A new line search method for barrier functions optimization with strong convergence properties

Emilie Chouzenoux, Saïd Moussaoui, J. Idier

A new line search method for barrier functions optimization with strong convergence properties

E. Chouzenoux[a], S. Moussaoui[a], J. Idier[a]

[a]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 24, fax: + 33 2 40 37 69 30,
Email: {emilie.chouzenoux,jerome.idier,said.moussaoui}@irccyn.ec-nantes.fr

---

Abstract

Descent direction algorithms are widely used for solving unconstrained minimization problems. After computing a descent direction, the algorithm performs a search along the line supported by this direction to get the value of the step size. Recently, several papers [2, 9] have introduced a new line search strategy founded on the theory of Majorization-Minimization algorithms [6]. An iterate of the line search algorithm is the minimizer of a quadratic majorant approximation of the criterion along the search direction. This approach, very easy to implement, provides good convergence properties to the overall algorithm, namely when combining with the conjugate gradient method. However, in the case of functions having a linear barrier, i. e. tending to infinity when some linear constraints are active, quadratic tangent majorant won't be well suited to approximate such functions. Then we suggest to appropriately modify the line search strategy by choosing a new form of tangent majorant. Since the majorant function is still not quadratic, precedent results of convergence analysis do not hold anymore. A complete convergence analysis is thus performed. This approach yields to a new efficient line search method for functions having a linear barrier.

Contents

1. Problem Statement

Iterative descent direction methods are widely used for solving unconstrained minimization problems. After computing a search direction, $\boldsymbol{d}$, such an algorithm performs a search along the line supported by the direction to find an adequate step size value, α. The determination of an acceptable step size is often based on verifying some sufficient convergence conditions, such as the Wolfe conditions. Some existing strategies for finding the step size are:

- Exact minimization (constrained or not) of the scalar function $f(\alpha) = F(\boldsymbol{x} + \alpha \boldsymbol{d})$,
- Backtracking or more generally dichotomy: the step is reduced until a stopping condition holds (for example Armijo rule),
- Approximation of the function using interpolation method [14, 11]. They consist in a succession of two phases : a *bracketing phase* that finds an interval containing at least one step verifying the stopping conditions (Wolfe conditions for example). Such a step is obtained by exploring the interval during a *selection phase*, using an interpolation (often polynomial) of $f(\alpha)$.
- Constant step size [1, Prop.1.2.3]. This method is based on a quadratic majorant approximation of $f(\alpha)$, assuming a Lipshitz continuity on the gradient of F with Lipshitz parameter L. Then the step size is simply the minimizer of this majorizing parabola (figure 1).

Recently, a generalization of this last line search method has been introduced by [2, 9]. Before presenting this methodology, let us introduce some necessary properties of the criterion $F(\boldsymbol{x})$.

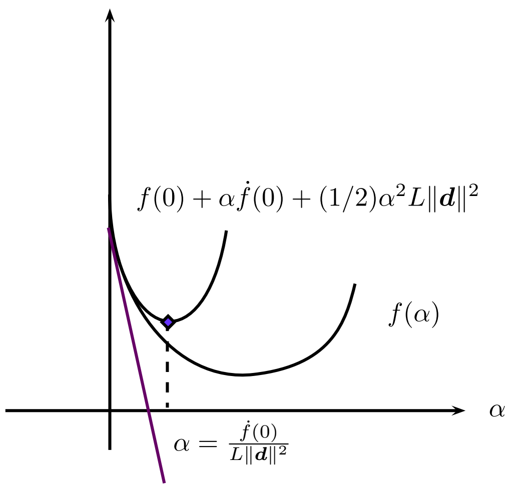

Figure 1: Constant stepsize idea from [1]

Assumption 1.1. *Gradient Lipschitz*
The level set $\mathcal{L}_0 = \{\boldsymbol{x} | F(\boldsymbol{x}) \leqslant F(\boldsymbol{x}_0)\}$ is assumed bounded. $F(\boldsymbol{x})$ is differentiable on a neighbourhood $\mathcal{V}$ of $\mathcal{L}_0$ and $\nabla F(\boldsymbol{x})$ is Lipschitz continuous on $\mathcal{V}$, i.e., there exists $0 < L < \infty$ such that

$$\|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\| \leqslant L\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} \tag{1.1}$$

Assumption 1.2. *There exists a series of semipositive definite (SPD) matrices $(\boldsymbol{M}_{\boldsymbol{x}})$ such that*

$$Q_{\boldsymbol{x}}(\boldsymbol{x}', \boldsymbol{x}) \geqslant F(\boldsymbol{x}'), \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{V}$$

for all $\boldsymbol{x} \in \mathcal{V}$*, where :*

$$Q_{\boldsymbol{x}}(\boldsymbol{x}', \boldsymbol{x}) = F(\boldsymbol{x}) + (\boldsymbol{x}' - \boldsymbol{x})^T \nabla F(\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{x}' - \boldsymbol{x})\boldsymbol{M}_{\boldsymbol{x}}(\boldsymbol{x}' - \boldsymbol{x}) \tag{1.2}$$

Moreover, matrices $M_{\boldsymbol{x}}$ *have a uniformly bounded spectrum, i. e. , there exist* $\nu_1, \nu_2 \in \mathbb{R}$ *,* $\nu_2 \geqslant \nu_1 > 0$ *such that*

$$\nu_1 \|v\|^2 \leqslant v^T M_{\boldsymbol{x}} v \leqslant \nu_2 \|v\|^2, \quad \forall \boldsymbol{x} \in \mathcal{V}$$

The quadratic function $Q_{\boldsymbol{x}}(\boldsymbol{x}', \boldsymbol{x})$ defined in (1.2) is a *tangent majorant* of $F(\boldsymbol{x})$ at $\boldsymbol{x}$ owing to the following properties:

$$\begin{cases} Q_{\boldsymbol{x}}(\boldsymbol{x}', \boldsymbol{x}) & \geqslant & F(\boldsymbol{x}'), \quad \forall \boldsymbol{x}, \, \boldsymbol{x}' \in \mathcal{V} \\ Q_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{x}) & = & F(\boldsymbol{x}) \\ \nabla Q_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{x}) & = & \nabla F(\boldsymbol{x}) \end{cases}$$

1.1. Line search using a quadratic majorant function

From this quadratic approximation of the criterion, a tangent majorant of the scalar function $f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$ can be deduced:

$$q_k(\alpha, \alpha_k) = f(\alpha_k) + (\alpha - \alpha_k)\dot{f}(\alpha_k) + \frac{1}{2} m_k (\alpha - \alpha_k)^2$$

with $m_k = \boldsymbol{d}_k^T M_{\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k} \boldsymbol{d}_k$. The minimizer of this scalar tangent majorant is known exactly. Starting from an initial value α_k^0, a sequence $\left\{\alpha_k^j\right\}_{j \geqslant 1}$ is computed by applying a Majorize-Minimize (MM) algorithm (figure 2) to the function $f(\alpha)$, according to:

$$\alpha_k^{j+1} = \operatorname{argmin} q_k^j(\alpha, \alpha_k^j)$$

This strategy leads to Algorithm 1. The convergence of this algorithm is proven in [9], whatever the value of J, for a family of non linear conjugate gradient methods covering classical conjugacy formulas such as Polak-Ribière or Hestenes-Stiefel. Experimentally the algorithm performed better when J is close to 1 although the higher J is, the closer α_k is to an exact minimizer of $f(\alpha)$. The convergence result [9, Th.4.1] is a generalization of the one described in [2], where J is fixed to 1 and θ must in a set of $]0;1[$, depending on the Lipschitz constant L.

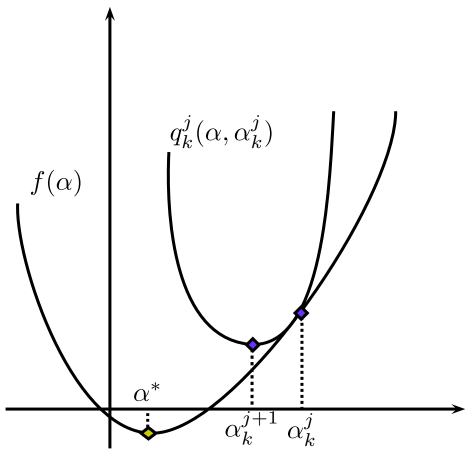

Figure 2: MM quadratic line search in the convex case

Algorithm 1 Descent direction algorithm with MM quadratic line search

Require: $\boldsymbol{x}_0$, $\boldsymbol{d}_0 = -\nabla F(\boldsymbol{x}_0)$, $J \geqslant 1$, $\theta \in]0,1[$
Ensure: $\boldsymbol{x}_k$ verifying a stopping condition SC
  while SC non verified **do**
    if $\boldsymbol{d}_k = 0$ **then**
      $\alpha_k = 0$
    else
      $\alpha_k^0 = 0$
      $\alpha_k^{j+1} = \alpha_k^j - \theta \frac{\boldsymbol{d}_k^T \nabla F(\boldsymbol{x}_k + \alpha_k^j \boldsymbol{d}_k)}{m_k^j}$, $0 \leqslant j \leqslant J-1$
      $\alpha_k = \alpha_k^J$
    end if
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
    $\boldsymbol{d}_{k+1}$ defined such that $\nabla F(\boldsymbol{x}_{k+1})^T \boldsymbol{d}_{k+1} \leqslant 0$
  end while

1.2. Line search in the case of barrier functions

In this paper, we consider the following optimization problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x}), \quad \mu \geqslant 0 \tag{1.3}$$

where P is differentiable and B is a *linear barrier* function, i.e. $B(\boldsymbol{x})$ tends to infinity when some of the constraints $c_i(\boldsymbol{x}) = [\boldsymbol{A}\boldsymbol{x}]_i + \theta_i$ are vanishing. For example, B could be the logarithmic barrier function:

$$B(\boldsymbol{x}) = -\sum_{i=1}^{\mathcal{I}} t_i \log([\boldsymbol{A}\boldsymbol{x}]_i + \theta_i), \quad t_i > 0, \forall i$$

The barrier function B is defined on the convex set $\mathcal{C}^*$:

$$\mathcal{C}^* = \{\boldsymbol{x} | [\boldsymbol{A}\boldsymbol{x}]_i + \theta_i > 0, \quad \forall i\} \tag{1.4}$$

Moreover, we assume that assumption 1.1 holds and that $B(\boldsymbol{x})$ is twice differentiable and convex on $\mathcal{C}^*$. These assumptions hold for the logarithmic barrier case.

The function $f(\alpha)$ reads:

$$f(\alpha) = p(\alpha) + \mu b(\alpha) \tag{1.5}$$

where the function $b(\alpha)$ is defined for $\alpha \in [0; \bar{\alpha}[$ with

$$\bar{\alpha} = \min_{i | \delta_i < 0} -\frac{a_i}{\delta_i}, \tag{1.6}$$

where $a_i = [\boldsymbol{A}\boldsymbol{x}]_i + \theta_i$, $\delta_i = [\boldsymbol{A}\boldsymbol{d}]_i$. If the set $\{i | \delta_i < 0\}$ is empty then $\bar{\alpha} = +\infty$.

