INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Superlinear Convergence of a Reduced BFGS Method with Piecewise Line-Search and Update Criterion

Jean Charles GILBERT

N° 2140
Décembre 1993

PROGRAMME 5

Traitement du signal,
automatique et
productique

Rapport de recherche

1993

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

SUPERLINEAR CONVERGENCE OF A REDUCED BFGS METHOD WITH PIECEWISE LINE-SEARCH AND UPDATE CRITERION

CONVERGENCE SUPERLINEAIRE D'UNE METHODE DE BFGS REDUITE AVEC RECHERCHE LINEAIRE BRISEE ET CRITERE DE MISE-A-JOUR

par

Jean Charles GILBERT

Rapport de Recherche n° 2140

13 Décembre 1993

SUPERLINEAR CONVERGENCE OF A REDUCED BFGS METHOD WITH PIECEWISE LINE-SEARCH AND UPDATE CRITERION

CONVERGENCE SUPERLINEAIRE D'UNE METHODE DE BFGS REDUITE AVEC RECHERCHE LINEAIRE BRISEE ET CRITERE DE MISE-A-JOUR

Jean Charles GILBERT[†]

ABSTRACT

We show the q-superlinear convergence of a reduced BFGS method for equality constrained problems, using eventually only one constraint linearization per iteration. The local method is globalized either with a standard arc-search or, when an update criterion is satisfied, with a piecewise line-search. The aim of the latter technique is to realize generalized Wolfe conditions, which allow the algorithm to maintain naturally the positive definiteness of the generated matrices. We show that if the sequence of iterates converges, the convergence is q-superlinear. No assumption is made on the speed of convergence of the sequence of iterates or on the boundedness of the sequence of generated matrices. The main difficulty is to show that the ideal step-size is accepted after finitely many steps.

RÉSUMÉ

Nous montrons la convergence q-superlinéaire d'une méthode de BFGS réduite pour minimiser une fonction sous contraintes d'égalité, n'utilisant asymptotiquement qu'une seule linéarisation des contraintes par itération. La méthode locale est globalisée soit par une recherche sur arc classique soit, lorsqu'un critère de mise à jour est vérifié, par une recherche linéaire par morceaux. Cette dernière technique vise à réaliser des conditions de Wolfe généralisées, qui permettent de maintenir de façon naturelle la définie positivité des matrices générées. Nous montrons que si la suite des itérés converge, la convergence est q-superlinéaire. Aucune hypothèse n'est faite sur la vitesse de convergence de la suite des itérés ni sur la bornitude de la suite des matrices générées. La difficulté principale est de montrer que le pas idéal est accepté après un nombre fini d'étapes.

Key words: BFGS formula, equality constrained optimization, piecewise line-search, reduced quasi-Newton, successive quadratic programming, superlinear convergence, update criterion.

Abbreviated title: Superlinear convergence of RQN methods.

AMS Subject Classification: primary: 65K05; secondary: 49M37, 90C30.

[†] INRIA, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France).

1 Introduction

In this paper we consider a particular algorithm for minimizing a function defined on an *open* set Ω of $\mathbb{R}^n$, in the presence of m ($m < n$) nonlinear equality constraints. We denote the real-valued function to minimize by $f : \Omega \to \mathbb{R}$ and the constraint function by $c : \Omega \to \mathbb{R}^m$, so that the problem can be written

$$\begin{cases} \min f(x) \\ c(x) = 0, \quad x \in \Omega. \end{cases} \tag{1.1}$$

Since Ω is supposed open, we do not take into account general or inequality constraints. This set is simply the collection of points where nice properties hold. For example, throughout, the $m \times n$ Jacobian matrix of the constraints, denoted by $A(x)$, is supposed surjective (i.e., to have full row rank) for all x in Ω.

The considered algorithm is in the family of *reduced quasi-Newton methods.* This means that the gradients of f and c_i ($1 \leq i \leq m$) have to be computed, while their second derivatives are conveniently approximated by a single matrix B_k of order $n - m$, representing the *reduced* Hessian of the Lagrangian function

$$\ell(x, \lambda) = f(x) + \lambda^\top c(x).$$

We have denoted by $\lambda \in \mathbb{R}^m$ the vector of Lagrange multipliers associated to the constraints. The Hessian (with respect to x) of the Lagrangian is denoted by

$$L(x, \lambda) = \nabla^2_{xx} \ell(x, \lambda).$$

The reduced Hessian is defined below.

The algorithm looks for a pair (x_*, λ_*), solution of the first order optimality conditions: $c(x_*) = 0$ and $\nabla_x \ell(x_*, \lambda_*) = 0$. For this, it generates a sequence of iterates y_k, which locally (i.e., near a solution) is expected to follow the recurrence

$$y_{k+1} = y_k + t_k + r_{k+1}. \tag{1.2}$$

In this formula, t_k is the *tangent* or *longitudinal* displacement and r_{k+1} is the *restoration* or *transversal* displacement. We also define a sequence $\{x_k\}_{k \geq 2}$ from the sequence $\{y_k\}_{k \geq 1}$ by

$$x_{k+1} = y_k + t_k. \tag{1.3}$$

Locally, this sequence should satisfy the recurrence $x_{k+1} = x_k + r_k + t_k$.

The tangent step t_k belongs to $N(A(y_k))$, the null space of $A(y_k)$, which is also the tangent space to the manifold $\mathcal{M}(y_k) = \{x \in \Omega : c(x) = c(y_k)\}$ at y_k. Let $Z^-(y_k)$ be a basis of this space, that is to say an $n \times (n - m)$ injective matrix (i.e., a matrix with full column rank) such that

$$A(y_k) Z^-(y_k) = 0.$$

We note $Z_*^- = Z^-(x_*)$, $L_* = L(x_*, \lambda_*)$, and

$$B_* = Z_*^{-\mathsf{T}} L_* Z_*^-,$$

the *reduced Hessian* of the Lagrangian at the solution. Then t_k has the form

$$t_k = -Z^-(y_k) B_k^{-1} g(y_k), \tag{1.4}$$

in which $g(y_k)$ is the *reduced gradient* of f at y_k, a vector of $\mathbb{R}^{n-m}$ defined by

$$g(y) = Z^-(y)^{\mathsf{T}} \nabla f(y),$$

and B_k is an updated matrix of order $n - m$, approximating B_*.

The restoration step r_{k+1} aims at reducing the norm of the constraint function c. It has the form

$$r_{k+1} = -A^-(y_k) c(y_k + t_k), \tag{1.5}$$

where $A^-(y_k)$ is an $n \times m$ injective matrix, which is a right inverse of $A(y_k)$:

$$A(y_k) A^-(y_k) = I.$$

Then, the range space $R(A^-(y_k))$ is complementary to $N(A(y_k))$, which implies that the displacement r_{k+1} is transversal with respect to the local tangent space. Note the useful identity

$$A^-(y_k) A(y_k) + Z^-(y_k) Z(y_k) = I,$$

which uniquely determines $Z(y_k)$ as an $(n-m) \times n$ surjective matrix. For additional information on the choice of the matrices $A^-(y)$, $Z^-(y)$, and $Z(y)$, see Gabay [12].

Basically, this local algorithm was proposed by Coleman and Conn [8]. Remark that the method differs from the so-called *reduced sequential quadratic programming* (SQP) *method*, in which the restoration step is obtained by evaluating c at y_k instead of $y_k + t_k$: see Murray and Wright [21], and Gabay [13]. This latter method is simpler but less efficient locally: see Byrd [2, 3], Gilbert [14, 16], Hoyer [19], and Yuan [28].

To make the method useful in practice, it is necessary to globalize it, that is to say to force its convergence when the starting pair (y_1, B_1) is not necessary close to (x_*, B_*). In this case, the recurrence (1.3) may not be appropriate. When the matrix B_k is computed from the second derivatives of f and c, it is usually satisfactory to get the next iterate by determining a step-size α_k along an arc originating from y_k in order to force the decrease of a penalty function such as

$$\Theta_\sigma(y) = f(y) + \sigma \|c(y)\|. \tag{1.6}$$

The *penalty parameter* is $\sigma > 0$ and $\|\cdot\|$ is a norm on $\mathbb{R}^m$. A simple Armijo [1] backtracking from a fixed value for the initial step-size is generally satisfactory.

In the quasi-Newton version of the algorithm, the matrix B_k is updated by a formula, using a pair (γ_k, δ_k) of vectors in $\mathbb{R}^{n-m}$. A possible good choice is to take for γ_k the change in the reduced gradient and for δ_k the corresponding reduced displacement (see [14, 15], and also Coleman and Conn [9], and Nocedal and Overton [22] for related choices). For guaranteeing a descent property of t_k on Θ_σ, it is desirable to maintain B_k positive definite, which for most update formulæ amounts to satisfy the condition

$$\gamma_k^\top \delta_k > 0. \tag{1.7}$$

Powell [26] has suggested to enforce this inequality by modifying the vector γ_k in a convenient manner if necessary. In our point of view, this technique is somehow artificial and it is probably more satisfactory to preserve the original geometrical meaning of inequality (1.7), which is that some function must look convex at the points considered. For this reason, it is attractive, as in unconstrained optimization with the Wolfe line-search, to design a search algorithm that can find a point such that inequality (1.7) holds. This raises a problem of compatibility with the necessity to decrease Θ_σ in the same time and it is no longer guaranteed that an appropriate point exists along a straight line. It has been proposed in [18] a piecewise line-search (PLS) algorithm that can satisfactorily face this difficulty by readjusting the search along new directions at some intermediate points. The method is presented in Section 2.

There is an additional difficulty however, which comes from a weak point of the algorithm when it is used in the quasi-Newton framework. Nocedal and Overton [22] and Gilbert [14, 15] have observed that an update of the matrix B_k is not always desirable. There are several ways of seeing this. For example, take a parametric representation of the optimal manifold $\mathcal{M}_* = \{x \in \Omega : c(x) = 0\}$ around x_*, which is a map $\psi : U \subset \mathbb{R}^{n-m} \to \mathbb{R}^n$ verifying in particular $c(\psi(u)) = 0$ for all u in the neighborhood U of 0. It is always possible to chose ψ such that $\psi(0) = x_*$ and $\psi'(u) = Z^-_\cdot(\psi(u))$ for all $u \in U$. Then, for $x = \psi(u)$, $g(x)$ is the gradient of $(f \circ \psi)$ at u and G_* is its Hessian at 0. This shows that to collect the proper information on G_*, it is suitable to use variations of the reduced gradient between points belonging to the same manifold of constant constraint, $\mathcal{M}(y_k)$ say, or to its tangent plane. This is safe but expensive, since it requires an extra evaluation of the reduced gradient at an intermediate point. The analysis made by Coleman and Conn [9], Byrd and Nocedal [5], and Gilbert [17] assumes that the user accepts this additional computation. Now, one could argue that the reduced gradient is also constant on manifolds $\mathcal{N}(y)$ complementary to $\mathcal{M}(y_k)$, so that the reduced gradient at the next iterate could also be used, provided one knows (at least at the first order) where $\mathcal{N}(y_{k+1})$ and $\mathcal{M}(y_k)$ intersect. In the framework of *reduced* quasi-Newton methods, however, there is apparently no way of knowing this, because the tangent space to $\mathcal{N}(y_{k+1})$ is a second order information, which is not stored in the reduced Hessian. The *full* Hessian contains information on the tangent space to the manifolds $\mathcal{N}$ and it is in this sense that *reduced* quasi-Newton methods have some weakness.

The idea proposed by Nocedal and Overton [22] and Gilbert [14, 15] to overcome this difficulty works surprisingly well, at least in theory. It consists in updating the matrix only when the restoration component r_{k+1} of the step is sufficiently small with respect to the tangential component t_k. In this case, the total step is almost tangent and the error introduced in B_{k+1} by taking $\gamma_k = g(y_{k+1}) - g(y_k)$ is minimized. As we shall see, this strategy and the BFGS theory guarantee that each step for which an update occurs provides a superlinear improvement. Now, the update criterion that selects the iteration at which an update is appropriate must not be too restrictive. Otherwise the matrix is almost never updated and the method has little chance of being superlinear at each step. A compromise has to be found. Nocedal and Overton [22] proposed an update criterion that turned out to be adequate locally and when the ideal step-size is taken. Their update criterion depends, however, on the iteration index, which is not very attractive in a global framework. From this point of view, the update criterion proposed in [14, 15] is more satisfactory, because it is the algorithm itself that decides to what extent r_{k+1} must be small with respect to t_k. There, the study is made in a global framework, without supposing that the ideal step-size is taken.

In this paper, we carry on with the update criterion of [14, 15] and we show how to integrate it in a method using the PLS technique introduced in [18]. This leads to an improvement of the criterion in the sense that there is a constant that need no longer be tuned by the algorithm. This is due to the fact that the search technique is always able to realize (1.7). Then, we show that the globalization technique proposed in [18] works well in the sense that the transition to the local method (1.2) is smooth and that the potential superlinear convergent property of the local method is actually achieved in normal circumstances. For this, we suppose that the algorithm generates a sequence $\{y_k\}$ that converges to a solution x_* and we show that the convergence of the sequence $\{x_k\}$ defined by (1.3) is q-superlinear. This requires to show that the ideal step-size is accepted after finitely many iterations. We stress the fact that we do not assume any speed of convergence for the sequences $\{y_k\}$ or $\{x_k\}$, or any hypotheses on the quality of the matrices B_k.

Other papers have analyzed the speed of convergence of quasi-Newton methods for equality constrained optimization. Let us review those not assuming that B_1 is close to B_*, which is also an assumption we avoid. The first contribution is probably due to Powell [27] who proves the r-superlinear convergence of SQP methods with the so-called Powell's modification of the BFGS formula. This is weaker than the q-superlinear rate of convergence proved here. He also assumes that the unit step-size is accepted, that the sequence of matrices approximating the Hessian of the Lagrangian is bounded, and that these matrices are uniformly positive definite on the tangent space to the constraint manifold at the solution. All these assumptions are not necessary here and are not used in the following papers. Byrd and Nocedal [5] consider the reduced SQP method with a possible correction involving the evaluation of the constraints after the reduced SQP step. They obtain results similar to ours but their algorithm requires two linearizations of the constraints per iteration. Byrd, Tapia, and Zhang [6] consider an augmented Lagrangian version of the SQP method and prove that convergence of the generated sequence implies its r-superlinear convergence. This is the strongest result proved so far for an algorithm linearizing the constraints only once per iteration. They also show that if the penalty parameter is eventually maintained fixed to a sufficiently large value, the convergence is q-superlinear. Unfortunately, the threshold value for the penalty parameter is usually unknown, which makes this result less attractive. Finally, Byrd and Xie [7] extend the work of Byrd and Nocedal [5] in the sense that they now consider the reduced SQP method with only one linearization of the constraints per iteration and various update criteria. They prove global and r-linear convergence with hypotheses similar to ours.

This review shows that what is proved in this paper is an improvement over known results since, as we said above, we show that the convergence of the sequence $\{y_k\}$ implies the q-superlinear convergence of the sequence $\{x_k\}$. Furthermore, our algorithm requires only one linearization of the constraints per iteration, asymptotically. Before this result, we believe it was an open question to know whether the q-superlinear convergence of quasi-Newton methods in constrained optimization was possible with only one constraint linearization per iteration. We have used three ingredients to achieve this goal: (i) the techniques developed by Powell [24, 25], and Byrd and Nocedal [4] for handling the contribution of the BFGS update scheme to the superlinear convergence, as well as their extension to constrained problems (see Byrd and Nocedal [5]); (ii) the update criterion introduced in [14, 15], although there are probably other update criteria leading to the same result; and (iii) the piecewise line-search introduced in [18], which guarantees (1.7) whenever desired.

The paper is organized as follows. In Section 2, we state the complete algorithm including the search technique and the update criterion. Then, in Section 3, we prove the q-superlinear convergence of the sequence generated by the algorithm, assuming its convergence to a point satisfying sufficient conditions of optimality. This is done by proving first the r-linear convergence of the sequence $\{y_k\}$ (Section 3.2), next by analyzing the conditions of admissibility of the ideal step-size and by proving their achievement (Sections 3.3 and 3.4), and finally by proving the superlinear convergence of the sequence $\{x_k\}$ (end of Section 3.4).

2 The algorithm

2.1 Some notation

Before getting to the heart of the matter, let us give some more notation. Even in a global framework, we note $x_{k+1} = y_k + t_k$,

$$e_k = y_{k+1} - y_k, \quad \text{and} \quad s_k = x_{k+1} - x_k. \tag{2.1}$$

Hence, because of the globalization technique, e_k may differ from $t_k + r_{k+1}$ and s_k may differ from $r_k + t_k$. We also introduce the following Lagrange multiplier estimate

$$\lambda(y) = -A^-(y)^\top \nabla f(y).$$

From the optimality conditions, $\lambda(x_*) = \lambda_*$. The following identity is used many times:

$$g'(x_*) = Z_*^{-\top} L_*. \tag{2.2}$$

It comes from the observation that $g(y) = Z^-(y)^\top \nabla_x \ell(y, \lambda_*)$ and from the optimality condition $\nabla_x \ell(x_*, \lambda_*) = 0$.

We denote by $\|\cdot\|_D$ the dual norm of the norm $\|\cdot\|$ used in the penalty function (1.6). It is defined by

$$\|v\|_D = \sup_{\|u\|=1} v^\top u.$$

If $\mathcal{K}$ is a subset of $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$, we denote by $\mathcal{K}^c = \mathbb{N}^*\backslash\mathcal{K}$ its complementary set in $\mathbb{N}^*$. We denote by $|\mathcal{S}|$ the number of elements in a finite set $\mathcal{S}$.

If $\{a_k\}_{k\geq 1}$ and $\{b_k\}_{k\geq 1}$ are two sequences in a normed space, we note $a_k = O(\|b_k\|)$ if there is a constant C such that $\|a_k\| \leq C\|b_k\|$ for all $k \geq 1$, we note $a_k = o(\|b_k\|)$ if for any $\epsilon > 0$ there is an index k_ϵ such that $\|a_k\| \leq \epsilon\|b_k\|$ for all $k \geq k_\epsilon$, and we note $a_k \sim b_k$ if $a_k = O(\|b_k\|)$ and $b_k = O(\|a_k\|)$.

Finally, $\xi'(x; d)$ denotes the directional derivative of a function ξ along a direction d.

