MATHEMATICS OF COMPUTATION
Volume 74, Number 251, Pages 1413–1423
S 0025-5718(04)01731-4
Article electronically published on December 30, 2004

GLOBAL CONVERGENCE OF SSM FOR MINIMIZING A QUADRATIC OVER A SPHERE

WILLIAM W. HAGER AND SOONCHUL PARK

Abstract. In an earlier paper [*Minimizing a quadratic over a sphere*, SIAM J. Optim., 12 (2001), 188–208], we presented the sequential subspace method (SSM) for minimizing a quadratic over a sphere. This method generates approximations to a minimizer by carrying out the minimization over a sequence of subspaces that are adjusted after each iterate is computed. We showed in this earlier paper that when the subspace contains a vector obtained by applying one step of Newton's method to the first-order optimality system, SSM is locally, quadratically convergent, even when the original problem is degenerate with multiple solutions and with a singular Jacobian in the optimality system. In this paper, we prove (nonlocal) convergence of SSM to a global minimizer whenever each SSM subspace contains the following three vectors: (i) the current iterate, (ii) the gradient of the cost function evaluated at the current iterate, and (iii) an eigenvector associated with the smallest eigenvalue of the cost function Hessian. For nondegenerate problems, the convergence rate is at least linear when vectors (i)–(iii) are included in the SSM subspace.

1. Introduction

We consider the problem of minimizing a quadratic over a sphere:

$$\text{minimize } \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} - \mathbf{b}^{\mathsf{T}}\mathbf{x} \quad \text{subject to } \mathbf{x} \in \mathbb{R}^n, \quad \|\mathbf{x}\| = 1. \tag{1}$$

Here $\mathbf{A}$ is a symmetric n by n matrix, $\mathbf{b} \in \mathbb{R}^n$, and $\|\cdot\|$ is the Euclidean norm; the problem has been scaled so that the sphere has unit radius. When n is small, the solution to (1) can be computed from a diagonalization of $\mathbf{A}$. But when n is large, it is not practical to compute a diagonalization. The sequential subpace method (SSM), introduced in [8], is an iterative method for solving (1); at step k, the associated iterate $\mathbf{x}_k$ is chosen to solve the problem

$$\text{minimize } \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} - \mathbf{b}^{\mathsf{T}}\mathbf{x} \quad \text{subject to } \mathbf{x} \in \mathcal{S}_k, \quad \|\mathbf{x}\| = 1, \tag{2}$$

where $\mathcal{S}_k$ is a subspace of $\mathbb{R}^n$.

In [8] we show that if $\mathcal{S}_k$ includes a point obtained by applying Newton's method to the first-order optimality system at the current iterate, then SSM is locally, quadratically convergent (cubically convergent when $\mathbf{b} = \mathbf{0}$) to a solution of (1),

Received by the editor August 12, 2003 and, in revised form, March 27, 2004.
2000 *Mathematics Subject Classification.* Primary 90C20, 65F10, 65Y20.
Key words and phrases. Quadratic optimization, quadratic programming, trust region subproblem, large-scale optimization, sparse optimization.
This material is based upon work supported by the National Science Foundation under Grant No. 0203270.

even when the original problem is degenerate with multiple solutions and with a singular Jacobian in the optimality system. The iterate obtained by applying Newton's method to the first-order optimality system is known as the SQP (sequential quadratic programming) iterate. In [8] we gave numerical comparisons to algorithms appearing in [6, 16, 19] for a specific implementation of SSM where $\mathcal{S}_k$ includes not only the SQP iterate, but also

(i) an estimate of an eigenvector of $\mathbf{A}$ for the smallest eigenvalue,
(ii) the current iterate $\mathbf{x}_k$,
(iii) $\mathbf{A}\mathbf{x}_k - \mathbf{b}$, the gradient of the cost function at $\mathbf{x}_k$.

If the SQP iterate is approximated by a Lanczos process, an estimate for an eigenvector associated with the smallest eigenvalue of $\mathbf{A}$ is obtained by computing an eigenpair of the tridiagonal system associated with the Lanczos process (see [5, 14]).

In [8] we proved local quadratic convergence when $\mathcal{S}_k$ contains the SQP iterate. Numerically, in a series of test problems, convergence to the global minimum was always obtained when the vectors (i)–(iii) were included in $\mathcal{S}_k$. In this paper, we prove global convergence of the SSM when $\mathcal{S}_k$ contains the vectors (ii) and (iii), along with an eigenvector associated with the smallest eigenvalue of $\mathbf{A}$. Moreover, the convergence rate is at least linear for nondegenerate problems.

The SSM is loosely related to the finite element for solving variational problems. In the finite element method, an approximate solution is computed by carrying out the minimization over a finite dimensional subspace. Convergence to the continuous solution is obtained by increasing the dimension of the subspace. In the SSM we obtain an approximate solution to the optimization problem by restricting the minimization to a low dimensional subspace. On the other hand, convergence is not achieved by increasing the dimension of the subspace; instead, we use the solution of the subpace problem to generate an even better low dimensional subspace in which to restrict the minimization.

Other approaches to (1) include the scheme of Golub and von Matt [4] based on a partial tridiagonalization of $\mathbf{A}$ using the Lanczos process, and the related implementation of Gould et al. [6], as well as the parametric eigenvalue approaches of Sorensen [19] (further developed by Rojas in [17]), and the related scheme of Rendl and Wolkowicz [16]. Numerical comparisons between these approaches are given in [8]. The quadratic minimization problem (1) is often called the trust region subproblem since it must be solved in each step of a trust region algorithm [1, 2, 3, 12, 15] in mathematical programming. Problems of this form arise in many other applications including regularization methods for ill-posed problems [11, 20] and graph partitioning problems [9, 10].