Convergence properties of algorithm 1 are obtained under the assumption that the tangent majorants are quadratic. The function $f(\alpha)$ can have an unbounded curvature due to the barrier term (figure 3). There doesn't exist a quadratic function majorizing $f(\alpha)$ along all its definition domain. It would be sufficient to majorize $f(\alpha)$ on the level set $\mathcal{L}_k = \{\alpha | F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leq F(\boldsymbol{x}_k)\}$ but this set is rarely explicit and is often very difficult to approximate.

This report is organized as follows : In section 2, we give some assumptions allowing us to construct such majorant function and compute the minimizer. Section 3 gives the properties of the step size that allowing us to establish convergence conditions such as Armijo or Zoutendijk conditions. We deduce from this properties the extension of [9] convergence result when this new form of tangent majorant is associated with a nonlinear conjugate gradient algorithm. Moreover, we show that this strategy can lead to the convergence of others large scale descent algorithms such as truncated Newton and L-BFGS. Then, in section 4, we propose a strategy for construction of this majorant function, in the case of a logarithmic barrier.

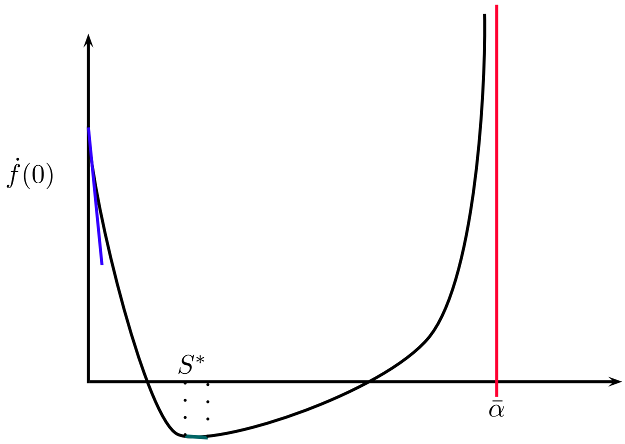

Figure 3: The function $f(\alpha)$ in the convex case and $\bar{\alpha} \neq \infty$. S^* is the set of stationnary points of f.

2. Proposed line search strategy

To overcome the limitation of the algorithm based on a quadratic majorant, we suggest a new tangent majorant of the form:

$$a_1\alpha^2 + a_2\alpha + a_3 - a_4 \log(\bar{\alpha} - \alpha) \tag{2.1}$$

This function can be seen as an extension of the quadratic form used in [9]. Such form has given good results in the case of interpolation-based line search applied to interior point methods([12, 13]). Moreover, the minimizer can still be computed exactly.

2.1. Construction of the majorant function

For all $\boldsymbol{x}_k$, $\boldsymbol{d}_k$ descent direction of F at $\boldsymbol{x}_k$, we assume that there exists a series of coefficients $(m_k^j, \gamma_k^j) \in \mathbb{R}^+$ such that

$$h_k^j(\alpha, \alpha_k^j) \geqslant f(\alpha), \quad \forall \alpha \in [0; \bar{\alpha}_k[\tag{2.2}$$

with

$$h_k^j(\alpha, \alpha_k^j) = f(\alpha_k^j) + (\alpha - \alpha_k^j)\dot{f}(\alpha_k^j) + \frac{1}{2} m_k^j (\alpha - \alpha_k^j)^2 + \gamma_k^j \left[(\bar{\alpha}_k - \alpha_k^j) \log \left(\frac{\bar{\alpha}_k - \alpha_k^j}{\bar{\alpha}_k - \alpha} \right) - \alpha + \alpha_k^j \right] \tag{2.3}$$

Moreover, we make the following assumption:

Assumption 2.1. *There exists some constants $0 < \nu_1 \leqslant \nu_2$ such that for each descent direction $\boldsymbol{d}_k \in \mathbb{R}^n$:*

$$\nu_1 \|\boldsymbol{d}_k\|^2 \leqslant m_k^j + \frac{\gamma_k^j}{\bar{\alpha}_k - \alpha_k^j} \leqslant \nu_2 \|\boldsymbol{d}_k\|^2 \tag{2.4}$$

Remark 1. *Notice that the special form of* (2.3) *leads to the following equalities, for any value of (m_k^j, γ_k^j):*

$$\begin{cases} h_k^j(\alpha_k^j, \alpha_k^j) &= f(\alpha_k^j) \\ \dot{h}_k^j(\alpha_k^j, \alpha_k^j) &= \dot{f}(\alpha_k^j) \end{cases}$$

Then if (2.2) *holds, $h_k^j(\alpha, \alpha_k^j)$ is a* tangent majorant *of $f(\alpha)$ at α_k^j.*

Remark 2. *If $\bar{\alpha} = +\infty$, then we assume that $\gamma_k^j = 0$. Therefore, in this case, $h_k^j(\alpha, \alpha_k^j)$ is a quadratic function.*

The construction of $h_k^j(\alpha, \alpha_k^j)$, tangent majorant of $f(\alpha)$ at α_k^j yields a MM algorithm for minimizing the function f:

$$\begin{array}{rcl} \alpha_k^0 & = & 0 \\ \alpha_k^{j+1} & = & \operatorname{argmin} h_k^j(\alpha, \alpha_k^j) \end{array} \tag{2.5}$$

In the rest of this section, for brevity, we will omit the index k except when it is necessary.

2.2. Computing the minimizer

Lemma 2.1. *The function $h^j(\alpha, \alpha^j)$ is strictly convex and its derivative is convex. Moreover, if γ_k^j is strictly positive, the derivative $\dot{h}^j(\alpha, \alpha^j)$ is strictly convex.*

Proof.

$$\ddot{h}^j(\alpha, \alpha^j) = m^j + \gamma^j \frac{\bar{\alpha}^j - \alpha^j}{\bar{\alpha}^j - \alpha}$$

According to (2.4),

$$m^j + \frac{\gamma^j}{\bar{\alpha} - \alpha^j} > 0$$

Furthermore, we have $(m^j, \gamma^j) \in \mathbb{R}^+$ and $\alpha^j < \bar{\alpha}^j$. Then, the second derivative of the tangent majorant is strictly positive.

$$\dddot{h}^j(\alpha, \alpha^j) = 2\gamma^j \frac{\bar{\alpha}^j - \alpha^j}{(\bar{\alpha}^j - \alpha)^2}$$

This quantity is positive for $\alpha^j < \bar{\alpha}^j$. Moreover, if $\gamma^j > 0$, $\dddot{h}^j(\alpha, \alpha^j)$ is strictly positive and thus the derivative of the tangent majorant is strictly convex. □

According to lemma 2.1, the tangent majorant has an unique minimize. Two cases are distinguished :

- Case 1. $\bar{\alpha} = +\infty$,

  α^{j+1} is the minimizer of a quadratic function :

$$\alpha^{j+1} = \alpha^j - \frac{\dot{f}(\alpha^j)}{m^j}$$

- Case 2. $\bar{\alpha} < +\infty$,

  α^{j+1} is the minimizer of the function:

$$h^j(\alpha, \alpha^j) = f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) + \frac{1}{2}(\alpha - \alpha^j)^2 m^j + \gamma^j \left[(\bar{\alpha} - \alpha^j) \log\left(\frac{\bar{\alpha} - \alpha^j}{\bar{\alpha} - \alpha} \right) - \alpha + \alpha^j \right] \tag{2.6}$$

which is also the unique root of the function $Q(\alpha)$, verifying $\alpha < \bar{\alpha}$, given by

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

whose complete expression is:

$$Q(\alpha) = A_1(\alpha - \alpha^j)^2 + A_2(\alpha - \alpha^j) + A_3$$

with

$$\left\{ \begin{array}{rcl} A_1 = \frac{1}{2}\ddot{Q}(\alpha^j) & = & -m^j \\ A_2 = \dot{Q}(\alpha^j) & = & \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha} - \alpha^j) \\ A_3 = Q(\alpha^j) & = & (\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j) \end{array} \right.$$

The calculation of this root depends on the value of A_1.

◦ Case where $A_1 = 0$

If $A_1 = -m^j$ is equal to zero, then Q is a linear function whose unique root is :

$$\begin{aligned}\alpha^{j+1} &= \alpha^j - \frac{A_3}{A_2} \\ &= \alpha^j - \frac{(\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j)}{\gamma^j - \dot{f}(\alpha^j)}\end{aligned}$$

◦ Case where $A_1 \neq 0$

If A_1 is non zero, Q is a second order polynomial vanishing at :

$$\alpha^j + \frac{-A_2 \pm \sqrt{A_2^2 - 4A_1A_3}}{2A_1}$$

The minimizer $h^j(\alpha, \alpha^j)$ satisfies the constraint $\alpha^{j+1} < \bar{\alpha}$. Then, it is equals to the smaller root of $Q(\alpha)$. The sign of A_2 then determines which root to choose:

$$\begin{aligned}\alpha^{j+1} &= \alpha^j + \frac{-|A_2| + \sqrt{A_2^2 - 4A_1A_3}}{2A_1} \\ &= \alpha^j + \frac{-2A_3}{|A_2| + \sqrt{A_2^2 - 4A_1A_3}}\end{aligned}$$

2.3. Optimization algorithm

The proposed form of the tangent majorant leads to the following iterative descent algorithm :

Algorithm 2 Descent direction algorithm with MM non quadratic line search

Require: $\boldsymbol{x}_0$ and $\boldsymbol{d}_0 = -\nabla F(\boldsymbol{x}_0)$, $J \geqslant 1$
Ensure: $\boldsymbol{x}_k$ verifying a stopping condition SC
  while SC non verified by $\boldsymbol{x}_k$ **do**
    if $\boldsymbol{d}_k = 0$ **then**
      $\alpha_k = 0$
    else
      $\alpha_k^0 = 0$
      $\alpha_k^{j+1} = \operatorname{argmin} h_k^j(\alpha, \alpha_k^j), \quad 0 \leqslant j \leqslant J-1$
      $\alpha_k = \alpha_k^J$
    end if
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
    $\boldsymbol{d}_{k+1}$ defined such that $\nabla F(\boldsymbol{x}_{k+1})^T \boldsymbol{d}_{k+1} \leqslant 0$
  end while

3. Convergence analysis

3.1. Properties of the stepsize series

The present section gathers technical results concerning the stepsize series generated by (2.5), which will be useful to derive the global convergence properties of the next section. We assume that assumption 1.1 holds.

Lemma 3.1. *For all $k \geqslant 0$ and $j \in [0, ..., J-1]$,*

$$\boldsymbol{x}_k + \alpha \boldsymbol{d}_k \in \mathcal{V}, \quad \forall \alpha \in [0, \alpha_k^j]$$

Lemma 3.2 is a corollary of [9, Lem.3.1] :

Lemma 3.2. *For all $k \geqslant 0$ and $j \in [0, ..., J-1]$,*

$$F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leqslant F(\boldsymbol{x}_k + \alpha_k^j \boldsymbol{d}_k), \quad \forall \alpha \in [\alpha_k^j, \alpha_k^{j+1}]$$

This could also be viewed as an interpretation of the 'MM capture property' given in [7, Th.6.4].