2.2 Description of the algorithm

At the beginning, the iteration index k is set to 1. When the kth iteration starts, an iterate $y_k = y_k^0 \in \Omega$ is known, as well as a symmetric positive definite matrix B_k approximating the reduced Hessian of the Lagrangian. The computation of the next iterate y_{k+1} is done by a 2-stage search, which differs somehow from the piecewise line-search (PLS) introduced in [18] because we want to recover algorithm (1.2) when the ideal step-size is accepted. In [18], the reduced SQP algorithm was chosen as local method, for simplicity. The first stage is mandatory, while the second stage is optional: its realization depends on an update criterion.

The first stage mainly aims at decreasing the penalty function Θ_{σ_k}, in which the penalty parameter $\sigma_k > 0$ may depend on the current iterate. Indeed, it is always required that $\sigma_k \geq \|\lambda(y_k)\|_D + \bar{\sigma}$. If this inequality does not hold for the current penalty parameter, this one is increased before the search begins. There are many possible ways of updating the penalty parameter σ_k. The important point is that the following properties be satisfied ($\bar{\sigma}$ is a fixed positive number):

$$\sigma_k \geq \|\lambda(y_k)\|_D + \bar{\sigma}, \quad \forall k \geq 1, \tag{2.3}$$

$$\exists \text{ an index } k_1, \ \forall k \geq k_1, \ \sigma_{k-1} \geq \|\lambda(y_k)\|_D + \bar{\sigma} \quad \Longrightarrow \quad \sigma_k = \sigma_{k-1}, \tag{2.4}$$

$$\{\sigma_k\} \text{ is bounded} \quad \Longrightarrow \quad \sigma_k \text{ is modified finitely often.} \tag{2.5}$$

We shall suppose that the rule for updating σ_k satisfies these properties. An example of such rule is

$$\text{if } \sigma_{k-1} \geq \|\lambda(y_k)\|_D + \bar{\sigma}, \text{ then } \sigma_k = \sigma_{k-1}, \text{ else } \sigma_k = \max(\tilde{\sigma}\sigma_{k-1}, \|\lambda(y_k)\|_D + \bar{\sigma}),$$

where $\tilde{\sigma} > 1$ is a constant (see Mayne and Polak [20]).

For decreasing Θ_{σ_k}, we introduce the reduced SQP direction, defined by

$$d_k^0 = -Z^-(y_k)B_k^{-1}g(y_k) - A^-(y_k)c(y_k). \tag{2.6}$$

If (2.3) holds and if y_k is not stationary, then

$$\Theta'_{\sigma_k}(y_k; d_k^0) = -g(y_k)^\top B_k^{-1} g(y_k) + \lambda(y_k)^\top c(y_k) - \sigma_k \|c(y_k)\|$$

is negative: d_k^0 is a descent direction of Θ_{σ_k} at y_k.

To recover the local method (1.2) for a unit step-size, a search is done along the path

$$\alpha \mapsto p_k^0(\alpha) = y_k + \alpha d_k^0 - \alpha^2 A^-(y_k)\Big(c(x_{k+1}) - c(y_k)\Big). \tag{2.7}$$

Note that $p_k^0(1) = y_k + t_k + r_{k+1}$ and that d_k^0 is tangent to p_k^0 at $\alpha = 0$. Because $y_k \in \Omega$ and d_k^0 is a descent direction of Θ_{σ_k} at y_k, it is always possible to find a step-size $\alpha_k^1 > 0$ such that

$$y_k^1 = p_k^0(\alpha_k^1) \in \Omega \tag{2.8}$$

and

$$\Theta_{\sigma_k}(y_k^1) \leq \Theta_{\sigma_k}(y_k) + \omega_1 \, \nu_k^0(\alpha_k^1), \tag{2.9}$$

where $\omega_1 \in (0, 1/2)$ and $\nu_k^0(\alpha) = \alpha \Theta'_{\sigma_k}(y_k; d_k^0)$.

Assumptions 2.1. It is supposed that the determination of α_k^1 obeys the following rules. First a trial is made with $\alpha_k^{0,1} = 1$ and as long as conditions (2.8) and (2.9) do not hold with $\alpha_k^1 = \alpha_k^{0,j}$ ($j \geq 1$), a new trial is made with $\alpha_k^{0,j+1} < \alpha_k^{0,j}$. The step-sizes $\alpha_k^{0,j}$ must be chosen such that $\alpha_k^{0,j} \to 0$ when $j \to \infty$ and $\alpha_k^{0,j+1} \geq \beta \alpha_k^{0,j}$, for $j \geq 1$ and some constant $\beta \in (0, 1)$.

For example, Armijo's [1] backtracking satisfies these rules by taking $\beta \in (0,1)$ and $\alpha_k^{0,j} = \beta^{j-1}$, but it is usually better to use safeguarded interpolation formulæ. A consequence of these assumptions is that $\alpha_k^1 \le 1$.

Next a test is performed to know whether an update of the matrix B_k is appropriate. The statement of the update criterion requires some notation. Let $\mathcal{K}$ denote the set of indices at which an update occurs in the considered run: when $k \in \mathcal{K}$, B_{k+1} is computed from B_k by the BFGS formula (see below), otherwise $B_{k+1} = B_k$. Then, for an index $k \in \mathbb{N}^*$, we define

$$k \ominus 1 = \begin{cases} 1 & \text{if } i \ge k \text{ for all } i \in \mathcal{K}, \\ \max\{i \in \mathcal{K} : i < k\} & \text{otherwise.} \end{cases}$$

Hence, if k is the index of the current iterate, $k \ominus 1$ is the index of the last iteration at which an update occurred. For $j > 1$, we define by induction $k \ominus j = (k \ominus (j-1)) \ominus 1$. Similarly, when $\mathcal{K}$ is unbounded, we define

$$k \oplus 1 = \min\{i \in \mathcal{K} : i > k\}$$

and, for $j > 1$, $k \oplus j = (k \oplus (j-1)) \oplus 1$. Figure 1 illustrates these notions. We have

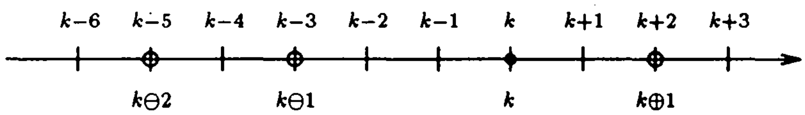

Figure 1: The meaning of $k \ominus j$ and $k \oplus j$.

circled the iterates at which an update occurs: $k-5$, $k-3$, and $k+2$.

We can now write the update criterion, which is

$$\Big(\|c(y_k)\| + \|c(x_{k+1})\|\Big) \le \mu \, \|e_{k\ominus 2}^1\| \, \|B_k^{-1} g(y_k)\|, \tag{2.10}$$

where $\mu > 0$ is an arbitrary constant and $e_k^1 = y_k^1 - y_k$. It is also possible to use $e_{k\ominus 2}$ instead of $e_{k\ominus 2}^1$ is (2.10), but this would require to memorize y_k during the PLS. When (2.10) is satisfied at iteration k, an update of B_k is desirable. Observe that, when $\mathcal{K}$ is unbounded, $e_{k\ominus 2}^1 \to 0$ and inequality (2.10) holds asymptotically when $c(y_k)$ and $c(y_{k+1})$ are sufficiently small with respect to the reduced tangent displacement $Z(y_k)t_k$. This is in the spirit of update criteria introduced by Nocedal and Overton [22] and Gilbert [14, 15]. In the former paper, the criterion is of the form

$$\|c(y_k)\| \le \frac{\mu_1}{k^{1+\nu}} \|B_k^{-1} g(y_k)\|,$$

where $\mu_1 > 0$ and $\nu > 0$ are constants. If this choice is convenient for a local analysis, it has some defects in practice. First, the factor $\mu_k = \mu_1 / k^{1+\nu}$ of $\|B_k^{-1} g(y_k)\|$

depends on the iteration index, which is not very appealing in a global framework. Secondly, if the constant ν is arbitrary, μ_1 has to be chosen sufficiently small for their Theorem 4.3 to apply. Also, no rule for tuning μ_1 correctly is provided in [22]. We believe it is better to let the algorithm choose itself the appropriate sequence $\{\mu_k\}$. The choice

$$\mu_k = \mu \,\|e^1_{k\ominus 2}\|$$

in (2.10) is an improvement on a criterion proposed in [14, 15], in the sense that the factor of $\|e^1_{k\ominus 2}\|$ (here μ) need no longer be adjusted by the algorithm. This is now possible because the PLS that follows is always able to satisfy (1.7). One property of the above formula is to decrease μ_k only when an update occurs. In particular, μ_k remains bounded away from zero when $\mathcal{K}$ is bounded. The choice of the index $k \ominus 2$ is motivated by a technique of proof that will be discussed after Lemma 3.22.

If the criterion (2.10) is not satisfied, the search is finished at y^1_k and we set

$$i_k = 1, \quad \alpha_k = \alpha^1_k, \quad y_{k+1} = y^1_k, \quad \text{and} \quad \nu_k = \nu^0_k.$$

In this case, the matrix B_k is not updated. Otherwise, the search is pursued with the aim of finding a point $y_{k+1} \in \Omega$ satisfying the conditions

$$\Theta_{\sigma_k}(y_{k+1}) \le \Theta_{\sigma_k}(y_k) + \omega_1\, \nu_k(\alpha_k), \tag{2.11}$$

$$g(y_{k+1})^\top Z(y_k) t_k \ge \omega_2\, g(y_k)^\top Z(y_k) t_k, \tag{2.12}$$

where $\omega_2 \in (0,1)$ and ν_k is a function that will be given in a moment. Note that the PLS algorithm do not need $\omega_2 > \omega_1$ to succeed. The search for the point y_{k+1} is done along a piecewise affine path originating from y^1_k and affine between the points y^1_k, y^2_k, ..., $y^{i_k}_k = y_{k+1}$. The determination of the intermediate points y^i_k, the corresponding step-sizes $0 = \alpha^0_k < \alpha^1_k < \cdots < \alpha^{i_k}_k = \alpha_k$ and the functions ν^i_k playing the role of ν_k in (2.11) is done recursively as follows.

Given α^i_k, y^i_k, and ν^{i-1}_k ($i \ge 1$), one checks whether the following *curvature condition* holds

$$g(y^i_k)^\top Z(y_k) t_k \ge \omega_2\, g(y_k)^\top Z(y_k) t_k. \tag{2.13}$$

If such is the case, the search is finished with

$$i_k = i, \quad \alpha_k = \alpha^i_k, \quad y_{k+1} = y^i_k, \quad \text{and} \quad \nu_k = \nu^{i-1}_k.$$

Otherwise, a step-size is taken from y^i_k along the direction

$$d^i_k = -\tau^i_k Z^-(y^i_k) B_k^{-1} g(y_k) - A^-(y^i_k) c(y^i_k), \tag{2.14}$$

in which a positive tangent scaling factor τ^i_k is introduced (we set $\tau^0_k = 1$). This scaling factor is important for practical reasons (see [18]). It is supposed to satisfy the following assumption.

Assumption 2.2. There are numbers $\tau_{\min}$ and $\tau_{\max}$, such that $0 < \tau_{\min} \le \tau_k^i \le \tau_{\max}$, for all $k \ge 1$ and $0 \le i \le i_k - 1$.

The direction d_k^i is the reduced SQP direction with a tangential reduced component $Z(y_k^i)d_k^i = -\tau_k^i B_k^{-1} g(y_k)$ inherited from the first stage of the search and scaled by τ_k^i.

The function ν_k^i that forces the decrease of Θ_{σ_k} is defined by

$$\nu_k^i(\alpha) = \begin{cases} \nu_k^{i-1}(\alpha), & \text{for } 0 \le \alpha < \alpha_k^i \\ (1-\rho_k^i)\nu_k^{i-1}(\alpha_k^i) + \rho_k^i \left(-\frac{T_k^i}{\omega_1}\right) + (\alpha - \alpha_k^i)\Theta'_{\sigma_k}(y_k^i; d_k^i), & \text{for } \alpha \ge \alpha_k^i, \end{cases}$$

where ρ_k^i is any number chosen in $[0,1]$ and

$$T_k^i = \Theta_{\sigma_k}(y_k) - \Theta_{\sigma_k}(y_k^i)$$

is the *total decrease* of Θ_{σ_k} from y_k to y_k^i. The choice of ρ_k^i is arbitrary in $[0,1]$ and will be discussed below.

Next, a step-size $(\alpha_k^{i+1} - \alpha_k^i)$ is chosen such that

$$\alpha_k^{i+1} \in (\alpha_k^i, \alpha_k^i + 1]$$

and such that the point $y_k^{i+1} = y_k^i + (\alpha^{i+1} - \alpha_k^i)d_k^i$ satisfies $y_k^{i+1} \in \Omega$ and the *descent condition*

$$\Theta_{\sigma_k}(y_k^{i+1}) \le \Theta_{\sigma_k}(y_k) + \omega_1\, \nu_k^i(\alpha_k^{i+1}). \tag{2.15}$$

Observe that, by taking $\rho_k^i = 0$ for all i, the function ν_k^i is continuous; while, with $\rho_k^i = 1$, (2.15) is more demanding and becomes

$$\Theta_{\sigma_k}(y_k^{i+1}) \le \Theta_{\sigma_k}(y_k^i) + \omega_1\,(\alpha_k^{i+1} - \alpha_k^i)\,\Theta'_{\sigma_k}(y_k^i; d_k^i),$$

which is the usual descent condition for a search starting from y_k^i along d_k^i. It is indeed more demanding, since the factor of ρ_k^i in the definition of ν_k^i is negative, by (2.15) at the preceding iteration. We study the PLS with non fixed $\rho_k^i \in [0,1]$, since it is not clear which of the options $\rho_k^i = 0$ or $\rho_k^i = 1$ is the best in practice.

The observation above indicates that condition (2.15) can be realized for α_k^{i+1} sufficiently closed to α_k^i, provided

$$\Theta'_{\sigma_k}(y_k^i; d_k^i) = -\tau_k^i g(y_k^i)^\top B_k^{-1} g(y_k) + \lambda(y_k^i)^\top c(y_k^i) - \sigma_k \|c(y_k^i)\|$$

is negative. This is the case if

$$\sigma_k \ge \|\lambda(y_k^i)\|_D + \bar{\sigma}. \tag{2.16}$$

We do not allow σ_k to change during the search. Therefore, the PLS is interrupted at $y_k^i = y_{k+1}$ if (2.16) does not hold. Then, this point may not satisfy the curvature condition (2.12). On the other hand, it is shown in [18] that if it is not interrupted because of the failure of (2.16) and if the determination of the step-size candidates α_k^i obeys some safeguard rules, then the search usually terminates in a finite number of trials i_k with a step-size $\alpha_k = \alpha_k^{i_k}$ and a point $y_{k+1} = y_k^{i_k}$ satisfying (2.11) with $\nu_k = \nu_k^{i_k-1}$ and (2.12).

If the search algorithm is interrupted after the first stage because the update criterion (2.10) is not satisfied or at y_k^i because condition (2.16) does not hold, then the matrix B_k is not updated. Otherwise, an update is done with the BFGS formula:

$$B_{k+1} = B_k - \frac{B_k\delta_k\delta_k^\top B_k}{\delta_k^\top B_k\delta_k} + \frac{\gamma_k\gamma_k^\top}{\gamma_k^\top\delta_k}, \tag{2.17}$$

where γ_k and δ_k are defined for all $k \geq 1$ by

$$\gamma_k = g(y_{k+1}) - g(y_k), \quad \delta_k = \bar{\alpha}_k Z(y_k)t_k, \tag{2.18}$$

and

$$\bar{\alpha}_k = \sum_{i=0}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i)\tau_k^i. \tag{2.19}$$

The use of this formula makes sense because, from (2.12), $\gamma_k^\top\delta_k > 0$ (more convincing geometrical considerations are discussed in [18]). Hence, the positive definiteness of B_k is transmitted to B_{k+1} (see for instance Dennis and Moré [10] or Dennis and Schnabel [11]). The aim of the described search is precisely to get $\gamma_k^\top\delta_k > 0$ when an update is desirable.

At the new point y_{k+1} a new iteration can begin.

Proposition 2.3. *When the search algorithm succeeds in finding a new point y_{k+1} (i.e., there is no cycling), one has*

$$\Theta_{\sigma_k}(y_{k+1}) \leq \Theta_{\sigma_k}(y_k) + \omega_1 \sum_{i=0}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i)\Theta'_{\sigma_k}(y_k^i; d_k^i).$$

Proof. By the descent condition (2.15), the total decrease T_k^i satisfies the inequality

$$T_k^i = \Theta_{\sigma_k}(y_k) - \Theta_{\sigma_k}(y_k^i) \geq -\omega_1\, \nu_k^{i-1}(\alpha_k^i), \quad \text{for } i = 1, \ldots, i_k.$$

Therefore, by definition of ν_k^i and because $\rho_k^i \geq 0$,

$$\nu_k^i(\alpha_k^{i+1}) \leq \nu_k^{i-1}(\alpha_k^i) + (\alpha_k^{i+1} - \alpha_k^i)\Theta'_{\sigma_k}(y_k^i; d_k^i), \quad \text{for } i = 1, \ldots, i_k - 1.$$

Adding up these inequalities gives

$$\nu_k^{i_k-1}(\alpha_k^{i_k}) \leq \nu_k^0(\alpha_k^1) + \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i)\Theta'_{\sigma_k}(y_k^i; d_k^i).$$

Now, $\nu_k^0(\alpha_k^1) = \alpha_k^1\Theta'_{\sigma_k}(y_k; d_k^0)$, $y_k = y_k^0$ and $\alpha_k^0 = 0$. Hence, the preceding inequality and the descent condition (2.11) give the result. □

2.3 Concise statement of the algorithm

At the beginning of a run, k is set to 1 and it is supposed given: a point $y_1 \in \Omega$; constants $\sigma_0 > 0$ and $\bar{\sigma} > 0$ for updating the penalty parameter (for example: $\sigma_0 = 2\|\lambda(y_1)\|_D$ if $\lambda(y_1) \neq 0$ and $\sigma_0 = 1$ otherwise, and $\bar{\sigma} = \sigma_0/100$); constants $\omega_1 \in (0, 1/2)$, $\omega_2 \in (0,1)$, and $\beta \in (0,1)$ for the search algorithm (possible values are $\omega_1 = 0.001$, $\omega_2 = 0.99$, and $\beta = 0.01$); a constant $\mu > 0$ for the update criterion (2.10); and an initial symmetric positive definite matrix B_1 (for example $B_1 = I$ with an Oren and Luenberger [23] scaling before the first update).

