Convergence of a short-step primal-dual algorithm based on the Gauss-Newton direction

Serge Kruk
sgkruk@acm.org
Department of Mathematics and Statistics
Oakland University
Rochester, MI, 48309

Henry Wolkowicz
hwolkowi@orion.math.uwaterloo.ca
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

April 9, 2003

Key words: Interior-point method, Semidefinite programming, Gauss-Newton.
MSC: 65K05,90C51,90C22.

Abstract

This short note proves the theoretical convergence of a short step, approximate path following, interior-point primal-dual algorithm for semidefinite programs based on the Gauss-Newton direction obtained from minimizing the norm of the perturbed optimality conditions. This is the first proof of convergence for the Gauss-Newton direction in this context. It assumes strict complementarity and uniqueness of the optimal solution as well as an estimate of the smallest singular value of the Jacobian.

1 The Gauss-Newton direction

The purpose of this short note is to develop a convergence proof for an infeasible interior-point algorithm based on the Gauss-Newton direction introduced in [3]. This is the first proof of convergence for this direction although an algorithm based on a projected and scaled Gauss-Newton direction was demonstrated in [1]. The approach is novel in that the proof relies only on classical results of nonlinear optimization. As a result, the iterates are not explicitly maintained feasible, nor even positive definite; rather, we maintain the weaker condition that the Jacobian of the optimality conditions is full rank. Moreover, our measure of distance to the central path combines feasibility and complementarity. The main result appears in Theorem 3.3.

The problem of interest is the semidefinite program pair

$$\text{Primal} \qquad \min\left\{\langle C, X\rangle \mid \mathcal{A}(X) = b, X \in \mathbb{S}^n_+\right\}, \tag{1}$$

and

$$\text{Dual} \qquad \max\left\{\langle b, y\rangle \mid \mathcal{A}^*(y) + Z = C, Z \in \mathbb{S}^n_+\right\}, \tag{2}$$

where $b \in \mathbb{R}^m$, $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$ is the vector space of symmetric matrices of order n equipped with the inner product $\langle X, Y\rangle := \operatorname{trace}(XY)$. For $M, N \in \mathbb{R}^{m\times n}$, the inner product is $\langle M, N\rangle := \operatorname{trace}(M^tN)$ and the corresponding (Frobenius) matrix norm is denoted $\|M\| = \|M\|_F = \sqrt{\operatorname{trace}(M^tM)}$. The operator $\mathcal{A}$ is linear and defined as

$$\mathcal{A}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix},$$

for matrices $A_1, \ldots, A_m \in \mathbb{S}^n$. Finally, $\mathbb{S}^n_+$ represents the cone of positive semidefinite matrices and $\mathbb{S}^n_{++}$, of positive definite matrices.

We assume the existence of a point (X_0, y_0, Z_0) such that

$$X_0 \in \mathbb{S}^n_{++}, \tag{3a}$$
$$Z_0 \in \mathbb{S}^n_{++}, \tag{3b}$$
$$\mathcal{A}(X_0) = b, \tag{3c}$$
$$\mathcal{A}^*(y_0) + Z_0 = C. \tag{3d}$$

If such a point exists, it is well-known that both the primal and dual problems have optimal solutions and that the optimal values are equal. We write the perturbed optimality conditions for the primal-dual pair (1-2) as a function of a continuation parameter $\mu \geq 0$,

$$\mathcal{A}^*(y) + Z - C = 0, \tag{4a}$$
$$\mathcal{A}(X) - b = 0, \tag{4b}$$
$$ZX - \mu I = 0, \tag{4c}$$
$$X, Z \in \mathbb{S}^n_+. \tag{4d}$$

To simplify the statements of the algorithm and of the following results we define

$$F_\mu(X, y, Z) := \begin{bmatrix} \mathcal{A}^*(y) + Z - C \\ \mathcal{A}(X) - b \\ ZX - \mu I \end{bmatrix}, \qquad \text{(central path defining function)} \tag{5a}$$

$$F_{\tau\mu}(X, y, Z) := \begin{bmatrix} \mathcal{A}^*(y) + Z - C \\ \mathcal{A}(X) - b \\ ZX - \tau\mu I \end{bmatrix}, 0 < \tau < 1. \qquad \text{(merit function)} \tag{5b}$$

Assumptions 1.1 *The following assumptions hold throughout the paper.*

- *There is a point* (X_0, y_0, Z_0) *satisfying conditions (3).*

- *The operator $\mathcal{A}$ is surjective.*
- *The optimal solution to the primal-dual pair (1-2) is unique and satisfies strict complementarity (i.e. $Z + X \in \mathbb{S}^n_{++}$).*

Under Assumptions 1.1, for every $\mu > 0$, there is a unique solution in $\mathbb{S}^n_{++} \times \mathbb{R}^m \times \mathbb{S}^n_{++}$ to $F_\mu(X, y, Z) = 0$, which we denote (X_μ, y_μ, Z_μ). This set of solutions is called the *central path*. The limit point of the central path corresponding to $\mu \to 0$ is the solution of the semidefinite pair (1-2).

The algorithm described in this paper approximately follows the central path by attempting to solve $F_\mu(X, y, Z) = 0$ for decreasing values of μ. This is common to all path-following algorithms. The novelty of the approach described here is to treat this approximation subproblem as a nonlinear equation and to apply classical tools.