If $\boldsymbol{g}_k^T \boldsymbol{d}_k = 0$, then $\alpha^j = ... = \alpha^0 = 0$. Let us consider the case when $\dot{f}(0)$ does not vanish for the current iteration k. Then we have the following lemma :

Lemma 3.3. *The series $\{\alpha^j\}$ is positive.*

$$\alpha^j > 0$$
$$f(\alpha^j) \leqslant h^0(\alpha^1, 0)$$

The relative position between α^{j+1} and α^j is given by the following proposition :

Lemma 3.4. *For all $j \geqslant 1$:*

$$(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \leqslant 0$$

Lemma 3.5. *Inequalities to approximate the step*

Let $h^j(\alpha, \alpha^j)$, the tangent majorant of $f(\alpha)$ at α^j. If $\dot{f}(\alpha^j)$ is negative and $\bar{\alpha} < +\infty$, then :

$$\alpha_{\min} \leqslant \alpha^{j+1} - \alpha^j \leqslant \alpha_{\max}$$

with :

$$\begin{cases} \alpha_{\min} &= -\dfrac{(\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j)}{\gamma^j + m^j(\bar{\alpha} - \alpha^j) - \dot{f}(\alpha^j)} \\ \alpha_{\max} &= -2\dfrac{(\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j)}{\gamma^j + m^j(\bar{\alpha} - \alpha^j) - \dot{f}(\alpha^j)} \end{cases}$$

Proof. The minimizer of $h^j(\alpha, \alpha^j)$ is a root of the polynomial $Q(\alpha)$:

$$\begin{aligned} Q(\alpha) &= (\bar{\alpha} - \alpha)\dot{h}^j(\alpha, \alpha^j) \\ &= a_1(\alpha - \alpha^j)^2 + a_2(\alpha - \alpha^j) + a_3 \end{aligned}$$

We will show that

$$-\frac{a_3}{a_2} \leqslant \alpha^{j+1} - \alpha^j \leqslant -2\frac{a_3}{a_2} \tag{3.1}$$

Let study the signs of a_1, a_2, a_3 :

$$a_1 = -m^j \leqslant 0$$

We suppose that $\dot{f}(\alpha^j) \leqslant 0$ then

$$\begin{array}{rcll} a_2 & = & \gamma^j - \dot{f}(\alpha^j) + m^j(\bar{\alpha} - \alpha^j) & \geqslant 0 \\ a_3 & = & (\bar{\alpha} - \alpha^j)\dot{f}(\alpha^j) & \leqslant 0 \end{array}$$

If $a_1 = 0$, we have

$$\alpha^{j+1} - \alpha^j = -\frac{a_3}{a_2}$$

Then (3.1) holds.

If $a_1 \neq 0$,

$$\alpha^{j+1} - \alpha^j = \frac{-2a_3}{a_2 + \sqrt{a_2^2 - 4a_1 a_3}}$$

Then (3.1) holds too. □

Theorem 3.1. *Minorization of the step size obtained after one MM iterate*

Let $\boldsymbol{x}_k$ and a descent direction $\boldsymbol{d}_k$. If the step α_k^1 is given by

$$\alpha_k^1 = \operatorname{argmin} h^0(\alpha, 0)$$

there exists $\nu > 0$ such that

$$\alpha_k^1 \geqslant \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\nu \|\boldsymbol{d}_k\|^2}$$

Proof. Let us study separately index k for whom $\bar{\alpha}_k$ is finite and those for whom there doesn't exist i such that $[\boldsymbol{A}\boldsymbol{d}_k]_i < 0$ and so $\bar{\alpha}_k = +\infty$.

- $\bar{\alpha}_k$ equals to infinity

If $\bar{\alpha}_k = +\infty$ then the tangent majorant of $F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) = f(\alpha)$ at $\alpha_0 = 0$ is a quadratic function with curvature m^0. This majorant is minimized at :

$$\alpha^1 = \frac{-\dot{f}(0)}{m^0}$$

According to assumption 2.1, we have :

$$\alpha_k^1 \geqslant \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\nu_2 \|\boldsymbol{d}_k\|^2}$$

- $\bar{\alpha}_k$ finite

According to lemma 3.5 :

$$\alpha_k^1 \geqslant -\frac{(\bar{\alpha}_k - 0)\boldsymbol{g}_k^T \boldsymbol{d}_k}{(\bar{\alpha}_k - 0)m_k^0 + \gamma_k^0 \boldsymbol{g}_k^T \boldsymbol{d}_k}$$

Hence :

$$\alpha_k^1 \geqslant \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{m_k^0 + \frac{\gamma_k^0}{\bar{\alpha}_k} - \frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\bar{\alpha}_k}}$$

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

$$m_k^0 + \frac{\gamma_k^0}{\bar{\alpha}_k} - \frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\bar{\alpha}_k} \leqslant \nu \|\boldsymbol{d}_k\|^2$$

We will use the following notations :

$$J(\boldsymbol{x}_k) = \underbrace{m_k^0 + \frac{\gamma_k^0}{\bar{\alpha}_k}}_{J_1(\boldsymbol{x}_k)} + \underbrace{\frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\bar{\alpha}_k}}_{J_2(\boldsymbol{x}_k)}$$

Each term J_1 and J_2 will be majorized separately.

◦ Majorizing $J_1(\boldsymbol{x}_k)$

Assumption 2.1 implies that there exists ν_2 such that

$$J_1(\boldsymbol{x}_k) \leqslant \nu_2 \|\boldsymbol{d}_k\|^2$$

◦ Majorizing $J_2(\boldsymbol{x}_k)$

The following lemma is due to the barrier term in the criterion $F(\boldsymbol{x})$:

Lemma 3.6. *Minorizing the distance to the constraint domain*
There exists $\epsilon_0 > 0$ such that for all $\boldsymbol{x} \in \mathcal{L}_0$, for all $i = 1, ..., m$,

$$c_i(\boldsymbol{x}) = [\boldsymbol{A}\boldsymbol{x}]_i + \theta_i \geqslant \epsilon_0$$

Then we have :

$$\bar{\alpha}_k \geqslant \frac{\epsilon_0}{\max\limits_{i|[\boldsymbol{A}\boldsymbol{d}_k]_i < 0} - [\boldsymbol{A}\boldsymbol{d}_k]_i} = \frac{\epsilon_0}{\max\limits_{i} - [\boldsymbol{A}\boldsymbol{d}_k]_i}$$

Let $\iota(k)$ be an index such that

$$\iota(k) = \underset{i}{\operatorname{argmax}} \; - [\boldsymbol{A}\boldsymbol{d}_k]_i$$

Given $\boldsymbol{a}_{\iota(k)}$ the $\iota(k)$th row of $\boldsymbol{A}$, we have

$$\bar{\alpha}_k \geqslant \frac{\epsilon_0}{|\boldsymbol{a}_{\iota(k)}^T \boldsymbol{d}_k|}$$

$$\frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\bar{\alpha}_k} = \frac{|\boldsymbol{g}_k^T \boldsymbol{d}_k|}{\bar{\alpha}_k} \leqslant |\boldsymbol{g}_k^T \boldsymbol{d}_k| . |\boldsymbol{a}_{\iota(k)}^T \boldsymbol{d}_k| \frac{1}{\epsilon_0}$$

According to Cauchy-Swartz inequality :

$$|\boldsymbol{g}_k^T \boldsymbol{d}_k||\boldsymbol{a}_{\iota(k)}^T \boldsymbol{d}_k| \leqslant \|\boldsymbol{g}_k\| \|\boldsymbol{a}_{\iota(k)}\| \|\boldsymbol{d}_k\|^2$$

$\|\boldsymbol{a}_{\iota(k)}\|$ is majorized by

$$\zeta = \max_i \|a_i\| = \max_i \sqrt{\sum_p \boldsymbol{A}_{ip}^2}$$

The matrix $\boldsymbol{A}$ contains at least a non zero row so ζ is strictly positive.
Now, let us study the value of the norm of $\boldsymbol{g}_k$.
According to assumption 1.1, there exists $\eta > 0$ such that

$$\|\nabla F(\boldsymbol{x})\| \leqslant \eta, \; \forall \boldsymbol{x} \in \mathcal{L}_0$$

Thus,

$$J_2(\boldsymbol{x}_k) \leqslant \nu_3 \|\boldsymbol{d}_k\|^2$$

with $\nu_3 = \eta\zeta$.

Then property 3.1 holds for all k, if we put :

$$\nu = \max(\nu_2, \nu_2 + \nu_3) = \nu_2 + \nu_3 = \nu_2 + \eta\zeta$$

$\square$

3.2. Armijo condition

Starting from a point $\boldsymbol{x}$ and a descent direction $\boldsymbol{d}$, the step size α has to reduce F 'enough' to ensure convergence of the overall algorithm. To measure this decrease, we often use the *Armijo condition*:

$$F(\boldsymbol{x} + \alpha \boldsymbol{d}) \leqslant F(\boldsymbol{x}) + c_1 \alpha \boldsymbol{g}^T \boldsymbol{d} \quad c_1 \in]0;1[\tag{3.2}$$

The condition (3.2) is equivalent to :

$$f(\alpha) - f(0) \leqslant c_1 \alpha \dot{f}(0)$$

3.2.1. Armijo condition for J=1

We will show that if for every iterate k, the step size α_k is given by :

$$\alpha_k = \alpha_k^1 = \operatorname{argmin} h_k^0(\alpha, 0)$$

then the Armijo condition holds for a well-chosen c_1 parameter.

Lemma 3.7. *Let $a_1 \in \mathbb{R}$ and $\psi(\alpha) = a_1 \left[\log(\frac{\bar{\alpha}}{\bar{\alpha}-\alpha}) - \frac{\alpha}{\bar{\alpha}}\right]$.*

$$\frac{\psi(\alpha)}{\alpha\dot{\psi}(\alpha)} \leqslant \frac{1}{2}, \quad \forall\alpha; 0 < \alpha < \bar{\alpha}$$

and

$$\lim_{\alpha \to 0} \frac{\psi(\alpha)}{\alpha\dot{\psi}(\alpha)} = \frac{1}{2}$$

Proof. We define :

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

The calculation of function g gives :

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

It could be noted that :

$$\lim_{\alpha \to 0} g = \frac{1}{2}$$

and

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

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 $\alpha\dot{g}(\alpha) = \rho(\frac{\alpha}{\bar{\alpha}}) = \rho(u)$.

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

After manipulating inequalities, we find that

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

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.