When an iteration begins it is supposed that a point $y_k \in \Omega$ is given, as well as a symmetric positive definite matrix B_k and a penalty parameter σ_{k-1}. It is also supposed that the constraints have been linearized at y_k and that the following quantities have been computed: $f(y_k)$, $c(y_k)$, $g(y_k)$ and $\lambda(y_k)$. Then, one iteration looks as follows.

ONE ITERATION OF THE ALGORITHM:

1. Check the convergence.
2. Adapt the penalty parameter $\sigma_{k-1} \to \sigma_k$ such that (2.3)–(2.5) hold.
3. Compute d_k^0 by (2.6), x_{k+1} by (1.3) and $c(x_{k+1})$.
4. First stage of the search: determine a step-size $\alpha_k^1 \in (0,1]$ along the path p_k^0 given by (2.7), such that $y_k^1 = p_k^0(\alpha_k^1) \in \Omega$ and the descent condition (2.9) holds. Assumptions 2.1 must be followed.
5. Update criterion: if the update criterion (2.10) does not hold, go to Step 8, skipping the update of B_k.
6. Second stage of the search (PLS): set $i = 1$.
   - 6.1. Linearize the constraints at y_k^i.
   - 6.2. If the curvature condition (2.13) holds, set $i_k = i$, $\alpha_k = \alpha_k^i$, $y_{k+1} = y_k^i$, $\nu_k = \nu_k^{i-1}$, and go to Step 7 to update B_k.
   - 6.3. If the penalty parameter σ_k is not sufficiently large to have (2.16), set $i_k = i$, $\alpha_k = \alpha_k^i$, $y_{k+1} = y_k^i$, $\nu_k = \nu_k^{i-1}$, and go to Step 8, skipping the update of B_k.
   - 6.4. Choose a tangent scaling factor $\tau_k^i > 0$ such that Assumption 2.2 holds and determine a step-size $\alpha_k^{i+1} \in (\alpha_k^i, \alpha_k^i + 1]$ from y_k^i along d_k^i given by (2.14), such that $y_k^{i+1} = y_k^i + (\alpha_k^{i+1} - \alpha_k^i) d_k^i \in \Omega$ and the descent condition (2.15) holds.
   - 6.5. Increase i by 1 and go to Step 6.1.
7. Update $B_k \to B_{k+1}$ by the BFGS formula (2.17), using γ_k and δ_k defined by (2.18) and (2.19).
8. Increase k by 1.

Note that no assumption is made on the way the step-sizes α_k^i ($i \geq 2$) are determined. Assumptions 2.1 only deal with α_k^1.

3 Superlinear convergence of the algorithm

3.1 Hypotheses and result

The superlinear convergence result needs the following assumptions.

Assumptions 3.1. *We suppose that there is an open convex neighborhood* Ω_0 *of a point* $x_* \in \mathbb{R}^n$, *with* $\Omega_0 \subset \Omega$, *and a multiplier* $\lambda_* \in \mathbb{R}^m$, *such that:*

(*i*) f *and* c *are twice differentiable on* Ω_0;

(*ii*) c *is a submersion on* Ω_0 *and the map*

$$\Omega_0 \ni x \mapsto (A^-(x), Z^-(x))$$

is Lipschitz continuous on Ω_0;

(*iii*) g *is differentiable on* Ω_0 *with Lipschitz continuous derivative;*

(*iv*) $c(x_*) = 0$, $g(x_*) = 0$, *and* $B_* = Z_*^{-\top} L_* Z_*^-$ *is positive definite.*

Our main result shows that if the algorithm described in Section 2 generates points $\{y_k^i\}_{k\geq 1, 0\leq i\leq i_k-1}$ converging to x_* in the sense that

$$\left(\max_{0\leq i\leq i_k-1} \|y_k^i - x_*\|\right) \to 0, \quad \text{when } k \to \infty, \tag{3.1}$$

then the sequence $\{x_k\}_{k\geq 2}$ defined by (1.3) converges q-superlinearly to x_*. As this result is asymptotic and no hypothesis is made on y_1 and B_1 is only supposed symmetric positive definite, we can consider that all the sequence is in Ω_0, which is assumed in the proofs.

Hypothesis (3.1) may look strong with respect to the more simple assumption that would only suppose the convergence of $\{y_k\}_{k\geq 1}$ to x_*. We have not found, however, a way of avoiding it because of the imaginated situation in which the manifold of the constraints would be like a sphere, allowing the iterates $y_k, y_k^1, \ldots, y_k^{i_k-1}$ to cycle and to visit periodically various stationary points, before coming back close to x_* with $y_{k+1} = y_k^{i_k}$. Now, from the definition of the points y_k^i, one sees that they look like "true" iterates (the search directions d_k^i look very much like reduced SQP directions), so that (3.1), which is equivalent to supposing that the sequence $y_1, \ldots,$ $y_1^{i_1-1}$, y_2, $\ldots$, $y_2^{i_2-1}$, y_3, $\ldots$, $y_3^{i_3-1}$, $\ldots$ converges to x_*, looks now more familiar. Also, we shall show that, under hypothesis (3.1), $i_k = 1$ for k large, so that there is no intermediate point asymptotically.

Alternatively hypothesis (3.1) could be replaced by the assumption that $y_k \to x_*$ and that the sequence $\{y_k^i\}_{k\geq 1, 0\leq i\leq i_k-1}$ remains sufficiently close to x_* when k is large. It is not difficult to see that these assumptions imply (3.1), since then, by Lemmas 3.3 and 3.9 below, we have for large k:

$$C\|y_k^i - x_*\|^2 \leq \Theta_\sigma(y_k^i) - \Theta_\sigma(x_*) \leq \Theta_\sigma(y_k) - \Theta_\sigma(x_*) \to 0.$$

The main result of the paper is the following.

Theorem 3.2. *Suppose that Assumptions* 3.1 *hold and that the algorithm described in Section* 2 *generates a sequence of points* $\{y_k^i\}_{k\geq 1, 0\leq i\leq i_k-1}$ *in* Ω *different from* x_* *and converging to* x_* *in the sense* (3.1). *Then, the following properties hold:*

(i) *the sequences of matrices* $\{B_k\}_{k\geq 1}$ *and* $\{B_k^{-1}\}_{k\geq 1}$ *are bounded;*

(ii) *the ideal step-size is accepted eventually:* $i_k = 1$ *and* $\alpha_k = 1$ *for* k *large;*

(iii) *the sequence* $\{y_k\}_{k\geq 1}$ *converges q-superlinearly in two steps to* x_**;*

(iv) *the sequence* $\{x_k\}_{k\geq 2}$ *converges q-superlinearly to* x_*.

This theorem assumes that $y_k^i \neq x_*$, i.e., that the algorithm does not terminate prematurely because a solution point is found (then there is nothing to prove). It also implicitly assumes that the piecewise line-search (PLS) does not fail when it is invoked. According to [18], the failure of the PLS should be rather rare. The 2 step q-superlinear convergence of the sequence $\{y_k\}$ means that $y_{k+2} - x_* = o(\|y_k - x_*\|)$.

The proof of the theorem is divided in lemmas grouped in three sets. We first prove the r-linear convergence of $\{y_k\}_{k\geq 1}$ (Section 3.2). Then we analyze the conditions of admissibility of the ideal step-size and prove their achievement (Sections 3.3 and 3.4). Finally, we show the q-superlinear convergence of the sequence $\{x_k\}_{k\geq 2}$ (end of Section 3.4).

The techniques used in the proofs have two sources of inspiration. On the one hand, the treatment of the contribution of the BFGS formula to get superlinear convergence is mainly inspired by the work of Powell [25] and Byrd and Nocedal [4, 5]. On the other hand, the study of the control realized by the update criterion is made as in [14, 15].

A comment on the notation. The indices $k_1 \leq k_2 \leq \ldots$ are determined in the intermediate results (the lemmas), while k_1', k_2', ... are temporary indices used in the proofs. Similarly, the positive constants (i.e., independent of k) C_1, C_2, ... are fixed by the intermediate results, while C_1', C_2', ... are temporarily defined in each proof, and C is an "absorbing" positive constant.

3.2 The r-linear convergence

A useful intermediate result leading to superlinear convergence is the r-linear convergence. In this section, we gather the lemmas leading to the r-linear convergence of the sequence $\{y_k\}_{k\geq 1}$.

This is done in two stages. First, we analyze the subsequence of iterates with update (i.e., with index in $\mathcal{K}$) and we show that a fixed proportion of them are "good" in the sense of inequalities (3.7) below. This implies the r-linear convergence to 0 of the sequence (3.11) (Lemma 3.12). Then, by the BFGS theory (Lemma 3.13), this result yields the boundedness of the sequences $\{B_k\}$ and $\{B_k^{-1}\}$ (Lemma 3.14). We deduce from this that all the iterates are "good" in the sense of inequalities (3.7). In the second stage of the proof (Lemma 3.15), we can then use the same technique (i.e., Lemma 3.10) as in the first stage, but now for all the sequence $\{y_k\}$. This yields its r-linear convergence.

The first lemma shows that the penalty parameter becomes fixed after a finite number of iterations. From that moment, the algorithm minimizes the same penalty function at each iteration. Another consequence of this result is that for large k, the second stage of the search algorithm is never interrupted because of the failure of (2.16).

Lemma 3.3. *Suppose that the hypotheses of Theorem 3.2 hold. Then, there is an index $k_2 \geq k_1$ (index given in (2.4)) such that for $k \geq k_2$, σ_k has a constant value σ. Furthermore, $\sigma \geq \|\lambda(y_k^i)\|_D + \bar{\sigma}$, for all $k \geq k_2$ and $0 \leq i \leq i_k - 1$.*

Proof. As $y_k \to x_*$, $\lambda(y_k) \to \lambda_*$. Then, for some index $k_1' \geq k_1$,

$$\|\lambda(y_k)\|_D \leq \|\lambda_*\|_D + \bar{\sigma}, \quad \forall k \geq k_1'.$$

This implies that $\{\sigma_k\}_{k\geq 1}$ is bounded. Indeed, if there is an index $k_2' \geq k_1'$ such that

$$\sigma_{k_2'} \geq \|\lambda_*\|_D + 2\bar{\sigma},$$

then

$$\sigma_{k_2'} \geq \|\lambda(y_k)\|_D + \bar{\sigma}, \quad \forall k \geq k_1'.$$

From property (2.4) and $k_2' \geq k_1$, one has $\sigma_k = \sigma_{k_2'}$, for all $k \geq k_2'$.

Then, by property (2.5), σ_k is modified finitely often, i.e., $\sigma_k = \sigma$ for $k \geq k_2$ say. By property (2.3) and the fact that $i = i_k$ when (2.16) fails, the conclusion of the lemma holds. □

Lemma 3.4 obtains from the descent condition (2.11) a technical result that is useful to show that $\delta_k \to 0$ (Lemma 3.5). The fact that $\delta_k \to 0$ is important in many places in the proofs. We recall that δ_k is defined by (2.18).

Lemma 3.4. *Suppose that the hypotheses of Theorem 3.2 hold. Then,*

$$\sum_{i=0}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i)\|c(y_k^i)\| \to 0, \quad \textit{when } k \to \infty.$$

Proof. Let $k \geq k_2$ (index given by Lemma 3.3), so that $\sigma_k = \sigma$. For $i = 1, \ldots, i_k - 1$, the curvature condition (2.13) does not hold, hence $g(y_k^i)^\top Z(y_k)t_k \leq 0$. Using also $g(y_k)^\top Z(y_k)t_k \leq 0$ and $\sigma \geq \|\lambda(y_k^i)\|_D + \bar{\sigma}$, we get for $i = 0, \ldots, i_k - 1$

$$\begin{aligned}
\Theta_\sigma'(y_k^i; d_k^i) &= \tau_k^i g(y_k^i)^\top Z(y_k)t_k + \lambda(y_k^i)^\top c(y_k^i) - \sigma\|c(y_k^i)\| \\
&\leq (\|\lambda(y_k^i)\|_D - \sigma)\|c(y_k^i)\| \\
&\leq -\bar{\sigma}\|c(y_k^i)\|.
\end{aligned}$$

Now, from Proposition 2.3

$$\omega_1 \bar{\sigma} \sum_{i=0}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \|c(y_k^i)\| \le \Theta_\sigma(y_k) - \Theta_\sigma(y_{k+1}).$$

The conclusion follows from the fact that $\Theta_\sigma(y_k) - \Theta_\sigma(y_{k+1}) \to 0$. □

Lemma 3.5. *Suppose that the hypotheses of Theorem* 3.2 *hold. Then,* $\delta_k \to 0$ *when* $k \to \infty$.

Proof. First, by (2.7)–(2.8), $Z(y_k)(y_k^1 - y_k) = \alpha_k^1 Z(y_k) t_k$, and by (3.1), $(y_k^1 - y_k)$ tends to 0. Hence

$$\alpha_k^1 Z(y_k) t_k \to 0. \tag{3.2}$$

On the other hand, denoting $\tilde{\delta}_k = \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \tau_k^i Z(y_k) t_k$, we have

$$\begin{aligned} Z^-(y_k)\tilde{\delta}_k &= \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \tau_k^i Z^-(y_k^i) Z(y_k) t_k \\ &\quad - \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \tau_k^i \left(Z^-(y_k^i) - Z^-(y_k) \right) Z(y_k) t_k \\ &= (y_{k+1} - y_k^1) + \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) A^-(y_k^i) c(y_k^i) \\ &\quad - \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \tau_k^i \left(Z^-(y_k^i) - Z^-(y_k) \right) Z(y_k) t_k. \end{aligned}$$

By (3.1), $(y_{k+1} - y_k^1) \to 0$. By Lemma 3.4 and the boundedness of $\{A^-(y_k^i)\}$, the first sum in the last right hand side tends to 0. Using the Lipschitz continuity of Z^-, we get

$$\|Z^-(y_k)\tilde{\delta}_k\| \le C \sum_{i=1}^{i_k-1} (\alpha_k^{i+1} - \alpha_k^i) \tau_k^i \|y_k^i - y_k\| \, \|Z(y_k) t_k\| + o(1).$$

But $\max_i \|y_k^i - y_k\| \to 0$ [by (3.1)] and $\|Z^-(y_k)\tilde{\delta}_k\| \ge C\|\tilde{\delta}_k\|$. Hence

$$\tilde{\delta}_k = o(\|\tilde{\delta}_k\|) + o(1).$$

From this we deduce that $\tilde{\delta}_k \to 0$, which with (3.2) gives the result. □

The next lemma gives a technical result that is only useful for proving Lemma 3.7. It says that, when the update criterion (2.10) holds, an inequality similar to (2.10) is verified: instead of $\|c(y_k)\|$ or $\|c(x_{k+1})\|$ in the left hand side, one can also put $\|c(y_k^i)\|$ for any $i = 0, \ldots, i_k$ and the right hand side is perturbated only by a term of order $O(\|\delta_k\| \, \|t_k\|)$.

Lemma 3.6. *Suppose that the hypotheses of Theorem 3.2 hold. Then, there is a constant $C_1 > 0$ such that*

$$\left(\max_{0\le i\le i_k} \|c(y_k^i)\|\right) \le C_1\left(\|e^1_{k\ominus 2}\| + \|\delta_k\|\right)\|t_k\|, \quad \textit{for } k \in \mathcal{K}. \tag{3.3}$$

Proof. The result clearly holds for $\|c(y_k)\|$ in the left hand side of (3.3): use the update criterion (2.10) and the fact that $Z(y_k)t_k = O(\|t_k\|)$.

Now, the definition (2.8) of y_k^1, the update criterion (2.10), and $\alpha_k^1 \le 1$ give for $k \in \mathcal{K}$

$$y_k^1 = y_k + \alpha_k^1 t_k - \alpha_k^1 A^-(y_k)c(y_k) - (\alpha_k^1)^2 A^-(y_k)\Big(c(x_{k+1}) - c(y_k)\Big),$$

$$\|y_k^1 - y_k\| \le \alpha_k^1\|t_k\| + C\alpha_k^1\|e^1_{k\ominus 2}\|\|t_k\|. \tag{3.4}$$

Let us expand $c(y_k^1)$ about y_k:

$$\begin{aligned} c(y_k^1) &= c(y_k) - \alpha_k^1 c(y_k) - (\alpha_k^1)^2\Big(c(x_{k+1}) - c(y_k)\Big) + O(\|y_k^1 - y_k\|^2) \\ &= \Big(1 - \alpha_k^1 + (\alpha_k^1)^2\Big)c(y_k) - (\alpha_k^1)^2 c(x_{k+1}) + O(\|y_k^1 - y_k\|^2). \end{aligned}$$

Now, $1 - \alpha_k^1 + (\alpha_k^1)^2 \le 1$, $(\alpha_k^1)^2 \le 1$, and $\|y_k^1 - y_k\| \le C\alpha_k^1\|t_k\|$, by (3.4). Therefore

$$\|c(y_k^1)\| \le \|c(y_k)\| + \|c(x_{k+1})\| + C(\alpha_k^1)^2\|t_k\|^2.$$

Using the update criterion (2.10) and the boundedness of $\{Z^-(y_k)\}$ to get $\alpha_k^1 t_k = O(\|\delta_k\|)$, we obtain

$$\|c(y_k^1)\| \le C(\|e^1_{k\ominus 2}\| + \|\delta_k\|)\|t_k\|. \tag{3.5}$$

This is the inequality of the result with $\|c(y_k^1)\|$ in the left hand side.

