



HAL
open science

A Majorize-Minimize line search algorithm for barrier function

Emilie Chouzenoux, Saïd Moussaoui, J. Idier

► **To cite this version:**

Emilie Chouzenoux, Saïd Moussaoui, J. Idier. A Majorize-Minimize line search algorithm for barrier function. 2009. hal-00362304v4

HAL Id: hal-00362304

<https://hal.science/hal-00362304v4>

Preprint submitted on 4 Jun 2009 (v4), last revised 7 Sep 2010 (v7)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MAJORIZE-MINIMIZE LINE SEARCH ALGORITHM FOR BARRIER FUNCTION

ÉMILIE CHOUZENOUX, SAÏD MOUSSAOUI, AND JÉRÔME IDIER

Abstract. Criteria containing a *barrier* function i.e., an unbounded function at the boundary of the feasible solution domain are frequently encountered in the optimization framework, in particular in interior point methods for constrained optimization. Barrier 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. The practical efficiency of the proposal scheme is illustrated by means of an example of constrained quadratic programming.

Key words.

AMS subject classifications.

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

$$\min_{\mathbf{x}} (F(\mathbf{x}) = P(\mathbf{x}) + \mu B(\mathbf{x})), \quad \mu \geq 0 \tag{1.1}$$

where $\mathbf{x} \in \mathbb{R}^n$, P is a differentiable function and B is a *barrier* function which makes the criterion unbounded at the boundary of the strictly feasible

$$\mathcal{C} = \{\mathbf{x} | c_i(\mathbf{x}) > 0, i = 1, \dots, m\}$$

defined by some concave constraints c_i so that the minimizers of F belong to \mathcal{C} . The most popular example is the logarithmic barrier

$$B(\mathbf{x}) = - \sum_{i=1}^m \log(c_i(\mathbf{x})). \tag{1.2}$$

The barrier property is used by interior point methods [28] to solve constrained optimization problems

$$\min F_0(\mathbf{x}) \quad \text{subject to} \quad c_i(\mathbf{x}) \geq 0, \tag{1.3}$$

a barrier function B being artificially introduced to keep the solution inside the feasible domain. The augmented criterion can be expressed as

$$F_\mu(\mathbf{x}) = F_0(\mathbf{x}) + \mu B(\mathbf{x}), \tag{1.4}$$

where $\mu \geq 0$ is the barrier parameter and B is the barrier function associated to the constraints c_i . The minimization of F_μ must be performed for a sequence of parameter values μ that decreases to 0 leading to the solution of (1.3).

A large family of optimization methods to solve (1.1) are based on iteratively decreasing the criterion by moving the current solution \mathbf{x}_k along a direction \mathbf{d}_k ,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \tag{1.5}$$

where $\alpha_k > 0$ is the *stepsize* and \mathbf{d}_k is a *descent direction* i.e., satisfying $\nabla F(\mathbf{x}_k)^T \mathbf{d}_k < 0$. In practice, such iterative descent direction methods consist in alternating the construction of \mathbf{d}_k and the determination of α_k . The process of choosing the stepsize α_k is called the *line search*. It is founded on minimizing the scalar function $f(\alpha) = F(\mathbf{x}_k + \alpha \mathbf{d}_k)$. Since the barrier function has a singularity at the boundary of \mathcal{C} , it causes the inefficiency of standard line search strategies often based on polynomial interpolation [19]. In this paper we propose an original line search procedure based on the majorize-minimize (MM) principle [12] by deriving an adequate form of a tangent majorant function well suited to approximate a criterion containing a barrier function. The convergence of classical descent algorithms is established when α_k is obtained with the proposed MM line search strategy.

This paper is organized as follows: In §2, we discuss the role of line search in optimization algorithms, and explain why special-purpose line search procedures may be helpful in dealing with barrier functions. A new MM line search strategy is proposed in §3 including methods to obtain a suitable form of tangent majorant function and its minimizer. §4 gives the properties of the step size that allowing us to establish convergence conditions such as Armijo relation. We deduce from this properties convergence results reported in §5 when the proposed MM line search is associated with classical descent algorithms to solve (1.1). §6 illustrates the efficiency of the proposed line search strategy through a numerical example in quadratically constrained quadratic programming.

2. Line search strategies.

2.1. Problem statement. In line search methods, the stepsize is frequently required to satisfy mathematical conditions ensuring the convergence of the descent algorithm. According to the Wolfe conditions, α_k is acceptable if there exists $(c_1, c_2) \in (0; 1)$ such that

$$F(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq F(\mathbf{x}_k) + c_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k \quad (2.1)$$

$$|\nabla F(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k| \leq c_2 |\mathbf{g}_k^T \mathbf{d}_k| \quad (2.2)$$

where $\mathbf{g}_k \triangleq \nabla F(\mathbf{x}_k)$. There exist several procedures [21] 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 [18, 21] or approximation of $f(\alpha)$ by a quadratic function [17].

However, the barrier term $B(\mathbf{x})$ implies that $f(\alpha)$ tends to infinity when α is equal to the smallest positive step $\bar{\alpha}$ cancelling some constraint at $\mathbf{x}_k + \bar{\alpha} \mathbf{d}_k$. 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 $\alpha \geq \bar{\alpha}$. Moreover, due to the vertical asymptote at $\bar{\alpha}$, methods using cubic interpolation or quadratic approximation are not suited [19].

The typical line search in barrier-related optimization methods chooses α_k as a fixed fraction close to unity of $\bar{\alpha}$ [24, 10]. However, this simple approach do not ensure the convergence of the optimization algorithm and can lead to a sequence of iterates ‘trapped’ near the singularity [19]. More elaborated line search strategies adapted to barrier function optimization have been proposed in [8, 19, 14]. In [14], a line search procedure for optimization of self-concordant functions [20] such as the logarithmic barrier function is developed and leads to convergence when \mathbf{d}_k is the direction of nonlinear conjugate gradient with Conjugate Descent (CD) conjugaison. However,

the computation of the step size requires the evaluation of the Hessian matrix which tends to be expensive or even impossible for large scale problems.

The main idea in [8, 19] is to approximate the one-dimensional barrier criterion along the search direction by function of the form

$$B(\mathbf{x} + \alpha \mathbf{d}) \approx p_0 + p_1 \alpha + p_2 \alpha^2 - \mu \log(p_3 - \alpha) \quad (2.3)$$

where the coefficients p_i are chosen to interpolate B and its gradient at two trial points. Making use of those specific interpolating functions, the line search strategy generates a stepsize fulfilling (2.1) and (2.2). Unfortunately, the resulting algorithms are not often used in practice, probably because the proposed interpolating functions are difficult to compute.

In contrast, our proposal is not based on interpolation, but rather on majorization, with a view to simplify the line search strategy and preserve the strong convergence properties.

2.2. Majoration-Minimization algorithms. In Majoration-Minimization (MM) algorithms [12, 13], the minimization of a function f is obtained by performing successive minimizations of *tangent majorant* functions for f . Function $h(\mathbf{u}, \mathbf{v})$ is said tangent majorant for $f(\mathbf{u})$ at \mathbf{v} if for all \mathbf{u} ,

$$\begin{cases} h(\mathbf{u}, \mathbf{v}) \geq f(\mathbf{u}) \\ h(\mathbf{v}, \mathbf{v}) = f(\mathbf{v}) \end{cases} \quad (2.4)$$

The initial optimization problem is then replaced by a sequence of easier subproblems, corresponding to the MM update rule

$$\mathbf{u}^{j+1} = \arg \min_{\mathbf{u}} h(\mathbf{u}, \mathbf{u}^j). \quad (2.5)$$

Recently, a line search procedure based on an MM strategy has been introduced [17]. The stepsize value α_k results from J successive minimizations of quadratic tangent majorant functions for the scalar function $f(\alpha)$. Let consider the family of parabolas

$$q^j(\alpha, \alpha^j) = f(\alpha^j) + (\alpha - \alpha^j) \dot{f}(\alpha^j) + \frac{1}{2} m^j (\alpha - \alpha^j)^2, \quad (2.6)$$

tangent majorants of $f(\alpha)$ at $\alpha^j, j = 1, \dots, J$. The stepsize α_k is given by the recurrence

$$\begin{aligned} \alpha^0 &= 0 \\ \alpha^{j+1} &= \alpha^j - \frac{\dot{f}(\alpha^j)}{m^j}, \quad j = 0, \dots, J-1 \\ \alpha_k &= \alpha^J \end{aligned} \quad (2.7)$$

The convergence of a family of non-linear conjugate gradient methods associated to this line search strategy is proved in [17] whatever the value of J .

However, since the function $f(\alpha)$ resulting from problem (1.1) is unbounded, there is no parameter m^j such that the quadratic $q^j(\alpha, \alpha^j)$ majorizes $f(\alpha)$ in the whole definition domain of α . Actually, it would be sufficient to majorize $f(\alpha)$ within the level set $\mathcal{L}_k = \{\alpha, F(\mathbf{x}_k + \alpha \mathbf{d}_k) \leq F(\mathbf{x}_k)\}$ but this set is difficult to determine or even to approximate.

3. Proposed line search strategy. Instead of a quadratic, we propose the following form of tangent majorant function

$$h(\alpha) = p_0 + p_1\alpha + p_2\alpha^2 - p_3 \log(\bar{\alpha} - \alpha), \quad (3.1)$$

which is reminiscent of interpolation functions (2.3).

According to MM theory, the stepsize α_k is defined by

$$\begin{aligned} \alpha^0 &= 0 \\ \alpha^{j+1} &= \arg \min_{\alpha} h^j(\alpha, \alpha^j), \quad j = 0, \dots, J-1 \\ \alpha_k &= \alpha^J \end{aligned} \quad (3.2)$$

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

$$\begin{aligned} h^j(\alpha, \alpha^j) &= f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}m^j(\alpha - \alpha^j)^2 \\ &\quad + \gamma^j \left[(\bar{\alpha} - \alpha^j) \log\left(\frac{\bar{\alpha} - \alpha^j}{\bar{\alpha} - \alpha}\right) - \alpha + \alpha^j \right] \end{aligned} \quad (3.3)$$

which depends on the value of f and its gradient at α^j and two design parameters m^j, γ^j .