We obtain :

$$\psi(\alpha) \leqslant \frac{1}{2}\alpha\dot{\psi}(\alpha);\ \forall\, \alpha \in]0;\bar{\alpha}[$$

□

Then, we have the following property 3.1 :

Property 3.1. *Let $c \leqslant \frac{1}{2}$. The sequence $\left\{\alpha^j\right\}_{j\geqslant 0}$ is defined by (2.5). If $\bar{\alpha} > \alpha^{j+1} \geqslant \alpha^j$, then :*

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

Proof. The property is trivial if $\alpha^{j+1} = \alpha^j$. Assume that $\alpha^{j+1} > \alpha^j$. According to lemma 3.4, for all $c \leqslant \frac{1}{2}$,

$$c(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j) \geqslant \frac{1}{2}(\alpha^{j+1} - \alpha^j)\dot{f}(\alpha^j)$$

The tangent majorant has the form :

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

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

$$\begin{array}{rcl} \tau(\alpha) & = & h^j(\alpha,\alpha^j) - f(\alpha^j) + (\alpha - \alpha^j)\dot{f}(\alpha^j) \\ & = & \frac{1}{2}m^j(\alpha - \alpha^j)^2 + \psi(\alpha - \alpha^j) \end{array}$$

According to lemma 3.7, for all $(\alpha - \alpha^j)$ in $]0;\bar{\alpha} - \alpha^j[$, hence for all $\alpha \in]\alpha^j;\bar{\alpha}[$:

$$\frac{\psi(\alpha - \alpha^j)}{(\alpha - \alpha^j)\dot{\psi}(\alpha - \alpha^j)} \leqslant \frac{1}{2}$$

Moreover, noting that :

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

we deduce :

$$\frac{\tau(\alpha - \alpha^j)}{(\alpha - \alpha^j)\dot{\tau}(\alpha - \alpha^j)} \leqslant \frac{1}{2} \tag{3.3}$$

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

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

And according to (3.3) :

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

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 :

$$\dot{\tau}(\alpha^{j+1} - \alpha^j) = -\dot{f}(\alpha^j) \tag{3.4}$$

This quantity is positive according to lemma 3.4. Then we have :

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

□

Corollary 1. *Consequence of property 3.1*

Suppose that $\boldsymbol{d}$ is a descent direction, i. e. $\dot{f}(0) \leqslant 0$. If $\alpha^0 = 0$ and $c_1 \leqslant \frac{1}{2}$, then the first Wolfe condition holds in α^1, the minimizer of $h(\alpha, 0)$:

$$f(0) - f(\alpha^1) + c_1\alpha^1\dot{f}(0) \geqslant 0$$

Proof. $\dot{f}(0) \leqslant 0$ implies that $\alpha^1 \geqslant 0 = \alpha^0$. The corollary is then a direct application of property 3.1 with $j = 0$. □

Remark 3. *Note that we never use informations concerning the function f in the proofs of property 3.1 and its corollary 1. The only assumption we make is on the majorizing quality of h^j. The property is deduced from the form of the tangent majorant only.*

According to corollary 1, the Armijo condition holds for every α_k if $J = 1$. The following results deals with the more general case when J can take any value.

3.2.2. Armijo condition for any J

We will show that if for every iterate k, the step size α_k is given by :

$$\alpha_k = \alpha_k^J$$

and α_k^J is the J th iterate of the (2.5) recurrence, then the Armijo condition holds for every value of J with c_1 a parameter depending on J.

Property 3.2. *Majorization of the step size*

Suppose that $\dot{f}(0) < 0$ and α^j is given by :

$$\begin{aligned} \alpha^0 &= 0 \\ \alpha^{j+1} &= \operatorname{argmin} h^j(\alpha, \alpha^j) \end{aligned}$$

Then, for all $j \in \mathbb{N} - \{0\}$:

$$\alpha^j \leqslant c_j^{\max}\alpha^1 \tag{3.5}$$

with

$$c_j^{\max} = \left(1 + \frac{2\nu_2 L}{\nu_1^2}\right)^{j-1}\left(1 + \frac{\nu}{L}\right) - \frac{\nu}{L} \geqslant 1 \tag{3.6}$$

Proof. This proof is similar to that of lemma 3.4 in [9].

Suppose $\dot{f}(\alpha^j) \leqslant 0$. If $\bar{\alpha}$ is finite then, according to equation (3.1) :

$$\alpha^{j+1} - \alpha^j \leqslant \frac{-2\dot{f}(\alpha^j)}{(\gamma^j - \dot{f}(\alpha^j))/(\bar{\alpha} - \alpha^j) + m^j} \tag{3.7}$$

Thereby :

$$\alpha^{j+1} - \alpha^j \leqslant \frac{-2\dot{f}(\alpha^j)}{\gamma^j/(\bar{\alpha} - \alpha^j) + m^j} \tag{3.8}$$

If $\bar{\alpha}$ is equal to $+\infty$, α^{j+1} is given by :

$$\alpha^{j+1} - \alpha^j = \frac{-\dot{f}(\alpha^j)}{m^j} < -\frac{2\dot{f}(\alpha^j)}{m^j} \tag{3.9}$$

The equation (3.8) still holds with $\gamma^j = 0$.

According to assumption 2.1 :

$$\|\boldsymbol{d}_k\|^2 \geqslant \frac{1}{\nu_2}\left(\gamma^0/\bar{\alpha} + m^0\right), \quad \forall k \tag{3.10}$$

and

$$\gamma^j/(\bar{\alpha} - \alpha^j) + m^j \geqslant \nu_1 \|\boldsymbol{d}_k\|^2, \quad \forall k$$

thus we have for all k:

$$\gamma^j/(\bar{\alpha} - \alpha^j) + m^j \geqslant \left(\gamma^0/\bar{\alpha} + m^0\right)\frac{\nu_1}{\nu_2} > 0 \tag{3.11}$$

Then, by (3.8):

$$\alpha^{j+1} \leqslant \alpha^j + |\dot{f}(\alpha^j)| \frac{2\nu_2}{(\gamma^0/\bar{\alpha} + m^0)\nu_1} \tag{3.12}$$

If $\dot{f}(\alpha^j) > 0$, α^{j+1} is smaller than α^j then (3.12) still holds.

According to assumption 1.1, ∇F is Lipschitz hence:

$$|\dot{f}(\alpha^j) - \dot{f}(0)| \leqslant L\|\boldsymbol{d}_k\|^2 \alpha^j \tag{3.13}$$

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

$$|\dot{f}(\alpha^j)| \leqslant L\alpha^j \|\boldsymbol{d}_k\|^2 - \dot{f}(0) \tag{3.14}$$

Using theorem 3.1 and (3.10)

$$\begin{aligned} -\dot{f}(0) &\leqslant \alpha^1 \nu \|\boldsymbol{d}_k\|^2 & (3.15)\\ &\leqslant \alpha^1 \frac{\nu}{\nu_1}(m^0 + \gamma_0/\bar{\alpha}) & (3.16) \end{aligned}$$

Given (3.14),(3.10) and (3.16) jointly with (3.12), we get :

$$\begin{aligned} \alpha^{j+1} &\leqslant \alpha^j + \frac{2\nu_2}{(m^0 + \gamma_0/\bar{\alpha})\nu_1}\left[L\alpha^j\left(\frac{m^0 + \gamma_0/\bar{\alpha}}{\nu_1}\right) + \alpha^1 \frac{\nu}{\nu_1}(m^0 + \gamma_0/\bar{\alpha})\right] & (3.17)\\ &\leqslant \alpha^j \left[1 + \frac{2\nu_2 L}{\nu_1^2}\right] + 2\alpha^1 \frac{\nu_2 \nu}{\nu_1^2} & (3.18) \end{aligned}$$

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

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

Given $c_1^{\max} = 1$, we could deduce the general term of the sequence and have (3.6). □

Remark 4. *Equation* (3.8) *is the key point of the proof. The new step* α^{j+1} *is either less than* α^j, *or less than the minimizer of a quadratic function tangent with* $f(\alpha)$ *at* α^j *and with curvature* $\frac{1}{2}\ddot{h}^j(\alpha^j, \alpha^j)$*. This result allows us to use the same kind of proof as in [9].*

Lemma 3.8. *Armijo condition*

The step corresponding to the J*-th element of the recurrence* (2.5) *verifies Armijo condition* (3.2) *with*

$$c_1 = c_1^J = \frac{1}{2c_J^{\max}} \in]0;1[$$

Proof. First, we have

$$f(0) - f(\alpha^J) \geqslant f(0) - f(\alpha^1)$$

According to corollary 1,

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

Moreover, according to property 3.2

$$\alpha^1 \geqslant \frac{\alpha^J}{c_J^{\max}}$$

Hence :

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

□

3.3. Zoutendijk condition

Global convergence of a descent direction method is ensure by a 'good choice' of the step but also by well-chosen search directions $\boldsymbol{d}_k$. A key property is the angle θ_k between $\boldsymbol{d}_k$ and steepest descent direction $-\boldsymbol{g}_k$. It is defined by :

$$\cos\theta_k = \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|\|\boldsymbol{d}_k\|} \tag{3.19}$$

Proofs of convergence of descent direction algorithms are very often based on the verification of the following condition :

Definition 1. *Zoutendijk Condition*
Consider $\{\boldsymbol{x}_k\}$, *iterations from a descent direction algorithm. Suppose that* F *is bounded below. The* Zoutendijk condition *is satisfied if*

$$\sum_{k=0}^{\infty} \|\boldsymbol{g}_k\|^2 \cos^2\theta_k < \infty \tag{3.20}$$

The inequality (3.20) implies that $\cos^2\theta_k \|\boldsymbol{g}_k\|^2$ tends to 0 as k tends to infinity. This property may be used to demonstrate global convergence of the algorithm. Suppose that the way we choose $\boldsymbol{d}_k$ ensures that θ_k is bounded away from $\frac{\pi}{2}$. Then $\cos\theta_k$ stays bounded away from zero :

$$\exists\delta \,\text{such that}\ \ \cos\theta_k \geq \delta > 0,\, \forall k \tag{3.21}$$

The condition (3.20) implies the convergence of the algorithm in the sense

$$\lim_{k\to\infty} \|\boldsymbol{g}_k\| = 0 \tag{3.22}$$

Lemma 3.8 and theorem 3.1 yield us to demonstrate that the Zoutendijk condition (3.20) holds by using our new line search.

Property 3.3. *Minorization of the step size*

Suppose $\dot{f}(0) < 0$ *and* α^j *given by :*

$$\begin{array}{lcl} \alpha^0 & = & 0 \\ \alpha^{j+1} & = & \underset{\alpha}{\operatorname{argmin}}\, h^j(\alpha, \alpha^j) \end{array}$$

Then for all $j \in \mathbb{N} - \{0\}$*, we have :*

$$\alpha^j \geqslant c^{\min} \alpha^1 \tag{3.23}$$

with

$$c^{\min} = \frac{-1 + \sqrt{1 + 2L/\nu_1}}{2L/\nu_1} \tag{3.24}$$

Proof. This proof is inspired from that of [9, Lem.3.3]

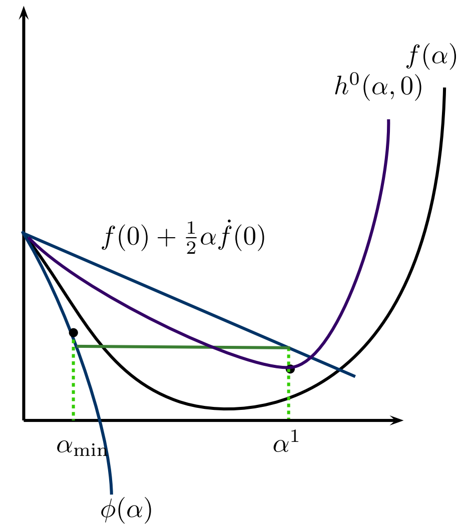

Figure 4: Minorizing α^j in the convex case : Illustration of inequality (3.26)

Let ϕ be the concave quadratic function :

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

with $m = -\frac{m^0 + \gamma^0/\bar{\alpha}}{\nu_1} L$

We have $\phi(0) = f(0)$ and $\dot{\phi}(0) = \dot{f}(0) < 0$. So ϕ is decreasing on $\mathbb{R}^+$.