We proceed similarly for $\|c(y_k^{i+1})\|$, when $1 \le i \le i_k - 1$. Then, $y_k^{i+1} = y_k^i + (\alpha_k^{i+1} - \alpha_k^i)d_k^i$ and the expansion of $c(y_k^{i+1})$ about y_k^i gives

$$c(y_k^{i+1}) = \Big(1 - (\alpha_k^{i+1} - \alpha_k^i)\Big)c(y_k^i) + O(\|y_k^{i+1} - y_k^i\|^2).$$

By choice $1 - (\alpha_k^{i+1} - \alpha_k^i) \ge 0$, then

$$\begin{aligned} \|c(y_k^{i+1})\| &\le \Big(1 - (\alpha_k^{i+1} - \alpha_k^i)\Big)\|c(y_k^i)\| + C_1'(\alpha_k^{i+1} - \alpha_k^i)^2(\tau_k^i)^2\|t_k\|^2 \\ &\quad + C_2'(\alpha_k^{i+1} - \alpha_k^i)^2\|c(y_k^i)\|^2 \\ &= \|c(y_k^i)\| + (\alpha_k^{i+1} - \alpha_k^i)\Big(C_2'(\alpha_k^{i+1} - \alpha_k^i)\|c(y_k^i)\| - 1\Big)\|c(y_k^i)\| \\ &\quad + C_1'(\alpha_k^{i+1} - \alpha_k^i)^2(\tau_k^i)^2\|t_k\|^2 \\ &\le \|c(y_k^i)\| + C(\alpha_k^{i+1} - \alpha_k^i)\tau_k^i\|t_k\|^2, \quad \text{for } k \text{ large,} \end{aligned}$$

because, by Lemma 3.4, $C_2'(\alpha_k^{i+1} - \alpha_k^i)\|c(y_k^i)\| \le 1$ for k large, $(\alpha_k^{i+1} - \alpha_k^i) \le 1$, and $\{\tau_k^i\}$ is bounded. Finally, using these inequalities recursively, one gets for $i = 2, \ldots, i_k$ and k large in $\mathcal{K}$:

$$\begin{aligned}
\|c(y_k^i)\| &\le \|c(y_k^1)\| + C\sum_{j=1}^{i-1}(\alpha_k^{j+1} - \alpha_k^j)\tau_k^j\|t_k\|^2 \\
&\le \|c(y_k^1)\| + C\bar{\alpha}_k\|t_k\|^2 \\
&\le C(\|e_{k\ominus 2}^1\| + \|\delta_k\|)\|t_k\|, \quad \text{by (3.5).}
\end{aligned}$$

This concludes the proof. □

The result of the next lemma allows us to measure in terms of δ_k and for indices in $\mathcal{K}$, the gap between the actual step from y_k to y_{k+1} and its tangential component $\bar{\alpha}_k t_k$. This is useful for the analysis of the subsequence of iterates with update. We recall that $\bar{\alpha}_k$ is defined by (2.19).

Lemma 3.7. *Suppose that the hypotheses of Theorem 3.2 hold. Then, there is a constant $C_2 > 0$ such that*

$$\|y_{k+1} - y_k - \bar{\alpha}_k t_k\| \le C_2\left(\|e_{k\ominus 2}^1\| + \max_{1\le i\le i_k-1}\|y_k^i - y_k\| + \|\delta_k\|\right)\|\delta_k\|, \quad \textit{for } k \in \mathcal{K}.$$

Proof. We have

$$\begin{aligned}
y_{k+1} - y_k - \bar{\alpha}_k t_k &= \alpha_k^1 t_k - \alpha_k^1 A^-(y_k)\Big((1-\alpha_k^1)c(y_k) + \alpha_k^1 c(x_{k+1})\Big) \\
&\quad + \sum_{i=1}^{i_k-1}(\alpha_k^{i+1} - \alpha_k^i)d_k^i - \bar{\alpha}_k t_k \\
&= X_1 + X_2 + X_3,
\end{aligned}$$

where

$$\begin{aligned}
X_1 &= -\alpha_k^1 A^-(y_k)\Big((1-\alpha_k^1)c(y_k) + \alpha_k^1 c(x_{k+1})\Big), \\
X_2 &= \sum_{i=1}^{i_k-1}(\alpha_k^{i+1} - \alpha_k^i)\tau_k^i\Big(Z^-(y_k^i) - Z^-(y_k)\Big)Z(y_k)t_k, \\
X_3 &= -\sum_{i=1}^{i_k-1}(\alpha_k^{i+1} - \alpha_k^i)A^-(y_k^i)c(y_k^i).
\end{aligned}$$

Using the boundedness of $\{A^-(y_k)\}$ and $\{\alpha_k^1\}$, the update criterion (2.10), $\alpha_k^1 \le \bar{\alpha}_k$, and the definition (2.18) of δ_k:

$$\|X_1\| \le C\alpha_k^1\|e_{k\ominus 2}^1\|\,\|Z(y_k)t_k\| \le C\|e_{k\ominus 2}^1\|\,\|\delta_k\|, \quad \text{for } k \in \mathcal{K}.$$

With the Lipschitz continuity of Z^-

$$\begin{aligned}
\|X_2\| &\le C\sum_{i=1}^{i_k-1}(\alpha_k^{i+1}-\alpha_k^i)\tau_k^i\|y_k^i-y_k\|\,\|Z(y_k)t_k\| \\
&\le C\left(\max_{1\le i\le i_k-1}\|y_k^i-y_k\|\right)\|\delta_k\|.
\end{aligned}$$

Finally, by the boundedness of $\{A^-(y_k^i)\}$, $\{Z^-(y_k)\}$, and $\{1/\tau_k^i\}$, and by Lemma 3.6

$$\begin{aligned}
\|X_3\| &\le C\sum_{i=1}^{i_k-1}(\alpha_k^{i+1}-\alpha_k^i)\|c(y_k^i)\| \\
&\le C\bar{\alpha}_k\left(\max_{1\le i\le i_k-1}\|c(y_k^i)\|\right) \\
&\le C(\|e_{k\ominus 2}^1\|+\|\delta_k\|)\|\delta_k\|, \quad \text{for } k\in\mathcal{K}.
\end{aligned}$$

The last three estimates and the first identity give the result. □

The next lemma is important for the control of the update of B_k with γ_k and δ_k. Inequalities (3.6) will be used to show that a fixed proportion of the steps with update are "good" in the sense of inequalities (3.7).

Lemma 3.8. *Suppose that the hypotheses of Theorem 3.2 hold. Then, there are constants $C_3>0$ and $C_4>0$ such that*

$$\gamma_k^{\mathsf{T}}\delta_k \ge C_3\|\delta_k\|^2 \quad \textit{and} \quad \gamma_k^{\mathsf{T}}\delta_k \ge C_4\|\gamma_k\|^2, \quad \textit{for } k\in\mathcal{K}. \tag{3.6}$$

Proof. If $\mathcal{K}$ is bounded, the result is clear since $\gamma_k^{\mathsf{T}}\delta_k>0$ for $k\in\mathcal{K}$. Suppose now that $\mathcal{K}$ is unbounded. Then, $e_{k\ominus 2}^1\to 0$ and Lemma 3.7, with Lemma 3.5 and (3.1), show that

$$y_{k+1}-y_k-\bar{\alpha}_k t_k = o(\|\delta_k\|), \quad \text{for } k\in\mathcal{K}.$$

This implies that $y_{k+1}-y_k=O(\|\delta_k\|)$ for $k\in\mathcal{K}$.

Using the fact that g is continuously differentiable around x_*, (2.2), and the estimates above

$$\begin{aligned}
\gamma_k &= g(y_{k+1})-g(y_k) \\
&= Z_*^{-\mathsf{T}}L_*(y_{k+1}-y_k)+o(\|y_{k+1}-y_k\|) \\
&= \bar{\alpha}_k Z_*^{-\mathsf{T}}L_* t_k + o(\|\delta_k\|), \quad \text{for } k\in\mathcal{K}.
\end{aligned}$$

Now, since $\bar{\alpha}_k t_k = Z^-(y_k)\delta_k = Z_*^-\delta_k + o(\|\delta_k\|)$

$$\gamma_k = B_*\delta_k + o(\|\delta_k\|), \quad \text{for } k\in\mathcal{K}.$$

Taking the inner product with δ_k in the estimate above and using the positive definiteness of B_* lead to the first inequality of the result. The second inequality of the result comes from the first one and the fact that $\gamma_k = O(\|\delta_k\|)$ for $k \in \mathcal{K}$. □

The next lemma, due to Byrd and Nocedal [5, Lemma 4.2], gives information on the map Θ_σ near x_*.

Lemma 3.9. *Suppose that Assumptions* 3.1 *hold. Then for any* $\sigma > \|\lambda_*\|_D$, *there exist two constants* $C_5 > 0$ *and* $C_6 > 0$ *and an open neighborhood* Ω_1 *of* x_* *such that for* $y \in \Omega_1$

$$C_5\|y - x_*\|^2 \le \Theta_\sigma(y) - \Theta_\sigma(x_*) \le C_6(\|g(y)\|^2 + \|c(y)\|).$$

We can suppose that $\Omega_1 \subset \Omega_0$. Then, since the points y_k^i tends to x_*, there is an index, say $k_3 \ge k_2$ (index given by Lemma 3.3), such that

$$y_k^i \in \Omega_1,$$

for $k \ge k_3$ and $0 \le i \le i_k - 1$.

We condense in the next lemma a reasoning that will be used for two subsequences $\mathcal{K}'$ of indices: for a subsequence of $\mathcal{K}$ and for $\mathbb{N}^*$. We denote by θ_k the angle between δ_k and $B_k\delta_k$, that is

$$\cos\theta_k = \frac{\delta_k^\top B_k \delta_k}{\|B_k\delta_k\|\,\|\delta_k\|}.$$

Lemma 3.10. *Suppose that the hypotheses of Theorem* 3.2 *hold. Let* $\mathcal{K}'$ *be a subsequence of indices and* $C' > 0$ *and* $C'' > 0$ *be two constants such that for all* k *in* $\mathcal{K}'$

$$\cos\theta_k \ge C' \quad and \quad \frac{1}{C''} \le \frac{\|B_k\delta_k\|}{\|\delta_k\|} \le C''. \tag{3.7}$$

Then there exist two constants $\breve{\alpha} > 0$ *and* $\zeta \in (0,1)$ *such that for* $k \in \mathcal{K}'$, $k \ge k_3$ *(index given after Lemma* 3.9*) and* $1 \le i \le i_k$

$$\alpha_k \ge \alpha_k^1 \ge \breve{\alpha},$$

$$\Theta_\sigma(y_k^i) - \Theta_\sigma(x_*) \le \zeta\Big(\Theta_\sigma(y_k) - \Theta_\sigma(x_*)\Big).$$

Proof. As $t_k = (1/\bar{\alpha}_k)Z^-(y_k)\delta_k$, we have using (3.7)

$$\|t_k\| \le \frac{C}{\bar{\alpha}_k}\|\delta_k\| \le \frac{C}{\bar{\alpha}_k}\|B_k\delta_k\| = C\|g(y_k)\|, \quad \text{for } k \in \mathcal{K}'. \tag{3.8}$$

We consider now large indices $k \geq k_2$, so that $\sigma_k = \sigma \geq \|\lambda(y_k)\|_D + \bar{\sigma}$. Then, using $g(y_k) = -(1/\bar{\alpha}_k)B_k\delta_k$ and (3.7), we have for large $k \in \mathcal{K}'$

$$\begin{aligned}
\Theta'_\sigma(y_k; d_k^0) &= g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k) - \sigma\|c(y_k)\| \\
&\leq -\frac{1}{\bar{\alpha}_k^2}\delta_k^\top B_k\delta_k - \bar{\sigma}\|c(y_k)\| \\
&= -\frac{1}{\bar{\alpha}_k^2}\|B_k\delta_k\|\,\|\delta_k\|\cos\theta_k - \bar{\sigma}\|c(y_k)\| \\
&\leq -\frac{C'}{C''\bar{\alpha}_k^2}\|B_k\delta_k\|^2 - \bar{\sigma}\|c(y_k)\| \\
&\leq -C(\|g(y_k)\|^2 + \|c(y_k)\|). \qquad (3.9)
\end{aligned}$$

To prove that α_k^1 is bounded away from 0, suppose that $\alpha_k^1 < 1$. Then, by Assumptions 2.1, there is a step-size $\tilde{\alpha}_k^1 \leq \min(1, \alpha_k^1/\beta)$ such that either

$$\tilde{y}_k^1 = p_k^0(\tilde{\alpha}_k^1) = y_k + \tilde{\alpha}_k^1 d_k^0 + (\tilde{\alpha}_k^1)^2 A^-(y_k)\Big(c(y_k) - c(x_{k+1})\Big)$$

is not in Ω or the descent condition (2.9) is not satisfied for y_k^1 and α_k^1 replaced by $\tilde{y}_k^1$ and $\tilde{\alpha}_k^1$ respectively. Clearly, it is the latter event that occurs eventually since for $k \to \infty$ in $\mathcal{K}'$, (3.8) shows that $t_k \to 0$ and, because $\tilde{\alpha}_k^1 \leq 1$, $\tilde{y}_k^1 \to x_* \in \Omega$. Therefore, for k large in $\mathcal{K}'$:

$$\Theta_\sigma(\tilde{y}_k^1) \geq \Theta_\sigma(y_k) + \omega_1\tilde{\alpha}_k^1\Theta'_\sigma(y_k; d_k^0). \qquad (3.10)$$

Let us expand the left hand side of this inequality about y_k. We have

$$c(x_{k+1}) = c(y_k) + O(\|t_k\|^2).$$

Using this estimate and the fact that $\{t_k\}_{k\in\mathcal{K}'}$ is bounded [by (3.8)]

$$f(\tilde{y}_k^1) = f(y_k) + \tilde{\alpha}_k^1 f'(y_k)\cdot d_k^0 + (\tilde{\alpha}_k^1)^2 O(\|t_k\|^2) + (\tilde{\alpha}_k^1)^2 O(\|c(y_k)\|^2).$$

Similarly

$$c(\tilde{y}_k^1) = (1 - \tilde{\alpha}_k^1)c(y_k) + (\tilde{\alpha}_k^1)^2 O(\|t_k\|^2) + (\tilde{\alpha}_k^1)^2 O(\|c(y_k)\|^2).$$

Since $1 - \tilde{\alpha}_k^1 \geq 0$, these two estimates give

$$\begin{aligned}
\Theta_\sigma(\tilde{y}_k^1) &= \Theta_\sigma(y_k) + \tilde{\alpha}_k^1\Big(f'(y_k)\cdot d_k^0 - \sigma\|c(y_k)\|\Big) \\
&\quad + (\tilde{\alpha}_k^1)^2 O(\|t_k\|^2) + (\tilde{\alpha}_k^1)^2 O(\|c(y_k)\|^2) \\
&= \Theta_\sigma(y_k) + \tilde{\alpha}_k^1\Theta'_\sigma(y_k; d_k^0) + (\tilde{\alpha}_k^1)^2 O(\|t_k\|^2) + (\tilde{\alpha}_k^1)^2 O(\|c(y_k)\|^2).
\end{aligned}$$

Let us go back to (3.10) with this identity. Because $\Theta'_\sigma(y_k; d_k^0)$ is negative, we have

$$-(1-\omega_1)\tilde{\alpha}_k^1\Theta'_\sigma(y_k; d_k^0) = (\tilde{\alpha}_k^1)^2 O(\|t_k\|^2) + (\tilde{\alpha}_k^1)^2 O(\|c(y_k)\|^2).$$

Dividing by $\tilde{\alpha}_k^1 > 0$ and using (3.8) and (3.9)

$$\|g(y_k)\|^2 + \|c(y_k)\| \le C_1' \tilde{\alpha}_k^1 \Big(\|g(y_k)\|^2 + \|c(y_k)\|^2 \Big),$$

where $C_1' > 0$ is a constant. Hence

$$\|g(y_k)\|^2 + \Big(1 - C_1' \tilde{\alpha}_k^1 \|c(y_k)\|\Big) \|c(y_k)\| \le C_1' \tilde{\alpha}_k^1 \|g(y_k)\|^2.$$

But $1 - C_1' \tilde{\alpha}_k^1 \|c(y_k)\| \ge 0$ for large k, therefore

$$\|g(y_k)\|^2 \le C_1' \tilde{\alpha}_k^1 \|g(y_k)\|^2, \quad \text{for large } k \in \mathcal{K}'.$$

For these large indices, $g(y_k) \neq 0$ otherwise the last but one inequality would give $c(y_k) = 0$ and y_k would be a stationary point, which we have discarded by hypothesis. Therefore $\tilde{\alpha}_k^1 \ge 1/C_1'$, showing that $\alpha_k^1 \ge \min(1, \beta/C_1')$ for large k in $\mathcal{K}'$. Hence $\{\alpha_k^1\}$ is bounded away from zero in $\mathcal{K}'$ and the first part of the lemma is proved.

For the second part, observe that for $i = 1, \ldots, i_k$

$$\Theta_\sigma(y_k^i) \le \Theta_\sigma(y_k) + \omega_1 \alpha_k^1 \Theta_\sigma'(y_k; d_k^0).$$

Using $\alpha_k^1 \ge \bar{\alpha}$, (3.9), and Lemma 3.9, we get for $i = 1, \ldots, i_k$ and $k \ge k_3$ in $\mathcal{K}'$:

$$\begin{aligned} \Theta_\sigma(y_k^i) - \Theta_\sigma(x_*) &\le \Theta_\sigma(y_k) - \Theta_\sigma(x_*) - C(\|g(y_k)\|^2 + \|c(y_k)\|) \\ &\le (1 - C_2')\Big(\Theta_\sigma(y_k) - \Theta_\sigma(x_*)\Big). \end{aligned}$$

Since $\Theta_\sigma(y_k^i)$ and $\Theta_\sigma(y_k)$ are greater than $\Theta_\sigma(x_*)$, and $C_2' \neq 0$, we see that $1 - C_2'$ is a number in $(0, 1)$. This concludes the proof. □

The next lemma shows that a given proportion of the steps in $\mathcal{K}$ are good in the sense that $\cos\theta_k$ does not approach 0 and the quantities $\|B_k\delta_k\|/\|\delta_k\|$ and their inverses form bounded sequences. This is exactly what is needed to apply Lemma 3.10, which is done in Lemma 3.12.

Lemma 3.11. *Suppose that the hypotheses of Theorem* 3.2 *hold and let* $r \in (0,1)$. *Then, there exist two constants* $C_7 > 0$ *and* $C_8 > 0$, *and a subsequence* $\mathcal{K}^r \subset \mathcal{K}$ *such that*

(*i*) $\cos\theta_k \ge C_7$, *for* $k \in \mathcal{K}^r$,
(*ii*) $1/C_8 \le \|B_k\delta_k\|/\|\delta_k\| \le C_8$, *for* $k \in \mathcal{K}^r$,
(*iii*) $|\mathcal{K}_k^r| \ge r|\mathcal{K}_k|$, *for* $k \ge 1$, *where* $\mathcal{K}_k = \mathcal{K} \cap \{1, \ldots, k\}$ *and* $\mathcal{K}_k^r = \mathcal{K}^r \cap \{1, \ldots, k\}$.