2. Convergence

Let $\mathcal{B}$ denote the unit ball in $\mathbb{R}^n$:

$$\mathcal{B} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\},$$

let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be the cost function in (1):

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{x},$$

and let $\mathcal{L} : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$ be the Lagrangian associated with (1):

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \frac{\mu}{2}(\mathbf{x}^\mathsf{T}\mathbf{x} - 1). \tag{3}$$

(We square the constraint, when forming the Lagrangian, to remove the square root from the constraint.) Recall that $\bar{\mathbf{x}}$ is a stationary point (maximizer, minimizer, or saddle point) of (1) if $\bar{\mathbf{x}} \in \mathcal{B}$ and there exists $\bar{\mu} \in \mathbb{R}$ such that the gradient of the Lagrangian vanishes; that is,

$$(\mathbf{A} + \bar{\mu}\mathbf{I})\bar{\mathbf{x}} = \mathbf{b}. \tag{4}$$

Lemma 1. *The vector $\bar{\mathbf{x}} \in \mathbb{R}^n$ is a stationary point of* (1) *if and only if the projection of the cost gradient onto the tangent plane of the constraint is zero and $\bar{\mathbf{x}} \in \mathcal{B}$.*

Proof. This connection between stationary points and projections is well known; we include the proof since we reference it later. Since $\bar{\mathbf{x}}$ is a unit vector perpendicular to the tangent plane of $\mathcal{B}$ at $\bar{\mathbf{x}}$, the projection $\mathbf{p}$ of the cost gradient at $\bar{\mathbf{x}}$ onto the tangent plane can be expressed as

$$\mathbf{p} = \left(\mathbf{I} - \bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathsf{T}}\right)(\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) = \mathbf{A}\bar{\mathbf{x}} - \mathbf{b} + \bar{\mathbf{x}}\left(\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b} - \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\bar{\mathbf{x}}\right). \tag{5}$$

If $\bar{\mathbf{x}}$ is a stationary point and $\bar{\mu}$ satisfies (4), then $\mathbf{A}\bar{\mathbf{x}} - \mathbf{b} = -\bar{\mu}\bar{\mathbf{x}}$, and by (5) we have

$$\begin{aligned} \mathbf{p} &= -\bar{\mu}\bar{\mathbf{x}} + \bar{\mathbf{x}}\left(\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b} - \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\bar{\mathbf{x}}\right) \\ &= \bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathsf{T}}\left(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}} - \bar{\mu}\bar{\mathbf{x}}\right) = \mathbf{0}. \end{aligned}$$

Conversely, if $\mathbf{p} = \mathbf{0}$, we have from (5) that

$$\left[\mathbf{A} + \bar{\mathbf{x}}^{\mathsf{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})\mathbf{I}\right]\bar{\mathbf{x}} = \mathbf{b}.$$

Hence, for $\bar{\mu} = \bar{\mathbf{x}}^{\mathsf{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})$, (4) holds. □

Lemma 2. *If $\bar{\mathbf{x}} \in \mathcal{B}$ and $\bar{\mathbf{x}}$ is not a stationary point, then for*

$$\mathcal{S} = \text{span}\,\{\bar{\mathbf{x}}, \mathbf{A}\bar{\mathbf{x}} - \mathbf{b}\},$$

we have

$$\min_{\mathbf{x} \in \mathcal{S} \cap \mathcal{B}} f(\mathbf{x}) < f(\bar{\mathbf{x}}).$$

Proof. By Lemma 1, the projected gradient $\mathbf{p}$ in (5) does not vanish since $\bar{\mathbf{x}}$ is not a stationary point. Observe that $\mathbf{p} \in \mathcal{S}$ since it is a linear combination of $\bar{\mathbf{x}}$ and $\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}$. Defining

$$\mathbf{x}(\alpha) = \frac{\bar{\mathbf{x}} - \alpha\mathbf{p}}{\|\bar{\mathbf{x}} - \alpha\mathbf{p}\|},$$

we have $\mathbf{x}(\alpha) \in \mathcal{B}$. Moreover, $\mathbf{x}(\alpha) \in \mathcal{S}$ since $\mathbf{p}$ and $\bar{\mathbf{x}} \in \mathcal{S}$. Since $\mathbf{p}$ is perpendicular to $\bar{\mathbf{x}}$, it follows that

$$\begin{aligned} \frac{d}{d\alpha} f(\mathbf{x}(\alpha))\bigg|_{\alpha=0} &= (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})^{\mathsf{T}} \left.\frac{d\mathbf{x}(\alpha)}{d\alpha}\right|_{\alpha=0} \\ &= (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})^{\mathsf{T}}\left(\bar{\mathbf{x}}(\mathbf{p}^{\mathsf{T}}\bar{\mathbf{x}}) - \mathbf{p}\right) \\ &= -(\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})^{\mathsf{T}}\mathbf{p}. \end{aligned}$$

Moreover, since $(\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) - \mathbf{p}$ is parallel to $\bar{\mathbf{x}}$, which is perpendicular to $\mathbf{p}$, we have

$$\frac{d}{d\alpha} f(\mathbf{x}(\alpha))\bigg|_{\alpha=0} = -\|\mathbf{p}\|^2 \neq 0. \tag{6}$$

