.9	2	35	9	2	40	7
10^{-4}	0.99	1	70	15	10^{-4}	0.99	2	35	9	3	40	7
10^{-3}	0.99	1	70	15	10^{-3}	0.99	2	35	9	4	40	7
10^{-2}	0.99	1	70	15	10^{-2}	0.99	2	35	9	5	40	7
10^{-1}	0.99	1	70	15	10^{-1}	0.99	2	35	9	10	40	8

Table 4. Comparison between MM, MTcubic and MTlog linesearch strategies for a maximum entropy NMR reconstruction problem solved with TN algorithm, in terms of iteration number and time (in s.) before convergence. Convergence is considered in the sense of (50).

5. Conclusion

This paper extends the linesearch strategy of [28] to the case of criteria containing barrier functions, by proposing a non-quadratic majorant approximation of the criterion in the linesearch direction. The proposed majorant has the same form as the interpolating function proposed in [23]. However, in the majorization approach, the construction of the approximation is easier and its minimization leads to a closed-form stepsize formula, which guarantees the convergence of several descent algorithms. Numerical experiments indicate that this linesearch strategy outperforms interpolating-based linesearch methods.

Two extensions of this work are envisaged. On the one hand, the analysis could be extended to additional forms of barrier functions, such as barriers for nonlinear constraints ([43]), roughness penalties ([22]) or inverse function ([44]). For the latter, the main difficulty will come from the fact that the inverse barrier grows faster than a logarithmic barrier near zero. Therefore, the proposed log-quadratic majorization will not be suited, and another form of majorant function should probably be envisaged.

On the other hand, the applicability of the proposed procedure for nonnegative matrix factorization (NMF) could be studied. NMF is usually based on the minimization of a Bregman divergence between two unknown matrices [45]. Particular Bregman divergences such as Kullback-Leibler and Itakuro-Saito contain barrier terms that fall within the scope of the present study. It would be interesting to analyse the performances of NMF iterative algorithms when the proposed linesearch is incorporated into the update schemes.

Appendix A. Proof of Theorem 1

Appendix A.1. Majorizing property

Let $\boldsymbol{x}_k \in \mathcal{C}$, $\boldsymbol{d}_k \in \mathbb{R}^N$ a search direction and $\alpha^j \in (\alpha_-, \alpha_+)$. Let us show that $h(\alpha, \alpha^j)$ whose parameters $(m^j, \gamma^j, \bar{\alpha}^j)$ are given by (19) and (20) is a tangent majorant for $F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) = f(\alpha)$ at α^j, over (α_-, α_+).

First, according to Assumption 1, $q(\alpha, \alpha^j) = p(\alpha^j) + (\alpha - \alpha^j)\dot{p}(\alpha^j) + \frac{1}{2} m_p^j (\alpha - \alpha^j)^2$ is a tangent majorant for $p(\alpha) = P(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$ at α^j for all $\alpha \in \mathbb{R}$. There remains to show that

$$\phi(\alpha, \alpha^j) = b(\alpha^j) + (\alpha - \alpha^j)\dot{b}(\alpha^j) + \frac{1}{2} m_b^j (\alpha - \alpha^j)^2 + \gamma^j \left[(\bar{\alpha}^j - \alpha^j) \log \frac{\bar{\alpha}^j - \alpha^j}{\bar{\alpha}^j - \alpha} + \alpha^j - \alpha \right], \tag{A.1}$$

with $m_b^j = m^j - m_p^j$ is a tangent majorant for $b(\alpha) = B(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$ at α^j. Let us define

$$b_1(\alpha) = \sum_{i \in \mathcal{I}_-} \psi_i(\theta_i + \alpha \delta_i), \qquad b_2(\alpha) = \sum_{i \in \mathcal{I}_+} \psi_i(\theta_i + \alpha \delta_i), \tag{A.2}$$

so that $b(\alpha) = b_1(\alpha) + b_2(\alpha) + b_0$ where b_0 is constant with respect to α. First, we will prove that

$$\begin{cases} \phi_1(\alpha, \alpha^j) = b_1(\alpha^j) + (\alpha - \alpha^j)\dot{b}_1(\alpha^j) + \frac{1}{2} m_b^j (\alpha - \alpha^j)^2 \\ \phi_2(\alpha, \alpha^j) = b_2(\alpha^j) + (\alpha - \alpha^j)\dot{b}_2(\alpha^j) + \gamma^j \left[(\alpha_+ - \alpha^j) \log \frac{\alpha_+ - \alpha^j}{\alpha_+ - \alpha} + \alpha^j - \alpha \right], \end{cases}$$

respectively majorize b_1 and b_2 for all $\alpha \in [\alpha^j, \alpha_+)$.