One major difference from standard practice resulting from this point of view is the relation between the iterates and the barrier parameter: The scalar μ is not updated using the iterates as is usually the case ($\mu = \tau\frac{\langle Z,X\rangle}{n}$), but rather it is reduced by a factor $\tau < 1$ at every step ($\mu \leftarrow \tau\mu$). In consequence, the initial point (X_0, y_0, Z_0) depends on μ_0, rather than the reverse. Another important difference is that no attempt is made to dampen the step to maintain the iterates within the cone of positive definite matrices. The algorithm only maintains the weaker full rank condition on the Jacobian.

The function F_μ is nonlinear. We can find its zeroes by transforming the problem into minimizing the Frobenius norm, namely

$$\min \|F_\mu(X, y, Z)\|^2 \quad \left(= \|\mathcal{A}^*(y) + Z - C\|_F^2 + \|\mathcal{A}(X) - b\|^2 + \|ZX - \mu I\|_F^2\right),$$

to which we apply the Gauss-Newton method: from a well-centered point (X, y, Z) with initial $\mu > 0$, we fix a target $\tau\mu$ for some $\tau \in (0, 1)$ and reduce $\|F_{\tau\mu}(X, y, Z)\|$ by finding the least squares solution of the Gauss-Newton equation namely the least squares solution of

$$F'_{\tau\mu}(X, y, Z) \begin{pmatrix} dX \\ dy \\ dZ \end{pmatrix} = \begin{bmatrix} \mathcal{A}^*(dy) + dZ \\ \mathcal{A}(dX) \\ ZdX + dZX \end{bmatrix} = -F_{\tau\mu}(X, y, Z) \tag{6}$$

for a direction (dX, dy, dZ). We use this direction as the step to obtain the next iterate. In more detail, see Algorithm 1. We explain later the requirement on the initial point and the choice of τ. Note that the Jacobian $F'_\mu(X, y, Z) : \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n \to \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$ where $\|F'_\mu(X, y, Z)\|$ is the operator norm on the underlying vector space.

The following result which, shown elsewhere [3], is stated here for convenience.

Lemma 1.1 *Under Assumptions (1.1), the Jacobian $F'_\mu(X, y, Z)$ is full column rank for all $X \in \mathbb{S}^n_{++}$, $Z \in \mathbb{S}^n_{++}$. Moreover it is full column rank at the optimal solution of (1-2).* □

For the sake of simplifying the expressions throughout, we define for any subscript ξ,

$$s_\xi := (X_\xi, y_\xi, Z_\xi), \qquad ds := (dX, dy, dZ).$$

We also define canonical central path points s_μ and $s_{\tau\mu}$ such that

$$F_\mu(s_\mu) = 0, \qquad F_{\tau\mu}(s_{\tau\mu}) = 0.$$

Algorithm 1 Gauss-Newton infeasible short-step

Given $\mu_0 > 0$	{Initial barrier parameter}
Given $\epsilon > 0$	{Merit function tolerance}
Find X_0, y_0, Z_0	{Must satisfy (21)}
$X = X_0, y = y_0, Z = Z_0$	{Initial iterate}
$\mu = \mu_0$	{Initial barrier parameter}
Choose $0 < \tau < 1$	{Chosen according to (18)}
while $\max\{\tau\mu, \|F_{\tau\mu}(s)\|\} > \epsilon$ **do**	
Find least squares solution of $[F'_{\tau\mu}(s)]ds = -F_{\tau\mu}(s)$	{Gauss-Newton direction}
$X = X + dX, y = y + dy, Z = Z + dZ$	{Update iterate}
Recompute $\mu \leftarrow \tau\mu$	{Update target}
end while	

2 Merit function and central path

This section describes some relations between the value of our chosen merit function $\|F_{\tau\mu}\|$ and the distance of the iterate to the central path. Note that we do not assume that the iterates are primal or dual feasible. Our measure of distance to the central path combines estimates of both infeasibility and complementarity. The section also describes the progress of the Gauss-Newton direction in minimizing $\|F_{\tau\mu}\|$. The results are of a technical nature and used as building blocks of the convergence proof given in the next section.

We begin this section with a well known result about approximations of inverses, often referred to as the Banach Lemma. For a proof see [2].

Lemma 2.1 *Suppose* $M \in \mathbb{R}^{n\times n}$ *and* $\|M\| < 1$. *Then* $I - M$ *is nonsingular and*

$$\|(I-M)^{-1}\| \le \frac{1}{1-\|M\|}.$$

$\square$

Since the Gauss-Newton direction is obtained from an overdetermined system of equations, pseudo-inverses allow succinct expressions of the solution. Namely, the least squares solution to $[F'_{\tau\mu}(s)]ds = -F_{\tau\mu}(s)$ is $ds = -[F'_{\tau\mu}(s)]^{\dagger}F_{\tau\mu}(s)$, where $(\cdot)^{\dagger}$ indicates the Moore-Penrose inverse.

To generalize to Gauss-Newton some results well-known about Newton's method we require a bound on the norm of the pseudo-inverse.