Hence, for positive α near 0,

$$f(\mathbf{x}(\alpha)) < f(\mathbf{x}(0)) = f(\bar{\mathbf{x}}).$$

□

Lemma 3. *If $\bar{\mathbf{x}}$ is a stationary point of (1) and $\bar{\mathbf{x}}$ is not a global minimizer, then for*

$$\mathcal{S} = \text{span } \{\boldsymbol{\phi}_1, \bar{\mathbf{x}}\},$$

where $\boldsymbol{\phi}_1$ is an eigenvector of $\mathbf{A}$ corresponding to the smallest eigenvalue, we have

$$\min_{\mathbf{x}\in\mathcal{S}\cap\mathcal{B}} f(\mathbf{x}) < f(\bar{\mathbf{x}}).$$

Proof. Since $\bar{\mathbf{x}}$ is a stationary point, there exists $\bar{\mu}$ satisfying (4). Expanding in a Taylor series around $\bar{\mathbf{x}}$,

$$f(\mathbf{x}) = \mathcal{L}(\mathbf{x}, \bar{\mu}) = f(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^{\mathsf{T}}(\mathbf{A} + \bar{\mu}\mathbf{I})(\mathbf{x} - \bar{\mathbf{x}}) \tag{7}$$

for any $\mathbf{x} \in \mathcal{B}$. We expand $\bar{\mathbf{x}}$ in terms of orthonormal eigenvectors $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, …, $\boldsymbol{\phi}_n$ of $\mathbf{A}$:

$$\bar{\mathbf{x}} = \sum_{i=1}^{n} \zeta_i \boldsymbol{\phi}_i. \tag{8}$$

Since $\bar{\mathbf{x}}$ is a unit vector,

$$\sum_{i=1}^{n} \zeta_i^2 = 1. \tag{9}$$

If $\zeta_1 \neq 0$, we consider the point

$$\mathbf{x} = \bar{\mathbf{x}} - 2\zeta_1\boldsymbol{\phi}_1 = -\zeta_1\boldsymbol{\phi}_1 + \sum_{i=2}^{n} \zeta_i \boldsymbol{\phi}_i,$$

which is a unit vector by (9). By (7), we have

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + 2\zeta_1^2(\lambda_1 + \bar{\mu}), \tag{10}$$

where λ_1 is the smallest eigenvalue of $\mathbf{A}$. By [18, Lemmas 2.4 and 2.8], a stationary point $\bar{\mathbf{x}}$ is the global minimizer of (1) if and only if $\mathbf{A} + \bar{\mu}\mathbf{I}$ is positive semidefinite. Since $\bar{\mathbf{x}}$ is not a global minimizer, it follows that $\mathbf{A} + \bar{\mu}\mathbf{I}$ is not positive definite or, equivalently, the smallest eigenvalue of $\mathbf{A} + \bar{\mu}\mathbf{I}$ is negative:

$$\lambda_1 + \bar{\mu} < 0. \tag{11}$$

Hence, (10) implies that $f(\mathbf{x}) < f(\bar{\mathbf{x}})$.

On the other hand, if $\zeta_1 = \bar{\mathbf{x}}^{\mathsf{T}}\boldsymbol{\phi}_1 = 0$, then we achieve a decrease in the cost function by taking

$$\mathbf{x} = k_1\boldsymbol{\phi}_1 + k_2\bar{\mathbf{x}} \tag{12}$$

for an appropriate choice of k_1 and k_2. If $\mathbf{x} \in \mathcal{B}$, k_1 and k_2 must satisfy the condition

$$\|\mathbf{x}\|^2 = \|k_1\boldsymbol{\phi}_1 + k_2\bar{\mathbf{x}}\|^2 = k_1^2 + k_2^2 = 1,$$

since $\bar{\mathbf{x}}^{\mathsf{T}}\bar{\mathbf{x}} = \boldsymbol{\phi}_1^{\mathsf{T}}\boldsymbol{\phi}_1 = 1$, and the cross product $\boldsymbol{\phi}_1^{\mathsf{T}}\bar{\mathbf{x}}$ vanishes. Hence,

$$k_1{}^2 = 1 - k_2{}^2. \tag{13}$$

By (8), (12), and the condition $\zeta_1 = 0$, we have

$$\mathbf{x} - \bar{\mathbf{x}} = k_1\boldsymbol{\phi}_1 + (k_2 - 1)\bar{\mathbf{x}} = k_1\boldsymbol{\phi}_1 + (k_2 - 1)\sum_{i=2}^{n} \zeta_i\boldsymbol{\phi}_i. \tag{14}$$

Combining (7), (13), and (14) gives

$$\begin{aligned} f(\mathbf{x}) - f(\bar{\mathbf{x}}) &= \frac{(1-k_2^2)}{2}(\lambda_1+\bar{\mu}) + \frac{(k_2-1)^2}{2}\sum_{i=2}^{n}\zeta_i^2(\lambda_i+\bar{\mu}) \\ &= (\lambda_1+\bar{\mu})R(k_2), \end{aligned} \tag{15}$$

where λ_i is the eigenvalue of $\mathbf{A}$ associated with $\boldsymbol{\phi}_i$, and

$$R(k) = \frac{(1-k^2)}{2} + \frac{(k-1)^2\sum_{i=2}^{n}\zeta_i^2(\lambda_i+\bar{\mu})}{2(\lambda_1+\bar{\mu})}.$$

Since $R(1) = 0$ and $R'(1) = -1$, it follows from (11) that $f(\mathbf{x}) < f(\bar{\mathbf{x}})$ when $k_2 < 1$ and k_2 is near 1. This completes the proof. □

Theorem 1. *If in each step of* SSM $\mathcal{S}_k$ *contains the vectors* $\mathbf{b} - \mathbf{A}\mathbf{x}_k$, $\mathbf{x}_k$, *and* $\boldsymbol{\phi}_1$, *an eigenvector associated with the smallest eigenvalue of* $\mathbf{A}$, *then* SSM *converges to a solution of* (1).