Let us assume that $\mathcal{I}_-$ is not empty. Then, according to the expression of b_1, $Z_1(\alpha) = \ddot{b}_1(\alpha)$ so $m_b^j = \ddot{b}_1(\alpha^j)$. The strict convexity of functions ψ_i, for all $i \in \{1, \ldots, I\}$, implies that b_1 is strictly convex and $\dot{b}_1$ is strictly concave. Then, for all $\alpha \in [\alpha_j, \alpha^+)$, $\ddot{b}_1(\alpha) \leqslant \ddot{b}_1(\alpha^j) = m_b^j$. Hence, $\phi_1(., \alpha^j)$ majorizes b_1 on $[\alpha_j, \alpha^+)$. If $\mathcal{I}_-$ is empty, both $b_1(.)$ and $\phi_1(., \alpha^j)$ equal zero so the latter majorizing property still holds.

Let us assume that $\mathcal{I}_+$ is not empty. The expression of b_2 leads to $Z_2(\alpha) = \ddot{b}_2(\alpha)$ so $\gamma^j = (\alpha_+ - \alpha^j)\ddot{b}_2(\alpha^j)$. Let us define $T(\alpha) = \dot{b}_2(\alpha)(\alpha_+ - \alpha)$ and $l(\alpha) = \dot{b}_2(\alpha^j)(\alpha_+ - \alpha) + \gamma^j(\alpha - \alpha^j)$. Given $\gamma^j = (\alpha_+ - \alpha^j)\ddot{b}_2(\alpha^j)$, the linear function l also reads:

$$l(\alpha) = \dot{\phi}_2(\alpha, \alpha^j)(\alpha_+ - \alpha). \tag{A.3}$$

Thus we have $l(\alpha^j) = T(\alpha^j)$ and $\dot{l}(\alpha^j) = \dot{T}(\alpha^j)$. Moreover:

$$\ddot{T}(\alpha) = \dddot{b}_2(\alpha)(\alpha_+ - \alpha) - 2\ddot{b}_2(\alpha) = \sum_{i \in \mathcal{I}_+} \delta_i^3 \dddot{\psi}_i(\theta_i + \alpha\delta_i)(\alpha_+ - \alpha) - 2\delta_i^2 \ddot{\psi}_i(\theta_i + \alpha\delta_i). \tag{A.4}$$

According to the definition of α_+:

$$\alpha_+ - \alpha < -\frac{\theta_i + \alpha\delta_i}{\delta_i}, \; \forall i \in \mathcal{I}_+. \tag{A.5}$$

According to (7), the third derivative of ψ_i is negative, so

$$\ddot{T}(\alpha) < \sum_{i \in \mathcal{I}_+} \delta_i^2 \left[-\dddot{\psi}_i(\theta_i + \alpha\delta_i)(\theta_i + \alpha\delta_i) - 2\ddot{\psi}_i(\theta_i + \alpha\delta_i) \right] < 0, \tag{A.6}$$

where the last inequality is a consequence of (7). Thus T is concave. Since l is a linear function tangent to T, we have

$$l(\alpha) \geqslant T(\alpha), \forall \alpha \in [\alpha_j, \alpha^+). \tag{A.7}$$

Given $\alpha_+ > \alpha$, (A.7) also reads:

$$\dot{\phi}_2(\alpha, \alpha^j) \geqslant \dot{b}_2(\alpha), \forall \alpha \in [\alpha_j, \alpha^+), \tag{A.8}$$

so $\phi_2(., \alpha^j)$ majorizes b_2 over $[\alpha_j, \alpha^+)$. This property still holds if $\mathcal{I}_+$ is empty, since both $b_2(.)$ and $\phi_2(., \alpha^j)$ equal zero in that case. Finally, $\phi(., \alpha^j) = \phi_1(., \alpha^j) + \phi_2(., \alpha^j)$ majorizes b for $\alpha \geqslant \alpha_j$. The same elements of proof apply to the case $\alpha \leqslant \alpha^j$. We can thus conclude that $h(\alpha, \alpha^j) = q(\alpha, \alpha^j) + \phi(\alpha, \alpha^j)$ is a tangent majorant for f at α^j.

Appendix B. Proof of Lemma 1

First, $h(., \alpha^j)$ is C^∞ over (α_-, α^j) and (α^j, α_+). Moreover, it is easy to check that h and its first two derivatives are continuous at α^j according to (19)-(20). Then, $h(., \alpha^j)$ is C^2 over (α_-, α_+). On the other hand, (19)-(20) imply, for all $\alpha \in (\alpha_-, \alpha^j]$,

$$\ddot{h}(\alpha, \alpha^j) = \begin{cases} m_p^j + Z_2(\alpha^j) + Z_1(\alpha^j)\frac{(\alpha_- - \alpha^j)^2}{(\alpha_- - \alpha)^2} & \text{if} \quad \mathcal{I}_- \neq \emptyset \\ m_p^j + Z_2(\alpha^j) & \text{otherwise} \end{cases} \tag{B.1}$$

and for all $\alpha \in [\alpha^j, \alpha_+)$,

$$\ddot{h}(\alpha, \alpha^j) = \begin{cases} m_p^j + Z_1(\alpha^j) + Z_2(\alpha^j)\frac{(\alpha_+ - \alpha^j)^2}{(\alpha_+ - \alpha)^2} & \text{if} \quad \mathcal{I}_+ \neq \emptyset \\ m_p^j + Z_1(\alpha^j) & \text{otherwise} \end{cases} \tag{B.2}$$

$Z_1(\cdot)$ and $Z_2(\cdot)$ are positive since ψ_i is strictly convex for all $i \in \{1, \dots, I\}$. Moreover, $m_p^j > 0$ according to Assumption 1. Thus, $h(., \alpha^j)$ is strictly convex.