3.1. Construction of the majorant function in the linear case. It is easy to check that $h^j(\alpha, \alpha) = f(\alpha)$ for all α . There remains to find values of m^j, γ^j such that $h^j(\alpha, \alpha^j) \geq f(\alpha)$ holds for all $\alpha \in [0; \bar{\alpha}]$. We assume that $p(\alpha) = P(\mathbf{x} + \alpha \mathbf{d})$ is majorized by the quadratic function

$$q^j(\alpha, \alpha^j) = p(\alpha^j) + (\alpha - \alpha^j)\dot{p}(\alpha^j) + \frac{1}{2}m_p^j(\alpha - \alpha^j)^2, \quad (3.4)$$

and we focus on the majoration of $B(\mathbf{x} + \alpha \mathbf{d})$ when this term can be written as linear combination of

$$\sum_t \psi(a_t + \alpha \delta_t) \quad \text{with} \quad \psi(u) = -\kappa \log(u), \quad \kappa > 0 \quad (3.5)$$

$$\text{or} \quad \psi(u) = u \log u \quad (3.6)$$

$$\text{or} \quad \psi(u) = -u^r, \quad r \in (0, 1) \quad (3.7)$$

Consider the logarithmic barrier (1.2) involved in the resolution of problem (1.3) by interior point method. In the following examples, $B(\mathbf{x} + \alpha \mathbf{d})$ can read $\sum_t -\kappa_t \log(a_t + \alpha \delta_t)$.

Linear programming: $c_i(\mathbf{x}) = [\mathbf{A}\mathbf{x}]_i + \rho_i$

Quadratic programming: $c_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T \mathbf{A}_i \mathbf{x} + a_i^T \mathbf{x} + \rho_i$ with \mathbf{A}_i symmetric definite positive.

Semidefinite programming [27]: Minimizing $P(\mathbf{x})$ under the matricial constraint $G(\mathbf{x}) \succ 0$ where $G(\mathbf{x})$ is the affine function

$$G(\mathbf{x}) = \mathbf{G}_0 + x_1 \mathbf{G}_1 + \dots + x_n \mathbf{G}_n.$$

The barrier associated to this constraint is

$$B(\mathbf{x}) = \log \det G(\mathbf{x})^{-1}$$

The scalar function $B(\mathbf{x} + \alpha \mathbf{d})$ reads $-\sum_i \log(1 + \alpha \lambda_i)$ when $\lambda_i, i = 1, \dots, \ell$ are the generalized eigen values of $(\sum_i \mathbf{d}_i \mathbf{G}_i, G(\mathbf{x}))$.

Second order conic programming [3]: Minimizing $P(\mathbf{x})$ subjected to the conic constraints $\|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i$. The resulting barrier function takes the form

$$B(\mathbf{x}) = -\sum_i \log(-\|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2^2 - (\mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i)^2) \quad (3.8)$$

In order to build a tangent majorant function of the barrier term $B(\mathbf{x} + \alpha \mathbf{d})$, let define $b_1(\alpha) = \sum_{t|\delta_t > 0} \psi(a_t + \alpha \delta_t)$ and $b_2(\alpha) = \sum_{t|\delta_t < 0} \psi(a_t + \alpha \delta_t)$. We will use the following lemma dealing with function ψ :

LEMMA 3.1. For all positive u , if ψ is given by (3.5), (3.6) or (3.7) then

$$-\ddot{\psi}(u)/\dot{\psi}(u) \leq 2/u \quad (3.9)$$

Proof.

1) $\psi(u) = -\kappa \log(u)$, $\kappa > 0$
 $\dot{\psi}(u) = \kappa/u^2$ and $\ddot{\psi}(u) = -2\kappa/u^3$ then

$$-\ddot{\psi}(u)/\dot{\psi}(u) = 2/u \quad (3.10)$$

2) $\psi(u) = u \log(u)$
 $\dot{\psi}(u) = 1/u$ and $\ddot{\psi}(u) = -1/u^2$ then

$$-\ddot{\psi}(u)/\dot{\psi}(u) = 1/u \leq 2/u \quad (3.11)$$

2) $\psi(u) = -u^r$, $r \in (0, 1)$
 $\dot{\psi}(u) = -r(r-1)u^{r-2}$ and $\ddot{\psi}(u) = -r(r-1)(r-2)u^{r-3}$ then since $r \in (0, 1)$

$$-\ddot{\psi}(u)/\dot{\psi}(u) = (2-r)/u \leq 2/u \quad (3.12)$$

□

On the one hand, b_1 is strictly convex and its derivative is strictly concave. According to [9, Th.1], $\phi_1^j(\alpha, \alpha^j) = b_1(\alpha^j) + (\alpha - \alpha^j)\dot{b}_1(\alpha^j) + \frac{1}{2}m_b^j(\alpha - \alpha^j)^2$ is a tangent majorant of $b_1(\alpha)$ at α^j if

$$m_b^j = \begin{cases} \ddot{b}_1(0) & \text{if } \alpha^j = 0 \\ \frac{b_1(0) - b_1(\alpha^j) + \alpha^j \dot{b}_1(\alpha^j)}{(\alpha^j)^2/2} & \text{elsewhere} \end{cases} \quad (3.13)$$

On the other hand, b_2 is strictly convex but its derivative is also strictly convex. Then the previous result cannot be used. b_2 has a vertical asymptote in

$$\alpha = \bar{\alpha} = \min_{t|\delta_t < 0} -a_t/\delta_t. \quad (3.14)$$

Therefore we look for a tangent majorant of b_2 of the form

$$\phi_2^j(\alpha, \alpha^j) = b_2(\alpha^j) + (\alpha - \alpha^j)\dot{b}_2(\alpha^j) + \gamma_b^j \left[(\bar{\alpha} - \alpha^j) \log \left(\frac{\bar{\alpha} - \alpha^j}{\bar{\alpha} - \alpha} \right) + \alpha^j - \alpha \right]$$

We define the following function, $T(\alpha) = \dot{b}_2(\alpha)(\bar{\alpha} - \alpha)$, whose derivatives are:

$$\dot{T}(\alpha) = \ddot{b}_2(\alpha)(\bar{\alpha} - \alpha) - \dot{b}_2(\alpha) \quad (3.15)$$

$$\ddot{T}(\alpha) = \dddot{b}_2(\alpha)(\bar{\alpha} - \alpha) - 2\ddot{b}_2(\alpha) \quad (3.16)$$

The second derivative of T reads also

$$\ddot{T}(\alpha) = \sum_t \delta_t^2 \left[\delta_t(\alpha - \bar{\alpha})\ddot{\psi}(a_t + \alpha\delta_t) - 2\dot{\psi}(a_t + \alpha\delta_t) \right] \quad (3.17)$$

$\alpha < \bar{\alpha}$ then $a_t + \alpha\delta_t > 0$ and $\delta_t(\alpha - \bar{\alpha}) < a_t + \alpha\delta_t$. Taking $u = a_t + \alpha\delta_t$, Lemma 3.1 implies $\ddot{T}(\alpha) < 0$, which shows that the function T is *strictly concave*. Moreover, let us consider the linear function $l(\alpha)$:

$$l(\alpha) = \phi_2^j(\alpha, \alpha^j)(\bar{\alpha} - \alpha) = \dot{b}_2(\alpha^j)(\bar{\alpha} - \alpha) + \gamma_b^j(\alpha - \alpha^j) \quad (3.18)$$

According to [9, Lem. 3], the strictly concave function $T(\alpha)$ intersects $l(\alpha)$ at most twice. Yet:

$$l(\alpha^j) = T(\alpha^j) \quad (3.19)$$

Let define

$$\gamma_b^j = \begin{cases} \bar{\alpha}\ddot{b}_2(0) & \text{if } \alpha^j = 0 \\ \frac{b_2(0) - b_2(\alpha^j) + \alpha^j\dot{b}_2(\alpha^j)}{(\bar{\alpha} - \alpha^j)\log(1 - \alpha^j/\bar{\alpha}) + \alpha^j} & \text{elsewhere} \end{cases} \quad (3.20)$$

First, let consider the particular case $\alpha^j = 0$. We have

$$T(0) = l(0), \quad \dot{T}(0) = \dot{l}(0) \quad (3.21)$$

Concavity of $T(\alpha)$ leads to

$$l(\alpha) \geq T(\alpha), \quad \alpha \in [0; \bar{\alpha}[\quad (3.22)$$

Hence, according to [9, Lem. 5], b_2 is majorized by $\phi_2^j(\alpha, 0)$ at $\alpha^j = 0$.

Assume that $\alpha^j > 0$. According to the value of γ_b^j ,

$$\begin{cases} b_2(0) = \phi_2^j(0, \alpha^j) \\ b_2(\alpha^j) = \phi_2^j(\alpha^j, \alpha^j) \end{cases} \quad (3.23)$$

In other words, the function $\Delta(\alpha, \alpha^j) = b_2(\alpha) - \phi_2^j(\alpha, \alpha^j)$ vanishes in 0 and in α^j . Then, there exists $\alpha_p \in [0; \alpha^j)$ such that the derivative $\dot{\Delta}(\alpha_p, \alpha^j)$ vanishes

$$\dot{\Delta}(\alpha_p, \alpha^j)(\bar{\alpha} - \alpha_p) = 0 \quad (3.24)$$

and equivalently,

$$T(\alpha_p) = l(\alpha_p). \quad (3.25)$$

α^j and α_p are the only intersection points between $l(\alpha)$ and $T(\alpha)$. Concavity of $T(\alpha)$ leads to

$$l(\alpha) < T(\alpha), \quad \alpha \in]\alpha^p; \alpha^j[\quad (3.26)$$

and

$$l(\alpha) > T(\alpha), \alpha \in [0; \alpha_p[\cup]\alpha^j; \bar{\alpha}[\quad (3.27)$$

Noticing that $\bar{\alpha} - \alpha > 0$, we could apply [9, Lem. 5] and then demonstrate that b_2 is majorized by $\phi_2^j(\alpha, \alpha^j)$.