Let us consider $\alpha \in [0, \alpha^j]$: $\boldsymbol{x}_k + \alpha \boldsymbol{d}_k \in \mathcal{V}$. According to assumption 1.1, we have

$$|\dot{f}(\alpha) - \dot{f}(0)| \leqslant \|\boldsymbol{d}_k\|^2 L |\alpha|$$

and according to assumption 2.1,

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

Then we obtain :

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

Hence :

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

Integrating (3.3) between 0 and α^j yields

$$\phi(\alpha^j) \leqslant f(\alpha^j) \tag{3.25}$$

According to corollary 1 :

$$h^0(\alpha^1, 0) \leqslant f(0) + \frac{1}{2}\alpha^1 \dot{f}(0)$$

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

According to (3.8) :

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

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

$$c_{\min} + c_{\min}^2 \frac{L}{\nu_1} = \frac{1}{2}$$

Then :

$$\phi(\alpha_{\min}) = \phi(c_{\min}\alpha^1) \geqslant f(0) + \frac{1}{2}\alpha^1 \dot{f}(0) \geqslant h^0(\alpha^1, 0) \tag{3.26}$$

On the other hand, α^j is positive. Assume that there exists j such that $0 \leqslant \alpha^j < \alpha_{\min}$. According to (3.25) and given that ϕ is decreasing on $\mathbb{R}^+$, we get :

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

wich is in contradiction with lemma 3.3. Figure 4 illustrated the proof when $f(\alpha)$ is convex.

□

Theorem 3.2. *Zoutendijk condition*

Suppose for all k that $\boldsymbol{d}_k$ is a descent direction and α_k is the Jth element of the series defined by (2.5)*. Then, Zoutendijk condition* (3.20) *holds.*

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

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

According to lemma 3.8, the first Wolfe condition (3.2) holds for $c_1 = \frac{1}{2c_J^{\max}}$:

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

$$\alpha_k \geqslant c^{\min} \alpha_k^1$$

According to theorem 3.1 :

$$\alpha_k^1 \geqslant \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\nu \|\boldsymbol{d}_k\|^2}$$

Hence :

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

with $c_0 = (c^{\min} c_1)/\nu > 0$. The assumption 1.1 and the boundedness of $\mathcal{L}_0$ implies that the limit $\lim_{k\to\infty} F(\boldsymbol{x}_k)$ is finite. Therefore :

$$\infty > \left[F(\boldsymbol{x}_0) - \lim_{k\to\infty} F(\boldsymbol{x}_k)\right]/c_0 \geqslant \sum_k \frac{(\boldsymbol{g}_k^T \boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2}$$

□

3.4. Convergence of Newton-like methods

Let us study the global convergence of algorithm 2 with different choices of the direction $\boldsymbol{d}_k$. The Zoutendijk condition holds. If the direction $\boldsymbol{d}_k$ is the steepest descent $-\boldsymbol{g}_k$, then the algorithm 2 converges in the sense :

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

As a general rule, assume that direction $\boldsymbol{d}_k$ is given by :

$$\boldsymbol{d}_k = -\boldsymbol{B}_k^{-1} \boldsymbol{g}_k \tag{3.27}$$

We have the following property :

Property 3.4. *[14]*
Assume that for all k, matrices $\boldsymbol{B}_k$ are positive definite and that there exists $M > 0$ such that :

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

If Zoutendijk condition holds, then the descent algorithm defined by

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

and (3.27) *converges in the sense :*

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

This result covers the following classical methods:

- Steepest descent: $\boldsymbol{d}_k = -\boldsymbol{g}_k$
- Newton: $\boldsymbol{d}_k = -\nabla^2 F(\boldsymbol{x}_k)^{-1} \boldsymbol{g}_k$
- Quasi Newton: $\boldsymbol{d}_k = -\boldsymbol{B}_k^{-1} \boldsymbol{g}_k$

3.5. Convergence of BFGS and L-BFGS methods

The BFGS algorithm is a Quasi-Newton algorithm where the inverse approximation of the Hessian $\boldsymbol{B}_k^{-1}$ is computed iteratively by accounting for the curvature measured during the precedent iteration. Then, the descent direction is given by :

$$\boldsymbol{d}_k = -\boldsymbol{H}_k \boldsymbol{g}_k$$

The udpate of the inverse Hessian approximation is given by

$$\boldsymbol{H}_{k+1} = (\boldsymbol{I} - \rho_k \boldsymbol{\delta}_k \boldsymbol{\gamma}_k^T) \boldsymbol{H}_k (\boldsymbol{I} - \rho_k \boldsymbol{\gamma}_k \boldsymbol{\delta}_k^T) + \rho_k \boldsymbol{\delta}_k \boldsymbol{\delta}_k^T$$

with $\boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$, $\boldsymbol{\gamma}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$ and $\rho_k = \frac{1}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k}$.

According to theorem 3.2, we have the following convergence result :

Assumption 3.1. *The level set $\mathcal{V}$ is convex, and there exist positive constants m and M such that*

$$m\|\boldsymbol{v}\|^2 \leq \boldsymbol{v}^T \nabla^2 F(\boldsymbol{x}) \boldsymbol{v} \leq M \|\boldsymbol{v}\|^2$$

for all $\boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{x} \in \mathcal{V}$

Theorem 3.3. *If F is twice continuously differentiable and assumption 3.1 holds, then the BFGS algorithm associated with the MM line search converges in the sense* $\liminf_{k\to\infty} \boldsymbol{g}_k = 0$*. Since under this assumption F is strongly convex, this result leads to the convergence of $\boldsymbol{x}_k$ to $\boldsymbol{x}^*$.*

The L-BFGS algorithm is a limited-memory method based on the BFGS updating formula. Instead of computing the full inverse Hessian approximation, only a few vectors representing implicitly the approximation are stored. The update strategy makes use of an initial Hessian approximation $\boldsymbol{B}_k^{(0)}$ which is allowed to vary at each iteration [14]. Since the convergence proof of L-BFGS algorithm is based on the fulfillment of Zoutendijk condition [10], we can establish the following result :

Theorem 3.4. *Let F a twice continuously differentiable function. If assumption 3.1 holds and if the matrices $\boldsymbol{B}_k^{(0)}$ are chosen so that $\|\boldsymbol{B}_k^{(0)}\|$ and $\|\boldsymbol{B}_k^{(0)^{-1}}\|$ are bounded then the L-BFGS algorithm associated with the MM line search converges in the sense* $\liminf_{k\to\infty} \boldsymbol{g}_k = 0$*. Since under this assumption F is strongly convex, this result leads to the convergence of $\boldsymbol{x}_k$ to $\boldsymbol{x}^*$.*

3.6. Convergence of the truncated Newton method

The Newton method is attractive because it converges rapidly from any sufficiently good initial guess $\boldsymbol{x}_0$. However, it is unpractical for large scale problems since it requires to solve a system of linear equations (Newton equations) at each stage. Therefore, an inexact form of the Newton method has been developped 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.

1. Choose a starting point $\boldsymbol{x}_0$
2. FOR all k UNTIL convergence DO
   Compute $\boldsymbol{d}_k$ by solving approximately the linear system $\nabla^2 F(\boldsymbol{x}_k)\boldsymbol{d} = -\boldsymbol{g}_k$ with I_k CG iterates
   Compute the step size α_k with some appropriate line search strategy
   $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$

Assumption 3.2. *For all $\boldsymbol{x} \in \mathcal{V}$, $\boldsymbol{H} = \nabla^2 F(\boldsymbol{x})$ is a symmetric positive definite (SPD) matrix. Let $\nu_1(\boldsymbol{H}) > 0$ and $\nu_2(\boldsymbol{H}) > 0$ denote the smallest and largest eigenvalues of $\boldsymbol{H}$. The matrix sequence $\mathcal{H} = \left\{\nabla^2 F(\boldsymbol{x}_k)\right\}$ has a uniformly bounded spectrum with a stricly positive lower bound i.e., there exist $\nu_1(\mathcal{H}), \nu_2(\mathcal{H}) \in \mathbb{R}$ such that*

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

Lemma 3.9. *[8]*
Let $\{\boldsymbol{x}_k\}$ be a sequence generated by the truncated Newton method and assume that assumption 3.2 holds. Then $\boldsymbol{d}_k$ is a descent direction i.e.,

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

Definition 2. *[1]*
The direction sequence $\{\boldsymbol{d}_k\}$ is gradient related *to $\{\boldsymbol{x}_k\}$ if 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*

$$\limsup_{k\leftarrow\infty, k\in\mathcal{K}} \boldsymbol{d}_k^T \boldsymbol{g}_k < 0$$

Lemma 3.10. *[8]*
Let $\{\boldsymbol{x}_k\}$ be a sequence generated by the truncated Newton method and assume that assumption 3.2 holds. Then there exists $\eta_1, \eta_2 > 0$ such that

$$\eta_1 \|\boldsymbol{g}_k\|^2 \leqslant -\boldsymbol{d}_k^T \boldsymbol{g}_k \tag{3.28}$$

$$\|\boldsymbol{d}_k\|^2 \leqslant \eta_2 \|\boldsymbol{g}_k\|^2 \tag{3.29}$$

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

Theorem 3.5. *Let $\{\boldsymbol{x}_k\}$ be a sequence generated by the truncated Newton method, α_k be defined by the recurrence* (2.5)*, and let assumptions 1.1,2.1 and 3.2 hold. Then we have convergence in the sense*

$$\lim_{k \leftarrow \infty} \boldsymbol{g}_k = 0$$

Proof. Let $\{\boldsymbol{x}_k\}$ be a sequence generated by the truncated Newton method. According to lemma 3.9, $\boldsymbol{d}_k$ is a descent direction. Then, according to lemma 3.8, there exists $c_1 \in (0,1)$ such that for all k,

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

According to theorem 3.1 and property 3.3,

$$\begin{aligned} \alpha_k &\geqslant c_{\min} \alpha_k^1 && (3.30) \\ &\geqslant \frac{c_{\min}}{\nu} \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|^2} && (3.31) \end{aligned}$$

Furthermore, according to lemma 3.10,

$$\frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|^2} \geqslant \frac{\eta_1}{\eta_2}$$

Let $\Omega = \frac{c_1 c_{\min} \eta_1^2}{\nu \eta_2}$. Then,

$$F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1}) \geqslant \Omega \|\boldsymbol{g}_k\|^2 \geqslant 0$$

On the other hand, the assumption 1.1 and the boundedness of $\mathcal{L}_0$ implies that

$$F(\boldsymbol{x}_l) \geqslant \inf_{\boldsymbol{x} \in \mathcal{V}} F(\boldsymbol{x}) > -\infty, \forall l$$

Then we deduce

$$\infty > F(\boldsymbol{x}_0) - \inf_{\boldsymbol{x} \in \mathcal{V}} F(\boldsymbol{x}) \geqslant F(\boldsymbol{x}_0) - F(\boldsymbol{x}_l) \geqslant \Omega \sum_{k=1}^{l-1} \|\boldsymbol{g}_k\|^2, \quad \forall l$$

Hence, $\lim_{k \to \infty} \boldsymbol{g}_k = 0$ □

3.7. Convergence of conjugate gradient methods

Let us consider the following family of conjugate gradient algorithms :