Appendix C. Proof of Property 1

Let us consider $\boldsymbol{x} \in \mathcal{V}_0$ and $\boldsymbol{d}$ a descent direction. First, we establish some preliminary results arising from the expression of the majorizing function. Then, some lower and upper bounds for the stepsize values are derived. Finally, Property 1 is proved.

Appendix C.1. Preliminary results

Lemma 3. *Let $j \in \{0, \dots, J-1\}$. If $\dot{f}(\alpha^j) \leqslant 0$ and $|\bar{\alpha}^j| < \infty$, then α^{j+1} fulfills:*

$$-\frac{q_3}{q_2} \leqslant \alpha^{j+1} - \alpha^j \leqslant -\frac{2q_3}{q_2}. \tag{C.1}$$

where q_1, q_2 and q_3 are given by (24).

Proof. If $\dot{f}(\alpha^j) \leqslant 0$ and $|\bar{\alpha}^j| < \infty$, then α^{j+1} reads

$$\alpha^{j+1} = \alpha^j - \frac{2q_3}{q_2 + \sqrt{q_2^2 - 4q_1q_3}}, \tag{C.2}$$

with

$$\begin{cases} q_1 = -m^j, \\ q_2 = \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha}^j - \alpha^j), \\ q_3 = (\bar{\alpha}^j - \alpha^j)\dot{f}(\alpha^j), \end{cases} \tag{C.3}$$

where parameters $(m^j, \gamma^j, \bar{\alpha}^j)$ are given by

$$\begin{cases} \bar{\alpha}^j = \alpha_+, \\ m^j = m_p^j + \sum_{i\in\mathcal{I}_-} \phi_i(\alpha^j), \\ \gamma^j = (\alpha_+ - \alpha^j) \sum_{i\in\mathcal{I}_+} \phi_i(\alpha^j), \end{cases} \tag{C.4}$$

with $\phi_i(\alpha) = \delta_i^2 \ddot{\psi}_i(\theta_i + \alpha\delta_i)$, a positive function. Therefore, we have $q_1 < 0$, $q_3 < 0$ and $q_2 > 0$, which yield (C.1). □

Lemma 4. *Let* $j \in \{0, \dots, J-1\}$. *If* $\dot{f}(\alpha^j) \leqslant 0$, *then:*

$$f(\alpha^j) - f(\alpha^{j+1}) + \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \geqslant 0. \tag{C.5}$$

Proof. The property is trivial if $\dot{f}(\alpha^j) = 0$. Let us assume that $\dot{f}(\alpha^j) < 0$ so that $\alpha_+ > \alpha^{j+1} > \alpha^j$. Let us define

$$\tau(\alpha) = h(\alpha, \alpha^j) - \Big(f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j)\Big). \tag{C.6}$$

If $\mathcal{I}_+$ is not empty, $\tau(\alpha) = Q(\alpha) + \gamma^j(\alpha_+ - \alpha^j)\varphi(\alpha)$ with $Q(\alpha) = \frac{1}{2}m^j(\alpha - \alpha^j)^2$ and $\varphi(\alpha) = \xi\left(\frac{\alpha-\alpha^j}{\alpha_+-\alpha^j}\right)$, where $\xi(u) = -\log(1-u) - u$ for all $u \in (0, 1)$. A straightforward analysis shows that

$$\frac{\xi(u)}{u\dot{\xi}(u)} \leqslant \frac{1}{2}, \quad \forall u \in (0, 1). \tag{C.7}$$

Taking $u = \frac{\alpha-\alpha^j}{\alpha_+-\alpha^j}$ in (C.7) leads to

$$\frac{\varphi(\alpha)}{(\alpha - \alpha^j)\dot{\varphi}(\alpha)} \leqslant \frac{1}{2}, \forall \alpha \in (\alpha^j, \alpha_+). \tag{C.8}$$

Furthermore, according to the expression of $Q(\alpha)$, we have

$$Q(\alpha) = \frac{1}{2}(\alpha - \alpha^j)\dot{Q}(\alpha). \tag{C.9}$$

Thus, using (C.8) and (C.9),

$$\frac{\tau(\alpha)}{(\alpha - \alpha^j)\dot{\tau}(\alpha)} \leqslant \frac{1}{2}, \forall \alpha \in (\alpha^j, \alpha_+). \tag{C.10}$$

If $\mathcal{I}_+$ is empty, $\tau(\alpha) = Q(\alpha)$ so (C.10) still holds, according to (C.9). $h(., \alpha^j)$ is a tangent majorant for f so

$$h(\alpha, \alpha^j) - f(\alpha) = f(\alpha^j) - f(\alpha) + (\alpha - \alpha^j)\dot{f}(\alpha^j) + \tau(\alpha) \geqslant 0. \tag{C.11}$$