Functions b_1 and b_2 are strictly convex then for $l = 1, 2$,

$$b_l(0) - b_l(\alpha^j) + \alpha^j \dot{b}_l(\alpha^j) > 0, \forall \alpha^j \in]0; \bar{\alpha}[\quad (3.28)$$

Hence $m_b^j > 0$. Moreover, for all $0 < u < 1$,

$$(1 - u) \log(1 - u) + u > 0. \quad (3.29)$$

Then $\gamma_b^j > 0$. Furthermore, m_b^j and γ_b^j are the smallest positive parameters such that ϕ_1^j and ϕ_2^j major respectively b_1 and b_2 since we have for $l = 1, 2$,

$$b_l(0) - \phi_l^j(0, \alpha^j) = 0, \forall \alpha \geq \alpha^j. \quad (3.30)$$

Finally, according to [9, Lem.6] and inequalities (3.22),(3.27), we have for $l = 1, 2$

$$\dot{b}_l(\alpha) - \dot{\phi}_l(\alpha, \alpha^j), \forall \alpha \geq \alpha^j. \quad (3.31)$$

This equation will be useful to perform convergence analysis for conjugate gradient algorithms. Results to design (m^j, γ^j) are summarized in table 3.1. Figure 3.1 illustrates an example of scalar criterion and the obtained majorant according to this method.

3.2. Computing the minimizer. The MM recurrence (3.2) involves the computation of the minimizer of $h^j(\alpha, \alpha^j)$. The positivity of parameters m^j, γ^j leads to the strict convexity of the tangent majorant. Hence, it has a unique minimizer, which takes an explicit form. The case when $\bar{\alpha} = +\infty$ is obvious since α^{j+1} is given by (2.7). Let assume that $\bar{\alpha} < +\infty$. α^{j+1} is the minimizer of the function $h^j(\alpha, \alpha^j)$ which is also the unique root of the function $Q(\alpha)$, verifying $\alpha < \bar{\alpha}$, given by

$$Q(\alpha) = (\bar{\alpha} - \alpha) \dot{h}^j(\alpha, \alpha^j) \quad (3.32)$$

whose complete expression is:

$$Q(\alpha) = q_1(\alpha - \alpha^j)^2 + q_2(\alpha - \alpha^j) + q_3 \quad (3.33)$$

with

$$\begin{cases} q_1 = -m^j \\ q_2 = \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha} - \alpha^j) \\ q_3 = (\bar{\alpha} - \alpha^j) \dot{f}(\alpha^j) \end{cases} \quad (3.34)$$

The calculation of this root depends on the value of q_1 . If q_1 equals zero, then Q is a linear function whose unique root is:

$$\alpha^{j+1} = \alpha^j - \frac{q_3}{q_2} = \alpha^j - \frac{(\bar{\alpha} - \alpha^j) \dot{f}(\alpha^j)}{\gamma^j - \dot{f}(\alpha^j)} \quad (3.35)$$

and if q_1 is non zero, Q is a second order polynomial vanishing at

$$\alpha^j + \frac{-q_2 \pm \sqrt{q_2^2 - 4q_1q_3}}{2q_1} \quad (3.36)$$

The minimizer of $h^j(\alpha, \alpha^j)$ satisfies the constraint $\alpha^{j+1} < \bar{\alpha}$. Then, it is equals to the smaller root of $Q(\alpha)$.

$$\alpha^{j+1} = \alpha^j + \frac{-q_2 + \sqrt{q_2^2 - 4q_1q_3}}{2q_1} = \alpha^j + \frac{-2q_3}{q_2 + \sqrt{q_2^2 - 4q_1q_3}} \quad (3.37)$$

Table 3.2 summarizes the expression of the minimizer of $h^j(\alpha, \alpha^j)$.

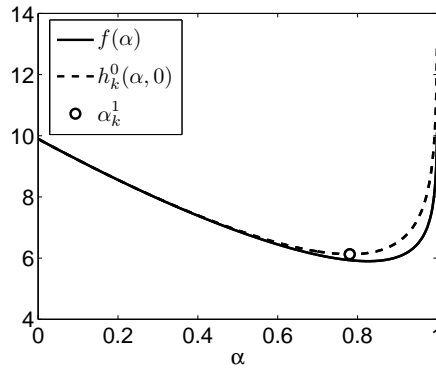


FIG. 3.1. Example of a tangent majorant function $h^0(\alpha, 0)$ for $f(\alpha) = (\alpha - 5)^2 - \sum_{i=1}^{10} \log(i - \alpha)$. $h^0(\alpha, 0)$ is defined by (3.3) with $m^0 = 2$, $\gamma^0 = 1.55$ and $\bar{\alpha} = 1$

$$\bar{\alpha} = \min_{t|\delta_t < 0} -a_t/\delta_t$$

m_p majorant curvature of $p(\alpha)$

$$m^j = m_p + \mu m_b^j \quad \gamma^j = \mu \gamma_b^j$$

with

$$m_b^j = \begin{cases} \ddot{b}_1(0) & \text{if } \alpha^j = 0 \\ \frac{b_1(0) - b_1(\alpha^j) + \alpha^j \dot{b}_1(\alpha^j)}{(\alpha^j)^2/2} & \text{elsewhere} \end{cases}$$

$$\gamma_b^j = \begin{cases} \bar{\alpha} \ddot{b}_2(0) & \text{if } \alpha^j = 0 \\ \frac{b_2(0) - b_2(\alpha^j) + \alpha^j \dot{b}_2(\alpha^j)}{(\bar{\alpha} - \alpha^j) \log(1 - \alpha^j/\bar{\alpha}) + \alpha^j} & \text{elsewhere} \end{cases}$$

where $b_1(\alpha) = \sum_{t|\delta_t > 0} b(a_t + \alpha \delta_t)$ and $b_2(\alpha) = \sum_{t|\delta_t < 0} b(a_t + \alpha \delta_t)$.

TABLE 3.1

Computation of the parameters (m^j, γ^j) when $B(\mathbf{x} + \alpha \mathbf{d})$ takes the form $-\sum_t \log(a_t + \alpha \delta_t)$

$$\alpha^{j+1} = \begin{cases} \alpha^j - \frac{\dot{f}(\alpha^j)}{m^j} & \text{if } \bar{\alpha} = +\infty \\ \alpha^j + \frac{-2q_3}{q_2 + \sqrt{q_2^2 - 4q_1q_3}} & \text{else} \end{cases} \quad (3.38)$$

where

$$q_1 = -m^j, \quad q_2 = \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha} - \alpha^j) \quad \text{and} \quad q_3 = (\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j).$$

TABLE 3.2

Computation of α^{j+1} , the minimizer of the tangent majorant (3.3)

4. Analysis of the stepsize series. The present section gathers technical results concerning the stepsize series generated by (3.2), which will be useful to derive the global convergence properties of §5. We made the following assumption on the criterion F :

ASSUMPTION 4.1. *Gradient Lipschitz*

The level set $\mathcal{L}_0 = \{\mathbf{x} | F(\mathbf{x}) \leq F(\mathbf{x}_0)\}$ is assumed bounded. $F(\mathbf{x})$ is differentiable on a neighbourhood \mathcal{V} of \mathcal{L}_0 and $\nabla F(\mathbf{x})$ is Lipschitz continuous on \mathcal{V} , i.e., there exists $0 < L < \infty$ such that

$$\|\nabla F(\mathbf{x}) - \nabla F(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{V} \quad (4.1)$$

Let consider $\mathbf{x} \in \mathcal{V}$ and \mathbf{d} a descent direction, so that $\dot{f}(0) = \mathbf{d}^T \mathbf{g} < 0$. The case $\dot{f}(0) = 0$ is excluded from this study since in this case the resulting stepsize is zero. The stepsize series is obtained from the recurrence (3.2). To build some properties of the stepsize series, we will need the following assumption on the curvature of the tangent majorant:

ASSUMPTION 4.2. *Majorant curvature*

There exists some constants $0 < \nu_1 \leq \nu_2$ such that:

$$\nu_1 \|\mathbf{d}\|^2 \leq \ddot{h}^j(\alpha^j, \alpha^j) \leq \nu_2 \|\mathbf{d}\|^2 \quad \forall j \quad (4.2)$$

4.1. Properties of the stepsize series. (3.2) produces monotonically decreasing values $\{f(\alpha^j)\}$ and the series $\{\alpha^j\}$ converges to a stationary point of $f(\alpha)$ [12]. The behaviour of the series is described by

$$(\alpha^{j+1} - \alpha^j) \dot{f}(\alpha^j) \leq 0, \quad \forall j \geq 1 \quad (4.3)$$

According to the capture property developed in [13],

$$F(\mathbf{x} + \alpha^j \mathbf{d}) \leq F(\mathbf{x}), \quad \forall j \geq 1 \quad (4.4)$$

Then

$$\mathbf{x} + \alpha^j \mathbf{d} \in \mathcal{V}, \quad \forall j \geq 1 \quad (4.5)$$

and since $\dot{f}(0) < 0$, it can easily be shown that

$$\alpha^j > 0, \quad \forall j \geq 1 \quad (4.6)$$

Furthermore, according to the expression of the minimizer of the tangent majorant, if $\dot{f}(\alpha^j) < 0$, then

$$\frac{-q_3}{q_2} \leq \alpha^{j+1} - \alpha^j \leq \frac{-2q_3}{q_2} \quad (4.7)$$

4.2. Lower and upper bounds for the stepsize. PROPERTY 4.1. *If the constraints $c_i(\mathbf{x})$ are linear, there exists $\nu, \nu' > 0$ such that*

$$\frac{-\mathbf{g}^T \mathbf{d}}{\nu \|\mathbf{d}\|^2} \leq \alpha^1 \leq \frac{-\mathbf{g}^T \mathbf{d}}{\nu' \|\mathbf{d}\|^2}. \quad (4.8)$$