Proof. Let $\{\mathbf{x}_k\}$ be the sequence of SSM iterates. Since $\mathcal{B}$ is compact, there exists a subsequence of $\{\mathbf{x}_k\}$, denoted $\{\mathbf{y}_j\}$, converging to $\mathbf{y}^*$. Since f is continuous, $f(\mathbf{y}_j)$ converges to $f(\mathbf{y}^*)$. Let $\mathcal{T}_j$ denote the $\mathcal{S}_k$ subspace associated with the $\mathbf{y}_j$. Since $\mathbf{x}_k \in \mathcal{S}_k$ for each k and $\mathbf{x}_{k+1}$ minimizes f over $\mathcal{S}_k \cap \mathcal{B}$, it follows that $f(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k)$ for each k. Furthermore, since $\{\mathbf{y}_j\}$ is a subsequence of $\{\mathbf{x}_k\}$, $f(\mathbf{y}_{j+1}) \le f(\mathbf{y}_j)$ for each j.

Now, suppose that $\mathbf{y}^*$ is not a stationary point of (1). We introduce the projections into the tangent plane appearing in Lemma 1:

$$\begin{aligned} \mathbf{p}^* &= \left(\mathbf{I} - \mathbf{y}^*\mathbf{y}^{*\mathsf{T}}\right)(\mathbf{A}\mathbf{y}^* - \mathbf{b}), \\ \mathbf{p}_j &= \left(\mathbf{I} - \mathbf{y}_j\mathbf{y}_j^{\mathsf{T}}\right)(\mathbf{A}\mathbf{y}_j - \mathbf{b}), \end{aligned}$$

and we let $\mathcal{T}^*$ be the space of dimension at most 3 given by

$$\mathcal{T}^* = \text{span}\,\{\mathbf{y}^*, \mathbf{A}\mathbf{y}^* - \mathbf{b}, \boldsymbol{\phi}_1\}.$$

As in the proof of Lemma 2, we form points $\mathbf{y}^*(\alpha) \in \mathcal{T}^* \cap \mathcal{B}$ and $\mathbf{y}_j(\alpha) \in \mathcal{T}_j \cap \mathcal{B}$:

$$\begin{aligned} \mathbf{y}^*(\alpha) &= \frac{\mathbf{y}^* - \alpha\mathbf{p}^*}{\|\mathbf{y}^* - \alpha\mathbf{p}^*\|}, \\ \mathbf{y}_j(\alpha) &= \frac{\mathbf{y}_j - \alpha\mathbf{p}_j}{\|\mathbf{y}_j - \alpha\mathbf{p}_j\|}. \end{aligned}$$

Since $\mathbf{y}_j$ converges to $\mathbf{y}^*$, we have $\lim_{j\to\infty}\mathbf{p}_j = \mathbf{p}^*$. Since $\mathbf{y}^*$ is not a stationary point of (1), $\mathbf{p}^* \neq \mathbf{0}$ by Lemma 1. Hence, there exists a positive constant c such that $\|\mathbf{p}_j\| \ge c$ for all j sufficiently large. By (6) in the proof of Lemma 2, it follows that

$$\frac{d}{d\alpha}f(\mathbf{y}_j(\alpha))\bigg|_{\alpha=0} = -\|\mathbf{p}_j\|^2 \le -c^2.$$

Henceforth, we use primes to denote derivatives of $f(\mathbf{y}_j(\alpha))$ with respect to α. Since f is a quadratic, $f''(\mathbf{y}_j(\alpha))$ is a polynomial in α and in $1/\|\mathbf{y}_j - \alpha\mathbf{p}_j\|$ of degree at most 6. Since $\|\mathbf{y}_j\| = 1$ and $\mathbf{y}_j^{\mathsf{T}}\mathbf{p}_j = 0$, we have $\|\mathbf{y}_j - \alpha\mathbf{p}_j\|^2 = \|\mathbf{y}_j\|^2 + \alpha^2\|\mathbf{p}_j\|^2 \ge 1$. Hence, there exist M and $\epsilon > 0$ such that

$$|f''(\mathbf{y}_j(\alpha))| \le M$$

whenever $|\alpha| \le \epsilon$. Expanding in a Taylor series around $\alpha = 0$,

$$f(\mathbf{y}_j(\alpha)) = f(\mathbf{y}_j) + f'(\mathbf{y}_j)\alpha + \frac{1}{2}f''(\mathbf{y}_j(\xi))\alpha^2$$

where $0 \le \xi \le \alpha$. Consequently, when $0 \le \alpha \le \epsilon$,

$$f(\mathbf{y}_j(\alpha)) \le f(\mathbf{y}_j) - c^2\alpha + \frac{1}{2}M\alpha^2.$$

Taking $\alpha = \min\{\epsilon, c^2/M\}$, we have

$$\begin{aligned} f(\mathbf{y}_{j+1}) &\le f(\mathbf{y}_j(\alpha)) \le f(\mathbf{y}_j) + \left(\frac{1}{2}M\alpha - c^2\right)\alpha \\ &\le f(\mathbf{y}_j) - \frac{1}{2}c^2\alpha \le f(\mathbf{y}_j) - \frac{1}{2}c^2\min\{\epsilon, c^2/M\}. \end{aligned} \tag{16}$$

The relation (16) contradicts the fact that $f(\mathbf{y}_j)$ converges to $f(\mathbf{y}^*)$. Hence, our initial supposition, that $\mathbf{y}^*$ is not a stationary point, cannot be valid, and we conclude that $\mathbf{y}^*$ is a stationary point of (1).