Lemma 2.2 *Suppose that* $A \in \mathbb{R}^{m\times n}$ *and* $B \in \mathbb{R}^{n\times m}$, *where* $m \ge n$; *and assume that* BA *is non-singular. Then*

$$\|A^{\dagger}\| \le \|(BA)^{-1}B\|.$$

Proof. Define the singular value decompositions $A = U_A\Sigma_A V_A^t$ and $B = U_B\Sigma_B V_B^t$ and let $\overline{\Sigma}_A, \overline{\Sigma}_B$

be the nonzero diagonal blocks of, respectively, Σ_A and Σ_B. Then

$$\begin{aligned} \|(BA)^{-1}B\| &= \|(U_B\Sigma_B V_B^t U_A \Sigma_A V_A^t)^{-1} U_B \Sigma_B V_B^t\| \\ &= \|(U_B[\overline{\Sigma}_B 0] V_B^t U_A \begin{bmatrix} \overline{\Sigma}_A \\ 0 \end{bmatrix} V_A^t)^{-1} U_B [\overline{\Sigma}_B 0] V_B^t\| \\ &= \|(U_B[\overline{\Sigma}_B 0] \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \begin{bmatrix} \overline{\Sigma}_A \\ 0 \end{bmatrix} V_A^t)^{-1} U_B [\overline{\Sigma}_B 0] V_B^t\| \\ &= \|V_A \overline{\Sigma}_A^{-1} Q_1^{-1} \overline{\Sigma}_B^{-1} U_B^t U_B \overline{\Sigma}_B\| \\ &= \|\overline{\Sigma}_A^{-1} Q_1^{-1} \overline{\Sigma}_B^{-1} \overline{\Sigma}_B\| \\ &= \|\overline{\Sigma}_A^{-1} Q_1^{-1}\|. \end{aligned}$$

Since $V_B^t U_A := \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix}$ is orthogonal we have $Q_1^t Q_1 + Q_3^t Q_3 = I$ and therefore $I \succeq Q_1^t Q_1$. This implies that all the singular values of Q_1 are at most 1; and all the singular values of Q_1^{-1} are at least 1. Therefore

$$\|\overline{\Sigma}_A^{-1} Q_1^{-1}\| \geq \|\overline{\Sigma}_A^{-1}\| = \|A^\dagger\|,$$

the required bound on the norm of the Moore-Penrose inverse. □

From Lemma 2.1 and Lemma 2.2, we can obtain the following result about approximation of pseudo-inverses.

Lemma 2.3 *Suppose that $\overline{A}$ is an approximation to the pseudo-inverse of A in the sense that $\|I - \overline{A}A\| < 1$. Then*

$$\|A^\dagger\| \leq \frac{\|\overline{A}\|}{1 - \|I - \overline{A}A\|}.$$

Proof. Consider that $\|I - \overline{A}A\| < 1$ is the required condition of Lemma 2.1. Therefore we can write

$$\|A^\dagger\| \leq \|(\overline{A}A)^{-1}\overline{A}\| \leq \|(\overline{A}A)^{-1}\|\|\overline{A}\| \leq \frac{\|\overline{A}\|}{1 - \|I - \overline{A}A\|},$$

where the first inequality is obtained from Lemma 2.2. □

Essentially from this bound on the norm of approximate pseudo-inverses we can establish a relation between the distance to the central path of an iterate (X, y, Z) and the current value of our merit function $\|F_{\tau\mu}(X, y, Z)\|$. To simplify the result we first establish Lipschitz continuity of the first derivative.

Lemma 2.4 *The operator $F'_{\tau\mu}(s)$ is Lipschitz continuous with constant 1.*

Proof. Direct calculations yield

$$\begin{aligned}
\|F'_{\tau\mu}(s+ds) - F'_{\tau\mu}(s)\| &= \left\| \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ d\mathcal{Z} & 0 & d\mathcal{X} \end{bmatrix} \right\| \\
&= \max_{\|s\|=1} \{\|d\mathcal{Z}s_x + d\mathcal{X}s_z\|\} \\
&\leq \max_{\|s\|=1} \{\|d\mathcal{Z}s_x\| + \|d\mathcal{X}s_z\|\} \\
&\leq \max_{\|s\|=1} \{\|dZ\|\|s_x\| + \|s_z\|\|dX\|\} \\
&\leq \|dZ\| + \|dX\| \\
&\leq \|ds\|.
\end{aligned}$$

Hence a constant of 1 will suffice. □

Lemma 2.5 *Under Assumptions 1.1, there is a* $\delta > 0$ *so that for all* s *such that* $\|s - s_{\tau\mu}\| < \delta$*,*

$$\|F'_{\tau\mu}(s)\| \leq 2\|F'_{\tau\mu}(s_{\tau\mu})\|, \tag{7a}$$

$$\|F'_{\tau\mu}(s)^\dagger\| \leq 2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|, \tag{7b}$$

$$\frac{\|s - s_{\tau\mu}\|}{2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|} \leq \|F_{\tau\mu}(s)\|, \tag{7c}$$

$$\|F_{\tau\mu}(s)\| \leq 2\|F'_{\tau\mu}(s_{\tau\mu})\|\|s - s_{\tau\mu}\|. \tag{7d}$$

Moreover, we can choose any δ *satisfying*

$$\delta < \frac{\sigma_{\min}}{2}, \tag{8}$$

where $\sigma_{\min}$ *denotes the smallest singular value of* $F'_{\tau\mu}(s_{\tau\mu})$*.*