Proof. If $\bar{\alpha} = +\infty$ then the tangent majorant of $F(\mathbf{x} + \alpha \mathbf{d}) = f(\alpha)$ at $\alpha_0 = 0$ is a quadratic function with curvature m^0 . This majorant is minimized at $\alpha^1 = -\dot{f}(0)/m^0$ and according to assumption 4.2, we have:

$$\frac{-\mathbf{g}^T \mathbf{d}}{\nu_2 \|\mathbf{d}\|^2} \leq \alpha^1 \leq \frac{-\mathbf{g}^T \mathbf{d}}{\nu_1 \|\mathbf{d}\|^2} \quad (4.9)$$

Let us assume for the sequel that $\bar{\alpha}$ is finite. According to equation 4.7:

$$\frac{-(\bar{\alpha} - 0)\mathbf{g}^T \mathbf{d}}{(\bar{\alpha} - 0)m^0 + \gamma^0 \mathbf{g}^T \mathbf{d}} \leq \alpha^1 \leq 2 \frac{-(\bar{\alpha} - 0)\mathbf{g}^T \mathbf{d}}{(\bar{\alpha} - 0)m^0 + \gamma^0 \mathbf{g}^T \mathbf{d}} \quad (4.10)$$

Hence:

$$\frac{-\mathbf{g}^T \mathbf{d}}{m^0 + \frac{\gamma^0}{\bar{\alpha}} - \frac{\mathbf{g}^T \mathbf{d}}{\bar{\alpha}}} \leq \alpha^1 \leq 2 \frac{-\mathbf{g}^T \mathbf{d}}{m^0 + \frac{\gamma^0}{\bar{\alpha}} - \frac{\mathbf{g}^T \mathbf{d}}{\bar{\alpha}}} \quad (4.11)$$

The quantity $-\mathbf{g}^T \mathbf{d}/\bar{\alpha}$ is positive. Moreover, according to assumption 4.2,

$$\nu_1 \|\mathbf{d}\|^2 \leq m^0 + \frac{\gamma^0}{\bar{\alpha}} \quad (4.12)$$

Then we establish the right part of inequality (4.8)

$$\alpha^1 \leq \frac{-\mathbf{g}^T \mathbf{d}}{\nu_1 \|\mathbf{d}\|^2}. \quad (4.13)$$

Let us show that there exists $\nu > 0$ such that:

$$m^0 + \frac{\gamma^0}{\bar{\alpha}} - \frac{\mathbf{g}^T \mathbf{d}}{\bar{\alpha}} \leq \nu \|\mathbf{d}\|^2 \quad (4.14)$$

Assumption 4.2 implies that there exists ν_2 such that

$$m^0 + \frac{\gamma^0}{\bar{\alpha}} \leq \nu_2 \|\mathbf{d}\|^2 \quad (4.15)$$

Assume that $c_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + \rho_i$ with for $i = 1, \dots, m$, $\mathbf{a}_i \in \mathbb{R}^{n \times 1}$, $\rho_i \in \mathbb{R}$. Since B is a barrier function, there exists $\epsilon_0 > 0$ such that for all $\mathbf{x} \in \mathcal{V}$, for all $i = 1, \dots, m$,

$$c_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + \rho_i \geq \epsilon_0 \quad (4.16)$$

Then we have:

$$\bar{\alpha} \geq \frac{\epsilon_0}{\max_{i|\mathbf{a}_i^T \mathbf{d} < 0} -\mathbf{a}_i^T \mathbf{d}} = \frac{\epsilon_0}{\max_i -\mathbf{a}_i^T \mathbf{d}} \quad (4.17)$$

Let ι be an index such that

$$\iota = \operatorname{argmax}_i -\mathbf{a}_i^T \mathbf{d} \quad (4.18)$$

We have

$$\bar{\alpha} \geq \frac{\epsilon_0}{|\mathbf{a}_\iota^T \mathbf{d}|} \quad (4.19)$$

Hence

$$\frac{-\mathbf{g}^T \mathbf{d}}{\bar{\alpha}} = \frac{|\mathbf{g}^T \mathbf{d}|}{\bar{\alpha}} \leq |\mathbf{g}^T \mathbf{d}| \cdot |\mathbf{a}_\iota^T \mathbf{d}| \frac{1}{\epsilon_0} \quad (4.20)$$

According to Cauchy-Swartz inequality:

$$|\mathbf{g}^T \mathbf{d}| |\mathbf{a}_\iota^T \mathbf{d}| \leq \|\mathbf{g}\| \|\mathbf{a}_\iota\| \|\mathbf{d}\|^2 \quad (4.21)$$

$\|\mathbf{a}_\iota\|$ is majorized by

$$\zeta = \max_i \|\mathbf{a}_i\| \quad (4.22)$$

Since $\{\mathbf{a}_i\}$ contains at least a non zero vector so ζ is strictly positive. According to assumption 4.1, there exists $\eta > 0$ such that

$$\|\mathbf{g}\| \leq \eta, \forall \mathbf{x} \in \mathcal{L}_0 \quad (4.23)$$

Thus,

$$\frac{-\mathbf{g}^T \mathbf{d}}{\bar{\alpha}} \leq \eta \zeta \|\mathbf{d}\|^2 \quad (4.24)$$

Then Property 4.1 holds for all (\mathbf{x}, \mathbf{d}) , if we put:

$$\nu = \max(\nu_2, \nu_2 + \eta \zeta) = \nu_2 + \eta \zeta, \quad \nu' = \nu_1 \quad (4.25)$$

□

REMARK 1. *The extension of Property (4.1) to the case of nonlinear constraints is not obvious. As a matter of fact, we were not able to establish an equivalent to inequality (4.19) for concave nonlinear constraints c_i .*

4.3. Sufficient decrease condition. Given a current solution \mathbf{x} and a current descent direction \mathbf{d} , the stepsize α must induce a sufficient decrease of F . The first Wolfe condition (2.1) measures this decrease. It is equivalent to

$$f(\alpha) - f(0) \leq c_1 \alpha \dot{f}(0). \quad (4.26)$$

First, we establish some technical lemmas which appear to be useful in the sequel.

LEMMA 4.1. *Let $a_1 \in \mathbb{R}$ and $\zeta(\alpha) = a_1 \left[\log\left(\frac{\bar{\alpha}}{\bar{\alpha}-\alpha}\right) - \frac{\alpha}{\bar{\alpha}} \right]$.*

$$\frac{\zeta(\alpha)}{\alpha \dot{\zeta}(\alpha)} \leq \frac{1}{2}, \quad \forall \alpha \in (0, \bar{\alpha})$$

and

$$\lim_{\alpha \rightarrow 0} \frac{\zeta(\alpha)}{\alpha \dot{\zeta}(\alpha)} = \frac{1}{2}$$

Proof. We define :

$$g(\alpha) := \frac{\psi(\alpha)}{\alpha \dot{\psi}(\alpha)} \quad (4.27)$$

The calculation of function g gives :

$$g(\alpha) = \left[\log \left(\frac{\bar{\alpha}}{\bar{\alpha} - \alpha} \right) - \frac{\alpha}{\bar{\alpha}} \right] \left[\frac{\bar{\alpha}(\bar{\alpha} - \alpha)}{\alpha^2} \right] \quad (4.28)$$

It could be noted that $\lim_{\alpha \rightarrow 0} g = \frac{1}{2}$ and

$$\dot{g}(\alpha) = \frac{1}{\alpha} + \left[\log \left(\frac{\bar{\alpha}}{\bar{\alpha} - \alpha} \right) - \frac{\alpha}{\bar{\alpha}} \right] \left[-2 \frac{\bar{\alpha}^2}{\alpha^3} + \frac{\bar{\alpha}}{\alpha^2} \right]; \forall \alpha \neq 0 \quad (4.29)$$

We introduce the reparametrization $u = \frac{\alpha}{\bar{\alpha}}$. $u \in (0, 1)$ for $\alpha \in (0, \bar{\alpha})$. So the function $\dot{g}(\alpha)$ has the same sign as $\alpha \dot{g}(\alpha) = \rho(\frac{\alpha}{\bar{\alpha}}) = \rho(u)$.

$$\rho(u) = 1 + (-\log(1 - u) - u) \left(-\frac{2}{u^2} + \frac{1}{u} \right) \quad (4.30)$$

After manipulating inequalities, we find that

$$\rho(u) \leq 0 \Leftrightarrow \log(1 - u) + \frac{2u}{2 - u} \leq 0 \quad (4.31)$$

The function $\log(1 - u) + \frac{2u}{2 - u}$ is decreasing on $[0; 1]$ and vanishes at $u = 0$. This implies that ρ is negative and then the function g is decreasing. Finally, we obtain :

$$\psi(\alpha) \leq \frac{1}{2} \alpha \dot{\psi}(\alpha); \forall \alpha \in]0; \bar{\alpha}[\quad (4.32)$$

. \square

LEMMA 4.2. $\forall j > 1$, there exists c_j^{\max} such that

$$\alpha^j \leq c_j^{\max} \alpha^1. \quad (4.33)$$

with

$$c_j^{\max} = \left(1 + \frac{2\nu_2 L}{\nu_1^2} \right)^{j-1} \left(1 + \frac{\nu}{L} \right) - \frac{\nu}{L} \geq 1 \quad (4.34)$$

Proof. Assume that $\dot{f}(\alpha^j) \leq 0$. According to equation 4.7

$$\alpha^{j+1} - \alpha^j \leq \frac{-2\dot{f}(\alpha^j)}{(\gamma^j - \dot{f}(\alpha^j))/(\bar{\alpha} - \alpha^j) + m^j}. \quad (4.35)$$

Thereby:

$$\alpha^{j+1} - \alpha^j \leq \frac{-2\dot{f}(\alpha^j)}{\gamma^j/(\bar{\alpha} - \alpha^j) + m^j} \quad (4.36)$$

According to assumption 4.2:

$$\|\mathbf{d}\|^2 \geq \frac{1}{\nu_2} (\gamma^0/\bar{\alpha} + m^0) \quad (4.37)$$

and

$$\gamma^j/(\bar{\alpha} - \alpha^j) + m^j \geq \nu_1 \|\mathbf{d}\|^2 \quad (4.38)$$

thus we have

$$\gamma^j/(\bar{\alpha} - \alpha^j) + m^j \geq (\gamma^0/\bar{\alpha} + m^0) \frac{\nu_1}{\nu_2} > 0 \quad (4.39)$$