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k && (3.32) \\ \boldsymbol{c}_k &= -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1} && (3.33) \\ \boldsymbol{d}_k &= -\boldsymbol{c}_k \operatorname{sign}(\boldsymbol{g}_k^T \boldsymbol{c}_k) && (3.34) \end{aligned}$$

with the conjugacy formulas :

$$\beta_0 = 0, \quad \beta_k = \beta_k^{\mu_k, \omega_k} = \nabla \boldsymbol{g}_k^T \boldsymbol{y}_{k-1} / D_k, \quad \forall k > 0 \qquad (3.35)$$

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

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

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

Expression (3.35) allows us to cover the following conjugate gradient methods :

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

Let us consider the following assumption :

Assumption 3.3. *Assumption 1.1 holds and F is strongly convex on $\mathcal{V}$: there exists $\lambda > 0$ such that*

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

We have the following convergence result :

Theorem 3.6. *Let α_k be defined by the recurrence* (2.5)*, and let assumptions 1.1 and 2.1 hold. Then, we have convergence in the sense* $\liminf_{k\to\infty} \boldsymbol{g}_k = 0$ *for the PRP and LS methods, and more generally for $\mu_k = 0$ and $\omega_k \in [0, 1]$. Moreover, if assumption 3.3 holds, then we have* $\liminf_{k\to\infty} \boldsymbol{g}_k = 0$ *in all cases.*

We have previously established :

- the step size minorization $\alpha_k \leqslant c_J^{\max} \alpha_k^1$ (property 3.3)
- the step size majorization $0 \leqslant c^{\min} \alpha_k^1 \leqslant \alpha_k$ (property 3.2)
- the verification of Zoutendijk condition (theorem 3.2)

Thus, the proof of 3.6 is identical to that in [9]. This result can be viewed as an extension of [9, Th.4.1] for an new form of tangent majorant.

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

Assumption 3.4. *Let $\Delta(\alpha, \alpha^j)$ the difference between f and its tangent majorant :*

$$\Delta(\alpha, \alpha^j) = f(\alpha) - h^j(\alpha, \alpha^j)$$

We assume that for all $\alpha \geqslant \alpha^j$:

$$\dot{\Delta}(\alpha, \alpha^j) \leqslant 0$$

Lemma 3.11. *Let us assume that $\dot{f}(0) < 0$. Then, if assumption 3.4 holds, the series defined by* (2.5) *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 :

$$\boldsymbol{g}_k^T \boldsymbol{d}_{k-1} \leqslant 0 \tag{3.36}$$

Lemma 3.11 leads to an important result for the convergence properties of conjugate gradient methods :

Lemma 3.12. *Sufficient descent condition*

Assume that assumption 3.4 holds and suppose that the successive directions $\boldsymbol{d}_k$ are given by the conjugate algorithm gradient :

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

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

$$\boldsymbol{g}_k^T \boldsymbol{d}_k \leqslant -c\|\boldsymbol{g}_k\|^2$$

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

Lemma 3.12 is a direct application of a remark made in [5, Part 4]. It is a consequency of inequality (3.36). We can directly use this result to proove global convergence of Fleetcher-Reeves(FR) method when :

$$\beta_k^{\mathrm{FR}} = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2} \geqslant 0$$

According to lemma 3.12, the FR method always generates descent directions with our choice of step, if we assume that assumption 3.4 holds. Then we can use [3] result : The Zoutendijk condition holds according to theorem 3.2 and $\boldsymbol{d}_k$ is a descent direction. So the FR method converges in the sense :

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

Let consider now the conjugacy formula PRP+, proposed in [5] :

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

According to [14], PRP+ method has led to the better convergence results during numerical tests when compared with others conjugacy formulas.

The convergence of our new algorithm with this conjugacy formula is a direct application of [5, Th.4.3]. The PRP+ method converges in the sense :

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

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

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

Let us make the following assumption :

Assumption 3.5. *Assumption 1.1 holds and F is convex on $\mathcal{V}$: For every $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{V}$ we have*

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

According to lemma 3.11, if assumption 3.5 holds, for all iteration k of the overall algorithm, we have the inequality :

$$|\boldsymbol{g}_{k+1}^T\boldsymbol{d}_k| \leqslant |\boldsymbol{g}_k^T\boldsymbol{d}_k| \tag{3.37}$$

Let us show recurrently on k that, with DY method, $\boldsymbol{d}_k$ is always a descent direction .

For $k = 0$, we have $\boldsymbol{d}_0 = -\boldsymbol{g}_0$, hence $\boldsymbol{d}_0$ is a descent direction. Consider an index k when $\boldsymbol{g}_k^T\boldsymbol{d}_k \leqslant 0$. If this quantity is zero, $\alpha_k = 0$ and then the algorithm will finish. Let us assume that $\boldsymbol{g}_k^T\boldsymbol{d}_k < 0$. Then, according to lemma 3.11, $\boldsymbol{g}_{k+1}^T\boldsymbol{d}_k \leqslant 0$ and moreover, according to inequality (3.37), $|\boldsymbol{g}_{k+1}^T\boldsymbol{d}_k| \leqslant |\boldsymbol{g}_k^T\boldsymbol{d}_k|$. Then, the coefficient β_k^{DY} is positive :

$$\beta_k^{\mathrm{DY}} = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\boldsymbol{d}_k^T(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)} \geqslant 0$$

And according to lemma 3.12, $\boldsymbol{d}_{k+1}$ is a descent direction.

Now we can use [3] result to show the global convergence of DY method : The Zoutendijk condition holds according to theorem 3.2 and $\boldsymbol{d}_k$ is a descent direction. If assumption 3.5 holds, then the DY method converges in the sense :

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

Theorem 3.7. *Let α_k be defined by the recurrence* (2.5)*, and let assumptions 1.1, 2.1 and 3.4 hold. Then, we have convergence in the sense* $\operatorname{liminf}_{k\to\infty}\boldsymbol{g}_k = 0$ *for the PRP+ and FR. Moreover, if assumption 3.5 holds, we have convergence in the sense* $\operatorname{liminf}_{k\to\infty}\boldsymbol{g}_k = 0$ *for the DY method.*

The following table 1 concentrates the convergence results of several nonlinear conjugate gradient methods, according to the assumptions made.

Conjugacy	β_k	Assumptions	Convergence result
Hestenes and Stiefel (HS)	$\dfrac{\boldsymbol{g}_{k+1}^T\boldsymbol{y}_k}{\boldsymbol{d}_k^T\boldsymbol{y}_k}$	Ass. 1.1, 3.3, 2.1	Th. 3.6
Fletcher and Reeves (FR)	$\dfrac{\Vert\boldsymbol{g}_{k+1}\Vert^2}{\Vert\boldsymbol{g}_k\Vert^2}$	Ass. 1.1, 2.1,3.4	Th. 3.7
Polak and Ribière (PRP)	$\dfrac{\boldsymbol{g}_{k+1}^T\boldsymbol{y}_k}{\Vert\boldsymbol{g}_k\Vert^2}$	Ass. 1.1, 2.1	Th. 3.6
Polak and Ribière modified (PRP+)	$\max(\beta_k^{\mathrm{PRP}}, 0)$	Ass. 1.1, 2.1,3.4	Th. 3.7
Liu and Storey (LS)	$\dfrac{\boldsymbol{g}_{k+1}^T\boldsymbol{y}_k}{-\boldsymbol{d}_k^T\boldsymbol{y}_k}$	Ass. 1.1, 2.1	Th. 3.6
Dai and Yan (DY)	$\dfrac{\Vert\boldsymbol{g}_{k+1}\Vert^2}{\boldsymbol{d}_k^T\boldsymbol{y}_k}$	Ass. 1.1, 2.1,3.4, 3.5	Th. 3.7

Table 1: Summary of convergence results for conjugate gradient methods

4. The logarithmic barrier case

The assumption 2.1 could appear restrictive. Moreover, it is not a constructive assumption because this doesn't lead directly to a method of constructing the tangent majorant function. In this section, we present a tangent majorant function verifying the assumption 2.1, for the case when $B(\boldsymbol{x})$ is a logarithmic barrier.

4.1. Calculation of the proposed tangent majorant parameters

Let us consider the following optimization problem :

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} F(\boldsymbol{x}) = P(\boldsymbol{x}) - \mu\sum_{i=1}^{\mathcal{I}} t_i \log([\boldsymbol{A}\boldsymbol{x}]_i + \theta_i) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x}), \quad \mu > 0,\ t_i > 0 \tag{4.1}$$

$\boldsymbol{A}$ is a matrix from $\mathbb{R}^{\mathcal{I}\times\mathcal{N}}$ and θ is a vector from $\mathbb{R}^{\mathcal{N}}$. We have :

$$f(\alpha) = p(\alpha) + \mu b(\alpha)$$

with

$$b(\alpha) = \sum_{i=1}^{\mathcal{I}} b_i(\alpha) = \sum_{i=1}^{\mathcal{I}} -t_i\log(a_i + \alpha\delta_i) \quad \boldsymbol{a} = A\boldsymbol{x} + \theta,\ \delta = \boldsymbol{A}\boldsymbol{d}$$

Let suppose that for all j, there exists m_p^j such that the following quadratic function is a tangent majorant of $p(\alpha)$ at α^j:

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

We assume that there exists $\nu_1^p, \nu_2^p \in \mathbb{R}$, $0 \leqslant \nu_1^p \leqslant \nu_2^p$ such that for all direction $\boldsymbol{d}_k \in \mathbb{R}^n$,

$$\nu_1^p\Vert\boldsymbol{d}_k\Vert^2 \leqslant m_p^j \leqslant \nu_2^p\Vert\boldsymbol{d}_k\Vert^2, \quad \forall k\in\mathbb{N}, \forall j \in [0; J-1]$$

This assumption holds if $P(\boldsymbol{x})$ verifies assumption 1.2. Notice that we accept $\nu_1^p = \nu_2^p = 0$. Actually, if $P(\boldsymbol{x})$ is a linear function, it is better to choose $m_p^j = 0$ for all j and k.

There is different ways of obtaining $q^j(\alpha,\alpha^j)$. We can use one of the strategies described in [6] to construct a quadratic tangent majorant of $p(\alpha)$. In the sequel, we will focus on constructing a tangent majorant of the term $b(\alpha)$ and we will use the following notations:

$$b(\alpha) \;=\; \underbrace{\sum_{i|\delta_i>0} -t_i\log(a_i+\alpha\delta_i)}_{b_1(\alpha)} + \underbrace{\sum_{i|\delta_i<0} -t_i\log(a_i+\alpha\delta_i)}_{b_2(\alpha)} + \underbrace{\sum_{i|\delta_i=0} -t_i\log(a_i+\alpha\delta_i)}_{b_3(0)} \tag{4.2}$$

Majorizing $b_1(\alpha)$. Let us assume that the set $\{i|\delta_i > 0\}$ is non empty.