Proof. Consider the sequence of matrices $\{B_k\}_{k\in\mathcal{K}}$. These matrices are updated with the BFGS formula using the pairs $\{(\gamma_k, \delta_k)\}_{k\in\mathcal{K}}$. As these pairs satisfy (3.6), Theorem 2.1 of Byrd and Nocedal [4] applies and gives the result. □

The next lemma proves the r-linear convergence of the sequence $\{\max_i \|y_k^i - x_*\| : k \in \mathcal{K}\}$ to zero. Recall that a sequence $\{u_k\}_{k\geq 1}$ converges r-linearly to a point u_* in a normed space, if

$$\limsup_{k\to\infty} \|u_k - u_*\|^{1/k} < 1.$$

Clearly, this implies that $\sum_{k\geq 1} \|u_k - u_*\| < \infty$.

Lemma 3.12. *Suppose that the hypotheses of Theorem* 3.2 *hold and that* $\mathcal{K}$ *is unbounded. Then the sequence*

$$\left\{\max_{0\leq i\leq i_k} \|y_k^i - x_*\|\right\}_{k\in\mathcal{K}} \tag{3.11}$$

converges r-linearly to 0. In particular

$$\sum_{k\in\mathcal{K}} \left(\max_{0\leq i\leq i_k} \|y_k^i - x_*\|\right) < \infty.$$

Proof. Let us choose $r \in (0,1)$ and let $\mathcal{K}^r$ be the subset of $\mathcal{K}$ given by Lemma 3.11. Then Lemma 3.10 with $\mathcal{K}' = \mathcal{K}^r$ can be applied because conditions (3.7) hold from Lemma 3.11. Changing $\zeta \in (0,1)$ into $\zeta^{2/r}$ for convenience, we get for $k \in \mathcal{K}^r$, $k \geq k_3$ and $i = 1,\ldots,i_k$

$$\Theta_\sigma(y_k^i) - \Theta_\sigma(x_*) \leq \zeta^{2/r}\Big(\Theta_\sigma(y_k) - \Theta_\sigma(x_*)\Big).$$

Recall that $y_k^{i_k} = y_{k+1}$. Since $\{\Theta_\sigma(y_k)\}_{k\geq k_3}$ is decreasing, we get now for *any* $k \geq k_3$ and $i = 1,\ldots,i_k$

$$\begin{aligned}\Theta_\sigma(y_k^i) - \Theta_\sigma(x_*) &\leq \zeta^{(2/r)|\mathcal{K}^r\cap\{k_3,\ldots,k\}|}\Big(\Theta_\sigma(y_{k_3}) - \Theta_\sigma(x_*)\Big)\\ &\leq \zeta^{2(|\mathcal{K}_k|-k_3/r)}\Big(\Theta_\sigma(y_{k_3}) - \Theta_\sigma(x_*)\Big).\end{aligned}$$

We have used the fact that $\zeta \in (0,1)$, $|\mathcal{K}^r\cap\{k_3,\ldots,k\}| \geq |\mathcal{K}_k^r| - k_3$, and $|\mathcal{K}_k^r| \geq r|\mathcal{K}_k|$ (by Lemma 3.11). The left hand side of this inequality can be minorized by using Lemma 3.9. Therefore

$$\left(\max_{1\leq i\leq i_k} \|y_k^i - x_*\|\right) \leq C\zeta^{|\mathcal{K}_k|-k_3/r}, \quad \text{for } k \geq k_3,$$

and

$$\left(\max_{0\leq i\leq i_k} \|y_k^i - x_*\|\right) \leq C\zeta^{|\mathcal{K}_{k-1}|-k_3/r}, \quad \text{for } k > k_3.$$

This estimate implies that the sequence (3.11) is r-linearly convergent to 0. Indeed, using the fact that $|\mathcal{K}_k| \to \infty$ and $|\mathcal{K}_{k-1}|/|\mathcal{K}_k| \to 1$, we have

$$\limsup_{\substack{k\to\infty\\k\in\mathcal{K}}} \left(\max_{0\leq i\leq i_k} \|y_k^i - x_*\|\right)^{1/|\mathcal{K}_k|} \leq \lim_{\substack{k\to\infty\\k\in\mathcal{K}}} C^{1/|\mathcal{K}_k|}\zeta^{(|\mathcal{K}_{k-1}|-k_3/r)/|\mathcal{K}_k|} = \zeta < 1.$$

□

The next lemma is a general result on BFGS updates. It can be found in the paper of Byrd and Nocedal [4, Theorem 3.2]. The first part of the result concerning the boundedness of $\{B_k\}$ and $\{B_k^{-1}\}$ was already proved by Powell [24, Theorem 3].

Lemma 3.13. *Let $\{B_k\}_{k\geq 1}$ be a sequence of matrices updated by the BFGS formula from a given symmetric positive definite matrix B_1 and pairs (γ_k, δ_k) of vectors verifying*

$$\gamma_k^\top \delta_k > 0, \quad \forall k \geq 1 \qquad \textit{and} \qquad \sum_{k\geq 1} \frac{\|\gamma_k - B_*\delta_k\|}{\|\delta_k\|} < \infty,$$

where B_* is a symmetric positive definite matrix. Then, the sequences $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded and

$$(B_k - B_*)\delta_k = o(\|\delta_k\|).$$

By applying this lemma, we get a result that will greatly simplify the analysis: the sequences $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded.

Lemma 3.14. *Suppose that the hypotheses of Theorem 3.2 hold. Then,*

$$(B_k - B_*)\delta_k = o(\|\delta_k\|), \quad \textit{for } k \in \mathcal{K}, \tag{3.12}$$

and the sequences $\{B_k\}_{k\geq 1}$ and $\{B_k^{-1}\}_{k\geq 1}$ are bounded.

Proof. When $\mathcal{K}$ is bounded, the estimate (3.12) is meaningless and the boundedness of the matrices is clear. Suppose now that $\mathcal{K}$ is unbounded. From (2.2) and the Lipschitz continuity of g', we have

$$\begin{aligned} \|\gamma_k - Z_*^{-\top} L_* e_k\| &\leq \left(\int_0^1 \|g'(y_k + \xi e_k) - g'(x_*)\| \, \mathrm{d}\xi\right) \|e_k\| \\ &\leq C(\|y_k - x_*\| + \|e_k\|)\|e_k\|. \end{aligned} \tag{3.13}$$

Consider only the indices $k \in \mathcal{K}$. From Lemma 3.7, $e_k = Z^-(y_k)\delta_k + o(\|\delta_k\|)$, for $k \in \mathcal{K}$. Hence

$$\|e_k\| \sim \|\delta_k\|, \quad \text{for } k \in \mathcal{K}.$$

On the other hand, because

$$\bar{\alpha}_k t_k = Z^-(y_k)\delta_k = Z_*^- \delta_k + O(\|y_k - x_*\|\|\delta_k\|)$$

and

$$\delta_k = O\left(\|e_k\|\right) = O(\|y_k - x_*\| + \|y_{k+1} - x_*\|),$$

Lemma 3.7 gives

$$\|e_k - Z_*^- \delta_k\| \leq C\left(\|e_{k\ominus 2}^1\| + \max_{0\leq i\leq i_k} \|y_k^i - x_*\|\right) \|\delta_k\|.$$

With (3.13)

$$\frac{\|\gamma_k - B_*\delta_k\|}{\|\delta_k\|} \le C\left(\|e^1_{k\ominus 2}\| + \max_{0\le i\le i_k}\|y^i_k - x_*\|\right), \quad \text{for } k \in \mathcal{K}.$$

But $k \ominus 2 \in \mathcal{K}$ and

$$\|e^1_{k\ominus 2}\| \le \|y^1_{k\ominus 2} - x_*\| + \|y_{k\ominus 2} - x_*\| \le 2 \max_{0\le i\le i_{k\ominus 2}} \|y^i_{k\ominus 2} - x_*\|.$$

Therefore, with Lemma 3.12

$$\sum_{k\in\mathcal{K}} \frac{\|\gamma_k - B_*\delta_k\|}{\|\delta_k\|} < \infty.$$

The conclusion now follows from Lemma 3.13. □

Lemma 3.15. *Suppose that the hypotheses of Theorem 3.2 hold. Then,*

(*i*) *the step-sizes are bounded away from zero: there exists a constant $\breve{\alpha} > 0$, such that $\alpha_k \ge \alpha^1_k \ge \breve{\alpha}$, for $k \ge 1$;*

(*ii*) *there is a constant $\zeta \in (0,1)$ such that for $k \ge k_3$*

$$\Theta_\sigma(y_{k+1}) - \Theta_\sigma(x_*) \le \zeta\Big(\Theta_\sigma(y_k) - \Theta_\sigma(x_*)\Big);$$

(*iii*) *$\{y_k\}_{k\ge 1}$ converges r-linearly to x_*; in particular, $\sum_{k\ge 1}\|y_k - x_*\| < \infty$.*

Proof. As $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded, we have

$$\cos\theta_k = \frac{\delta_k^\top B_k \delta_k}{\|B_k\delta_k\|\,\|\delta_k\|} \ge \frac{\lambda_{\min}(B_k)}{\lambda_{\max}(B_k)} \ge C > 0,$$

$$\|B_k\delta_k\| \le C\|\delta_k\|,$$

$$\|B_k\delta_k\| \ge \frac{\|\delta_k\|}{\|B_k^{-1}\|} \ge C\|\delta_k\|.$$

By this, we can apply Lemma 3.10 with $\mathcal{K}' = \mathbb{N}^*$. The first consequence is that the step-sizes α^1_k are bounded away from zero.

This lemma also gives a constant $\zeta \in (0,1)$ such that

$$\Theta_\sigma(y_{k+1}) - \Theta_\sigma(x_*) \le \zeta^2\Big(\Theta_\sigma(y_k) - \Theta_\sigma(x_*)\Big), \quad \text{for } k \ge k_3.$$

Hence, by induction

$$\Theta_\sigma(y_k) - \Theta_\sigma(x_*) \le \zeta^{2(k-k_3)}\Big(\Theta_\sigma(y_{k_3}) - \Theta_\sigma(x_*)\Big), \quad \text{for } k \ge k_3.$$

Now with Lemma 3.9,

$$\|y_k - x_*\|^{1/k} \le C^{1/k}\zeta^{1-k_3/k}, \quad \text{for } k \ge k_3.$$

This implies the r-linear convergence of $\{y_k\}$. □

With Lemma 3.14 we have proved point (i) of Theorem 3.2 and this last lemma gives some information on points (ii) and (iii). The next sections concentrate on points (ii)–(iv) of the theorem.

3.3 Admissibility of the ideal step-size

We say that an *ideal step-size* is taken if the search algorithm terminates with $\alpha_k = \alpha_k^1 = 1$. According to the description and the assumptions of Section 2, this occurs when the first step-size trial is successful, i.e., when

$$\Theta_{\sigma_k}(y_k + t_k + r_{k+1}) \le \Theta_{\sigma_k}(y_k) + \omega_1 \Theta'_{\sigma_k}(y_k; d_k^0) \tag{3.14}$$

and either the update criterion (2.10) is not satisfied or it is satisfied together with

$$g(y_k + t_k + r_{k+1})^\top Z(y_k)t_k \ge \omega_2 g(y_k)^\top Z(y_k)t_k. \tag{3.15}$$

The result that gives conditions under which (3.14) and possibly (3.15) hold is based on expansions of $\Theta_{\sigma_k}(y_k + t_k + r_{k+1})$ and $g(y_k + t_k + r_{k+1})$ about y_k. Thinking of possible future uses, we state them in a general context, which even does not depend on the algorithm given in Section 2. In particular, it will be licit to use these results for subsequences of $\{y_k\}_{k\ge 1}$.

Proposition 3.16. *Suppose that Assumptions* 3.1 *hold and that we are given a sequence* $\{y_k\}$ *in* Ω_0 *converging to* x_*, *a sequence of tangent directions* $\{t_k \in N(A(y_k))\}$ *converging to zero, and a bounded sequence of positive numbers* $\{\sigma_k\}$. *Define* $r_{k+1} = -A^-(y_k)c(y_k + t_k)$ *and* $d_k^0 = t_k - A^-(y_k)c(y_k)$. *Then*

$$\Theta_{\sigma_k}(y_k + t_k + r_{k+1}) = \Theta_{\sigma_k}(y_k) + \Theta'_{\sigma_k}(y_k; d_k^0) + \frac{1}{2} t_k^\top L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|)$$

and

$$g(y_k + t_k + r_{k+1}) = g(y_k) + B_* Z(y_k)t_k + o(\|t_k\|) + O(\|r_{k+1}\|).$$

Proof. Expanding $c(y_k + t_k)$ about y_k and using $c'(y_k) \cdot t_k = 0$ give

$$c(y_k + t_k) = c(y_k) + \frac{1}{2} c''(x_*) \cdot t_k^2 + o(\|t_k\|^2).$$

Note that $t_k = Z^-(y_k)Z(y_k)t_k$. Let us now expand $f(y_k + t_k + r_{k+1})$ and $c(y_k + t_k + r_{k+1})$ about y_k at the second order in t_k and the first order r_{k+1}, using the estimate above:

$$\begin{aligned} f(y_k + t_k + r_{k+1}) &= f(y_k) + g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k + t_k) \\ &\quad + \frac{1}{2}f''(x_*) \cdot (t_k + r_{k+1})^2 + o(\|t_k + r_{k+1}\|^2) \\ &= f(y_k) + g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k) \\ &\quad + \frac{1}{2}t_k^\top L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|) \end{aligned}$$

and

$$\begin{aligned} c(y_k + t_k + r_{k+1}) &= c(y_k) - c(y_k + t_k) + \frac{1}{2}c''(x_*) \cdot (t_k + r_{k+1})^2 + o(\|t_k + r_{k+1}\|^2) \\ &= o(\|t_k\|^2) + o(\|r_{k+1}\|). \end{aligned}$$

Hence, because $\{\sigma_k\}$ is bounded

$$\begin{aligned} \Theta_{\sigma_k}(y_k + t_k + r_{k+1}) &= \Theta_{\sigma_k}(y_k) - \sigma_k\|c(y_k)\| + g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k) \\ &\quad + \frac{1}{2}t_k^\top L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|). \end{aligned}$$

As $\Theta'_{\sigma_k}(y_k; d_k^0) = g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k) - \sigma_k\|c(y_k)\|$, the first part of the proposition is proved.

For the second part, we have with (2.2) and $t_k = Z_*^- Z(y_k)t_k + o(\|t_k\|)$

$$\begin{aligned} g(y_k + t_k + r_{k+1}) &= g(y_k) + Z_*^{-\top} L_*(t_k + r_{k+1}) + o(\|t_k + r_{k+1}\|) \\ &= g(y_k) + Z_*^{-\top} L_* t_k + o(\|t_k\|) + O(\|r_{k+1}\|) \\ &= g(y_k) + B_* Z(y_k)t_k + o(\|t_k\|) + O(\|r_{k+1}\|). \end{aligned}$$

This concludes the proof. □

The next proposition analyzes conditions of admissibility of the ideal step-size that will be useful for the steps with update ($k \in \mathcal{K}$). For these steps, we shall see that the estimate (3.16) holds.

Proposition 3.17. *Suppose that the hypotheses of Proposition* 3.16 *hold and that*

$$g(y_k) + B_* Z(y_k)t_k = o(\|t_k\|) + O(\|r_{k+1}\|). \tag{3.16}$$

Then,

(*i*) *if* $\omega_1 < 1/2$ *and the bounded sequence* $\{\sigma_k\}$ *satisfies* $\sigma_k \geq \|\lambda(y_k)\|_D + \bar{\sigma}$ *for some constant* $\bar{\sigma} > 0$, *then* (3.14) *holds for large* k;

(*ii*) *if* $\omega_2 > 0$ *and* $r_{k+1} = o(\|t_k\|)$, *then* (3.15) *holds for large* k.

Proof. First, consider part (i) with its hypotheses. Then

$$\Theta'_{\sigma_k}(y_k; d_k^0) = g(y_k)^{\top} Z(y_k)t_k + \lambda(y_k)^{\top} c(y_k) - \sigma_k \|c(y_k)\| \leq g(y_k)^{\top} Z(y_k)t_k.$$

Using Proposition 3.16, $t_k = Z_*^- Z(y_k)t_k + o(\|t_k\|)$, and (3.16), we have

$$\begin{aligned}
&\Theta_{\sigma_k}(y_k + t_k + r_{k+1}) - \Theta_{\sigma_k}(y_k) - \omega_1 \Theta'_{\sigma_k}(y_k; d_k^0) \\
&= (1-\omega_1)\Theta'_{\sigma_k}(y_k; d_k^0) + \frac{1}{2} t_k^{\top} L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|) \\
&\leq \left(\frac{1}{2} - \omega_1\right) \Theta'_{\sigma_k}(y_k; d_k^0) + \frac{1}{2} g(y_k)^{\top} Z(y_k)t_k + \frac{1}{2} t_k^{\top} L_* t_k \\
&\quad + o(\|t_k\|^2) + o(\|r_{k+1}\|) \\
&= \left(\frac{1}{2} - \omega_1\right) \Theta'_{\sigma_k}(y_k; d_k^0) + \frac{1}{2} \left(g(y_k) + B_* Z(y_k)t_k\right)^{\top} Z(y_k)t_k \\
&\quad + o(\|t_k\|^2) + o(\|r_{k+1}\|) \\
&= \left(\frac{1}{2} - \omega_1\right) \Theta'_{\sigma_k}(y_k; d_k^0) + o(\|t_k\|^2) + o(\|r_{k+1}\|).
\end{aligned}$$

On the other hand, with (3.16) and $\sigma_k \geq \|\lambda(y_k)\|_D + \bar{\sigma}$

$$\Theta'_{\sigma_k}(y_k; d_k^0) \leq -t_k^{\top} Z(y_k)^{\top} B_* Z(y_k)t_k - \bar{\sigma}\|c(y_k)\| + o(\|t_k\|^2) + o(\|r_{k+1}\|).$$

Since $c(y_k + t_k) = c(y_k) + O(\|t_k\|^2)$, $r_{k+1} = O(\|c(y_k)\|) + O(\|t_k\|^2)$. Finally

$$\begin{aligned}
&\Theta_{\sigma_k}(y_k + t_k + r_{k+1}) - \Theta_{\sigma_k}(y_k) - \omega_1 \Theta'_{\sigma_k}(y_k; d_k^0) \\
&\leq -\left(\frac{1}{2} - \omega_1\right) t_k^{\top} Z(y_k)^{\top} B_* Z(y_k)t_k - \left(\frac{1}{2} - \omega_1\right) \bar{\sigma}\|c(y_k)\| \\
&\quad + o(\|Z(y_k)t_k\|^2) + o(\|c(y_k)\|).
\end{aligned}$$

Part (i) of the proposition is proved because B_* is positive definite and $\omega_1 < 1/2$.