Proof. Since $F'_{\tau\mu}$ is Lipschitz continuous with constant 1,

$$\|F'_{\tau\mu}(s)\| \leq \|F'_{\tau\mu}(s_{\tau\mu})\| + \|s - s_{\tau\mu}\|.$$

Take δ small enough so that

$$\delta < \|F'_{\tau\mu}(s_{\tau\mu})\| \tag{9}$$

to obtain (7a). For the second result (7b), take δ small enough so that

$$\delta < \frac{1}{2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|}, \tag{10}$$

which implies $\|s - s_{\tau\mu}\| \leq \frac{1}{2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|}$. Now we can write (since $[F'_{\tau\mu}(s_{\tau\mu})]^\dagger F'_{\tau\mu}(s_{\tau\mu}) = I$),

$$\begin{aligned}
\|I - F'_{\tau\mu}(s_{\tau\mu})^\dagger F'_{\tau\mu}(s)\| &= \|F'_{\tau\mu}(s_{\tau\mu})^\dagger [F'_{\tau\mu}(s_{\tau\mu}) - F'_{\tau\mu}(s)]\| \\
&\leq \|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|\|F'_{\tau\mu}(s_{\tau\mu}) - F'_{\tau\mu}(s)\| \\
&\leq \|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|\|s_{\tau\mu} - s\| \\
&\leq \frac{\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|}{2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|}.
\end{aligned}$$

From the last inequality we get

$$||I - F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F'_{\tau\mu}(s)|| \leq \frac{1}{2}. \tag{11}$$

Then, from Lemma 2.3 with the identification $A = F'_{\tau\mu}(s)$ and $\overline{A} = F'_{\tau\mu}(s_{\tau\mu})^{\dagger}$, and from (11) we obtain

$$||F'_{\tau\mu}(s)^{\dagger}|| \leq \frac{||F'_{\tau\mu}(s_{\tau\mu})^{\dagger}||}{1 - ||I - F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F_{\tau\mu}(s)||} \leq 2||F'_{\tau\mu}(s_{\tau\mu})^{\dagger}||,$$

our second required inequality. For the third inequality (7c), we use the Fundamental theorem of calculus to express

$$F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F_{\tau\mu}(s) = F'_{\tau\mu}(s_{\tau\mu})^{\dagger} \int_0^1 F'_{\tau\mu}(s_{\tau\mu} + t(s - s_{\tau\mu}))(s - s_{\tau\mu})dt.$$

Take norms on both sides to get

$$\begin{aligned}
||F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F_{\tau\mu}(s)|| &= ||F'_{\tau\mu}(s_{\tau\mu})^{\dagger} \int_0^1 F'_{\tau\mu}(s_{\tau\mu} + t(s - s_{\tau\mu}))(s - s_{\tau\mu})dt|| \\
&= ||(s - s_{\tau\mu}) - \int_0^1 [I - F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F'_{\tau\mu}(s_{\tau\mu} + t(s - s_{\tau\mu}))](s - s_{\tau\mu})dt|| \\
&\geq ||s - s_{\tau\mu}|| - \int_0^1 ||I - F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F'_{\tau\mu}(s_{\tau\mu} + t(s - s_{\tau\mu}))||||s - s_{\tau\mu}||dt \\
&\geq ||s - s_{\tau\mu}|| - ||s - s_{\tau\mu}||\frac{1}{2} \\
&= ||s - s_{\tau\mu}||\frac{1}{2}.
\end{aligned}$$

Therefore

$$\frac{||s - s_{\tau\mu}||}{2} \leq ||F'_{\tau\mu}(s_{\tau\mu})^{\dagger} F_{\tau\mu}(s)|| \leq ||F'_{\tau\mu}(s_{\tau\mu})^{\dagger}||||F_{\tau\mu}(s)||.$$

The fourth inequality (7d), is obtained similarly. We use the assumption $F_{\tau\mu}(s_{\tau\mu}) = 0$ and the bound (7a) to get

$$\begin{aligned}
||F_{\tau\mu}(s)|| &\leq \int_0^1 ||F'_{\tau\mu}(s_{\tau\mu} + t(s - s_{\tau\mu}))||||s - s_{\tau\mu}||dt \\
&\leq \int_0^1 2||F'_{\tau\mu}(s_{\tau\mu})||||s - s_{\tau\mu}||dt \\
&= 2||F'_{\tau\mu}(s_{\tau\mu})||||s - s_{\tau\mu}||.
\end{aligned}$$

Now we need to restrict δ using (9) and (10). Take

$$\delta = \min\left\{\frac{1}{2||F'_{\tau\mu}(s_{\tau\mu})^{\dagger}||}, ||F'_{\tau\mu}(s_{\tau\mu})||\right\} = \frac{\sigma_{\min}}{2}$$

to complete the result. □

Corollary 2.6 *Suppose that the hypotheses of Lemma 2.5 hold. Then for all* s *as defined in the Lemma,* $F'_{\tau\mu}(s)$ *is full column rank.*

Proof. From (7b), we see that the smallest nonzero singular value of $F'_{\tau\mu}(s)$ is bounded below on the entire neighbourhood about $s_{\tau\mu}$. Therefore, no nonzero singular value can approach 0. □

From these relations between the central path and our merit function, we obtain a radius of quadratic convergence to a point on the central path as well as a decrease of the merit function.