Taking $\alpha = \alpha^{j+1} > \alpha^j$ in (C.10) and (C.11), we obtain

$$f(\alpha^j) - f(\alpha^{j+1}) + (\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{\tau}(\alpha^{j+1}) \geqslant 0. \tag{C.12}$$

Finally, the result holds since $\dot{\tau}(\alpha^{j+1}) = \dot{h}(\alpha^{j+1}, \alpha^j) - \dot{f}(\alpha^j) = -\dot{f}(\alpha^j)$. □

Lemma 5. *Let* $j \in \{0, \ldots, J-1\}$. *Under Assumptions 1 and 2, there exists* $\nu_{\min}$, $\nu_{\max}$, $0 < \nu_{\min} \leqslant \nu_{\max}$, *such that for all* $\boldsymbol{x} \in \mathcal{V}_0$ *and for all descent direction* $\boldsymbol{d}$ *at* $\boldsymbol{x}$*:*

$$\nu_{\min}\|\boldsymbol{d}\|^2 \leqslant \ddot{h}(\alpha^j, \alpha^j) \leqslant \nu_{\max}\|\boldsymbol{d}\|^2, \; \forall j \geqslant 0. \tag{C.13}$$

Proof. Let us first remark that, according to Assumption 2, there exists $\eta > 0$ such that

$$\|\nabla F(\boldsymbol{x})\| \leqslant \eta, \; \forall \boldsymbol{x} \in \mathcal{V}_0. \tag{C.14}$$

Moreover, because the gradient of B is unbounded at the boundary of $\mathcal{C}$, (C.14) leads to the existence of $C_{\min} > 0$ such that

$$C_i(\boldsymbol{x}) \geqslant C_{\min}, \; \forall \boldsymbol{x} \in \mathcal{V}_0, \; \forall i \in \{1, \ldots, I\}\,,, \tag{C.15}$$

and the boundedness assumption on $\mathcal{V}_0$ implies that there exists $C_{\max} > 0$ such that

$$C_i(\boldsymbol{x}) \leqslant C_{\max}, \; \forall \boldsymbol{x} \in \mathcal{V}_0, \; \forall i \in \{1, \ldots, I\}\,.. \tag{C.16}$$

According to Lemma 1,

$$\ddot{h}(\alpha^j, \alpha^j) = m_p^j + \ddot{b}(\alpha^j), \tag{C.17}$$

where $b(\alpha) = B(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$, so that

$$\ddot{b}(\alpha^j) = \boldsymbol{d}^T \nabla^2 B(\boldsymbol{x} + \alpha^j \boldsymbol{d})\boldsymbol{d}. \tag{C.18}$$

On the other hand,

$$m_p^j = \boldsymbol{d}^T \boldsymbol{A}(\boldsymbol{x} + \alpha^j \boldsymbol{d})\boldsymbol{d}. \tag{C.19}$$

Since $\boldsymbol{x} + \alpha^j \boldsymbol{d} \in \mathcal{V}_0$, it is sufficient to show that the set $\left\{\boldsymbol{A}(\boldsymbol{x}) + \nabla^2 B(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{V}_0\right\}$ has a positive bounded spectrum.

$$\nabla^2 B(\boldsymbol{x}) = \boldsymbol{C}^T \mathrm{Diag}(\ddot{\psi}_i(C_i(\boldsymbol{x})))\boldsymbol{C}, \tag{C.20}$$

with $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_I]^T$. According to (7), for all $i \in \{1, \ldots, I\}$, $\ddot{\psi}_i$ is decreasing on $\mathbb{R}^+$. Therefore, (C.16) and (C.15) yield

$$\boldsymbol{d}^T \boldsymbol{H}(C_{\max})\boldsymbol{d} \leqslant \boldsymbol{d}^T \nabla^2 B(\boldsymbol{x})\boldsymbol{d} \leqslant \boldsymbol{d}^T \boldsymbol{H}(C_{\min})\boldsymbol{d}, \quad \forall \boldsymbol{x} \in \mathcal{V}_0, \tag{C.21}$$

with $\boldsymbol{H}(c) = \boldsymbol{C}^T \mathrm{Diag}(\ddot{\psi}_i(c))\boldsymbol{C}$. Since $\ddot{\psi}_i$ is strictly convex, matrix $\boldsymbol{H}(c)$ is symmetric and has a nonnegative bounded spectrum with bounds $(\nu_{\min}^{\mathcal{H}}(c), \nu_{\max}^{\mathcal{H}}(c))$. Moreover, according to Assumption 1, $\boldsymbol{A}(\boldsymbol{x})$ has a positive bounded spectrum with bounds $(\nu_{\min}^{\mathcal{A}}, \nu_{\max}^{\mathcal{A}})$ on $\mathcal{V}_0$. Thus, Lemma 5 holds with $\nu_{\min} = \nu_{\min}^{\mathcal{A}} + \nu_{\min}^{\mathcal{H}}(C_{\max}) > 0$ and $\nu_{\max} = \nu_{\max}^{\mathcal{A}} + \nu_{\max}^{\mathcal{H}}(C_{\min})$. □