By Proposition 3.16, (3.16), and $r_{k+1} = o(\|t_k\|)$

$$g(y_k + t_k + r_{k+1}) = o(\|t_k\|).$$

Hence, using again (3.16)

$$\begin{aligned}
&g(y_k + t_k + r_{k+1})^{\top} Z(y_k)t_k - \omega_2 g(y_k)^{\top} Z(y_k)t_k \\
&= \omega_2 t_k^{\top} Z(y_k)^{\top} B_* Z(y_k)t_k + o(\|t_k\|^2).
\end{aligned}$$

The conclusion follows from the positive definiteness of B_* and $\omega_2 > 0$. □

The next proposition is useful to analyze the admissibility of the ideal step-size for iterations at which there is no update. In this case, a condition like (3.16) does not necessarily hold. This one is indeed given by the BFGS theory, which does not give any information when the matrix is not updated. Now, when $\mathcal{K}$ is bounded, $\|e^1_{k\ominus 2}\|$ is constant for large k and it is not difficult to show that the negation of the update criterion (2.10) implies that $t_k = O(\|r_{k+1}\|)$ (see the proof of Lemma 3.26). In this case, part (i) of the next proposition can be applied. The case when $\mathcal{K}$ is unbounded is more difficult, but the result obtained in part (ii) of the next proposition will be useful for treating this case with an argument by contradiction (proof of Lemma 3.26).

Proposition 3.18. *Suppose that the hypotheses of Proposition* 3.16 *hold. Suppose also that there are constants* $C' > 0$ *and* $\bar{\sigma} > 0$ *such that* $g(y_k)^\top Z(y_k)t_k \leq -C'\|t_k\|^2$ *and* $\sigma_k \geq \|\lambda(y_k)\|_D + \bar{\sigma}$, *and that* $\omega_1 < 1$. *Then,*
(i) *if* $t_k = O(\|r_{k+1}\|)$, *then* (3.14) *holds for large* k;
(ii) *if* (3.14) *does not hold, then* $c(y_k)$ *and* $c(y_k + t_k)$ *are of order* $O(\|t_k\|^2)$.

Proof. With the assumptions common to (i) and (ii), we can write

$$\begin{aligned}\Theta'_{\sigma_k}(y_k; d_k^0) &= g(y_k)^\top Z(y_k)t_k + \lambda(y_k)^\top c(y_k) - \sigma_k\|c(y_k)\| \\ &\leq -C'\|t_k\|^2 - \bar{\sigma}\|c(y_k)\|. \end{aligned} \tag{3.17}$$

It follows from Proposition 3.16, (3.17), and $\omega_1 < 1$

$$\begin{aligned} X_k &= \Theta_{\sigma_k}(y_k + t_k + r_{k+1}) - \Theta_{\sigma_k}(y_k) - \omega_1\Theta'_{\sigma_k}(y_k; d_k^0) \\ &= (1-\omega_1)\Theta'_{\sigma_k}(y_k; d_k^0) + \frac{1}{2}t_k^\top L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|) \\ &\leq -C(\|t_k\|^2 + \|c(y_k)\|) + \frac{1}{2}t_k^\top L_* t_k + o(\|t_k\|^2) + o(\|r_{k+1}\|). \end{aligned}$$

Consider case (i). Because $t_k \to 0$ and $t_k = O(\|r_{k+1}\|)$, we have $t_k^\top L_* t_k = o(\|r_{k+1}\|)$ and, because $r_{k+1} = O(\|c(y_k + t_k)\|) = O(\|c(y_k)\|) + O(\|t_k\|^2)$, we have

$$X_k \leq -C(\|t_k\|^2 + \|c(y_k)\|) + o(\|t_k\|^2) + o(\|c(y_k)\|).$$

This shows that $X_k \leq 0$ for large k, i.e., point (i).

Consider now case (ii). Since (3.14) does not hold, $X_k \geq 0$. Hence

$$\|t_k\|^2 + \|c(y_k)\| = O(\|t_k\|^2) + o(\|c(y_k)\|).$$

From this we deduce that $c(y_k) = O(\|t_k\|^2)$. But $c(y_k + t_k) = c(y_k) + O(\|t_k\|^2)$, so that the same estimate holds for $c(y_k + t_k)$. □

Proposition 3.17 allows us to show that the algorithm has the property to take an ideal step-size for k large in $\mathcal{K}$. Proving that $\alpha_k = \alpha_k^1 = 1$ for large k in $\mathcal{K}^c$ is a much harder task. In fact, when we shall have proved this, the q-superlinear convergence will follow immediately. Therefore, we leave the study of the sequence $\mathcal{K}^c$ for the next section.

Lemma 3.19. *Suppose that the hypotheses of Theorem 3.2 hold. If $\mathcal{K}$ is unbounded, we have*

$$g(y_k) + B_* Z(y_k)t_k = o(\|t_k\|), \quad \textit{for } k \in \mathcal{K}$$

and the ideal step-size ($\alpha_k = \alpha_k^1 = 1$) is taken for k large in $\mathcal{K}$, say $k \geq k_4 \geq k_3$ (index given after Lemma 3.9).

Proof. From Lemma 3.14 and the fact that $\delta_k = \bar{\alpha}_k Z(y_k)t_k$, we have

$$(B_k - B_*)Z(y_k)t_k = o(\|Z(y_k)t_k\|) = o(\|t_k\|), \quad \text{for } k \in \mathcal{K}.$$

Now $B_k Z(y_k)t_k = -g(y_k)$, hence

$$g(y_k) + B_* Z(y_k)t_k = o(\|t_k\|), \quad \text{for } k \in \mathcal{K},$$

which is the first claim of the lemma.

Note that $t_k \to 0$ (since the matrices B_k^{-1} are bounded, by Lemma 3.14). Then, we can apply Proposition 3.17. As $\omega_1 < 1/2$, point (i) of that proposition states that (3.14) holds for large k. By the update criterion (2.10), $r_{k+1} = o(\|t_k\|)$ for $k \in \mathcal{K}$ and point (ii) of the same proposition claims that (3.15) holds for large k. Therefore, the ideal step-size is taken for k large in $\mathcal{K}$. □

Since the ideal step-size is accepted for large $k \in \mathcal{K}$ (supposed unbounded), $e_{k\ominus 2}^1 = e_{k\ominus 2}$ for large k, so that, from now on, the update criterion (2.10) can also be written with $e_{k\ominus 2}$ instead of $e_{k\ominus 2}^1$.

3.4 The q-superlinear convergence

The proof of the superlinear improvement of the steps with update ($k \in \mathcal{K}$) is standard and is based on the following theorem. As for the propositions of Section 3.3, we state the theorem independently of the algorithm of Section 2, so that it can be used for any subsequence of $\{y_k\}_{k\geq 1}$. For a proof, see [16, Theorem 5.1] (the statement of the theorem is slightly different here, but the proof is similar).

Theorem 3.20. *Suppose that Assumptions 3.1 hold. Let $\{y_k\}_{k\geq 1} \subset \Omega_0$ be a sequence converging to x_* and $\{t_k\}_{k\geq 1}$ be a sequence of steps $t_k \in N(A(y_k))$ converging to 0. Define the sequence $\{x_k\}_{k\geq 2}$ by $x_{k+1} = y_k + t_k$ and define also $r_{k+1} = -A^-(y_k)c(y_k + t_k)$. If $\mathcal{K}'$ is a subsequence of indices such that $y_k = x_k + r_k$ for $k \in \mathcal{K}'$, then the following conditions are equivalent:*
(i) $x_{k+1} - x_* = o(\|x_k - x_*\|)$, for $k \in \mathcal{K}'$;
(ii) $g(y_k) + B_* Z(y_k)t_k = o(\|x_k - x_*\|)$, for $k \in \mathcal{K}'$.

When condition (i) is satisfied, we say that $x_{k+1} - x_k$ is a superlinear step. Note that condition (ii) is not as strong as condition (3.16), which is given by the update scheme (see Lemma 3.14 or 3.19).

We cannot immediately apply this theorem to the sequence $\{x_k\}_{k\in\mathcal{K}}$, since we still do not know that $\alpha_{k-1} = \alpha_{k-1}^1 = 1$ for large $k \in \mathcal{K}$, hence y_k may differ from $x_k + r_k$ infinitely often in $\mathcal{K}$. We can show however that when $k \in \mathcal{K}$ is sufficiently large (in particular to have admissibility of the ideal step-size, by Lemma 3.19) the step from y_k to y_{k+1} is superlinear. This result may look surprising since the whole sequence $\{y_k\}_{k\geq 1}$ does not converge q-superlinearly in general: see Byrd [2] and Yuan [28]. Of course the estimate $r_{k+1} = o(\|t_k\|)$ for $k \in \mathcal{K}$, coming from the update criterion (2.10), plays a crucial role in obtaining the result and is generally not true at each iteration.

Lemma 3.21. *Suppose that the hypotheses of Theorem 3.2 hold and that $\mathcal{K}$ is unbounded. Then $y_{k+1} - x_* = o(\|y_k - x_*\|)$, for $k \in \mathcal{K}$.*

Proof. Since $\{B_k^{-1}\}$ is bounded (Lemma 3.14),

$$t_k = O(\|y_k - x_*\|). \tag{3.18}$$

Therefore, using (3.18)

$$\begin{aligned} r_{k+1} &= -A^-(y_k)c(x_{k+1}) \\ &= -A^-(y_k)c(y_k) + o(\|t_k\|) \\ &= -A_*^- A_*(y_k - x_*) + o(\|y_k - x_*\|). \end{aligned} \tag{3.19}$$

Now, as $\mathcal{K}$ is unbounded, the update criterion (2.10) gives

$$r_{k+1} = O(\|c(x_{k+1})\|) = o(\|t_k\|) = o(\|y_k - x_*\|), \quad \text{for } k \in \mathcal{K}.$$

We deduce from this and (3.19) that

$$A_*(y_k - x_*) = o(\|y_k - x_*\|), \quad \text{for } k \in \mathcal{K}. \tag{3.20}$$

By Lemma 3.19 and (3.18), $g(y_k) + B_* Z(y_k) t_k = o(\|t_k\|) = o(\|y_k - x_*\|)$, for $k \in \mathcal{K}$. Hence

$$B_k^{-1} g(y_k) = B_*^{-1} g(y_k) + o(\|y_k - x_*\|), \quad \text{for } k \in \mathcal{K}. \tag{3.21}$$

Let us now express t_k, using successively (3.21), (2.2) and (3.20)

$$\begin{aligned} t_k &= -Z^-(y_k) B_k^{-1} g(y_k) \\ &= -Z^-(y_k) B_*^{-1} g(y_k) + o(\|y_k - x_*\|) \\ &= -Z^-(y_k) B_*^{-1} Z_*^{-\mathsf{T}} L_*(y_k - x_*) + o(\|y_k - x_*\|) \\ &= -Z_*^- B_*^{-1} Z_*^{-\mathsf{T}} L_* Z_*^- Z_*(y_k - x_*) + o(\|y_k - x_*\|), \quad \text{for } k \in \mathcal{K}. \end{aligned}$$

Since $B_* = Z_*^{-\top} L_* Z_*^-$

$$t_k = -Z_*^- Z_*(y_k - x_*) + o(\|y_k - x_*\|), \quad \text{for } k \in \mathcal{K}. \tag{3.22}$$

Using (3.19) and (3.22), we can now conclude that for $k \in \mathcal{K}$ and $k \geq k_4$ (note that $\alpha_k = \alpha_k^1 = 1$, by Lemma 3.19):

$$y_{k+1} - x_* = y_k - x_* + t_k + r_{k+1} = o(\|y_k - x_*\|).$$

This ends the proof. □

For $k \notin \mathcal{K}$, $y_{k+1} - x_*$ and $y_k - x_*$ can still be compared.

Lemma 3.22. *Suppose that the hypotheses of Theorem 3.2 hold. Then* $y_{k+1} - x_* = O(\|y_k - x_*\|)$, *for* $k \geq 1$.

Proof. By Lemma 3.21, we only have to consider the indices $k \in \mathcal{K}^c$. For k large in $\mathcal{K}^c$, the search is interrupted at $y_{k+1} = y_k^1$, because the update criterion (2.10) does not hold. Then, the result follows from the definition (2.8) of y_k^1, $\alpha_k^1 \leq 1$, the boundedness of $\{B_k^{-1}\}$, and the smoothness of c and g. □

Lemma 3.21 and the point (*ii-b*) in the next lemma were crucial for the design of the update criterion. These results are important for the treatment of the iterations without update. Indeed, for $k \in \mathcal{K}^c$, the negation of the update criterion (2.10) and the estimates (*i*) and (*ii-b*) in the next lemma (assuming that $\alpha_k = \alpha_k^1 = 1$ and $\alpha_{k-1} = \alpha_{k-1}^1 = 1$ in this discussion) give

$$\|t_k\| \leq C \frac{\|y_{k-1} - x_*\|}{\|y_{k\ominus 2} - x_*\|} \|x_k - x_*\|.$$

Suppose also that there is a positive constant C such that $\|y_l - x_*\| \leq C\|y_k - x_*\|$ for all $l \geq k \geq 1$ (we shall discuss this below). Then, the last inequality gives

$$\|t_k\| \leq C \frac{\|y_{(k\ominus 2)+1} - x_*\|}{\|y_{k\ominus 2} - x_*\|} \|x_k - x_*\|,$$

because $k - 1 \geq (k \ominus 2) + 1$. By Lemma 3.21, the ratio in the inequality above goes to zero when $\mathcal{K}$ is unbounded (since $k \ominus 2 \in \mathcal{K}$). Hence $t_k = o(\|x_k - x_*\|)$ and, by point (*ii-c*) of the next lemma, $x_{k+1} - x_* = o(\|x_k - x_*\|)$.

This discussion summarizes the reasoning that is followed in the rest of the paper. The path to the final goal is still long, however, because we still do not know that the ideal step-size is accepted eventually and we may not have $\|y_l - x_*\| \leq C\|y_k - x_*\|$, for all $l \geq k \geq 1$.

Recall that s_k is defined in (2.1) by $s_k = x_{k+1} - x_k$.

Lemma 3.23. *Suppose that the hypotheses of Theorem 3.2 hold and let $\mathcal{K}'$ be a subsequence of indices. Then,*
(*i*) *if $\alpha_k = \alpha_k^1 = 1$ for $k \in \mathcal{K}'$, we have $\|e_k\| \sim \|y_k - x_*\|$, for $k \in \mathcal{K}'$;*
(*ii*) *if $\alpha_{k-1} = \alpha_{k-1}^1 = 1$ for $k \in \mathcal{K}'$, we have*
  (*a*) *$\|s_k\| \sim \|x_k - x_*\|$, for $k \in \mathcal{K}'$,*
  (*b*) *$c(y_k)$ and $c(x_{k+1})$ are $O(\|y_{k-1} - x_*\| \, \|x_k - x_*\|)$, for $k \in \mathcal{K}'$,*
  (*c*) *$(y_k - x_*)$ and $(x_{k+1} - x_*)$ are $O(\|t_k\|) + o(\|x_k - x_*\|)$, for $k \in \mathcal{K}'$.*

Proof. The fact that $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded is used throughout the proof. The proof of (*i*) and (*ii-a*) can be found in [16, Proposition 4.4]. These statements are based on the estimates

$$e_k = -(A_*^- A_* + Z_*^- B_k^{-1} Z_*^{-\mathsf{T}} L_*)(y_k - x_*) + o(\|y_k - x_*\|),$$

$$s_k = -(A_*^- A_* + Z_*^- B_k^{-1} B_* Z_*)(x_k - x_*) + o(\|x_k - x_*\|),$$

which hold when $\alpha_k = \alpha_k^1 = 1$ for $k \in \mathcal{K}'$ or when $\alpha_{k-1} = \alpha_{k-1}^1 = 1$ for $k \in \mathcal{K}'$, respectively.

Consider now the point (*ii-b*). We have $c(y_k) = A_*(y_k - x_*) + O(\|y_k - x_*\|^2)$ and the same estimate is true for $c(x_{k+1}) = c(y_k) + O(\|t_k\|^2)$, since $t_k = O(\|y_k - x_*\|)$. Hence

$$\left\{ \begin{array}{c} c(y_k) \\ c(x_{k+1}) \end{array} \right\} = A_*(y_k - x_*) + O(\|y_k - x_*\|^2). \tag{3.23}$$

If $\alpha_{k-1} = \alpha_{k-1}^1 = 1$, we have

$$\begin{aligned} & y_k - x_* \\ &= x_k - x_* - A^-(y_{k-1})c(x_k) \\ &= x_k - x_* - A^-(y_{k-1})A_*(x_k - x_*) + O(\|x_k - x_*\|^2) \\ &= Z_*^- Z_*(x_k - x_*) + O(\|y_{k-1} - x_*\| \, \|x_k - x_*\|) + O(\|x_k - x_*\|^2). \end{aligned} \tag{3.24}$$

This injected in (3.23) gives

$$\left\{ \begin{array}{c} c(y_k) \\ c(x_{k+1}) \end{array} \right\} = O(\|y_{k-1} - x_*\| \, \|x_k - x_*\|) + O(\|x_k - x_*\|^2) + O(\|y_k - x_*\|^2).$$

From (3.24), $y_k - x_* = O(\|x_k - x_*\|)$ and $x_k - x_* = y_{k-1} - x_* + t_{k-1} = O(\|y_{k-1} - x_*\|)$. The point (*ii-b*) then follows from the last estimate.