Then, by (4.36):

$$\alpha^{j+1} \leq \alpha^j + |\dot{f}(\alpha^j)| \frac{2\nu_2}{(\gamma^0/\bar{\alpha} + m^0)\nu_1} \quad (4.40)$$

If $\dot{f}(\alpha^j) > 0$, α^{j+1} is smaller than α^j then (4.40) still holds. According to assumption 4.1, ∇F is Lipschitz hence:

$$|\dot{f}(\alpha^j) - \dot{f}(0)| \leq L\|\mathbf{d}\|^2 \alpha^j \quad (4.41)$$

Using the fact that $|\dot{f}(\alpha^j)| \leq |\dot{f}(\alpha^j) - \dot{f}(0)| + |\dot{f}(0)|$, and $\dot{f}(0)$ negative, we get:

$$|\dot{f}(\alpha^j)| \leq L\alpha^j\|\mathbf{d}\|^2 - \dot{f}(0) \quad (4.42)$$

Using theorem 4.1 and (4.37)

$$-\dot{f}(0) \leq \alpha^1 \nu \|\mathbf{d}\|^2 \quad (4.43)$$

$$\leq \alpha^1 \frac{\nu}{\nu_1} (m^0 + \gamma_0/\bar{\alpha}) \quad (4.44)$$

Given (4.42), (4.37) and (4.44) jointly with (4.40), we get:

$$\alpha^{j+1} \leq \alpha^j + \frac{2\nu_2}{(m^0 + \gamma_0/\bar{\alpha})\nu_1} \left[L\alpha^j \left(\frac{m^0 + \gamma_0/\bar{\alpha}}{\nu_1} \right) + \alpha^1 \frac{\nu}{\nu_1} (m^0 + \gamma_0/\bar{\alpha}) \right] \quad (4.45)$$

Hence

$$\alpha^{j+1} \leq \alpha^j \left[1 + \frac{2\nu_2 L}{\nu_1^2} \right] + 2\alpha^1 \frac{\nu_2 \nu}{\nu_1^2} \quad (4.46)$$

This corresponds to a recursive definition of the series $\{c_j^{\max}\}$ with:

$$c_{j+1}^{\max} = c_j^{\max} \left[1 + 2\frac{\nu_2 L}{\nu_1^2} \right] + 2\frac{\nu \nu_2}{\nu_1^2} \quad (4.47)$$

Given $c_1^{\max} = 1$, we could deduce the general term of the sequence and have (4.34). \square

LEMMA 4.3. *Let $c \leq \frac{1}{2}$. The sequence $\{\alpha^j\}_{j \geq 0}$ is defined by (3.2). If $\bar{\alpha} > \alpha^{j+1} \geq \alpha^j$, then:*

$$f(\alpha^j) - f(\alpha^{j+1}) + c(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \geq 0 \quad (4.48)$$

Proof. The property is trivial if $\alpha^{j+1} = \alpha^j$. Assume that $\alpha^{j+1} > \alpha^j$. According to equation (4.3), for all $c \leq \frac{1}{2}$,

$$c(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \geq \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \leq 0 \quad (4.49)$$

The tangent majorant has the form:

$$h^j(\alpha, \alpha^j) = f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}m^j(\alpha - \alpha^j)^2 + \psi(\alpha - \alpha^j) \quad (4.50)$$

if we set $a_1 = \gamma^j$. Let us define:

$$\tau(\alpha) = h^j(\alpha, \alpha^j) - f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) \quad (4.51)$$

$$= \frac{1}{2}m^j(\alpha - \alpha^j)^2 + \psi(\alpha - \alpha^j) \quad (4.52)$$

According to Lemma 4.1, for all $(\alpha - \alpha^j)$ in $(0; \bar{\alpha} - \alpha^j)$, hence for all $\alpha \in]\alpha^j; \bar{\alpha}[$:

$$\frac{\psi(\alpha - \alpha^j)}{(\alpha - \alpha^j)\dot{\psi}(\alpha - \alpha^j)} \leq \frac{1}{2} \quad (4.53)$$

Moreover, noting that:

$$\frac{1}{2}m^j(\alpha - \alpha^j)^2 = \frac{1}{2}(\alpha - \alpha^j) [m^j(\alpha - \alpha^j)] \quad (4.54)$$

we deduce:

$$\frac{\tau(\alpha - \alpha^j)}{(\alpha - \alpha^j)\dot{\tau}(\alpha - \alpha^j)} \leq \frac{1}{2} \quad (4.55)$$

$h^j(\alpha, \alpha^j)$ is a tangent majorant of f in α^j :

$$h^j(\alpha^{j+1}, \alpha^j) - f(\alpha^{j+1}) = f(\alpha^j) - f(\alpha^{j+1}) + (\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) + \tau(\alpha^{j+1} - \alpha^j) \geq 0 \quad (4.56)$$

And according to (4.55):

$$f(\alpha^j) - f(\alpha^{j+1}) + (\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{\tau}(\alpha^{j+1} - \alpha^j) \geq 0 \quad (4.57)$$

We made the assumption that α^{j+1} is strictly higher than α^j . So, $\alpha^{j+1} > 0$. The derivative of $h^j(\alpha, \alpha^j)$ is canceling at this point:

$$\dot{\tau}(\alpha^{j+1} - \alpha^j) = -\dot{f}(\alpha^j) \quad (4.58)$$

This quantity is positive according to inequality 4.3. Then we have

$$f(\alpha^j) - f(\alpha^{j+1}) + \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \geq 0.$$

□

We are now able to establish that the proposed MM line search procedure generates steps fulfilling the first Wolfe condition (4.26):

PROPERTY 4.2. *The stepsize (3.2) fulfills (4.26) with*

$$c_1 = (2c_J^{\max})^{-1} \in (0; 1) \quad (4.59)$$

Proof. First, taking $j = 1$ in lemma 4.3, we obtain that the first Wolfe condition holds in α^1 , the minimizer of $h(\alpha, 0)$ if $c_1 \leq \frac{1}{2}$ i.e,

$$f(0) - f(\alpha^1) + c_1\alpha^1\dot{f}(0) \geq 0 \quad (4.60)$$

We have

$$f(0) - f(\alpha^j) \geq f(0) - f(\alpha^1) \quad (4.61)$$

According to equation (4.60),

$$f(0) - f(\alpha^1) + \frac{1}{2}\alpha^1 \dot{f}(0) \geq 0 \quad (4.62)$$

Moreover, according to Property 4.2

$$\alpha^1 \geq \frac{\alpha^J}{c_J^{\max}} \quad (4.63)$$

Hence:

$$f(0) - f(\alpha^J) + \frac{1}{2c_J^{\max}}\alpha^J \dot{f}(0) \geq 0 \quad (4.64)$$

□

The sufficient decrease condition (2.1) holds for arbitrary small values for α and can conduce to ‘false’ convergence to a non-stationary point. In order to avoid too short steps, a second condition is required, for example the second Wolfe condition (2.2). It turned out difficult or even impossible to fulfill the curvature condition (2.2) for any value of J . Fortunately, the stepsize series ensures a condition that is nonetheless sufficient to lead us to convergence results. Lemma 4.4 will be useful to obtain a minoration of the stepsize values.

LEMMA 4.4. $\forall j \geq 1$, there exists c^{\min} such that

$$\alpha^j \geq c^{\min} \alpha^1 \quad (4.65)$$

with

$$c^{\min} = \frac{-1 + \sqrt{1 + 2L/\nu_1}}{2L/\nu_1} \in (0, 1/2) \quad (4.66)$$

Proof. Let ϕ be the concave quadratic function:

$$\phi(\alpha) = f(0) + \alpha \dot{f}(0) + m \frac{\alpha^2}{2} \quad (4.67)$$

with $m = -(m^0 + \gamma^0/\bar{\alpha})/\nu_1 L$. We have $\phi(0) = f(0)$ and $\dot{\phi}(0) = \dot{f}(0) < 0$. So ϕ is decreasing on \mathbb{R}^+ . Let us consider $\alpha \in [0, \alpha^j]$: $\mathbf{x} + \alpha \mathbf{d} \in \mathcal{V}$. According to assumption 4.1, we have

$$|\dot{f}(\alpha) - \dot{f}(0)| \leq \|\mathbf{d}\|^2 L |\alpha| \quad (4.68)$$

and according to assumption 4.2,

$$|\dot{f}(\alpha) - \dot{f}(0)| \leq (m^0 + \gamma^0/\bar{\alpha}) L \alpha / \nu_1 \quad (4.69)$$

Then we obtain:

$$|\dot{f}(\alpha)| \leq (m^0 + \gamma^0/\bar{\alpha}) L \alpha / \nu_1 - \dot{f}(0) \quad (4.70)$$

Hence:

$$\dot{\phi}(\alpha) \leq \dot{f}(\alpha), \quad \forall \alpha \in [0, \alpha^j] \quad (4.71)$$

Integrating (4.71) between 0 and α^j yields

$$\phi(\alpha^j) \leq f(\alpha^j) \quad (4.72)$$

The value of ϕ in α_{\min} reads

$$\begin{aligned} \phi(\alpha_{\min}) &= f(0) + c_{\min} \alpha^1 \dot{f}(0) - \frac{m^0 + \gamma^0 / \bar{\alpha}}{\nu_1} L \frac{(c_{\min} \alpha^1)^2}{2} \\ &= f(0) + \alpha^1 \dot{f}(0) \left(c_{\min} + c_{\min}^2 \frac{m^0 + \gamma^0 / \bar{\alpha}}{-\dot{f}(0) 2\nu_1} L \alpha^1 \right) \end{aligned} \quad (4.73)$$

According to (4.36):

$$\alpha^1 \leq -\frac{2\dot{f}(0)}{m^0 + \gamma^0 / \bar{\alpha}} \quad (4.74)$$

Choosing $c_{\min} = \frac{-1 + \sqrt{1 + 2L/\nu_1}}{2L/\nu_1}$, we have

$$c_{\min} + c_{\min}^2 \frac{L}{\nu_1} = \frac{1}{2} \quad (4.75)$$

Then according to equation (4.60):

$$\phi(\alpha_{\min}) = \phi(c_{\min} \alpha^1) \geq f(0) + \frac{1}{2} \alpha^1 \dot{f}(0) \geq h^0(\alpha^1, 0) \quad (4.76)$$

On the other hand, α^j is positive. Assume that there exists j such that $0 \leq \alpha^j < \alpha_{\min}$. According to (4.72) and given that ϕ is decreasing on \mathbb{R}^+ , we get:

$$f(\alpha^j) \geq \phi(\alpha^j) > \phi(\alpha_{\min}) \geq h^0(\alpha^1, 0) \quad (4.77)$$

which is in contradiction with the majorant property of h . \square

Finally we obtain a minoration of the stepsize values:

PROPERTY 4.3. For all $j \geq 1$,

$$\alpha^j \geq c^{\min} \frac{-\mathbf{g}^T \mathbf{d}}{\nu \|\mathbf{d}\|^2} \quad (4.78)$$

Proof. According to property 4.1, the following minoration holds for α^1 :

$$\alpha^1 \geq \frac{-\mathbf{g}^T \mathbf{d}}{\nu \|\mathbf{d}\|^2} \quad (4.79)$$

Then, using lemma 4.4, we obtain relation (4.78). \square

5. Convergence results. This section discusses the convergence of the iterative descent algorithm

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \quad k = 1, \dots, K \quad (5.1)$$

when \mathbf{d}_k satisfies $\mathbf{g}_k^T \mathbf{d}_k < 0$ and the line search is performed using the proposed MM strategy.