Appendix C.2. Upper and lower bound for the stepsize series

Lemma 6. *Under Assumptions 1 and 2, there exist $\nu, \nu' > 0$ such that for all $\boldsymbol{x} \in \mathcal{V}_0$ and for all descent direction $\boldsymbol{d}$ at $\boldsymbol{x}$,*

$$\frac{-\boldsymbol{g}^T\boldsymbol{d}}{\nu\,\|\boldsymbol{d}\|^2} \leqslant \alpha^1 \leqslant \frac{-\boldsymbol{g}^T\boldsymbol{d}}{\nu'\,\|\boldsymbol{d}\|^2}, \tag{C.22}$$

where $\boldsymbol{g}$ denotes the gradient of $F(.)$ at $\boldsymbol{x}$.

Proof. $\boldsymbol{d}$ is a descent direction, so $\dot{f}(0) < 0$ and $h(.,0)$ has a barrier at $\bar{\alpha}^0 = \alpha_+$.

If $\alpha_+ = +\infty$ then $h(.,0)$ is a quadratic function with curvature m^0. This majorant is minimized at $\alpha^1 = \frac{-\dot{f}(0)}{m^0}$ and according to Lemma 5, we have:

$$\frac{-\boldsymbol{g}^T\boldsymbol{d}}{\nu_{\max}\|\boldsymbol{d}\|^2} \leqslant \alpha^1 \leqslant \frac{-\boldsymbol{g}^T\boldsymbol{d}}{\nu_{\min}\|\boldsymbol{d}\|^2}. \tag{C.23}$$

If $\alpha_+ < +\infty$, according to Lemma 3:

$$\frac{-\boldsymbol{g}^T\boldsymbol{d}}{\frac{\gamma^0}{\alpha_+} - \frac{\boldsymbol{g}^T\boldsymbol{d}}{\alpha_+} + m^0} \leqslant \alpha^1 \leqslant \frac{-2\boldsymbol{g}^T\boldsymbol{d}}{\frac{\gamma^0}{\alpha_+} - \frac{\boldsymbol{g}^T\boldsymbol{d}}{\alpha_+} + m^0}. \tag{C.24}$$

Using Lemma 5 and the positivity of $-\boldsymbol{g}^T\boldsymbol{d}$, we obtain

$$\nu_{\min}\|\boldsymbol{d}\|^2 \leqslant \frac{\gamma^0}{\alpha_+} - \frac{\boldsymbol{g}^T\boldsymbol{d}}{\alpha_+} + m^0. \tag{C.25}$$

On the other hand, taking $\iota = \arg\max_{i\in\{1,\ldots,I\}} -\boldsymbol{c}_i^T\boldsymbol{d}$, we deduce from (C.15) that

$$\alpha^+ \geqslant \frac{C_{\min}}{|\boldsymbol{c}_\iota^T\boldsymbol{d}|}. \tag{C.26}$$

Thus, using Cauchy-Schwartz inequality and (C.14),

$$\frac{-\boldsymbol{g}^T\boldsymbol{d}}{\alpha_+} = \frac{|\boldsymbol{g}^T\boldsymbol{d}|}{\alpha_+} \leqslant |\boldsymbol{g}^T\boldsymbol{d}|\,|\boldsymbol{c}_\iota^T\boldsymbol{d}|\frac{1}{C_{\min}} \leqslant \|\boldsymbol{g}\|\,\|\boldsymbol{c}_\iota\|\,\|\boldsymbol{d}\|^2\frac{1}{C_{\min}} \leqslant \frac{\eta\bar{C}}{C_{\min}}\|\boldsymbol{d}\|^2, \tag{C.27}$$

with $\bar{C} = \max_{i\in\{1,\ldots,I\}} \|\boldsymbol{c}_i\| > 0$. Moreover, there exists $\nu_{\max}$ such that

$$m^0 + \frac{\gamma^0}{\alpha_+} \leqslant \nu_{\max}\|\boldsymbol{d}\|^2, \tag{C.28}$$

according to Lemma 5. Therefore, (C.25)-(C.28) allow to check that Lemma 6 holds for $\nu = \nu_{\max} + \frac{\eta\bar{C}}{C_{\min}}$ and $\nu' = \frac{\nu_{\min}}{2}$. □

Property 5. *Under Assumptions 1 and 2, for all $j \in \{1,\ldots,J\}$,*

$$\alpha^j \leqslant \sigma_{\max}^j \alpha^1, \tag{C.29}$$

where

$$\sigma_{\max}^j = \left(1 + \frac{2\nu_{\max}L}{\nu_{\min}^2}\right)^{j-1}\left(1 + \frac{\nu}{L}\right) - \frac{\nu}{L} \geqslant 1. \tag{C.30}$$