Finally, consider the point (*ii-c*). An easy expansion gives with (3.24)

$$\begin{aligned} t_k &= -Z^-(y_k) B_k^{-1} g(y_k) \\ &= -Z_*^- B_k^{-1} Z_*^{-\mathsf{T}} L_*(y_k - x_*) + o(\|y_k - x_*\|) \\ &= -Z_*^- B_k^{-1} B_* Z_*(x_k - x_*) + o(\|x_k - x_*\|). \end{aligned}$$

Hence

$$Z_*(x_k - x_*) = -B_*^{-1} B_k Z_* t_k + o(\|x_k - x_*\|) = O(\|t_k\|) + o(\|x_k - x_*\|). \quad (3.25)$$

We deduce from (3.24) and (3.25)

$$y_k - x_* = O(\|t_k\|) + o(\|x_k - x_*\|)$$

and also

$$x_{k+1} - x_* = y_k - x_* + t_k = O(\|t_k\|) + o(\|x_k - x_*\|),$$

which are the expected estimates. □

The next two lemmas deal with the problem of comparing iterates y_{k+j} and y_k whose indices may be far from each other ($j \geq 1$ may be large). This comparison is necessary if we want to relate two iterates whose indices follow one another in $\mathcal{K}$ but not in $\mathbb{N}^*$. In [15] and in the reasoning given above Lemma 3.23, this difficulty was bypassed by supposing that there is a constant $C > 0$ such that

$$\|y_{k+j} - x_*\| \leq C\|y_k - x_*\|, \quad \forall k \geq 1, \quad \forall j \geq 1.$$

This is true for $j = 1$ (Lemma 3.22) and hence for bounded j, but we were not able to prove this in all generality. It is instructive to note, however, that such a result would be easily obtained in unconstrained optimization by observing that, when $f''(x_*)$ is positive definite,

$$\|y_k - x_*\|^2 \sim f(y_k) - f(x_*).$$

Hence

$$C_1'\|y_{k+j} - x_*\|^2 \leq f(y_{k+j}) - f(x_*) \leq f(y_k) - f(x_*) \leq C_2'\|y_k - x_*\|^2,$$

because $\{f(y_k)\}$ decreases. This argument can no longer be used for constrained problems when f is replaced by Θ_σ, since if Lemma 3.9 shows that $\|y_k - x_*\|^2 = O(\Theta_\sigma(y_k) - \Theta_\sigma(x_*))$, the converse $\Theta_\sigma(y_k) - \Theta_\sigma(x_*) = O(\|y_k - x_*\|^2)$ cannot be true, because of the non-differentiability of Θ_σ at x_*. Now, if $y_k - x_* \in N(A_*)$, then $c(y_k) = O(\|y_k - x_*\|^2)$ and $\Theta_\sigma(y_k) - \Theta_\sigma(x_*) = O(\|y_k - x_*\|^2)$ (by Lemma 3.9), so that the argument above can be applied. This situation almost occurs when the update criterion is satisfied ($k \in \mathcal{K}$), because then the step is almost tangent to the current local manifold. The next lemma exploits this idea.

Lemma 3.24. *Suppose that the hypotheses of Theorem 3.2 hold and that $\mathcal{K}$ is unbounded. Let $\eta > 0$ and $r > 0$ be fixed constants. Then, there is an index $k_{\eta,r} \geq k_4$ (index defined in Lemma 3.19) such that if $k \geq k_{\eta,r}$ and*

$$\|y_{k\ominus 2} - x_*\| \geq \eta\|y_{k\ominus 4} - x_*\|, \quad (3.26)$$

then

$$\|y_{k\ominus 1} - x_*\| \leq r\|y_{k\ominus 2} - x_*\|.$$

Proof. Fix $\eta > 0$ and $r > 0$. Let $k'_1 = k_4 \oplus 5$, where k_4 is fixed in Lemma 3.19. Then

$$\alpha_{k\ominus 4} = \alpha^1_{k\ominus 4} = 1, \quad \forall k \geq k'_1.$$

For $k \geq k'_1$, we can apply Lemma 3.9 with $y = y_{k\ominus 1}$, Lemma 3.15 [from which comes the constant $\zeta^2 \in (0,1)$, we note $\Delta_k = (k\ominus 1) - (k\ominus 2) \geq 1$], Lemma 3.9 again with $y = y_{k\ominus 2}$, and finally $g(y_k) = O(\|y_k - x_*\|)$:

$$\begin{aligned}
\|y_{k\ominus 1} - x_*\|^2 &\leq C\Big(\Theta_\sigma(y_{k\ominus 1}) - \Theta_\sigma(x_*)\Big) \\
&\leq C\zeta^{2\Delta_k}\Big(\Theta_\sigma(y_{k\ominus 2}) - \Theta_\sigma(x_*)\Big) \\
&\leq C\zeta^{2\Delta_k}\Big(\|g(y_{k\ominus 2})\|^2 + \|c(y_{k\ominus 2})\|\Big) \\
&\leq C\zeta^{2\Delta_k}\Big(\|y_{k\ominus 2} - x_*\|^2 + \|c(y_{k\ominus 2})\|\Big).
\end{aligned}$$

Now, because $k\ominus 2 \in \mathcal{K}$, the update criterion (2.10) gives

$$\begin{aligned}
\|c(y_{k\ominus 2})\| &\leq C\|e_{k\ominus 4}\|\,\|g(y_{k\ominus 2})\| \\
&\leq C\|y_{k\ominus 4} - x_*\|\,\|y_{k\ominus 2} - x_*\|,
\end{aligned}$$

We have used Lemma 3.23 (i) with $\mathcal{K}' = \mathcal{K} \cap \{k : k \geq k_4\}$, which is licit because the ideal step-size is accepted in this subsequence $\mathcal{K}'$. Grouping the inequalities obtained above, we see that there is a constant $C'_1 > 0$ such that

$$\frac{\|y_{k\ominus 1} - x_*\|}{\|y_{k\ominus 2} - x_*\|} \leq C'_1\zeta^{\Delta_k}\left(1 + \frac{\|y_{k\ominus 4} - x_*\|}{\|y_{k\ominus 2} - x_*\|}\right)^{1/2}, \quad \text{for } k \geq k'_1. \tag{3.27}$$

We set

$$C'_2 = C'_1\left(1 + \frac{1}{\eta}\right)^{1/2}$$

and remembering that $0 < \zeta < 1$, we determine $q \in \mathbb{N}^*$ such that

$$C'_2\zeta^q \leq r. \tag{3.28}$$

By Lemma 3.22, there is a constant $C'_3 \geq 1$ such that

$$\|y_{j+1} - x_*\| \leq C'_3\|y_j - x_*\|, \quad \forall j \geq 1, \tag{3.29}$$

and from Lemma 3.21 ($\mathcal{K}$ is unbounded), there is an index k'_2 such that

$$\|y_{j+1} - x_*\| \leq \frac{r}{(C'_3)^{q-1}}\|y_j - x_*\|, \quad j \geq k'_2, \quad j \in \mathcal{K}. \tag{3.30}$$

Then, take

$$k_{\eta,r} = \max(k'_1, k'_2 \oplus 3) \geq k_4.$$

We can now conclude. Let $k \geq k_{\eta,r}$ be an index such that (3.26) holds. As $k \geq k_1'$, (3.27) gives

$$\|y_{k\ominus 1} - x_*\| \leq C_2' \zeta^{\Delta_k} \|y_{k\ominus 2} - x_*\|.$$

If $\Delta_k > q$, the result is proved, due to (3.28). Otherwise $\Delta_k \leq q$ and we have by (3.29), $C_3' \geq 1$ and (3.30) (because $k \ominus 2 \in \mathcal{K}$ and $k \ominus 2 \geq k_2'$)

$$\begin{aligned} \|y_{k\ominus 1} - x_*\| &\leq (C_3')^{\Delta_k - 1} \|y_{(k\ominus 2)+1} - x_*\| \\ &\leq (C_3')^{q-1} \|y_{(k\ominus 2)+1} - x_*\| \\ &\leq r \|y_{k\ominus 2} - x_*\|. \end{aligned}$$

Hence the result is also proved when $\Delta_k \leq k$. □

Condition (3.26) used in the preceding lemma is not very desirable. It is just the opposite. It is expected that $\|y_{k\ominus 2} - x_*\|$ be smaller and smaller with respect to $\|y_{k\ominus 4} - x_*\|$ when the algorithm works well. The next lemma analyzes the situation when the progress from $y_{k\ominus 2}$ to $y_{k\ominus 1}$ is good. This result is therefore complementary to Lemma 3.24 and will be used as such.

Lemma 3.25. *Suppose that the hypotheses of Theorem 3.2 hold and that $\mathcal{K}$ and $\mathcal{K}^c$ are unbounded. Then, there is a constant $\bar{r} \in (0,1]$ such that for all $r \in (0, \bar{r}]$, there is an index k_r such that if $k \notin \mathcal{K}$, $k \geq k_r$ and*

$$\alpha_\ell = \alpha_\ell^1 = 1, \quad \textit{for } \ell = k \ominus 1 \ldots, k-1,$$

$$\|y_{k\ominus 1} - x_*\| \leq r \|y_{k\ominus 2} - x_*\|,$$

then

$$\|y_\ell - x_*\| \leq r \|y_{k\ominus 1} - x_*\|, \quad \textit{for } \ell = (k\ominus 1)+1, \ldots, k+1, \tag{3.31}$$

$$\|y_\ell - x_*\| \leq r \|y_{k\ominus 2} - x_*\|, \quad \textit{for } \ell = k\ominus 1, \ldots, k+1. \tag{3.32}$$

Proof. The second set of inequalities (3.32) is a clear consequence of the first set of inequalities, the fact that $\|y_{k\ominus 1} - x_*\| \leq r \|y_{k\ominus 2} - x_*\|$, and $r \leq 1$.

1) *Let us first determine* $\bar{r} \in (0,1]$. By Lemma 3.19, if we set $k_1' = k_4 \oplus 3$, then

$$\alpha_{k\ominus 2} = \alpha_{k\ominus 2}^1 = 1, \quad \text{for } k \geq k_1'. \tag{3.33}$$

Consider the subset of indices

$$\mathcal{L} = \{\ell \in \mathbb{N}^* : \ell \geq k_1',\ \ell \notin \mathcal{K},\ \alpha_{\ell-1} = \alpha_{\ell-1}^1 = 1\}.$$

This set is unbounded, since $\mathcal{K}$ and $\mathcal{K}^c$ are unbounded, and $\alpha_k = \alpha_k^1 = 1$ for k large in $\mathcal{K}$. By taking $\mathcal{K}' = \mathcal{L}$ in Lemma 3.23 (*ii-c*), we have

$$y_\ell - x_* = O(\|t_\ell\|) + o(\|x_\ell - x_*\|), \quad \text{for } \ell \in \mathcal{L}.$$

As $\ell \notin \mathcal{K}$ for $\ell \in \mathcal{L}$, the negation of the update criterion (2.10) gives

$$\begin{aligned}
\|t_\ell\| &\leq C\frac{\|c(y_\ell)\| + \|c(x_{\ell+1})\|}{\|e_{\ell\ominus 2}\|} \\
&\leq C\frac{\|y_{\ell-1} - x_*\|}{\|y_{\ell\ominus 2} - x_*\|}\|x_\ell - x_*\|, \quad \text{for } \ell \in \mathcal{L},
\end{aligned}$$

where we used Lemma 3.23 (*ii-b*) (valid because $\alpha_{\ell-1} = \alpha^1_{\ell-1} = 1$) and (*i*) (valid because of (3.33)). As $x_\ell - x_* = O(\|y_{\ell-1} - x_*\|)$, the preceding estimations allow us to write

$$\|y_\ell - x_*\| \leq \left(\varepsilon_\ell + C_1'\frac{\|y_{\ell-1} - x_*\|}{\|y_{\ell\ominus 2} - x_*\|}\right)\|y_{\ell-1} - x_*\|, \quad \text{for } \ell \in \mathcal{L}, \tag{3.34}$$

where $C_1' > 0$ is a constant and $\varepsilon_\ell \to 0$ when $\ell \to \infty$ in $\mathcal{L}$. Now we can choose

$$\bar{r} = \min\left(\frac{1}{2C_1'}, 1\right).$$

2) *Let $r \in (0, \bar{r}]$ and let us determine the index k_r.* By Lemma 3.22, there is a constant $C_2' \geq 1$ such that

$$\|y_{k+1} - x_*\| \leq C_2'\|y_k - x_*\|, \quad \forall k \geq 1. \tag{3.35}$$

On the other hand, from Lemma 3.21, there is an index k_2' such that

$$\|y_{(k\ominus 1)+1} - x_*\| \leq \frac{r}{C_2'}\|y_{k\ominus 1} - x_*\|, \quad \forall k \geq k_2'. \tag{3.36}$$

We fix the index k_r such that

$$\begin{cases} k_r \geq \max(k_1' + 1, k_2') \\ \varepsilon_\ell \leq \dfrac{1}{2}, \text{ for } \ell \in \mathcal{L} \text{ and } \ell \geq k_r \ominus 1. \end{cases}$$

3) *Let k be an index as in the statement of the lemma and let us show that*

$$\|y_\ell - x_*\| \leq \frac{r}{C_2'}\|y_{k\ominus 1} - x_*\|, \quad \text{for } \ell = (k\ominus 1) + 1, \ldots, k. \tag{3.37}$$

This is done by induction. The inequality is true for $\ell = (k \ominus 1) + 1$ by (3.36) and $k \geq k_2'$. Suppose now that it is true for some $\ell \in \{(k\ominus 1) + 1, \ldots, k-1\}$ and let us prove it for $\ell + 1$. Since $C_2' \geq 1$ and $r \leq \bar{r} \leq 1$, the inequality in (3.37) for the given index ℓ and the properties assumed for index k give

$$\|y_\ell - x_*\| \leq \|y_{k\ominus 1} - x_*\| \leq r\|y_{k\ominus 2} - x_*\|. \tag{3.38}$$

Clearly, $\ell+1 \in \mathcal{L}$ [because $\ell+1 \geq (k\ominus 1)+2 \geq (k_r \ominus 1)+2 \geq k_1'+2 \Rightarrow \ell+1 \geq k_1'$; $k \geq \ell+1 \geq (k\ominus 1)+2 \Rightarrow \ell+1 \notin \mathcal{K}$; and $k-1 \geq \ell \geq (k\ominus 1)+1 \Rightarrow \alpha_\ell = \alpha_\ell^1 = 1$] and $(\ell+1)\ominus 2 = k\ominus 2$ [because $(k\ominus 1)+2 \leq \ell+1 \leq k \notin \mathcal{K}$]. Therefore from (3.34)

$$\begin{aligned} \|y_{\ell+1} - x_*\| &\leq \left(\varepsilon_{\ell+1} + C_1' \frac{\|y_\ell - x_*\|}{\|y_{k\ominus 2} - x_*\|}\right) \|y_\ell - x_*\| \\ &\leq \left(\frac{1}{2} + C_1' r\right) \|y_\ell - x_*\| \\ &\leq \frac{r}{C_2'} \|y_{k\ominus 1} - x_*\|, \end{aligned}$$

where we used the fact that $\varepsilon_{\ell+1} \leq 1/2$ [because $\ell+1 \geq k\ominus 1 \geq k_r \ominus 1$], (3.38), $C_1' r \leq C_1' \bar{r} \leq 1/2$, and $\|y_\ell - x_*\| \leq (r/C_2')\|y_{k\ominus 1} - x_*\|$ (induction hypothesis). The proof by induction is finished.

4) *We conclude.* From Step 3 and the fact that $C_2' \geq 1$, we see that it remains to show

$$\|y_{k+1} - x_*\| \leq r\|y_{k\ominus 1} - x_*\|.$$

This follows from (3.35) and the inequality corresponding to $\ell = k$ in (3.37). □

We can now show that the ideal step-size is taken for any sufficiently large step.

Lemma 3.26. *Suppose that the hypotheses of Theorem 3.2 hold. Then, the ideal step-size ($\alpha_k = \alpha_k^1 = 1$) is taken for k sufficiently large.*

Proof. 1) *The case when $\mathcal{K}$ is bounded* is easy to treat. In this case, $\|e_{k\ominus 2}\|$ is constant and the update criterion (2.10) does not hold, for k large. Then, we have by negation of (2.10):

$$t_k = O(\|c(y_k)\|) + O(\|c(x_{k+1})\|).$$

But $c(x_{k+1}) = c(y_k) + o(\|t_k\|)$, hence for k large:

$$t_k = O(\|c(x_{k+1})\|) = O(\|r_{k+1}\|).$$

The hypotheses of Proposition 3.18 are satisfied since $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded. Then, its point (i) claims that condition (3.14) holds for large k. Since the update criterion is not satisfied the search algorithm finishes with $\alpha_k = \alpha_k^1 = 1$.

2) If $\mathcal{K}^c$ is bounded the result follows from Lemma 3.19. Therefore, it remains to consider the case when $\mathcal{K}$ and $\mathcal{K}^c$ are unbounded, which is assumed from now on. We prove the result by contradiction. Let us show first the following claim.

2a) CLAIM C: *If the ideal step-size is not accepted for all large k then there exists a subsequence $\mathcal{K}' \subset \mathcal{K}^c$ whose indices k have the following properties:*

$$\begin{cases} (a) & k \ominus 2 \geq k_4; \\ (b) & \text{the ideal step-size is not accepted;} \\ (c) & \alpha_\ell = \alpha_\ell^1 = 1, \ \text{for } \ell = k \ominus 1, \ldots, k-1; \\ (d) & y_{k-1} - x_* = o(\|y_{k\ominus 2} - x_*\|). \end{cases} \tag{3.39}$$

As the ideal step-size is accepted for k large in $\mathcal{K}$, it is clear that a sequence $\mathcal{K}' \subset \mathcal{K}^c$ verifying properties (3.39a), (3.39b), and (3.39c) exists. The fact that it can also satisfy property (3.39d) is less obvious. We proceed as follows.