Theorem 2.7 *Let* $\sigma_{\min}$ *and* $\sigma_{\max}$ *be, respectively, the smallest and largest singular value of* $F'_{\tau\mu}(s_{\tau\mu})$*. Under Assumptions 1.1 there is a* $\delta > 0$ *such that for all* s_c *such that* $\|s_c - s_{\tau\mu}\| < \delta$*, the Gauss-Newton step*

$$s_+ = s_c - F'_{\tau\mu}(s_c)^\dagger F_{\tau\mu}(s_c)$$

is well-defined and converges to $s_{\tau\mu}$ *at a rate such that*

$$\|s_+ - s_{\tau\mu}\| \le \frac{1}{\sigma_{\min}} \|s_c - s_{\tau\mu}\|^2.$$

Moreover, we can choose δ *as long as* $\delta < \frac{\sigma_{\min}}{2}$*.*

Proof. Let δ be small enough so that the hypothesis of Lemma 2.5 holds, i.e. $\delta < \frac{\sigma_{\min}}{2}$. First we express the error on the iterate both before and after the step, then by the fundamental Theorem of calculus and the fact that $F'_{\tau\mu}(s_c)$ is full column rank (and hence that $[F'_{\tau\mu}(s_c)]^\dagger F'_{\tau\mu}(s_c) = I$),

$$\begin{aligned}(s_+ - s_{\tau\mu}) &= (s_c - s_{\tau\mu}) - F'_{\tau\mu}(s_c)^\dagger F_{\tau\mu}(s_c) \\ &= F'_{\tau\mu}(s_c)^\dagger \int_0^1 (F'_{\tau\mu}(s_c) - F'_{\tau\mu}(s_{\tau\mu} + t(s_c - s_{\tau\mu})))(s_c - s_{\tau\mu})dt.\end{aligned}$$

Take norms on both sides and use the Lipschitz continuity of $F'_{\tau\mu}$ to get

$$\|s_+ - s_{\tau\mu}\| \le \frac{1}{2}\|F'_{\tau\mu}(s_c)^\dagger\| \|s_c - s_{\tau\mu}\|^2.$$

Now use Lemma 2.5, inequality (7b) to get

$$\|s_+ - s_{\tau\mu}\| \le \|F'_{\tau\mu}(s_{\tau\mu})^\dagger\| \|s_c - s_{\tau\mu}\|^2,$$

the required reduction of the error. □

The next result relates the reduction in the error to the reduction in the merit function.

Corollary 2.8 *Let* $\sigma_{\min}$ *and* $\sigma_{\max}$ *be, respectively, the smallest and largest singular value of* $F'_{\tau\mu}(s_{\tau\mu})$*. Under Assumptions 1.1 there is a* $\delta > 0$ *where for all* s_c *such that* $\|s_c - s_{\tau\mu}\| < \delta$*,*

$$\|F_{\tau\mu}(s_+)\| \le \frac{1}{2}\|F_{\tau\mu}(s_c)\|.$$

Moreover, we can choose any δ *such that*

$$\delta < \frac{\sigma_{\min}^2}{8\sigma_{\max}}. \tag{12}$$

Proof. Consider the inequality (7d) at the point s_+ to obtain

$$\|F_{\tau\mu}(s_+)\| \leq 2\sigma_{\max}\|s_+ - s_{\tau\mu}\|.$$

Now assume that δ satisfies the condition of Theorem 2.7 and apply the result as well as inequality (7c) at the point s_c to get

$$\begin{aligned}
\|F_{\tau\mu}(s_+)\| &\leq 2\frac{\sigma_{\max}}{\sigma_{\min}}\|s_c - s_{\tau\mu}\|^2, \\
&\leq 2\frac{\sigma_{\max}}{\sigma_{\min}}\|s_c - s_{\tau\mu}\|\frac{2}{\sigma_{\min}}\|F_{\tau\mu}(s_c)\| \\
&= 4\frac{\sigma_{\max}}{\sigma_{\min}^2}\|F_{\tau\mu}(s_c)\|\|s_c - s_{\tau\mu}\|.
\end{aligned}$$

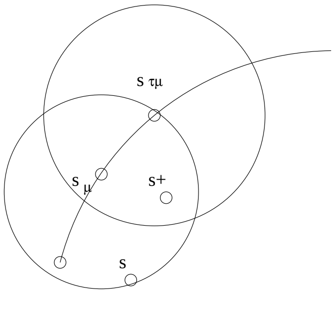
Therefore we need $\|(s_c - s_{\tau\mu})\| < \delta$, with δ as defined in (12), to obtain the required decrease. $\square$

3 Convergence of the algorithm

At this point we have established all the necessary relations between our merit function and the distance between an iterate and the central path. The current section describes the convergence of Algorithm 1. For easy reference, we repeat the definitions of the two canonical points s_μ and $s_{\tau\mu}$ on the central path. They satisfy

$$F_\mu(s_\mu) = 0, \qquad F_{\tau\mu}(s_{\tau\mu}) = 0. \tag{13}$$

The general idea of the algorithm is that, from a iterate s_k , "close enough" to s_μ, we can choose a target on the central path $s_{\tau\mu}$ in such a way that the next iterate s_{k+1}, obtained from the Gauss-Newton direction, is now "close enough" to $s_{\tau\mu}$ for the process to be repeated.