Proof. It is easy to check (C.29) for $j = 1$, with $\sigma^1_{\max} = 1$. Let us prove that (C.29) holds for $j > 1$. Assume that $\dot{f}(\alpha^j) < 0$. Then $\bar{\alpha}^j = \alpha_+$. Let us denote

$$\hat{m}^j = \begin{cases} m^j + \frac{\gamma^j}{\alpha_+ - \alpha^j} & \text{if} \quad \mathcal{I}_+ \neq \emptyset, \\ m^j & \text{otherwise.} \end{cases} \tag{C.31}$$

If $\mathcal{I}_+$ is not empty, i.e. $\alpha_+ < +\infty$, we deduce from Lemma 3 that

$$\alpha^{j+1} - \alpha^j \leqslant \frac{-2\dot{f}(\alpha^j)}{\frac{\gamma^j - \dot{f}(\alpha^j)}{\alpha_+ - \alpha^j} + m^j}. \tag{C.32}$$

Since $\dot{f}(\alpha^j)$ is negative,

$$\alpha^{j+1} - \alpha^j \leqslant \frac{-2\dot{f}(\alpha^j)}{\hat{m}^j}. \tag{C.33}$$

If $\mathcal{I}_+$ is empty, then $\alpha^{j+1} - \alpha^j = \frac{-\dot{f}(\alpha^j)}{m^j}$, so (C.33) also holds. According to Lemma 5:

$$\|\boldsymbol{d}\|^2 \geqslant \frac{\hat{m}^0}{\nu_{\max}}, \tag{C.34}$$

and

$$\hat{m}^j \geqslant \nu_{\min} \|\boldsymbol{d}\|^2, \tag{C.35}$$

thus we have

$$\hat{m}^j \geqslant \left(\hat{m}^0\right) \frac{\nu_{\min}}{\nu_{\max}} > 0. \tag{C.36}$$

Then, from (C.33):

$$\alpha^{j+1} \leqslant \alpha^j + \frac{2|\dot{f}(\alpha^j)|}{\hat{m}^0} \frac{\nu_{\max}}{\nu_{\min}}. \tag{C.37}$$

If $\dot{f}(\alpha^j) \geqslant 0$, α^{j+1} is lower than α^j so (C.37) still holds. According to Assumption 2, ∇F is Lipschitz, so that $|\dot{f}(\alpha^j) - \dot{f}(0)| \leqslant L\|\boldsymbol{d}\|^2 \alpha^j$. Using the fact that $|\dot{f}(\alpha^j)| \leqslant |\dot{f}(\alpha^j) - \dot{f}(0)| + |\dot{f}(0)|$, and $\dot{f}(0) < 0$, we get:

$$|\dot{f}(\alpha^j)| \leqslant L\alpha^j \|\boldsymbol{d}\|^2 - \dot{f}(0). \tag{C.38}$$

Using Lemma 6 and (C.34):

$$-\dot{f}(0) \leqslant \alpha^1 \nu \|\boldsymbol{d}\|^2 \leqslant \alpha^1 \frac{\nu}{\nu_{\min}} \hat{m}^0. \tag{C.39}$$

Given (C.34)- (C.39), we get:

$$\alpha^{j+1} \leqslant \alpha^j + \frac{2\nu_{\max}}{\nu_{\min}} \frac{1}{\hat{m}^0} \left[L\alpha^j \left(\frac{\hat{m}^0}{\nu_{\min}} \right) + \alpha^1 \frac{\nu}{\nu_{\min}} \hat{m}^0 \right]. \tag{C.40}$$

Hence

$$\alpha^{j+1} \leqslant \alpha^j \left(1 + \frac{2\nu_{\max} L}{\nu^2_{\min}} \right) + 2\alpha^1 \frac{\nu_{\max}\nu}{\nu^2_{\min}}. \tag{C.41}$$

This corresponds to a recursive definition of the series $(\sigma^j_{\max})$ with:

$$\sigma^{j+1}_{\max} = \sigma^j_{\max} \left(1 + 2\frac{\nu_{\max} L}{\nu^2_{\min}} \right) + 2\frac{\nu\nu_{\max}}{\nu^2_{\min}}. \tag{C.42}$$

Given $\sigma^1_{\max} = 1$, (C.30) is the general term of the series. □

Appendix C.3. First Wolfe condition

First, for $j = 1$, the first Wolfe condition (28) holds according to Lemma 4, since it identifies with (C.5) when $j = 0$, given $\sigma^1_{\max} = 1$. For all $j > 1$, (28) holds by immediate recurrence, given Property 5, hence the result.