Let $\bar{r} \in (0, 1]$ be fixed by Lemma 3.25, r be any number in $(0, \bar{r}]$, and k_1' be any index in $\mathbb{N}^*$. Let k_2' be the index given by Lemma 3.24 (there called $k_{\eta,r}$) corresponding to $\eta = r^2$ and the same r as here (hence $k_2' = k_{r^2,r}$). Let k_3' be the index given by Lemma 3.25 (there called k_r) for the same r as here. By Lemma 3.21, there is an index k_4' such that

$$\|y_{k+1} - x_*\| \leq r\|y_k - x_*\|, \quad \text{for all } k \in \mathcal{K} \text{ with } k \geq k_4'. \tag{3.40}$$

Then, we set

$$k_5' = \max(k_1', k_2' \oplus 2, k_3', k_4').$$

Claim C will be proved if we can show that:

CLAIM C': *If the ideal step-size is not accepted for all large k, and if r and k_5' are fixed as above, there is an index $k \notin \mathcal{K}$ such that $k > k_5'$, verifying* (3.39b), (3.39c), *and*

$$\|y_{k\ominus 1} - x_*\| \leq r\|y_{k\ominus 2} - x_*\|. \tag{3.41}$$

Indeed, observe first that the index k given by Claim C' satisfies (3.39a), because $k > k_5' \geq k_2' \oplus 2 \geq k_4 \oplus 2$, hence $k \ominus 2 \geq k_4$. On the other hand, the hypotheses of Lemma 3.25 are satisfied with this k. Hence, using (3.32) with $\ell = k - 1$

$$\|y_{k-1} - x_*\| \leq r\|y_{k\ominus 2} - x_*\|.$$

As r is arbitrarily small and k_1' is arbitrarily large, we can use Claim C' repetitively, each time with r replaced by $r/2$ and k_1' replaced by the index k just obtained. This allows us to build a sequence $\mathcal{K}'$ verifying Claim C.

Let us now prove Claim C'. One can find $k' \in \mathcal{K}^c$ such that $k' \ominus 3 \geq k_5'$, and (3.39b) and (3.39c) hold. If (3.41) is also true for $k = k'$, Claim C' is proved with $k = k'$. Otherwise

$$\|y_{k'\ominus 1} - x_*\| > r\|y_{k'\ominus 2} - x_*\|. \tag{3.42}$$

This implies that

$$\|y_{k'\ominus 2} - x_*\| \leq r^2 \|y_{k'\ominus 4} - x_*\|, \tag{3.43}$$

because otherwise, Lemma 3.24 with $\eta = r^2$ and $k = k'$ would lead to (3.41) with $k = k'$ and not (3.42). Inequality (3.43) can hold only if one of the following two situations occurs.

Situation A: $\|y_{k'\ominus 2} - x_*\| \leq r\|y_{k'\ominus 3} - x_*\|$. By (3.40), (3.42), and $k' \ominus 2 \geq k'_4$, we see that $(k'\ominus 2)+1 < k'\ominus 1$ and we can consider the nonvoid set $J_{k'\ominus 2} = \{(k'\ominus 2)+1, \ldots, (k'\ominus 1) - 1\} \subset \mathcal{K}^c$. If there was no index $j \in J_{k'\ominus 2}$ corresponding to a non-ideal step-size, one could apply Lemma 3.25 [with one shift back in the indices]: the index k of this lemma would be set to $(k' \ominus 1) - 1$ and inequality (3.31) with $\ell = k' \ominus 1$ would show that (3.42) does not hold. Therefore, there is an index $j \in J_{k'\ominus 2}$ to which corresponds a non-ideal step-size. The smallest of these indices j is an index k satisfying the desired conditions in Claim C' with one shift back in the indices.

Situation B: $\|y_{k'\ominus 2} - x_*\| > r\|y_{k'\ominus 3} - x_*\|$ and $\|y_{k'\ominus 3} - x_*\| \leq r\|y_{k'\ominus 4} - x_*\|$. Proceeding as in Situation A, one sees that $(k'\ominus 3)+1 < k'\ominus 2$ and, using Lemma 3.25, that there is a smallest index $j \in \{(k'\ominus 3)+1, \ldots, (k'\ominus 2) - 1\}$ to which corresponds a non-ideal step-size. This index is an index k satisfying the desired conditions in Claim C' with two shifts back in the indices.

2b) *Let us show that, with the subsequence $\mathcal{K}'$ given by Claim C, we have*

$$t_k = o(\|x_k - x_*\|), \quad \text{for } k \in \mathcal{K}'. \tag{3.44}$$

Let $k \in \mathcal{K}'$. Since $\mathcal{K}' \subset \mathcal{K}^c$, the negation of the update criterion (2.10) gives

$$\begin{aligned} \|t_k\| &\leq C\frac{\|c(y_k)\| + \|c(x_{k+1})\|}{\|e_{k\ominus 2}\|}, \quad \text{for } k \in \mathcal{K}' \\ &\leq C\frac{\|y_{k-1} - x_*\|}{\|y_{k\ominus 2} - x_*\|}\|x_k - x_*\|, \quad \text{for } k \in \mathcal{K}', \end{aligned} \tag{3.45}$$

where we have used Lemma 3.23 (i) $[\alpha_{k\ominus 2} = \alpha^1_{k\ominus 2} = 1$, since $k \ominus 2 \geq k_4]$ and $(ii\text{-}b)$ $[\alpha_{k-1} = \alpha^1_{k-1} = 1]$. Then (3.44) follows from the last inequality and (3.39d).

2c) *We show a contradiction*, which concludes the proof of the lemma. From Lemma 3.23 $(ii\text{-}a)$ $[\alpha_{k-1} = \alpha^1_{k-1} = 1$ for $k \in \mathcal{K}']$ and (3.44), we have

$$t_k = o(\|s_k\|), \quad \text{for } k \in \mathcal{K}'.$$

As $\alpha_{k-1} = \alpha^1_{k-1} = 1$, $s_k = r_k + t_k$, so that the estimate above implies

$$t_k = o(\|r_k\|), \quad \text{for } k \in \mathcal{K}'. \tag{3.46}$$

For $k \in \mathcal{K}'$, the ideal step-size is not accepted and the update criterion does not hold. This can only occur when the descent condition (3.14) does not hold. Then, we can apply Proposition 3.18 (*ii*). We get, with (3.46) and the fact that $r_k = O(\|x_k - x_*\|) = O(\|y_{k-1} - x_*\|)$:

$$\left\{ \begin{array}{c} c(y_k) \\ c(x_{k+1}) \end{array} \right\} = O(\|t_k\|^2) = o(\|r_k\| \, \|t_k\|) = o(\|y_{k-1} - x_*\| \, \|t_k\|), \quad \text{for } k \in \mathcal{K}'.$$

If we use this estimate in (3.45), we now get

$$t_k = o\left(\frac{\|y_{k-1} - x_*\|}{\|y_{k\ominus 2} - x_*\|} \|t_k\| \right) = o(\|t_k\|), \quad \text{for } k \in \mathcal{K}'.$$

Hence $t_k = 0$ for k large in $\mathcal{K}'$. But from Proposition 3.18 (*i*), this would imply that $\alpha_k = \alpha_k^1 = 1$ for k large in $\mathcal{K}'$, which is in contradiction with property (3.39*b*) of the subsequence $\mathcal{K}'$. □

Lemma 3.26 proves point (*ii*) of Theorem 3.2. The next lemma is useful to conclude the proof of points (*iii*) and (*iv*) of this theorem, which will be done in Lemma 3.28. It claims that an estimate similar to (3.44) holds, but now for all the sequence $\mathcal{K}^c$.

Lemma 3.27. *Suppose that the hypotheses of Theorem 3.2 hold and that $\mathcal{K}^c$ is unbounded. Then $t_k = o(\|x_k - x_*\|)$ for $k \in \mathcal{K}^c$.*

Proof. Now, we know that the ideal step-size is accepted for large k, so that the conclusions of Lemma 3.23 hold with $\mathcal{K}' = \mathbb{N}^*$. Using this lemma, we have for k large in $\mathcal{K}^c$

$$\|t_k\| \le C \frac{\|y_{k-1} - x_*\|}{\|y_{k\ominus 2} - x_*\|} \|x_k - x_*\|.$$

Therefore, the result will hold if we show that

$$y_{k-1} - x_* = o(\|y_{k\ominus 2} - x_*\|), \quad \text{for } k \in \mathcal{K}^c.$$

If $\mathcal{K}$ is bounded, this estimate is clear. Therefore, we can suppose that $\mathcal{K}$ and $\mathcal{K}^c$ are unbounded. Let $\bar{r} \in (0, 1]$ be given by Lemma 3.25, let us choose $r \in (0, \bar{r}]$, and let k_r be set by Lemma 3.25. Since r is arbitrary, it is now enough to prove that

$$\|y_{k-1} - x_*\| \le r \|y_{k\ominus 2} - x_*\|, \quad \text{for large } k \in \mathcal{K}^c. \tag{3.47}$$

Note that, by Lemma 3.21, there is an index k_1' such that

$$\|y_{k+1} - x_*\| \le r \|y_k - x_*\|, \quad \text{for } k \in \mathcal{K},\ k \ge k_1'. \tag{3.48}$$

Now, there is at least one index $k_2' \in \mathcal{K}^c$ with $k_2' \ge \max(k_r, k_1' \oplus 2)$, such that for $k = k_2'$

$$\|y_{k\ominus 1} - x_*\| \le r \|y_{k\ominus 2} - x_*\|. \tag{3.49}$$

Indeed, otherwise, for k large in $\mathcal{K}^c$: $\|y_{k\ominus 1} - x_*\| > r\|y_{k\ominus 2} - x_*\|$. Due to (3.48), $k \ominus 1 > (k \ominus 2) + 1$, hence $k' = (k \ominus 2) + 1 \in \mathcal{K}^c$ and (3.49) does not hold for $k = k'$, i.e., $\|y_{k'\ominus 1} - x_*\| > r\|y_{k'\ominus 2} - x_*\|$ or $\|y_{k\ominus 2} - x_*\| > r\|y_{k\ominus 3} - x_*\|$. For a similar reason, $\|y_{k\ominus 3} - x_*\| > r\|y_{k\ominus 4} - x_*\|$. Therefore, we have shown that $\|y_{k\ominus 2} - x_*\| \geq r^2\|y_{k\ominus 4} - x_*\|$, for k large in $\mathcal{K}^c$. But then, applying Lemma 3.24 with $\eta = r^2$ would show that we do have indeed (3.49) for some large $k \in \mathcal{K}^c$.

Let us now show by induction that (3.49) is verified for all $k \in \mathcal{K}^c$, with $k \geq k_2'$. We have just seen that (3.49) holds for $k = k_2' \in \mathcal{K}^c$. Now, fix $k \in \mathcal{K}^c$ with $k > k_2'$ and suppose that (3.49) holds for any index in $[k_2', k-1] \cap \mathcal{K}^c$. We have to prove that (3.49) holds for k. In fact, one of the following situations occurs.

S_1 : $k \ominus 1 = (k \ominus 2) + 1$. As $k \ominus 2 \geq k_1' \oplus 1 \geq k_1'$, (3.49) comes from (3.48).

S_2 : $k \ominus 1 > (k \ominus 2) + 1$ and $k \ominus 1 = k_2' \ominus 1$. Then (3.49) follows from the definition of k_2'.

S_3 : $k \ominus 1 > (k \ominus 2) + 1$ and $k \ominus 1 > k_2' \ominus 1$. Then, $k' = (k \ominus 1) - 1 \in \mathcal{K}^c$ and $k_2' \leq k' < k$, so that (3.49) holds for $k = k'$ (induction hypothesis):

$$\|y_{k'\ominus 1} - x_*\| \leq r\|y_{k'\ominus 2} - x_*\|.$$

As $k' \geq k_2' \geq k_r$ and $k' \in \mathcal{K}^c$, this inequality allows us to apply Lemma 3.25 with the index k of this lemma set to k'. Then, (3.31) with $\ell = k' + 1$ gives

$$\|y_{k'+1} - x_*\| \leq r\|y_{k'\ominus 1} - x_*\|.$$

But $k' + 1 = k \ominus 1$ and $k' \ominus 1 = k \ominus 2$, so that (3.49) holds for k.

Finally, since (3.49) holds for all large $k \in \mathcal{K}^c$, inequality (3.32) in Lemma 3.25 gives (3.47). □

The next lemma concludes the proof of Theorem 3.2.

Lemma 3.28. *Suppose that the hypotheses of Theorem 3.2 hold. Then, we have*

$$y_{k+2} - x_* = o(\|y_k - x_*\|) \quad \textit{and} \quad x_{k+1} - x_* = o(\|x_k - x_*\|).$$

Proof. 1) Let us first show that $x_{k+1} - x_* = o(\|x_k - x_*\|)$. Since $\alpha_k = \alpha_k^1 = 1$ for large k, we have

$$t_k = O(\|y_k - x_*\|) = O(\|x_k - x_*\|).$$

Then from Lemma 3.19,

$$g(y_k) + B_* Z(y_k) t_k = o(\|x_k - x_*\|), \quad \text{for } k \in \mathcal{K}.$$

By Theorem 3.20, this implies that $x_{k+1} - x_* = o(\|x_k - x_*\|)$ for $k \in \mathcal{K}$.

For the indices $k \notin \mathcal{K}$, Lemma 3.27 says that $t_k = o(\|x_k - x_*\|)$. Then, by Lemma 3.23 (*ii-c*), $x_{k+1} - x_* = o(\|x_k - x_*\|)$, for $k \in \mathcal{K}^c$.

2) To prove that $y_{k+2} - x_* = o(\|y_k - x_*\|)$, observe first that

$$x_{k+1} - x_* = y_k - x_* + t_k = O(\|y_k - x_*\|).$$

Hence, using the q-superlinear convergence of $\{x_k\}$, we have

$$\begin{aligned} y_{k+2} - x_* &= x_{k+2} - x_* + r_{k+2} \\ &= O(\|x_{k+2} - x_*\|) \\ &= o(\|x_{k+1} - x_*\|) \\ &= o(\|y_k - x_*\|). \end{aligned}$$

This concludes the proof. □

References

[1] L. Armijo (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16, 1–3.

[2] R.H. Byrd (1985). An example of irregular convergence in some constrained optimization methods that use the projected Hessian. *Mathematical Programming*, 32, 232–237.

[3] R.H. Byrd (1990). On the convergence of constrained optimization methods with accurate Hessian information on a subspace. *SIAM Journal on Numerical Analysis*, 27, 141–153.

[4] R.H. Byrd, J. Nocedal (1989). A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis*, 26, 727–739.

[5] R.H. Byrd, J. Nocedal (1991). An analysis of reduced Hessian methods for constrained optimization. *Mathematical Programming*, 49, 285–323.

[6] R.H. Byrd, R.A. Tapia, Y. Zhang (1992). An SQP augmented Lagrangian BFGS algorithm for constrained optimization. *SIAM Journal on Optimization*, 2, 210–241.

[7] R.H. Byrd, Y. Xie (1993). A global and R-linear convergence analysis of the reduced Hessian successive quadratic programming algorithms. Technical report, Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado 80309.

[8] T.F. Coleman, A.R. Conn (1982). Nonlinear programming via an exact penalty function: asymptotic analysis. *Mathematical Programming*, 24, 123–136.

[9] T.F. Coleman, A.R. Conn (1984). On the local convergence of a quasi-Newton method for the nonlinear programming problem. *SIAM Journal on Numerical Analysis*, 21, 755–769.

[10] J.E. Dennis, J.J. Moré (1977). Quasi-Newton methods, motivation and theory. *SIAM Review*, 19, 46–89.

[11] J.E. Dennis, R.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs.

[12] D. Gabay (1982). Minimizing a differentiable function over a differential manifold. *Journal of Optimization Theory and Applications*, 37, 177–219.

[13] D. Gabay (1982). Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization. *Mathematical Programming Study*, 16, 18–44.

[14] J.Ch. Gilbert (1986). Une méthode à métrique variable réduite en optimisation avec contraintes d'égalité non linéaires. Rapport de Recherche 482, INRIA, BP 105, 78153 Le Chesnay, France.

[15] J.Ch. Gilbert (1988). Mise à jour de la métrique dans les méthodes de quasi-Newton réduites en optimisation avec contraintes d'égalité. *Modélisation Mathématique et Analyse Numérique*, 22, 251–288.

[16] J.Ch. Gilbert (1989). On the local and global convergence of a reduced quasi-Newton method. *Optimization*, 20, 421–450.

[17] J.Ch. Gilbert (1991). Maintaining the positive definiteness of the matrices in reduced secant methods for equality constrained optimization. *Mathematical Programming*, 50, 1–28.

[18] J.Ch. Gilbert (1993). On the realization of the Wolfe conditions in reduced quasi-Newton methods for equality constrained optimization. Rapport de Recherche 2127, INRIA, BP 105, 78153 Le Chesnay, France.

[19] W. Hoyer (1986). Variants of the reduced Newton method for nonlinear equality constrained optimization problems. *Optimization*, 17, 757–774.

[20] D.Q. Mayne, E. Polak (1982). A superlinearly convergent algorithm for constrained optimization problems. *Mathematical Programming Study*, 16, 45–61.

[21] W. Murray, M.H. Wright (1978). Projected Lagrangian methods based on the trajectories of penalty and barrier functions. Technical Report SOL-78-23, Department of Operations Research, Stanford University, Stanford, CA 94305.

[22] J. Nocedal, M.L. Overton (1985). Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 22, 821–850.

[23] S.S. Oren, D.G. Luenberger (1974). Self-scaling variable metric (SSVM) algorithms, part I: criteria and sufficient consditions for scaling a class of algorithms. *Management Science*, 20, 845–872.

[24] M.J.D. Powell (1971). On the convergence of the variable metric algorithm. *Journal of the Institute of Mathematics and its Applications*, 7, 21–36.

[25] M.J.D. Powell (1976). Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In R.W. Cottle, C.E. Lemke (editors), *Nonlinear Programming*, SIAM-AMS Proceedings 9. American Mathematical Society, Providence, RI.

[26] M.J.D. Powell (1978). Algorithms for nonlinear constraints that use Lagrangian functions. *Mathematical Programming*, 14, 224–248.

[27] M.J.D. Powell (1978). The convergence of variable metric methods for nonlinearly constrained optimization calculations. In O.L. Mangasarian, R.R. Meyer, S.M. Robinson (editors), *Nonlinear Programming 3*, pages 27–63.

[28] Y. Yuan (1985). An only 2-step Q-superlinearly convergence example for some algorithms that use reduced Hessian informations. *Mathematical Programming*, 32, 224–231.

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR_2148