Suppose that $\mathbf{y}^*$ is not a global minimizer of (1). Expand $\mathbf{y}^*$ in terms of the eigenvectors of $\mathbf{A}$:

$$\mathbf{y}^* = \sum_{i=1}^{n} \zeta_i \boldsymbol{\phi}_i,$$

where $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, …, $\boldsymbol{\phi}_n$ are orthonormal eigenvectors of $\mathbf{A}$. Since $\mathbf{y}^*$ is not a global minimizer of (1), f can be decreased by choosing $\mathbf{y}$ as in the proof of Lemma 3:

$$\mathbf{y} = \begin{cases} \mathbf{y}^* - 2\zeta_1\boldsymbol{\phi}_1 & \text{if } \zeta_1 \ne 0, \\ k_1\boldsymbol{\phi}_1 + k_2\mathbf{y}^* & \text{if } \zeta_1 = 0. \end{cases} \tag{17}$$

Recall that for this choice of $\mathbf{y}$, for $k_2 < 1$ near 1 and for $k_1 = \sqrt{1-k_2^2}$, we have

$$f(\mathbf{y}) - f(\mathbf{y}^*) = -\delta, \quad \delta > 0, \tag{18}$$

where

$$-\delta = \begin{cases} 2\zeta_1^2(\lambda_1 + \bar{\mu}) & \text{if } \zeta_1 \ne 0 \ \text{(see (10))}, \\ (\lambda_1 + \bar{\mu})R(k_2) & \text{if } \zeta_1 = 0 \ \text{(see (15))}. \end{cases}$$

With the same ζ_1, k_1, and k_2 associated with $\mathbf{y}$ in (17), define

$$\mathbf{z}_j = \begin{cases} \frac{\mathbf{y}_j - 2\zeta_1\boldsymbol{\phi}_1}{\|\mathbf{y}_j - 2\zeta_1\boldsymbol{\phi}_1\|} & \text{if } \zeta_1 \ne 0, \\ \frac{k_1\boldsymbol{\phi}_1 + k_2\mathbf{y}_j}{\|k_1\boldsymbol{\phi}_1 + k_2\mathbf{y}_j\|} & \text{if } \zeta_1 = 0. \end{cases} \tag{19}$$

Since $\mathbf{y}_j$ converges to $\mathbf{y}^*$ and $\|\mathbf{y}^* - 2\zeta_1\boldsymbol{\phi}_1\| = \|k_1\mathbf{y}^* + k_2\boldsymbol{\phi}_1\| = 1$, it follows that the denominators in (19) converge to 1 and $\mathbf{z}_j$ converges to $\mathbf{y}$. Since f is continuous, $f(\mathbf{z}_j)$ converges to $f(\mathbf{y})$. Choose K large enough that

$$f(\mathbf{z}_j) - f(\mathbf{y}) < \frac{\delta}{2} \ \text{ for all } j \ge K.$$

This bound combined with (18) gives

$$\begin{aligned} f(\mathbf{z}_j) - f(\mathbf{y}_j) &= f(\mathbf{z}_j) - f(\mathbf{y}) + f(\mathbf{y}) - f(\mathbf{y}^*) + f(\mathbf{y}^*) - f(\mathbf{y}_j) \\ &\le -\frac{\delta}{2} + f(\mathbf{y}^*) - f(\mathbf{y}_j). \end{aligned} \tag{20}$$

Since the $\mathbf{y}_j$ decrease monotonically to $\mathbf{y}^*$, we have $f(\mathbf{y}^*) \le f(\mathbf{y}_j)$ and by (20),

$$f(\mathbf{z}_j) - f(\mathbf{y}_j) \le -\delta/2. \tag{21}$$

Since $\mathbf{z}_j \in \mathcal{T}_j \cap \mathcal{B}$ and

$$f(\mathbf{y}_{j+1}) \leq \min_{\mathbf{x} \in \mathcal{T}_j \cap \mathcal{B}} f(\mathbf{x}),$$

we have $f(\mathbf{y}_{j+1}) \leq f(\mathbf{z}_j)$. By (21),

$$f(\mathbf{y}_{j+1}) - f(\mathbf{y}_j) \leq f(\mathbf{z}_j) - f(\mathbf{y}_j) \leq -\delta/2 < 0,$$

which contradicts the fact that $f(\mathbf{y}_j)$ converges to $f(\mathbf{y}^*)$. Hence, $\mathbf{y}^*$ is a global minimizer of (1). □

Theorem 1 combined with the results of [8] imply that if each SSM subspace contains $\boldsymbol{\phi}_1$, $\mathbf{x}_k$, $\mathbf{A}\mathbf{x}_k - \mathbf{b}$, and an SQP iterate associated with $\mathbf{x}_k$, then SSM is quadratically convergent to a global minimizer of (1) from any starting guess $\mathbf{x}_0$.