The proof is in three parts. First we estimate the distance between two points on the central paths in terms of the required radius of convergence.

Lemma 3.1 *Let $\sigma_{\min}$ and $\sigma_{\max}$ be, respectively, the smallest and largest singular value of $F'_{\tau\mu}(s_{\tau\mu})$. Let s_μ and $s_{\tau\mu}$ satisfy (13).*

1. *If we choose* $0 < \tau < 1$ *such that*

$$1 - \tau \leq \frac{\sigma_{\min}^2}{8\sqrt{n}\mu}, \tag{14}$$

then

$$\|s_\mu - s_{\tau\mu}\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}}{2}\right), \tag{15}$$

which implies s_μ *is within half of the radius of quadratic convergence of* $s_{\tau\mu}$*.*

2. *If we choose* $0 < \tau < 1$ *such that*

$$1 - \tau \leq \frac{\sigma_{\min}^3}{32\sqrt{n}\mu\sigma_{\max}}, \tag{16}$$

then

$$\|s_\mu - s_{\tau\mu}\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}^2}{8\sigma_{\max}}\right). \tag{17}$$

In this case s_μ *is within half of the radius of guaranteed constant decrease of the merit function in (12) in Corollary 2.8.*

Proof. First note that a straightforward calculation based on the definition of s_μ (13) yields

$$\|F_{\tau\mu}(s_\mu)\| = \sqrt{n}(1-\tau)\mu.$$

By Lemma 2.5, inequality (7d)

$$\begin{aligned} \|s_\mu - s_{\tau\mu}\| &\leq 2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|\|F_{\tau\mu}(s_\mu)\| \\ &= 2\|F'_{\tau\mu}(s_{\tau\mu})^\dagger\|(1-\tau)\sqrt{n}\mu. \end{aligned}$$

Let τ satisfy (14) to get

$$\|s_\mu - s_{\tau\mu}\| \leq \frac{\sigma_{\min}}{4},$$

which, by Theorem 2.7, yields one half of the quadratic radius of convergence. The proof of part 2 of the lemma is similar. □

We now estimate the distance to the new target after a Gauss-Newton step.

Lemma 3.2 *Let* $\sigma_{\min}$ *and* $\sigma_{\max}$ *be, respectively, the smallest and largest singular value of* $F'_{\tau\mu}(s_{\tau\mu})$*. Let* s_μ *and* $s_{\tau\mu}$ *satisfy (13). Suppose that the point* s_c *is well-centered in the sense that*

$$\|s_\mu - s_c\| \leq \min\left\{\frac{\sigma_{\min}}{4}, \frac{\sigma_{\min}^2}{16\sigma_{\max}}\right\},$$

and we choose τ *to satisfy*

$$0 < \tau < 1, \quad 1 - \tau \leq \min\left\{\frac{\sigma_{\min}^2}{8\sqrt{n}\mu}, \frac{\sigma_{\min}^3}{32\sqrt{n}\mu}\right\}, \tag{18}$$

as in Lemma 3.1. Then, after one Gauss-Newton step, the new point s_+ *will be within half the radius of convergence of* $s_{\tau\mu}$*, i.e.*

$$\|s_{\tau\mu} - s_+\| \quad \le \quad \frac{\sigma_{\min}}{4}. \tag{19}$$

Moreover, the merit function is reduced

$$\|F_{\tau\mu}(s_+)\| \le \frac{1}{2}\|F_{\tau\mu}(s_c)\|. \tag{20}$$

Proof.

By hypothesis and by Lemma 3.1,

$$\|s_c - s_\mu\| \le \frac{\sigma_{\min}}{4}, \quad \|s_\mu - s_{\tau\mu}\| \le \frac{\sigma_{\min}}{4}.$$

Therefore

$$\begin{aligned} \|s_c - s_{\tau\mu}\| &= \|s_c - s_\mu + s_\mu - s_{\tau\mu}\| \\ &\le \|s_c - s_\mu\| + \|s_\mu - s_{\tau\mu}\| \\ &\le \frac{\sigma_{\min}}{2}, \end{aligned}$$

which is within the radius of quadratic convergence of $s_{\tau\mu}$. After one Gauss-Newton step, by Theorem 2.7, we get

$$\begin{aligned} \|s_+ - s_{\tau\mu}\| &\le \frac{1}{\sigma_{\min}}\|s_c - s_{\tau\mu}\|^2 \\ &\le \frac{1}{\sigma_{\min}}(\frac{\sigma_{\min}}{2})^2 \\ &= \frac{\sigma_{\min}}{4}. \end{aligned}$$

Therefore the new point is within half the radius of convergence of $s_{\tau\mu}$ and the procedure can be repeated.

The constant reduction of the merit function follows from Corollary 2.8. □

We now present the main result of the paper, the convergence proof for Algorithm 1.