5.1. 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 \mathbf{d}_k . Convergence proofs are often based on the fulfillment of Zoutendijk condition

$$\sum_{k=0}^{\infty} \|\mathbf{g}_k\|^2 \cos^2 \theta_k < \infty, \quad (5.2)$$

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

$$\cos \theta_k = -\mathbf{g}_k^T \mathbf{d}_k / (\|\mathbf{g}_k\| \|\mathbf{d}_k\|). \quad (5.3)$$

Inequality (5.2) implies that $\cos \theta_k \|\mathbf{g}_k\|$ vanishes for large values of k . Moreover, provided that \mathbf{d}_k is not orthogonal to $-\mathbf{g}_k$ (i.e., $\cos \theta_k > 0$), condition (5.2) implies the convergence of the algorithm in the sense

$$\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0. \quad (5.4)$$

The property holds for classical line search methods such as backtracking or Wolfe line search [21]. In the case of the proposed line search, the following result holds [5].

PROPERTY 5.1. *Let α_k be defined by (3.2). Then Zoutendijk condition (5.2) holds.*

Proof. First, note that for all k , $\mathbf{d}_k \neq 0$, because we make the assumption:

$$\mathbf{g}_k^T \mathbf{d}_k < 0$$

According to Lemma 4.2, the first Wolfe condition holds for $c_1 = \frac{1}{2c_J^{\max}}$:

$$F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}) \geq -c_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k$$

According to equation 4.78:

$$\alpha_k \geq c^{\min} \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\nu \|\mathbf{d}_k\|^2}$$

Hence:

$$F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}) \geq c_0 \frac{(\mathbf{g}_k^T \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \geq 0$$

with $c_0 = (c^{\min} c_1) / \nu > 0$. Assumption 4.1 and the boundedness of \mathcal{L}_0 implies that the limit $\lim_{k \rightarrow \infty} F(\mathbf{x}_k)$ is finite. Therefore:

$$\infty > [F(\mathbf{x}_0) - \lim_{k \rightarrow \infty} F(\mathbf{x}_k)] / c_0 \geq \sum_k \frac{(\mathbf{g}_k^T \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \quad (5.5)$$

□

5.2. Gradient related algorithms. A general convergence result can be established by using the concept of *gradient related* direction [1]. A direction sequence $\{\mathbf{d}_k\}$ is said gradient related to $\{\mathbf{x}_k\}$ if the following property can be shown:

For any subsequence $\{\mathbf{x}_k\}_{\mathcal{K}}$ that converges to a nonstationary point, the corresponding subsequence $\{\mathbf{d}_k\}_{\mathcal{K}}$ is bounded and satisfies

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} \mathbf{g}_k^T \mathbf{d}_k < 0 \quad (5.6)$$

Roughly, this means that \mathbf{d}_k does not become 'too small' or 'too large' relative to \mathbf{g}_k , and that the angle between \mathbf{d}_k and \mathbf{g}_k does not get 'too close' to 90 degrees.

The following convergence result holds:

THEOREM 5.1. [25]

Let $\{\mathbf{x}_k\}$ a sequence generated by a descent method $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$. Assume that the sequence $\{\mathbf{d}_k\}$ is gradient related to $\{\mathbf{x}_k\}$ and that Zoutendijk condition (5.2) holds then the descent algorithm converges in the sense $\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$.

As we will show, Theorem 5.1 leads to the convergence of the truncated Newton method and the projected gradient method for constrained optimization.

5.2.1. Preconditioned gradient algorithms. Let consider the family of descent algorithms when the search direction has the form

$$\mathbf{d}_k = -\mathbf{B}_k \mathbf{g}_k$$

with \mathbf{B}_k a positive definite symmetric matrix. In the steepest descent method \mathbf{B}_k is simply the identity matrix \mathbf{I} , while in Newton's method \mathbf{B}_k is the exact inverse Hessian $\nabla^2 F(\mathbf{x}_k)$. In quasi-Newton methods, \mathbf{B}_k is an approximation to the inverse Hessian that is updated at every iteration by means of a low rank formula. Since \mathbf{B}_k is positive definite, \mathbf{d}_k is a descent direction. Moreover, if we assume that the matrices \mathbf{B}_k have a uniformly bounded condition number, that is, there exists $M_1, M_2 > 0$ such that

$$M_1 \|\mathbf{z}\|^2 \leq \mathbf{z}^T \mathbf{B}_k \mathbf{z} \leq M_2 \|\mathbf{z}\|^2, \quad (5.7)$$

it can be seen that $\{\mathbf{d}_k\}$ is gradient related (provided \mathbf{x}_k is nonstationary for all k) [2] and the descent algorithm converges in the sense $\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$ according to Theorem 5.1.

5.2.2. Truncated Newton method. The truncated Newton method corresponds to an inexact form of the Newton method where the search direction is computed by applying the conjugate gradient method to the Newton equations. Since the conjugate gradient iterations are stopped before convergence, this method is known as *truncated Newton* method. The direction \mathbf{d}_k is computed by solving approximately the linear system $\nabla^2 F(\mathbf{x}_k) \mathbf{d} = -\mathbf{g}_k$ with I_k CG iterates. Let make the following assumption:

ASSUMPTION 5.1. For all $\mathbf{x} \in \mathcal{V}$, $\mathbf{H} = \nabla^2 F(\mathbf{x})$ is a symmetric positive definite (SPD) matrix. Let $\nu_1(\mathbf{H}) > 0$ and $\nu_2(\mathbf{H}) > 0$ denote the smallest and largest eigenvalues of \mathbf{H} . The matrix sequence $\mathcal{H} = \{\nabla^2 F(\mathbf{x}_k)\}$ has a uniformly bounded spectrum with a strictly positive lower bound i.e., there exist $\nu_1(\mathcal{H}), \nu_2(\mathcal{H}) \in \mathbb{R}$ such that

$$\nu_2(\mathcal{H}) \geq \nu_2(\mathbf{H}_k) \geq \nu_1(\mathbf{H}_k) \geq \nu_1(\mathcal{H}) > 0, \quad \forall k$$

LEMMA 5.1. [16]

Let $\{\mathbf{x}_k\}$ be a sequence generated by the truncated Newton method and assume that assumption 5.1 holds. Then there exists $\eta_1, \eta_2 > 0$ such that

$$\eta_1 \|\mathbf{g}_k\|^2 \leq -\mathbf{d}_k^T \mathbf{g}_k \quad (5.8)$$

$$\|\mathbf{d}_k\|^2 \leq \eta_2 \|\mathbf{g}_k\|^2 \quad (5.9)$$

Thus, the direction sequence $\{\mathbf{d}_k\}$ is gradient related to $\{\mathbf{x}_k\}$ [2] and Theorem 5.1 implies the convergence of the algorithm.

5.2.3. Feasible directions methods for constrained optimization. Consider the constrained problem

$$\begin{aligned} & \text{minimize} && F(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \mathcal{C} \end{aligned} \quad (5.10)$$

where \mathcal{C} is a nonempty, closed, and convex set. We examine here the properties of convergence of algorithms belonging to the class of the feasible direction methods.

Given a feasible vector \mathbf{x} , a feasible direction at \mathbf{x} is a vector $\mathbf{d} \neq 0$ such that $\mathbf{x} + \alpha \mathbf{d}$ is feasible for all $\alpha > 0$ that are sufficiently small [2]. Starting with a feasible vector \mathbf{x}_0 , the method generates a sequence of feasible vectors according to

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

where \mathbf{d}_k is a feasible direction at \mathbf{x}_k and also a descent direction, and the step size is chosen positive and such that $\mathbf{x}_{k+1} \in \mathbf{X}$. In the case where \mathbf{X} is a convex set, a feasible direction method can be written in the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k (\mathbf{x}'_k - \mathbf{x}_k)$$

where $\alpha_k \in [0, 1)$ and if \mathbf{x}_k is nonstationnary

$$\mathbf{x}'_k \in \mathbf{X}, \quad \mathbf{g}_k^T (\mathbf{x}'_k - \mathbf{x}_k) < 0.$$

The gradient projection method is a particular case of feasible direction method where

$$\mathbf{x}'_k = \mathcal{P}_{\mathbf{X}} [\mathbf{x}_k - s_k \mathbf{g}_k]$$

with $\mathcal{P}_{\mathbf{X}}[\cdot]$ denoting projection on the set \mathbf{X} and s_k is a positive scalar.

Convergence analysis of feasible direction methods is very close to those of descent direction methods in the unconstrained case. In particular, we have the following property

PROPERTY 5.2. [2]

Let $\{\mathbf{x}_k\}$ a sequence generated by any of the following feasible direction methods

- conditionnal gradient
- gradient projection with parameter $s_k = s$ constant
- scaled gradient projection with $s_k = s$ constant and uniformly bounded scaling matrices

Then, the direction sequence $\{\mathbf{d}_k\}$ is gradient related to $\{\mathbf{x}_k\}$.