3. Linear convergence

We now establish linear convergence for nondegenerate problems, under the hypotheses of Theorem 1. Given any $\mathbf{x} \in \mathcal{B}$, let $\mu(\mathbf{x})$ be the solution of the problem:

$$\min_{\mu} \|\mathbf{b} - (\mathbf{A} - \mu\mathbf{I})\mathbf{x}\|.$$

The solution of this least squares problem is

$$\mu(\mathbf{x}) = \mathbf{x}^{\mathsf{T}}(\mathbf{b} - \mathbf{A}\mathbf{x}).$$

Define $\mu_k = \mu(\mathbf{x}_k)$ and

$$\mathbf{p}_k = (\mathbf{A} + \mu_k\mathbf{I})\mathbf{x}_k - \mathbf{b}.$$

Referring to (5), we see that $\mathbf{p}_k$ is the projection of the gradient $\mathbf{A}\mathbf{x}_k - \mathbf{b}$ at the current iterate onto the tangent plane of the constraint.

Lemma 4. *If in each step of* SSM $\mathcal{S}_k$ *contains the vectors* $\mathbf{A}\mathbf{x}_k - \mathbf{b}$ *and* $\mathbf{x}_k$*, then there exist* $\tau > 0$*, independent of* k*, such that*

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \tau\|\mathbf{p}_k\|^2. \tag{22}$$

Proof. Consider the vector

$$\mathbf{x}(\alpha) = \frac{\mathbf{x}_k - \alpha\mathbf{q}_k}{\|\mathbf{x}_k - \alpha\mathbf{q}_k\|}, \quad \mathbf{q}_k = \mathbf{p}_k/\|\mathbf{p}_k\|.$$

Similarly to (6), we have

$$\left.\frac{d}{d\alpha} f(\mathbf{x}_k(\alpha))\right|_{\alpha=0} = -\|\mathbf{p}_k\|.$$

Choose M large enough that

$$\left|\frac{d^2}{d\alpha^2} f(\mathbf{x}(\alpha))\right| \leq M$$

for all possible choices of $\mathbf{x}_k \in \mathcal{B}$ and for all $\alpha \in [0, 1/2]$. By a Taylor expansion around $\alpha = 0$,

$$f(\mathbf{x}(\alpha)) \leq f(\mathbf{x}_k) - \|\mathbf{p}_k\|\alpha + \frac{1}{2}M\alpha^2, \tag{23}$$

whenever $0 \leq \alpha \leq 1/2$. The quadratic in (23) is minimized by taking $\alpha = \|\mathbf{p}_k\|/M$. Since $\|\mathbf{p}_k\|$ is uniformly bounded over all choices of $\mathbf{x}_k \in \mathcal{B}$, it follows that M can be chosen large enough to ensure that $\|\mathbf{p}_k\|/M \leq 1/2$. With the choice $\alpha = \|\mathbf{p}_k\|/M$ in (23), we have

$$f(\mathbf{x}(\alpha)) \leq f(\mathbf{x}_k) - \|\mathbf{p}_k\|^2/(2M),$$

which corresponds to (22) and $\tau = 1/(2M)$. □

Recall that the stationary points of (1) consist of pairs $(\mu, \mathbf{x}) \in \mathbb{R} \times \mathcal{B}$ satisfying the equation

$$(\mathbf{A} + \mu\mathbf{I})\mathbf{x} = \mathbf{b}. \tag{24}$$

By [18, Lemmas 2.4 and 2.8], the global minimizers of (1) correspond to the unique multiplier $\mu = \nu_1$ satisfying (24) for some $\mathbf{x} \in \mathcal{B}$ with $\lambda_1 + \nu_1 \geq 0$. Let $\mathcal{S}^*$ denote the set of solutions to (24) associated with $\mu = \nu_1$. By Theorem 1, the SSM iterates $\mathbf{x}_k$ converge to $\mathcal{S}^*$. Since $\mu(\mathbf{x}) = \nu_1$ for all $\mathbf{x} \in \mathcal{S}^*$, it follows that μ_k converges to ν_1 as $k \to \infty$. We now prove linear convergence of $\mathbf{x}_k$ when the global minimizer is unique (more precisely, when $\lambda_1 + \nu_1 > 0$).

Theorem 2. *Suppose that in each step of* SSM $\mathcal{S}_k$ *contains the vectors* $\mathbf{A}\mathbf{x}_k - \mathbf{b}$, $\mathbf{x}_k$, *and* $\boldsymbol{\phi}_1$, *an eigenvector associated with the smallest eigenvalue of* $\mathbf{A}$. *If* $\nu_1 + \lambda_1 > 0$ *and* σ *and* K *are chosen so that* $\mu_k + \lambda_1 \geq \sigma > 0$ *for all* $k \geq K$, *then we have*

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_A \leq \left(1 - \frac{\tau\sigma}{\lambda_n + \nu_1}\right)^{1/2} \|\mathbf{x}_k - \mathbf{x}^*\|_A$$

for each $k \geq K$, *where* $\mathbf{x}^*$ *is the unique global minimizer of* (1), τ *is the parameter given in Lemma* 4, λ_1 *and* λ_n *are the smallest and largest eigenvalues of* $\mathbf{A}$, *respectively, and the norm* $\|\cdot\|_A$ *is defined by*

$$\|\mathbf{x}\|_A^2 = \mathbf{x}^\mathsf{T}(\mathbf{A} + \nu_1\mathbf{I})\mathbf{x}.$$