Theorem 3.3 *Suppose that we are given a tolerance* $\epsilon > 0$*, an initial barrier parameter* $\mu_0 > \epsilon$*, and* $Z_0, X_0 \in \mathbb{S}^n_{++}$ *such that* $s_0 = (X_0, y_0, Z_0)$ *is a well-centered starting point:* s_0 *is within half the quadratic convergence radius of* s_{μ_0} *in Theorem 2.7,*

$$\|s_{\mu_0} - s_0\| \le \frac{1}{2}\left(\frac{\sigma_{\min}}{2}\right). \tag{21}$$

Suppose moreover that s_0 *is within half the radius for guaranteed constant decrease of the merit function given in Corollary 2.8,*

$$\|s_{\mu_0} - s_0\| \le \frac{1}{2}\left(\frac{\sigma_{\min}^2}{8\sigma_{\max}}\right),$$

where $0 < \sigma_{\min}$ *(respectively* $\sigma_{\max}$*) is smaller than the smallest (respectively larger than the largest) singular value of* $F'_{\omega\mu_0}(s_{\omega\mu_0})$*, for all* $\frac{\epsilon}{\mu_0} < \omega < 1$.

If we choose τ *(*small*) satisfying (18) in Lemma 3.2, i.e.*

$$\alpha = \min\left\{\frac{\sigma_{\min}^2}{8\sqrt{n}\mu_0}, \frac{\sigma_{\min}^3}{32\sqrt{n}\mu_0}\right\},$$

and $\tau \geq \max\{0, 1-\alpha\}$, $0 < \tau < 1$*, then Algorithm 1 produces a sequence* s_k *converging to* $\bar{s}$*,* ϵ*-optimal in the following sense,*

$$\tau^k\mu_0 \leq \epsilon, \quad \|F_{\tau^k\mu_0}(\bar{s})\| \leq \epsilon, \quad \|\bar{s} - s_{\tau^k\mu_0}\| \leq 2\frac{\epsilon}{\sigma_{\min}};$$

and the number of iterations, k*, depends on* τ*:*

1. $$\mathcal{O}\left(\max\left\{\log\left(\frac{\|F_{\tau\mu_0}(s_0)\|}{\epsilon}\right), \ \log\left(\frac{\mu_0\sqrt{n}}{\epsilon}\right)\right\}\right) \tag{22}$$

   iterations, if $0 < \tau \leq \frac{1}{2}$*;*

2. $$\mathcal{O}\left(\max\left\{\log\left(\frac{\|F_{\tau\mu_0}(s_0)\|}{\epsilon}\right), \ \left(\frac{\log\frac{(2\tau-1)\epsilon}{(1-\tau)2\mu_0\sqrt{n}}}{\log\tau}\right), \ \left(\frac{\log\frac{\epsilon}{\mu_0}}{\log\tau}\right)\right\}\right) \tag{23}$$

   iterations, if $\frac{1}{2} < \tau < 1$.

Proof. First we note that the required constant $\sigma_{\min}$ exists by Corollary 2.6. By Lemma 2.5,

$$\|s_k - s_{\tau^k\mu_0}\| \leq 2\|F'_{\tau^k\mu_0}(s_{\tau^k\mu_0})^\dagger\|\|F_{\tau^k\mu_0}(s_k)\|,$$

which results in the desired bound on $\|s_k - s_{\tau^k\mu_0}\|$, if $\|F_{\tau^k\mu_0}(s_k)\| \leq \epsilon$. From the constant decrease guarantee we get (we add and subtract the multiple of the identity in the third term in the norm)

$$\begin{aligned}
\|F_{\tau^k\mu_0}(s_k)\| &\leq \frac{1}{2}\|F_{\tau^k\mu_0}(s_{k-1})\| \\
&\leq \frac{1}{2}\|F_{\tau^{k-1}\mu_0}(s_{k-1})\| + \frac{1}{2}\tau^{k-1}(1-\tau)\mu_0\sqrt{n} \\
&\leq \frac{1}{2^2}\|F_{\tau^{k-2}\mu_0}(s_{k-2})\| + \frac{1}{2^2}\left\{\tau^{k-2}(1-\tau)\mu_0\sqrt{n} + 2\tau^{k-1}(1-\tau)\mu_0\sqrt{n}\right\} \\
&\leq \left\{\frac{1}{2^k}\|F_{\tau\mu_0}(s_0)\|\right\} + \left\{(1-\tau)\mu_0\sqrt{n}\left(\frac{\tau^0}{2^k} + \frac{\tau}{2^{k-1}} + \frac{\tau^2}{2^{k-2}} + \ldots + \frac{\tau^{k-1}}{2}\right)\right\} \qquad (24) \\
&= \frac{1}{2^k}\|F_{\tau\mu_0}(s_0)\| + (1-\tau)\mu_0\sqrt{n}\tau^k\left(\frac{\tau^{0-k}}{2^k} + \frac{\tau^{1-k}}{2^{k-1}} + \frac{\tau^{2-k}}{2^{k-2}} + \ldots + \frac{\tau^{-1}}{2}\right) \\
&= \frac{1}{2^k}\|F_{\tau\mu_0}(s_0)\| + (1-\tau)\mu_0\sqrt{n}\tau^k\left(\left(\frac{1}{2\tau}\right)^k + \left(\frac{1}{2\tau}\right)^{k-1} + \left(\frac{1}{2\tau}\right)^{k-2} + \ldots + \frac{1}{2\tau}\right) \\
&= \left\{\frac{1}{2^k}\|F_{\tau\mu_0}(s_0)\|\right\} + \left\{(1-\tau)\mu_0\sqrt{n}\tau^k\left(\frac{1-(\frac{1}{2\tau})^k}{2\tau-1}\right)\right\}. \qquad (25)
\end{aligned}$$