Thus, since the MM linesearch ensures Zoutendijk condition, Theorem 5.1 implies the convergence of those three constrained optimization algorithms.

5.3. Convergence of conjugate gradient methods. This section is focused on the convergence of the nonlinear conjugate gradient algorithm (NLCG) defined by the following recurrence

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{d}_k \\ \mathbf{d}_k &= -\mathbf{c}_k \operatorname{sign}(\mathbf{g}_k^T \mathbf{c}_k) \\ \mathbf{c}_k &= -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}\end{aligned}\tag{5.11}$$

First, let us consider the conjugacy formulas of the form:

$$\beta_0 = 0, \quad \beta_k = \beta_k^{\mu_k, \omega_k} = \mathbf{g}_k^T \mathbf{y}_{k-1} / D_k, \quad \forall k > 0\tag{5.12}$$

with

$$D_k = (1 - \mu_k - \omega_k) \|\mathbf{g}_{k-1}\|^2 + \mu_k \mathbf{d}_{k-1}^T \mathbf{y}_{k-1} - \omega_k \mathbf{d}_{k-1}^T \mathbf{g}_{k-1}$$

$$\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}$$

$$\mu_k \in [0, 1], \quad \omega_k \in [0, 1 - \mu_k]$$

Expression (5.12) allows us to cover the following conjugate gradient methods:

$$\begin{aligned}\beta_k^{1,0} &= \mathbf{g}_k^T \mathbf{y}_{k-1} / \mathbf{d}_{k-1}^T \mathbf{y}_{k-1} && \text{Hestenes-Stiefel (HS)} \\ \beta_k^{0,0} &= \mathbf{g}_k^T \mathbf{y}_{k-1} / \|\mathbf{g}_{k-1}\|^2 && \text{Polak-Ribire-Polyak (PRP)} \\ \beta_k^{0,1} &= -\mathbf{g}_k^T \mathbf{y}_{k-1} / \mathbf{d}_{k-1}^T \mathbf{g}_{k-1} && \text{Liu-Storey (LS)}\end{aligned}$$

In the sequel, Assumption 4.1 will appear to be sufficient for the global convergence of the CG method when $k = 0$ and $\omega_k \in [0, 1]$, which encompasses the PRP and the LS cases, but not the HS case. Let us consider the following stronger assumption to deal with the more general case $k \in [0, 1], \omega_k \in [0, 1 - k]$.

ASSUMPTION 5.2. *Assumption 4.1 holds and F is strongly convex on \mathcal{V} : there exists $\lambda > 0$ such that*

$$[\nabla F(\mathbf{x}) - \nabla F(\mathbf{x}')]^T (\mathbf{x} - \mathbf{x}') \geq \lambda \|\mathbf{x} - \mathbf{x}'\|^2, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathcal{N}$$

We have the following convergence result:

THEOREM 5.2. *The NLCG algorithm is convergent in the sense $\liminf_{k \rightarrow \infty} \mathbf{g}_k = 0$ when α_k is defined by (3.2) and β_k is chosen according to the PRP and LS methods, and more generally for $\mu_k = 0$ and $\omega_k \in [0, 1]$. Moreover, if assumption 5.2 holds, then we have $\liminf_{k \rightarrow \infty} \mathbf{g}_k = 0$ in all cases.*

Proof. We have previously established:

- the inequality on α_k^1 (Property 4.78)
- the step size minorization $\alpha_k \leq c_f^{\max} \alpha_k^1$ (Property 4.2)
- the step size majorization $0 \leq c_f^{\min} \alpha_k^1 \leq \alpha_k$ (Property 4.4)
- the verification of Zoutendijk condition (Property 5.1)

Thus, the proof of Property 5.2 is identical to that in [17]. This result can be viewed as an extension of [17, Th.4.1] for a new form of tangent majorant. \square

The convergence results can be extended to others conjugacy formulas if we make an additional assumption on the tangent majorant:

ASSUMPTION 5.3. For all $\alpha \geq \alpha^j$:

$$\dot{f}(\alpha) - \dot{h}^j(\alpha, \alpha^j) \leq 0$$

REMARK 2. The assumption holds if m_k^j and γ_k^j are chosen according to the strategy developed in §3 assuming that the parameter m_p^j ensures

$$\dot{p}(\alpha) - \dot{q}^j(\alpha, \alpha^j) \leq 0, \forall \alpha \geq \alpha^j$$

LEMMA 5.2. Let us assume that $\dot{f}(0) < 0$. Then, if assumption 5.3 holds, the series defined by (3.2) is increasing. Moreover, the derivative of f at α^j is negative for all j .

This lemma means that for all iteration k of the overall algorithm, we have the inequality:

$$\mathbf{g}_k^T \mathbf{d}_{k-1} \leq 0 \quad (5.13)$$

Lemma 5.2 leads to an important result for the convergence properties of conjugate gradient methods:

LEMMA 5.3. *Sufficient descent condition*

Assume that assumption 5.3 holds and suppose that the successive directions \mathbf{d}_k are given by the conjugate algorithm gradient:

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$$

If β_k is non negative, the sufficient descent condition holds for each iteration k i.e., there exists some $0 < c \leq 1$ such that for all k :

$$\mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2$$

In particular, \mathbf{d}_k is a descent direction.

Lemma 5.3 is a direct application of a remark made in [11, Part 4]. It is a consequence of inequality (5.13). We can directly use this result to prove global convergence of Fletcher-Reeves (FR) method when:

$$\beta_k^{\text{FR}} = \frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2} \geq 0$$

According to Lemma 5.3, the FR method always generates descent directions with our choice of step, if we assume that assumption 5.3 holds. Then we can use [6] result: The Zoutendijk condition holds according to theorem 5.1 and \mathbf{d}_k is a descent direction. So the FR method converges in the sense:

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$$

Let consider now the conjugacy formula PRP+, proposed in [11]:

$$\beta_k = \max(\beta_k^{\text{PRP}}, 0) \quad \text{with } \beta_k^{\text{PRP}} = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\|\mathbf{g}_k\|}$$

According to [21], PRP+ method has led to the better convergence results during numerical tests when compared with others conjugacy formulas. The convergence of the PRP+ method associated with the MM line search is a direct application of [11, Th.4.3]. The PRP+ method converges in the sense:

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$$

Finally, we can state a similar convergence result in the convex case for the Dai and Yuan (DY) conjugacy formula:

$$\beta_k^{\text{DY}} = \frac{\|\mathbf{g}_{k+1}\|^2}{\mathbf{d}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}$$

Let us make the following assumption:

ASSUMPTION 5.4. *Assumption 4.1 holds and F is convex on \mathcal{V} : For every $(\mathbf{x}, \mathbf{y}) \in \mathcal{V}$ we have*

$$F(\omega\mathbf{x} + (1 - \omega)\mathbf{y}) \leq \omega F(\mathbf{x}) + (1 - \omega)F(\mathbf{y}), \quad \forall \omega \in [0, 1]$$

According to Lemma 5.2, if assumption 5.4 holds, for all iteration k of the overall algorithm, we have the inequality:

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq |\mathbf{g}_k^T \mathbf{d}_k| \quad (5.14)$$

Let us show recurrently on k that, with DY method, \mathbf{d}_k is always a descent direction. For $k = 0$, we have $\mathbf{d}_0 = -\mathbf{g}_0$, hence \mathbf{d}_0 is a descent direction. Consider an index k when $\mathbf{g}_k^T \mathbf{d}_k \leq 0$. If this quantity is zero, $\alpha_k = 0$ and then the algorithm will finish. Let us assume that $\mathbf{g}_k^T \mathbf{d}_k < 0$. Then, according to Lemma 5.2, $\mathbf{g}_{k+1}^T \mathbf{d}_k \leq 0$ and moreover, according to inequality (5.14), $|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq |\mathbf{g}_k^T \mathbf{d}_k|$. Then, the coefficient β_k^{DY} is positive:

$$\beta_k^{\text{DY}} = \frac{\|\mathbf{g}_{k+1}\|^2}{\mathbf{d}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)} \geq 0$$

And according to Lemma 5.3, \mathbf{d}_{k+1} is a descent direction. Now we can use [6] result to show the global convergence of DY method: The Zoutendijk condition holds according to theorem 5.1 and \mathbf{d}_k is a descent direction. If assumption 5.4 holds, then the DY method converges in the sense:

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$$

THEOREM 5.3. *Let α_k be defined by the recurrence (3.2), and let assumptions 4.1, 4.2 and 5.3 hold. Then, we have convergence in the sense $\liminf_{k \rightarrow \infty} \mathbf{g}_k = 0$ for the PRP+ and FR. Moreover, if assumption 5.4 holds, we have convergence in the sense $\liminf_{k \rightarrow \infty} \mathbf{g}_k = 0$ for the DY method.*

6. Numerical example. The goal of this section is to analyse the performance of descent optimization algorithms when the step size is obtained by the proposed MM line search procedure.