Proof. Expanding the Lagrangian in a Taylor series around $\mathbf{x}_k$, we see that for any $\mathbf{x} \in \mathcal{B}$,

$$f(\mathbf{x}) = \mathcal{L}(\mathbf{x}, \mu_k) = f(\mathbf{x}_k) + \mathbf{p}_k^\mathsf{T}(\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^\mathsf{T}(\mathbf{A} + \mu_k\mathbf{I})(\mathbf{x} - \mathbf{x}_k).$$

Since $f(\mathbf{x}_k) \geq f(\mathbf{x}^*)$, it follows that

$$\frac{1}{2}(\mathbf{x}^* - \mathbf{x}_k)^\mathsf{T}(\mathbf{A} + \mu_k\mathbf{I})(\mathbf{x}^* - \mathbf{x}_k) \leq -\mathbf{p}_k^\mathsf{T}(\mathbf{x}^* - \mathbf{x}_k). \tag{25}$$

Using the relation $\lambda_1 + \mu_k \geq \sigma$ on the left side of (25), we have

$$\frac{\sigma}{2}\|\mathbf{x}^* - \mathbf{x}_k\|^2 \leq \|\mathbf{p}_k\|\|\mathbf{x}^* - \mathbf{x}_k\|,$$

which gives

$$\|\mathbf{x}^* - \mathbf{x}_k\| \leq \frac{2}{\sigma}\|\mathbf{p}_k\|.$$

Inserting this bound for $\mathbf{p}_k$ in (22) and exploiting the relation

$$\|\mathbf{x}\|_A^2 \leq (\lambda_n + \nu_1)\|\mathbf{x}\|^2$$

yields

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) - \frac{\tau\sigma}{2}\|\mathbf{x}^* - \mathbf{x}_k\|^2 \\ &\leq f(\mathbf{x}_k) - \frac{\tau\sigma}{2(\lambda_n + \nu_1)}\|\mathbf{x}^* - \mathbf{x}_k\|_A^2. \end{aligned} \tag{26}$$

Expanding in a Taylor series around $\mathbf{x}^*$, we obtain

$$f(\mathbf{x}) = \mathcal{L}(\mathbf{x}, \nu_1) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\mathsf{T}(\mathbf{A} + \nu_1\mathbf{I})(\mathbf{x} - \mathbf{x}^*).$$

Combining this with (26), the proof is complete. □

Theorem 2 gives a linear convergence estimate for k sufficiently large. For smaller k, the convergence analysis is more complicated since the iterates may begin to converge towards a saddle point, before sliding off the saddle and continuing their progress to the global minimum. Estimates for the decrease in cost when the iterates are far from the minimum can be obtained using expansions similar to those in Lemma 3. For example, if $\zeta_1 = \mathbf{x}_k^\mathsf{T}\boldsymbol{\phi}_1$, then it can be shown that

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - 2\zeta_1 \mathbf{p}_k^\mathsf{T}\boldsymbol{\phi}_1 + 2(\lambda_1 + \mu_k)\zeta_1^2. \tag{27}$$

Lemma 4 gives an estimate for the cost decay when $\mathbf{p}_k$ is large. When $\|\mathbf{p}_k\|$ is small, μ_k must be close to the discrete set of multipliers satisfying (24), and in this case, (27) can be used to estimate the decay in the cost function.

4. Numerical illustration

To illustrate the convergence results, we consider the first test problem in [19] in which $\mathbf{A} = \mathbf{A}_0 - 5\mathbf{I}$, where $\mathbf{A}_0$ is the standard 2D discrete Laplacian on the unit square based on a 5-point stencil with equally spaced mesh points. The radius of the sphere constraint is 100, and $\mathbf{b}$ is a vector with elements uniformly distributed on $[0, 1]$. Our starting guess is $\mathbf{x}_0 = \mathbf{b}(100/\|\mathbf{b}\|)$. The solid curve in Figure 1 corresponds to the choice $\mathcal{S}_k = \text{span}\,\{\mathbf{x}_k, \mathbf{p}_k, \boldsymbol{\phi}_1\}$. In the dashed curve, we augment $\mathcal{S}_k$ with the SQP vector $\mathbf{z}$ obtained by solving the following linear system for the unknowns $\mathbf{z}$ and ν:

$$\begin{aligned} (\mathbf{A} + \mu_k\mathbf{I})\mathbf{z} + \mathbf{x}_k\nu &= \mathbf{b} - (\mathbf{A} + \mu_k\mathbf{I})\mathbf{x}_k, \\ \mathbf{x}_k^\mathsf{T}\mathbf{z} &= 0. \end{aligned}$$

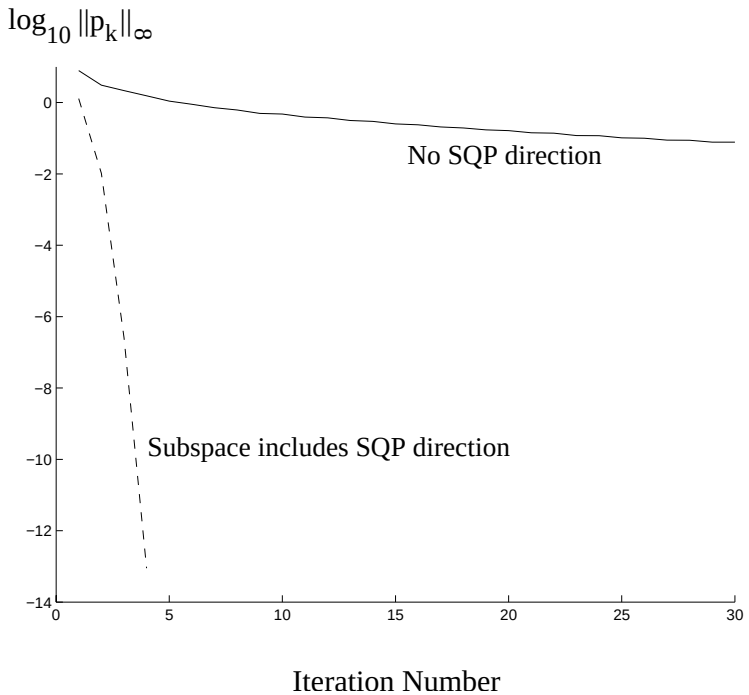

Figure 1. $\log_{10}\|\mathbf{p}_k\|_\infty$ versus iteration, $\|\cdot\|_\infty$ = max absolute component