Appendix D. Proof of Property 2

First, let us show that (29) holds for all $j \geqslant 1$ with

$$\sigma_{\min} = \frac{\sqrt{1 + 2\frac{L}{\nu_{\min}}} - 1}{2\frac{L}{\nu_{\min}}} \in \left(0, \frac{1}{2}\right). \tag{D.1}$$

Let ϕ be the concave quadratic function $\phi(\alpha) = f(0) + \alpha \dot{f}(0) + m\frac{\alpha^2}{2}$, with $m = -\frac{L}{\nu_{\min}}\hat{m}^0$, where $\hat{m}^0$ is defined in (C.31). We have $\phi(0) = f(0)$ and $\dot{\phi}(0) = \dot{f}(0) < 0$, so ϕ is decreasing on $\mathbb{R}^+$. Let us consider $\alpha \in [0, \alpha^j]$, so that $\boldsymbol{x} + \alpha\boldsymbol{d} \in \mathcal{V}_0$. According to Assumption 2, we have $|\dot{f}(\alpha) - \dot{f}(0)| \leqslant \|\boldsymbol{d}\|^2 L|\alpha|$, and according to Lemma 5,

$$|\dot{f}(\alpha) - \dot{f}(0)| \leqslant \frac{L\alpha}{\nu_{\min}}\hat{m}^0. \tag{D.2}$$

Then we obtain:

$$|\dot{f}(\alpha)| \leqslant \frac{L\alpha}{\nu_{\min}}\hat{m}^0 - \dot{f}(0). \tag{D.3}$$

Hence:

$$\dot{\phi}(\alpha) \leqslant \dot{f}(\alpha), \quad \forall\alpha \in [0, \alpha^j]. \tag{D.4}$$

Integrating (D.4) between 0 and α^j yields

$$\phi(\alpha^j) \leqslant f(\alpha^j). \tag{D.5}$$

On the other hand, the expression of ϕ at $\alpha_{\min} = \sigma_{\min}\alpha^1$ reads $\phi(\alpha_{\min}) = f(0) + S\alpha^1\dot{f}(0)$, where

$$S = \sigma_{\min} - \sigma^2_{\min} L\alpha^1 \frac{\hat{m}^0}{2\dot{f}(0)\nu_{\min}}. \tag{D.6}$$

According to (C.33):

$$\alpha^1 \leqslant \frac{-2\dot{f}(0)}{\hat{m}^0}, \tag{D.7}$$

so that

$$S \leqslant \sigma_{\min} + \sigma^2_{\min}\frac{L}{\nu_{\min}} = \frac{1}{2}, \tag{D.8}$$

where the latter equality directly stems from the expression of $\sigma_{\min}$. Since ϕ is decreasing on $\mathbb{R}^+$, we get

$$\phi(\alpha_{\min}) \geqslant f(0) + \frac{1}{2}\alpha^1\dot{f}(0) \geqslant f(\alpha^1), \tag{D.9}$$

where the last inequality is the first Wolfe condition (28) for $j = 1$.

Finally, $\alpha^j > 0$ for all $j \geqslant 1$. Assume that there exists j such that $\alpha^j < \alpha_{\min}$. According to (D.5) and given that ϕ is decreasing on $\mathbb{R}^+$, we get:

$$f(\alpha^j) \geqslant \phi(\alpha^j) > \phi(\alpha_{\min}) \geqslant f(\alpha^1), \tag{D.10}$$

which contradicts the fact that $f(\alpha^j)$ is nonincreasing. Thus, (29) holds. So does (30), according to Lemma 6.

Appendix E. Proof of Property 3

Let us first remark that for all k, $\boldsymbol{d}_k \neq \boldsymbol{0}$, since $\boldsymbol{g}_k^T \boldsymbol{d}_k < 0$. According to Property 1, the first Wolfe condition holds for $\sigma_1 = \sigma_1^J$:

$$F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1}) \geqslant -\sigma_1^J \alpha_k \boldsymbol{g}_k^T \boldsymbol{d}_k. \tag{E.1}$$

According to Property 2,

$$\alpha_k \geqslant \sigma_{\min} \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\nu \|\boldsymbol{d}_k\|^2}, \tag{E.2}$$

so

$$0 \leqslant \sigma_0 \frac{(\boldsymbol{g}_k^T \boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} \leqslant F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1}), \tag{E.3}$$

with $\sigma_0 = \frac{1}{\nu} \sigma_{\min} \sigma_1^J > 0$. According to Assumption 2, the level set $\mathcal{L}_0$ is bounded, so $\lim_{k \to \infty} F(\boldsymbol{x}_k)$ is finite. Therefore:

$$\sum_{k=0}^{\infty} \frac{(\boldsymbol{g}_k^T \boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} \leqslant \frac{1}{\sigma_0} \left[F(\boldsymbol{x}_0) - \lim_{k \to \infty} F(\boldsymbol{x}_k) \right] < \infty. \tag{E.4}$$

References

[1] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* IOP Publishing Ltd., London, UK, 1998.

[2] W. M. Boerner, H. Brand, L. A. Cram, A. K. Jordan, W. D. T. G., Keydel, G. Schwierz, and M. Vogel. *Inverse methods in electromagnetic imaging, Part I.* Number 143 in NATO ASI Series. D. Reidel (ed.), 1985.

[3] G. Demoment. Image reconstruction and restoration: Overview of common estimation structure and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):2024–2036, December 1989.

[4] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* John Wiley, New York, NY, 1977.

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

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

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

[8] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, December 2007.

[9] C. A. Johnson and A. Sofer. A primal-dual method for large-scale image reconstruction in emission tomography. *SIAM Journal on Optimization*, 11:691–715, March 2000.