Property 4.1. *Let* $b_1(\alpha) = -\sum_{i|\delta_i>0} t_i \log(a_i + \alpha\delta_i)$*. Then, for all* $(\alpha, \alpha^j) \in [0; \bar{\alpha}[$*, the following function* $\phi_1^j(\alpha, \alpha^j)$ *is a tangent majorant of* $b_1(\alpha)$ *at* α^j*.*

$$\phi_1^j(\alpha, \alpha^j) = b_1(\alpha^j) + (\alpha - \alpha^j)\dot{b}_1(\alpha^j) + \frac{1}{2}m_b^j(\alpha - \alpha^j)^2$$

if

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

Furthermore, this is the best tangent majorant of this form, in the sense:

$$m_b^j = \min\left\{m \geqslant 0 | b_1(\alpha) \leqslant \phi_1^j(\alpha, \alpha^j)\right\}$$

Finally, let $\Delta(\alpha, \alpha^j) = b_1(\alpha) - \phi_1^j(\alpha, \alpha^j)$*, then for all* $\alpha \geqslant \alpha^j$*, we have :*

$$\dot{\Delta}(\alpha, \alpha^j) \leqslant 0$$

$$\Delta(0, \alpha^j) = b_1(0)$$

Proof.

$$\dot{b}_1(\alpha) = \sum_{i|\delta_i>0} -\frac{t_i\delta_i}{a_i + \alpha\delta_i}$$

$$\ddot{b}_1(\alpha) = \sum_{i|\delta_i>0} \frac{t_i\delta_i^2}{(a_i + \alpha\delta_i)^2}$$

$$\dddot{b}_1(\alpha) = \sum_{i|\delta_i>0} \frac{-2t_i\delta_i^3}{(a_i + \alpha\delta_i)^3}$$

For all i, $t_i > 0$, then b_1 is strictly convex and its derivative is strictly concave. According to [4, 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. Moreover,

$$m_b^j = \min\left\{m \geqslant 0 | h(l) \leqslant h(l^j) + \dot{h}(l^j)(l - l^j) + \frac{1}{2}m(l - l^j)^2, \forall l \geqslant 0\right\}$$

Coefficient m_b^j insures :

$$\Delta(0, \alpha^j) = b_1(0)$$

Finally, according to [4, Lem. 6], for all $\alpha \geqslant \alpha^j$:

$$\dot{\Delta}(\alpha, \alpha^j) \leqslant 0$$

□

Majorizing $b_2(\alpha)$. Let us assume that the set $\{i|\delta_i < 0\}$ is non empty, so $\bar{\alpha}$ is finite.

Property 4.2. *Let* $b_2(\alpha) = -\sum_{i|\delta_i<0} t_i \log(a_i + \alpha\delta_i)$. *Then, for all* $(\alpha, \alpha^j) \in [0; \bar{\alpha}[$, *the following function* $\phi_2^j(\alpha, \alpha^j)$ *is a tangent majorant of* $b_2(\alpha)$ *at* α^j.

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

with

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

Furthermore, this is the best tangent majorant of this form, in the sense :

$$\gamma_b^j = \min \left\{ \gamma \geqslant 0 | b_2(\alpha) \leqslant \phi_2^j(\alpha, \alpha^j) \right\}$$

Finally, let $\Delta(\alpha, \alpha^j) = b_2(\alpha) - \phi_2^j(\alpha, \alpha^j)$, *then for all* $\alpha \geqslant \alpha^j$ *we have :*

$$\dot{\Delta}(\alpha, \alpha^j) \leqslant 0$$

$$\Delta(0, \alpha^j) = b_2(0)$$

Proof. Let us define the following function,

$$T(\alpha) = \dot{b}_2(\alpha)(\bar{\alpha} - \alpha) \tag{4.3}$$

and its derivatives:

$$\dot{T}(\alpha) = \ddot{b}_2(\alpha)(\bar{\alpha} - \alpha) - \dot{b}_2(\alpha) \tag{4.4}$$

$$\ddot{T}(\alpha) = \dddot{b}_2(\alpha)(\bar{\alpha} - \alpha) - 2\ddot{b}_2(\alpha) \tag{4.5}$$

Formulating the successive derivatives of $b(\alpha)$,

$$\ddot{T}(\alpha) = (\bar{\alpha} - \alpha) \sum_{i|\delta_i<0} \frac{-2t_i\delta_i^3}{(a_i + \alpha\delta_i)^3} - 2 \sum_{i|\delta_i<0} \frac{t_i\delta_i^2}{(a_i + \alpha\delta_i)^2}.$$

For each i, $\delta_i < 0$:

$$\ddot{T}(\alpha) = (\bar{\alpha} - \alpha) \sum_{i|\delta_i<0} \frac{2t_i}{(-\frac{a_i}{\delta_i} - \alpha)^3} - 2 \sum_{i|\delta_i<0} \frac{t_i}{(-\frac{a_i}{\delta_i} - \alpha)^2}$$

$$\ddot{T}(\alpha) = \sum_{i|\delta_i<0} \frac{2t_i}{(-\frac{a_i}{\delta_i} - \alpha)^2} \left(\frac{\bar{\alpha} - \alpha}{-\frac{a_i}{\delta_i} - \alpha} - 1 \right)$$

then $\ddot{T}(\alpha) < 0$, which shows that the function T is *strictly concave*. Moreover, let us consider the linear function $l(\alpha)$:

$$l(\alpha) = \dot{\phi}_2^j(\alpha, \alpha^j)(\bar{\alpha} - \alpha) = \dot{b}_2(\alpha^j)(\bar{\alpha} - \alpha) + \gamma^j(\alpha - \alpha^j) \tag{4.6}$$

According to [4, Lem. 3], the function $T(\alpha)$ intersects $l(\alpha)$ at most twice. Yet :

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

and, because of our choice for γ_b^j,

$$\begin{cases} b_2(0) = \phi_2^j(0, \alpha^j) \\ b_2(\alpha^j) = \phi_2^j(\alpha^j, \alpha^j) \end{cases} \tag{4.7}$$

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

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

and equivalently,

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

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

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

and

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

Noticing that $\bar{\alpha} - \alpha > 0$, we could apply [4, Lem. 5] and then demonstrate the first part of the property.

Let us proove that there doesn't exist another tangent majorant of the same form, with γ less than γ_b^j. Assume $\tilde{\gamma}_b^j < \gamma_b^j$ and let

$$\tilde{\phi}_2^j(\alpha, \alpha^j) = b_2(\alpha^j) + (\alpha - \alpha^j)\dot{b_2}(\alpha^j) + \tilde{\gamma}_b^j \left[(\bar{\alpha} - \alpha^j) \log\left(\frac{\bar{\alpha} - \alpha^j}{\bar{\alpha} - \alpha} \right) + \alpha^j - \alpha \right]$$

We have $\phi_2^j(0, \alpha^j) = b_2(0)$ and

$$\left[(\bar{\alpha} - \alpha^j) \log\left(\frac{\bar{\alpha} - \alpha^j}{\bar{\alpha}} \right) + \alpha^j \right] > 0, \forall \alpha^j > 0$$

then for $\alpha^j > 0$,$\tilde{\phi}_2^j(0, \alpha^j) < \phi_2^j(0, \alpha^j) = b_2(0)$. That is $\tilde{\phi}_2^j$ doesn't majorize b_2. If $\alpha^j = 0$, then $\tilde{\gamma}_b^j < \gamma_b^j$ would lead to $\ddot{b}_2(0) > \ddot{\tilde{\phi}}_2^j(0, 0)$. Then, there exists $\epsilon > 0$ such that for $\epsilon > \alpha > 0$, $\tilde{\phi}_2^j(\alpha, \alpha^j) < b_2(\alpha)$. This proof is inspired from that of [4, Th. 1]. $b_2(\alpha)$ and the tangent majorant $\phi_2^j(\alpha, \alpha_j)$, as well as $T(\alpha)$ and $l(\alpha)$ are illustrated on figure 5(a) and (b). Variations of $T(\alpha)$ do not mean to be realistic.

□

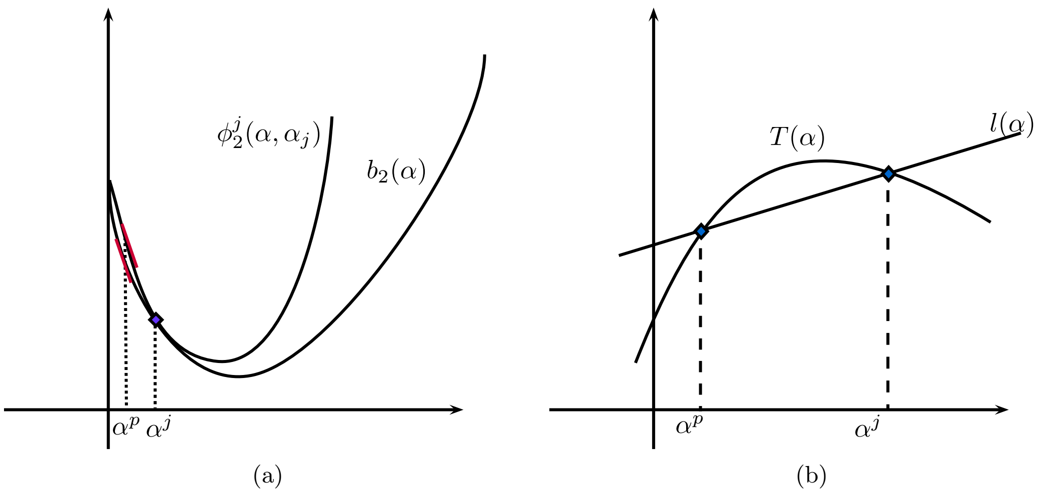

Figure 5: Construction of the tangent majorant

Tangent majorant of $f(\alpha)$.

Theorem 4.1. *For each $0 \leqslant (\alpha, \alpha^j) < \bar{\alpha}$, the following function $h^j(\alpha, \alpha^j)$ is a tangent majorant at α^j of*

$$f(\alpha) = p(\alpha) + \mu b(\alpha) = p(\alpha) + \mu \sum_i -t_i \log(a_i + \alpha\delta_i)$$

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

with :

$$\bar{\alpha} = \min_{i|\delta_i<0} -\frac{a_i}{\delta_i}$$

- $m^j = m_p^j + \mu m_b^j$, $\gamma^j = \mu \gamma_b^j$
- m_p^j *assumed to be known*
- m_b^j *described in property 4.1*
- γ_b^j *described in property 4.2*

4.2. Properties of the tangent majorant

Let us give some properties of the tangent majorant $\phi^j(\alpha, \alpha_j)$ of $b(\alpha)$ at α^j.

$$\phi^j(\alpha, \alpha_j) = \phi_1^j(\alpha, \alpha_j) + \phi_2^j(\alpha, \alpha_j)$$

with $\phi_1^j(\alpha, \alpha_j)$ described in property 4.1 and $\phi_2^j(\alpha, \alpha_j)$ in property 4.2.