Let consider the following quadratically constrained quadratic optimization problem

$$\begin{aligned} \min F_0(\mathbf{x}) &= \rho_0 + a_0^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{A}_0 \mathbf{x} \\ \text{s.t. : } q_i(\mathbf{x}) &= -\frac{1}{2} \mathbf{x}^T \mathbf{A}_i \mathbf{x} + a_i^T \mathbf{x} + \rho_i \geq 0 \end{aligned} \quad (6.1)$$

where $\mathbf{A}_i, i = 0, \dots, m$ are positive semidefinite $(n \times n)$ matrices. (6.1) can be solved with an interior point method. The augmented criterion has the form (1.4) where the barrier function is

$$B(\mathbf{x}) = -\sum_{i=1}^m \log q_i(\mathbf{x}) \quad (6.2)$$

For a decreasing sequence of μ , the augmented criterion $F_\mu(\mathbf{x})$ is minimized using Newton method:

$$\mathbf{d}_k = -\nabla^2 F_\mu^{-1}(\mathbf{x}_k) \nabla F_\mu(\mathbf{x}_k) \quad (6.3)$$

The stepsize α_k must be inferior to $\bar{\alpha}$ defining the limit of feasibility of $F_\mu(\mathbf{x}_k + \alpha \mathbf{d}_k)$. In this example, $\bar{\alpha}$ can be computed exactly. Given \mathbf{x}_k and a descent direction \mathbf{d}_k ,

$$q_i(\mathbf{x} + \alpha \mathbf{d}) = q_1 \alpha^2 + q_2 \alpha + q_3 \quad (6.4)$$

with $q_1 = -\frac{1}{2} \mathbf{d}_k^T \mathbf{A}_i \mathbf{d}_k$, $q_2 = -\frac{1}{2} \mathbf{d}_k^T \mathbf{A}_i \mathbf{x}_k - \frac{1}{2} \mathbf{x}_k^T \mathbf{A}_i \mathbf{d}_k + a_i^T \mathbf{d}_k$ and $q_3 = -\frac{1}{2} \mathbf{x}_k^T \mathbf{A}_i \mathbf{x}_k + a_i^T \mathbf{x}_k + \rho_i$. Hence $\bar{\alpha}$ is the smallest positive root of the concave polynomes $q_i(\mathbf{x}_k + \alpha \mathbf{d}_k)$.

A classical choice for α_k is ensured by backtracking line search. Starting with the feasible step $\alpha = 0.99 \bar{\alpha}$, the stepsize is reduced until it fulfills Armijo condition (2.1).

In the context of interior point methods, [20] developed a *damped Newton* method to minimize the augmented criterion F_μ . The Newton direction \mathbf{d}_k is damped by a factor $\alpha_k \in (0, 1]$ ensuring that $\mathbf{x}_k + \alpha_k \mathbf{d}_k$ is feasible and the criterion decreases by at least some fixed amount. Defining the *Hessian norm*

$$\|\mathbf{u}\|_{\mathbf{x}} = \sqrt{\mathbf{u}^T \nabla^2 F_\mu(\mathbf{x}) \mathbf{u}}, \quad (6.5)$$

the damping factor is given by

$$\alpha_k = \frac{1}{1 + \|\mathbf{d}_k\|_{\mathbf{x}_k}} \quad (6.6)$$

Its convergence properties are based on the self concordancy of F_μ [20]. The interior point method for the resolution of the constrained problem (6.1) is given on table 6.1.

We propose to compare the backtracking and the damping procedure with our MM line search. In particular, we take 20 random problems of size $n = 400$, $m = 200$. The line search parameters are fixed to $c_1 = 0.01$ and $J = 1$. Table 6.2 reports the performances of the overall interior point algorithm associated with different line search procedures for the minimization of F_μ . The algorithm is initialized with a

- 1) Set $\mu = 1$, $t = 0$, $K = 0$, and a tolerance ϵ and select a feasible point \mathbf{x}_0
- 2) **WHILE** (6.7) not ensured,
 - Compute Newton direction of F_μ
 - Compute step size
 - $K(t) \leftarrow K(t) + 1$
- 3) $\ell \leftarrow 0$
- 4) **IF** $\mu < \epsilon$, **RETURN**
 ELSE Set $\mu = \theta\mu$, $t = t + 1$ and go to step 2.

TABLE 6.1

Interior point algorithm

	Backtracking	MM	Damping
$E(K)$	283.5	66.2	134.9
$\sigma(K)$	26.33	1.13	2.73
$E(T)$	1538.2	16.07	40.13
$\sigma(T)$	879.98	1.61	1.33

TABLE 6.2

Means and standard deviations of the total inner iteration numbers $\sum_i K(i)$ and the time before convergence T with $\epsilon = 10^{-8}$ over 10 random problems.

uniform null object and $\mu = 1$. The convergence of the inner iterations is checked using the following stopping rule [3]

$$\frac{1}{2}(\mathbf{d}_k^T \mathbf{g}_k)^2 \leq 10^{-5} \tag{6.7}$$

It can be noted that the Newton algorithm with MM line search requires less iterations than the backtracking or damped Newton approaches. Moreover, although it requires the exact computation of $\bar{\alpha}$, it is faster than the two other procedures. Let us emphasize that the damping strategy is dedicated to the particular case when \mathbf{d} is the Newton direction. It must be modified when the minimization of F_μ is obtained by means of others algorithms (See [14] for the conjugate gradient case). On the opposite, the design of the proposed MM line search does not depend on the descent algorithm used.

Conclusion. In [17], 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 that is encountered for example in interior point methods for constrained optimization. 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.

Then, it can be used in several constrained programming applications using interior point methods [27, 4, 15] and more generally in optimization problems when the criteria contains a barrier term such as maximum entropy reconstruction [26] or image processing under Poisson noise [22].

Since the proposal scheme requires the exact computation of the step to the boundary of the feasible region, it is restricted to problems when the constraints are simple such as linear and quadratic programming. In case of nonlinear constraints, we advise to proceed with the procedure described in [19] to build the stepsize.

In this paper, the analysis is limited to the logarithmic barrier function and the entropy function. The extension of the present results to other forms of barriers such as cross-entropy [23] or inverse function [7] should be discussed.

REFERENCES

- [1] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, Belmont, MA, 2nd ed., 1996.
- [2] ———, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 2nd ed., 1999.
- [3] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, New York, 1st ed., 2004.
- [4] E. CANDÈS AND J. ROMBERG, *L1-magic: Recovery of sparse signals via convex programming*, tech. report, California Institute of Technology, October 2005. <http://www.acm.caltech.edu/l1magic/downloads/l1magic.pdf>.
- [5] E. CHOZENOUX, S. MOUSSAOUI, AND J. IDIER, *A new line search method for barrier functions with strong convergence properties*, tech. report, IRCCyN, 2009. <http://hal.archives-ouvertes.fr/IRCCYN-ADTSI>.
- [6] Y. DAI AND Y. YUAN, *A three-parameter family of nonlinear conjugate gradient methods*, *Mathematics of Computation*, 70 (2001), pp. 1155–1167.
- [7] D. DEN HERTOOG, C. ROOS, AND T. TERLAKY, *Inverse barrier methods for linear programming*, *Revue française d'automatique, informatique, recherche opérationnelle*, 28 (1994), pp. 135–163.
- [8] M. DOYLE, *A Barrier Algorithm for Large Nonlinear Optimization Problems*, PhD thesis, University of Stanford, 2003. www.stanford.edu/group/SOL/dissertations/maurenthesis.pdf.
- [9] H. ERDOGAN AND J. FESSLER, *Monotonic algorithms for transmission tomography*, *IEEE Transactions on Medical Imaging*, 18 (1999), pp. 801–814.
- [10] A. FORSGREN, P. GILL, AND M. WRIGHT, *Interior methods for nonlinear optimization*, *SIAM Review*, 44 (2002), pp. 525–597.
- [11] J. C. GILBERT AND J. NOCEDAL, *Global convergence properties of conjugate gradient methods for optimization*, *SIAM Journal on Optimization*, 2 (1992), pp. 21–42.
- [12] D. R. HUNTER AND K. L., *A tutorial on MM algorithms*, *The American Statistician*, 58 (2004), pp. 30–37.
- [13] M. JACOBSON AND J. FESSLER, *An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms*, *IEEE Transactions on Image Processing*, 16 (2007), pp. 2411–2422.
- [14] H. JI, M. HUANG, J. MOORE, AND J. MANTON, *A globally convergent conjugate gradient method for minimizing self-concordant functions with application to constrained optimisation problems*, in *American Control Conference*, July 2007, pp. 540–545.
- [15] S.-J. KIM, K. KOH, M. LUSTIG, S. BOYD, AND D. GORINEVSKY, *An interior-point method for large-scale L1-regularized least squares*, *IEEE Journal of Selected Topics in Signal Processing*, 1 (2007), pp. 606–617.
- [16] C. LABAT AND J. IDIER, *Convergence of truncated half-quadratic and Newton algorithms, with application to image restoration*, technical report, IRCCyN, June 2007.
- [17] ———, *Convergence of conjugate gradient methods with a closed-form stepsize formula*, *Journal of Optimization Theory and Applications*, 136 (2008), pp. 43–60.
- [18] J. J. MORÉ AND D. J. THUENTE, *Line search algorithms with guaranteed sufficient decrease*, *ACM Transactions on Mathematical Software*, 20 (1994), pp. 286–307.
- [19] W. MURRAY AND M. H. WRIGHT, *Line search procedures for the logarithmic barrier function*, *SIAM Journal on Optimization*, 4 (1994), pp. 229–246.
- [20] Y. NESTEROV AND A. NEMIROVSKII, *Interior point polynomial algorithms in convex programming*, no. 13 in *Studies in Applied and Numerical Mathematics*, SIAM, Philadelphia, Pennsylvania, 1994.
- [21] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, NY, 1999.
- [22] J. M. OLLINGER AND J. A. FESSLER, *Positron-emission tomography*, *IEEE Signal Processing Magazine*, 14 (1997), pp. 43–55.
- [23] J. O’SULLIVAN, *Roughness penalties on finite domains*, *IEEE Transactions on Image Processing*, 4 (1995).
- [24] C. ROOS, T. TERLAKY, AND J. VIAL, *Interior Point Methods for Linear Optimization*, Springer-Verlag, New York, NY, 2nd ed., 2006.
- [25] Z.-J. SHI, *Convergence of line search methods for unconstrained optimization*, *Applied Mathematics and Computations*, 157 (2004), pp. 393–405.
- [26] J. SKILLING AND R. K. BRYAN, *Maximum entropy image reconstruction: General algorithm*, *Monthly Notices of the Royal Astronomical Society*, 211 (1984), pp. 111–124.
- [27] L. VANDENBERGHE, S. BOYD, AND S. WU, *Determinant maximisation with linear matrix in-*

- equality constraints*, SIAM Journal on Mathematical Analysis, 19 (1998), pp. 499–533.
- [28] M. H. WRIGHT, *Interior methods for constrained optimization*, in Acta Numerica 1992, Cambridge University Press, 1991, pp. 341–407.