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

[11] C. A. Johnson and D. McGarry. Maximum entropy reconstruction methods in electron paramagnetic resonance imaging. *Annals of Operations Research*, 119:101–118, 2003.

[12] M. Bertero, P. Boccacci, G. Desidera, and G. Vicidomini. Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12), 2009.

[13] Y. Cao, P. P. B. Eggermont, and S. Terebey. Cross Burg entropy maximization and its application to ringing suppression in image reconstruction. *IEEE Transactions on Image Processing*, 8(2):286–292, 1999.

[14] G. Aubert and J.-F. Aujol. A variational approach to removing multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4):925–946, 2008.

[15] I. T. Hsiao, A. Rangarajan, and G. Gindi. Joint-MAP Bayesian tomographic reconstruction with a gamma-mixture prior. *IEEE Transactions on Image Processing*, 11(12):1466–1477, December 2002.

[16] G. Le Besnerais, J.-F. Bercher, and G. Demoment. A new look at entropy for solving linear inverse problems. *IEEE Transactions on Information Theory*, 45(5):1565–1578, 1999.

[17] A. Mohammad-Djafari. Maximum d'entropie et problèmes inverses en imagerie. *Traitement du Signal*, pages 87–116, 1994.

[18] S. F. Gull and J. Skilling. Maximum entropy method in image processing. *Proceedings of the IEEE*, 131:646–659, October 1984.

[19] R. Nityananda and R. Narayan. Maximum entropy image reconstruction-a practical non-information-theoretic approach. *Journal of Astrophysics and Astronomy*, 3(4), December 1982.

[20] C. L. Byrne. Iterative image reconstruction algorithms based on cross-entropy minimization. *IEEE Transactions on Image Processing*, 2:96–103, 1993.

[21] M. V. Narayanan, C. L. Byrne, and M. A. King. An interior point iterative maximum-likelihood reconstruction algorithm incorporating upper and lower bounds with application to SPECT transmission imaging. *IEEE Transactions on Medical Imaging*, 20(4):342 –353, April 2001.

[22] J. A. O'Sullivan. Roughness penalties on finite domains. *IEEE Transactions on Image Processing*, 4(9), September 1995.

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

[24] C. H. Lin. *A Null-Space Primal-Dual Algorithm for NonLinear Network Optimization*. PhD thesis, University of Stanford, 2002.

[25] S. G. Nash and A. Sofer. A barrier method for large-scale constrained optimization. *ORSA Journal on Computing*, 5(41), 1993.

[26] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Transactions on Image Processing*, 8(5):688–699, May 1999.

[27] J. Sun and J. Zhang. Global convergence of conjugate gradient methods without line search. *Annals of Operations Research*, 103:161–173, March 2001.

[28] 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, January 2008.

[29] C. Labat and J. Idier. Convergence of truncated half-quadratic and Newton algorithms, with application to image restoration. Technical report, IRCCyN, June 2007.

[30] M. Allain, J. Idier, and Y. Goussard. On global and local convergence of half-quadratic algorithms. *IEEE Transactions on Image Processing*, 15(5):1130–1142, 2006.

[31] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, February 2004.

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

[33] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1999.

[34] Z.-J. Shi. Convergence of line search methods for unconstrained optimization. *Applied Mathematics and Computations*, 157:393–405, 2004.

[35] E. Chouzenoux, S. Moussaoui, and J. Idier. A Majorize-Minimize line search algorithm for barrier functions. Technical report, IRCCyN, November 2009.

[36] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6:298–311, 1997.

[37] H. Lanteri, M. Roche, and C. Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Problems*, 18(5):1397–1419, 2002.

[38] G. Vicidomini, P. Boccacci, A. Diaspro, and M. Bertero. Application of the split-gradient method to 3D image deconvolution in fluorescence microscopy. *Journal of Microscopy*, 234(1):47–61, April 2009.

[39] Y. Zhang. Interior-point gradient methods with diagonal-scalings for simple-bound constrained optimization. Technical Report TR04-06, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 2004.

[40] W. W. Hager and H. Zhang. Recent advances in bound constrained optimization. In *System Modeling and Optimization*, volume 199 of *IFIP International Federation for Information Processing*, pages 67–82. Springer, 2006.

[41] J. P. Butler, J. A. Reeds, and S. V. Dawson. Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing. *SIAM Journal on Numerical Analysis*, 18(3):381–397, June 1981.

[42] S. G. Nash. A survey of truncated-Newton methods. *Journal of Computational and Applied Mathematics*, 124:45–59, 2000.

[43] A. Nemirovski. Advances in convex optimization : Conic programming. In *Proc. of International Congress of Mathematicians*, volume 1, pages 413–444, Madrid, Spain, August 2006.

[44] D. Den Hertog, C. Roos, and T. Terlaky. Inverse barrier methods for linear programming. *Revue française d'automatique, informatique, recherche opérationnelle*, 28(2):135–163, 1994.

[45] I. S. Dhillon and Sra. S. Generalized nonnegative matrix approximations with Bregman divergences. *Advances in Neural Information Processing*, 19:283–290, 2005.