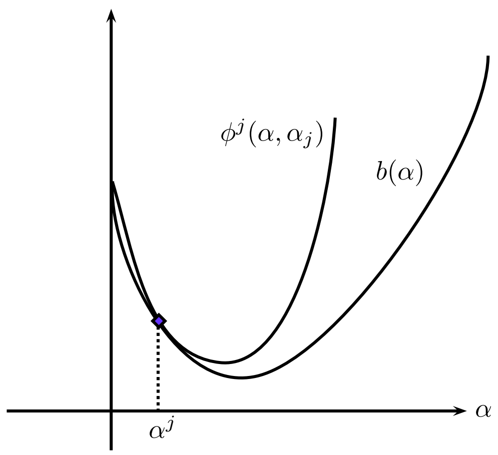

Figure 6: The construction of the tangent majorant of $b(\alpha)$ at α^j

Lemma 4.1. *According to properties 4.1 and 4.2 , we have, as illustrated on figure 6, for every $\alpha \geqslant \alpha^j$:*

$$\dot{b}(\alpha) - \dot{\phi}^j(\alpha, \alpha^j) \leqslant 0$$

and

$$b(0) - \phi^j(0, \alpha^j) = 0$$

Then, the assumption 3.4 holds if we use the tangent majorant described in this section and if m_p^j is chosen such that :

$$\dot{p}(\alpha) - \dot{q}^j(\alpha, \alpha^j) \leqslant 0, \quad \forall \alpha \geqslant \alpha^j \tag{4.8}$$

For instance, (4.8) *holds if m_p^j is the maximum of curvature of $p(\alpha)$ on the set* $[0; \bar{\alpha}[$.

Lemma 4.2. *Let $\{m_j^b\}_{0 \leqslant j \leqslant J}$ the series described in property 4.1. Then, for all* $j \in [0; J]$,

$$m_j^b \leqslant m_0^b$$

Proof. Let j an index in $[0; J]$. We need to prove that $m_j^b \leqslant m_0^b$ with :

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

Assume that $m_j^b > m_0^b$. We will prove that the following function is a tangent majorant for $b_1(\alpha)$ at α^j :

$$\tilde{\phi}_1^j(\alpha, \alpha^j) = b_1(\alpha^j) + (\alpha - \alpha^j)\dot{b}_1(\alpha^j) + \frac{1}{2} m_b^0 (\alpha - \alpha^j)^2$$

The function $b_1(\alpha)$ is convexe and its derivative is concave. Then $\ddot{b}_1(\alpha)$ is maximised on $[0; \bar{\alpha}[$ at $\alpha = 0$. $m_0^b = \ddot{b}_1(0)$ is the maximum of curvature of $b_1(\alpha)$ on $[0; \bar{\alpha}[$, so $\tilde{\phi}_1^j(\alpha, \alpha^j)$ is a tangent majorant of $b_1(\alpha)$ at α^j. This contradicts the following property :

$$m_b^j = \min \left\{ m \geqslant 0 | h(l) \leqslant h(l^j) + \dot{h}(l^j)(l - l^j) + \frac{1}{2} m (l - l^j)^2, \forall l \geqslant 0 \right\}$$

Then $m_j^b \leqslant m_0^b$ for all $j \in [0; J]$. ∎

Lemma 4.3. *For all index $j \in [0; J]$, for all α^j, there exists $\tilde{\alpha}^j \in [0; \alpha_j[$ such that :*

$$\dot{\phi}_2^j(\alpha^j, \alpha^j) = \frac{\gamma_b^j}{\bar{\alpha} - \alpha_j} \leqslant \ddot{b}(\tilde{\alpha}_j)$$

Proof. If $\alpha^j = 0$, then by construction we have

$$\dot{\phi}_2^j(0, 0) = \ddot{b}(0)$$

Then lemma 4.3 holds with $\tilde{\alpha}_j = 0$. Let consider the case when $\alpha^j > 0$. Let $l(\alpha)$ and $T(\alpha)$ be defined in (4.3) and (4.6). We have :

$$\dot{l}(\alpha^j) = \dot{l}(0) \leqslant \dot{T}(0)$$

Thus,

$$\ddot{\phi}_2(\alpha^j, \alpha^j)(\bar{\alpha} - \alpha^j) - \dot{b}_2(\alpha^j) \leqslant \ddot{b}_2(0)\bar{\alpha} - \dot{b}_2(0)$$

Using $\ddot{\phi}_2(\alpha^j, \alpha^j) = \frac{\gamma_b^j}{\bar{\alpha} - \alpha_j}$, we deduce :

$$\gamma_b^j \leqslant \dot{b}_2(\alpha^j) + \ddot{b}_2(0)\bar{\alpha} - \dot{b}_2(0)$$

$\dot{b}_2(\alpha)$ is continuous on $[0; \alpha^j]$, then there exists $\tilde{\alpha}^j \in]0; \alpha^j[$ such that :

$$\ddot{b}_2(\tilde{\alpha}^j) = \frac{\dot{b}_2(0) - \dot{b}_2(\alpha^j)}{\alpha^j}$$

Hence :

$$\gamma_b^j \leqslant -\alpha^j \ddot{b}_2(\tilde{\alpha}^j) + \ddot{b}_2(0)\bar{\alpha}$$

Moreover, the derivative of b_2 is convexe then $\ddot{b}_2(\alpha)$ is increasing :

$$\gamma_b^j \leqslant -\alpha^j \ddot{b}_2(\tilde{\alpha}^j) + \ddot{b}_2(\tilde{\alpha}^j)\bar{\alpha} = (\bar{\alpha} - \alpha^j)\ddot{b}_2(\tilde{\alpha}^j)$$

□

Property 4.3. *If $\nu_1^p > 0$ or if $\boldsymbol{A}$ is non degenerate i.e,* $\mathrm{Ker}(\boldsymbol{A}) = \{0\}$, *then the tangent majorant parameters γ_k^j and m_k^j described in theorem 4.1 fulfill* (4.9) *for all j, k. Thus, assumption 2.1 holds.*

Proof. Let us show that there exists some constants $0 < \nu_1 \leqslant \nu_2$ such that for each descent direction $\boldsymbol{d}_k \in \mathbb{R}^n$:

$$\nu_1 \|\boldsymbol{d}_k\|^2 \leqslant m_k^j + \frac{\gamma_k^j}{\bar{\alpha}_k - \alpha_k^j} \leqslant \nu_2 \|\boldsymbol{d}_k\|^2 \tag{4.9}$$

or equivalently :

$$\nu_1 \|\boldsymbol{d}_k\|^2 \leqslant \ddot{h}^j(\alpha^j, \alpha^j) \leqslant \nu_2 \|\boldsymbol{d}_k\|^2$$
$$\ddot{h}^j(\alpha^j, \alpha^j) = m_p^j + \mu(\ddot{\phi}_1^j(\alpha^j, \alpha^j) + \ddot{\phi}_2^j(\alpha^j, \alpha^j))$$

$\phi_1^j(\alpha, \alpha^j) + \phi_1^j(\alpha, \alpha^j)$ is a tangent majorant of $b(\alpha)$ at α^j then we have the following inequality :

$$\ddot{h}^j(\alpha^j, \alpha^j) \geqslant m_p^j + \mu \ddot{b}(\alpha^j)$$

According to lemmas 4.2 and 4.3 and since b_1 and b_2 are convexe,

$$\ddot{h}^j(\alpha^j, \alpha^j) \leqslant m_p^j + \mu\left(\ddot{b}_1(0) + \ddot{b}_2(\tilde{\alpha}^j)\right) \tag{4.10}$$
$$\leqslant m_p^j + \mu\left(\ddot{b}(0) + \ddot{b}(\tilde{\alpha}^j)\right) \tag{4.11}$$

For each iteration k and for all $\alpha \in [0; \alpha_k^j]$, $\boldsymbol{x}_k + \alpha \boldsymbol{d}_k$ is in $\mathcal{V}$ according to lemma 3.1 and according to assumption 1.1, there exists a positive bound M such that for all i and for all $\boldsymbol{x} \in \mathcal{V}$:

$$[\boldsymbol{A}\boldsymbol{x}]_i + \theta_i + \alpha[\boldsymbol{A}\boldsymbol{d}]_i \leqslant M, \quad \forall \alpha$$

Moreover, according to lemma 3.6,

$$[\boldsymbol{A}\boldsymbol{x}]_i + \theta_i + \alpha[\boldsymbol{A}\boldsymbol{d}]_i \geqslant \epsilon_0, \quad \forall \alpha$$

Hence for $\alpha = \tilde{\alpha}_k^j$:

$$\frac{1}{M^2}(\boldsymbol{A}\boldsymbol{d}_k)^T diag(\boldsymbol{t})(\boldsymbol{A}\boldsymbol{d}_k) \leqslant \boldsymbol{d}_k^T \nabla^2 B(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)\boldsymbol{d}_k \leqslant \frac{1}{\epsilon_0^2}(\boldsymbol{A}\boldsymbol{d}_k)^T diag(\boldsymbol{t})(\boldsymbol{A}\boldsymbol{d}_k)$$

$\boldsymbol{t}$ contains strictly positive terms so $\boldsymbol{T} = diag(\boldsymbol{t})$ is a symmetric positive definite matrix. Then $\boldsymbol{A}^T\boldsymbol{T}\boldsymbol{A}$ is symmetric positive semidefinite and there exists $0 \leqslant \eta \leqslant \bar{\eta}$ such that

$$\eta\|\boldsymbol{v}\|^2 \leqslant \boldsymbol{v}^T\boldsymbol{A}^T\boldsymbol{T}\boldsymbol{A}\boldsymbol{v} \leqslant \bar{\eta}\|\boldsymbol{v}\|^2$$

If we assume either $nu_1^p > 0$ or $\eta > 0$ i.e., $\boldsymbol{A}$ is non degenerate, then 2.1 holds with $\nu_1 = \nu_1^p + \mu\frac{\eta}{M^2}$ and $\nu_2 = \nu_2^p + \mu\frac{2\bar{\eta}}{\epsilon_0^2}$. $\square$

5. Conclusion

The quadratic MM line search method proposed in [9] is simple and efficient, but it is restricted to gradient-Lipschitz criteria, which excludes the cases of barrier functions that are frequently encountered in signal and image reconstruction. Thoses cases might be treated with the MM line search presented in this paper, which benefits from strong convergence results and is still very easy to implement. Then we eluded the More-Thuente algorithm often coupled with a tedious adjustement of Wolfe parameters by replacing it with an analytical iterative method, prooved to lead convergence of the overall algorithm for any value of the number of iterates J. We have demonstrated that the convergence theorem of [9] still applied with this new form of tangent majorant, leading to the convergence of nonlinear conjugate gradient with conjugacy PRP, LS and also HS for the strictly convex case. Moreover, the step size strategy leads to the convergence of others large scale methods such as the L-BFGS and truncated Newton algorithms. Finally, under an additional assumption on the tangent majorant, the convergence is proved for NLCG with conjugacy formulas PRP+, FR and DY.

References

[1] Bertsekas, D. P., 1999. Nonlinear Programming, 2nd Edition. Athena Scientific, Belmont, MA.

[2] Chen, X., Sun, J., 2002. Global convergence of two-parameter family of conjugate gradient methods without line search. Journal of Computational and Applied Mathematics 146, 37–45.

[3] Dai, Y., Yuan, Y., 2001. A three-parameter family of nonlinear conjugate gradient methods. Mathematics of Computation 70, 1155–1167.

[4] Erdogan, H., Fessler, J., September 1999. Monotonic algorithms for transmission tomography. IEEE Transactions on Medical Imaging 18 (9), 801–814.

[5] Gilbert, J. C., Nocedal, J., 1992. Global convergence properties of conjugate gradient methods for optimization. SIAM Journal on Optimization 2 (1), 21–42.

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

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

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

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

[10] Liu, D. C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Mathematical Programming 45 (3), 503–528.

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

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

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

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