We will bound each of the two terms in brackets in the last line above by $\frac{\epsilon}{2}$. From here onward log will indicate $\log_2$. For the first term we get

$$k \geq \left\lceil \log\left(\frac{2||F_{\tau\mu_0}(s_0)||}{\epsilon}\right)\right\rceil \tag{26}$$

where $\lceil x \rceil$ is the ceiling operator. It produces the smallest integer larger than or equal to x. For the second term in the brackets, we use the form in (24) while considering the case $\tau \leq \frac{1}{2}$. We get

$$(1-\tau)\mu_0\sqrt{n}\left(\frac{\tau^0}{2^k} + \frac{\tau}{2^{k-1}} + \frac{\tau^2}{2^{k-2}} + \ldots + \frac{\tau^{k-1}}{2}\right) \leq \mu_0\sqrt{n}\frac{1}{2^k}k \leq \frac{\epsilon}{2}, \tag{27}$$

or equivalently

$$\log\left(\frac{2\mu_0\sqrt{n}}{\epsilon}\right) \leq k - \log k \leq k. \tag{28}$$

Thus the case $\tau \leq \frac{1}{2}$ for the second term is bounded by $\frac{\epsilon}{2}$ if

$$k \geq \left\lceil \log\left(\frac{2\mu_0\sqrt{n}}{\epsilon}\right)\right\rceil. \tag{29}$$

For the case $\tau > \frac{1}{2}$ we use the form (25) to get

$$\begin{aligned}(1-\tau)\mu_0\sqrt{n}\tau^k\left(\frac{1-(\frac{1}{2\tau})^k}{2\tau-1}\right) &\leq \frac{(1-\tau)}{2\tau-1}\mu_0\sqrt{n}\tau^k \\ &\leq \frac{\epsilon}{2}\end{aligned} \tag{30}$$

is implied by

$$k \geq \left\lceil \frac{\log\left(\frac{(2\tau-1)\epsilon}{(1-\tau)2\mu_0\sqrt{n}}\right)}{\log\tau}\right\rceil, \tag{31}$$

where the direction of the inequality changed since $\tau < 1$.

Therefore, we can obtain $||F_{\tau^k\mu_0}(s_k)|| \leq \epsilon$ by choosing k using each of the lower bounds given in (26), (29), and (31). This guarantees that we are close to the central path.

We finally need to be close to optimality, $\mu_0\tau^k \leq \epsilon$. This is equivalent to

$$k \geq \frac{\log\frac{\epsilon}{\mu_0}}{\log\tau}. \tag{32}$$

The dependence on τ can be be eliminated, in the case $0 < \tau < \frac{1}{2}$, by

$$k \geq -\log\frac{\epsilon}{\mu_0},$$

which is implied by (29). The final $\mathcal{O}$ expression bounding the number of iterations is a simplification of (26), (29), (31), and (32). $\square$

4 Towards a long-step algorithm

The algorithm, as presented, is not practical; the assumptions that the initial iterate satisfies the conditions of Theorem 3.3 and that we need an estimate of the smallest singular values are significant. But the singular values are used, throughout the paper, only to show the existence of a radius of convergence. A practical version of the algorithm would more likely try some value for τ, compute the step and the value of the merit function, then reduce τ if the merit function reduction is not sufficient. Since we have shown the existence of a radius where the merit function is halved (20), such a scheme will necessarily converge. We presented the algorithm without these practical encumbrances to clarify the presentation.

The Gauss-Newton direction for solving semidefinite programs was introduced in [3] without a proof of convergence but with experimental results that warranted more research. Then, in [1], a scaled version of the direction was used in an algorithm shown to be polynomially convergent. The algorithm and the convergence proof presented in this paper are new in that the direction is used without any scaling and the algorithm never explicitly forces the iterates to remain within the positive definite cone. Moreover, the measure used to quantify the distance of the iterates to the central path (5b) estimates both the infeasibility and the complementarity and seems perfectly adapted to infeasible interior-point algorithms. It would be interesting to see how this measure can be used for different directions.

The dependence on the smallest singular value of the Jacobian for choosing τ, though unsurprising in the context, should be relaxed to some other, more easily estimated function of the data (possibly some condition measure [4]). But the ultimate goal of this avenue of research is to establish polynomial convergence of an infeasible algorithm using long steps, that is, not restricted to a narrow neighbourhood of the central path. Both experimental data and preliminary results suggests the possibility of such an algorithm.

References

[1] E. de Klerk, J. Peng, C. Roos, and T. Terlaky. A scaled Gauss-Newton primal-dual search direction for semidefinite optimization. *SIAM J. Optim.*, 11(4):870–888 (electronic), 2001.

[2] L. V. Kantorovich and G. P. Akilov. *Functional analysis.* Pergamon Press, Oxford, second edition, 1982. Translated from the Russian by Howard L. Silcock.

[3] Serge Kruk, Masakazu Muramatsu, Franz Rendl, Robert J. Vanderbei, and Henry Wolkowicz. The Gauss-Newton direction in semidefinite programming. *Optim. Methods Softw.*, 15(1):1–28, 2001.

[4] James Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.