As seen in Figure 1, including the SQP vector in the subspace yields much faster convergence; in [8] we prove quadratic convergence when the SQP vector is included in the subspace, in contrast to the linear convergence established in this paper. Note though that the SQP iteration by itself typically converges to a stationary point. Hence, it is important to include the additional vectors $\mathbf{x}_k$, $\mathbf{p}_k$, and $\boldsymbol{\phi}_1$ in the subspace to ensure convergence to the global minimum. SSM can be implemented efficiently using the iterative techniques of [7] to approximately solve the SQP system (see [8]).

Acknowledgments

The authors gratefully acknowledge the constructive comments and suggestions of the associate editor and the editor, which led to the addition of Sections 3 and 4.

References

[1] R. H. Byrd, R. B. Schnabel, and G. A. Schultz, *A trust region algorithm for nonlinearly constrained optimization*, SIAM J. Numer. Anal., **24** (1987), pp. 1152–1170. MR0909071 (89f:65069)

[2] M. Celis, J. E. Dennis, and R. A. Tapia, *A trust region strategy for nonlinear equality constrained optimization*, in Numerical Optimization 1984, SIAM, Philadelphia, PA, 1985, pp. 71–82. MR0802084 (87c:90175)

[3] M. El-Alem, *Celis-Dennis-Tapia trust region algorithm for constrained optimization*, SIAM J. Numer. Anal., **28** (1991), pp. 266–290. MR1083336 (91k:90161)

[4] G. H. Golub and U. von Matt, *Quadratically constrained least squares and quadratic problems*, Numer. Math., **59** (1991), pp. 561–580. MR1124128 (92f:65049)

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989. MR1002570 (90d:65055)

[6] N. I. M. Gould, S. Lucidi, M. Roma, and P. L. Toint, *Solving the trust-region subproblem using the Lanczos Method*, SIAM J. Optim, **9** (1999), pp. 504–525. MR1686795 (2000b:90046)

[7] W. W. Hager, *Iterative methods for nearly singular linear systems*, SIAM J. Sci. Comp., **22** (2000), pp. 747–766. MR1780623 (2001g:65031)

[8] W. W. Hager, *Minimizing a quadratic over a sphere*, SIAM J. Optim., **12** (2001), pp. 188–208. MR1870591 (2002m:90054)

[9] W. W. Hager and Y. Krylyuk, *Graph partitioning and continuous quadratic programming*, SIAM J. Discrete Math., **12** (1999), pp. 500–523. MR1720400 (2000k:90066)

[10] W. W. Hager and Y. Krylyuk, *Multiset graph partitioning*, Math. Methods Oper. Res., **55** (2002), pp. 1–10. MR1892714 (2003k:90079)

[11] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego, 1989.

[12] J. J. Moré, *Recent developments in algorithms and software for trust region methods*, in A. Bachem, M. Grotschel, and B. Korte, editors, Mathematical Programming: State of the Art, Springer-Verlag, Berlin, 1983, pp. 258–287. MR0717404 (85b:90066)

[13] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Stat. Comput., **4** (1983), pp. 553–572. MR0723110 (86b:65063)

[14] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliff, NJ, 1980. MR0570116 (81j:65063)

[15] M. J. D. Powell and Y. Yuan, *A trust region algorithm for equality constrained optimization*, Math. Programming, **49** (1991), pp. 189–211. MR1087453 (91m:90162)

[16] R. Rendl and H. Wolkowicz, *A semidefinite framework for trust region subproblems with applications to large scale minimization*, Math. Programming, **77** (1997), pp. 273–299. MR1461384 (98i:90063)

[17] M. Rojas, *A Large-scale Trust-region Approach to the Regularization of Discrete Ill-posed Problems*, PhD Dissertation, Computational and Applied Mathematics, Rice University, Houston, TX, May, 1998.

[18] D. C. Sorensen, *Newton's method with a model trust region modification*, SIAM J. Numer. Anal., **16** (1982), pp. 409–426. MR0650060 (84h:49061)
[19] D. C. Sorensen, *Minimization of a large-scale quadratic function subject to a spherical constraint*, SIAM J. Optim., **7** (1997), pp. 141–161. MR1430561 (98b:90102)
[20] A. Tarantola, *Inverse Problem Theory*, Elsevier, Amsterdam, 1987. MR0930881 (89b:65007)

Department of Mathematics, P.O. Box 118105, University of Florida, Gainesville, Florida 32611-8105
E-mail address: hager@math.ufl.edu
URL: http://www.math.ufl.edu/~hager

Department of Mathematics, P.O. Box 118105, University of Florida, Gainesville, Florida 32611-8105
E-mail address: scp@math.ufl.edu
URL: http://www.math.ufl.edu/~scp