Optimization Methods and Software

ISSN: 1055-6788 (Print) 1029-4937 (Online) Journal homepage: https://www.tandfonline.com/loi/goms20

The Gauss-Newton direction in semidefinite programming

Serge Kruk , Masakazu Muramatsu , Franz Rendl , Robert. J. Vanderbei & Henry Wolkowicz

To cite this article: Serge Kruk , Masakazu Muramatsu , Franz Rendl , Robert. J. Vanderbei & Henry Wolkowicz (2001) The Gauss-Newton direction in semidefinite programming , Optimization Methods and Software, 15:1, 1-28, DOI: 10.1080/10556780108805808

To link to this article: https://doi.org/10.1080/10556780108805808

Optimization Meth. & Soft., 2001, Vol. 15, pp. 1–28
Reprints available directly from the publisher
Photocopying permitted by license only

THE GAUSS-NEWTON DIRECTION IN SEMIDEFINITE PROGRAMMING*

SERGE KRUK[a,†], MASAKAZU MURAMATSU[b], FRANZ RENDL[c,‡], ROBERT J. VANDERBEI[d,§] and HENRY WOLKOWICZ[a,¶]

[a]*Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada;*
[b]*Department of Mechanical Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102, Japan;*
[c]*University of Klagenfurt, Institute of Mathematics, A-9020 Klagenfurt, Austria;*
[d]*ACE-42 E-Quad, Princeton University, Princeton NJ 08544*

(Received 22 October 1999; In final form 14 September 2000)

Most of the directions used in practical interior-point methods for semidefinite programming try to follow the approach used in linear programming, *i.e.*, they are defined using the optimality conditions which are modified with a symmetrization of the perturbed complementarity conditions to allow for application of Newton's method. It is now understood that all the Monteiro-Zhang family, which include, among others, the popular AHO, NT, HKM, Gu, and Toh directions, can be expressed as a scaling of the problem data and of the iterate followed by the solution of the AHO system of equations, followed by the inverse scaling. All these directions therefore share a defining system of equations.

The focus of this work is to propose a defining system of equations that is essentially different from the AHO system: the over-determined system obtained from the minimization of a nonlinear equation. The resulting solution is called the Gauss-Newton

---

*This report is available by anonymous ftp at orion.uwaterloo.ca, in directory pub/henry/reports and also with URL http://orion.uwaterloo.ca/~hwolkowi/henry/reports/newtondir.ps.gz

†Research supported by the Natural Sciences and Engineering Research Council of Canada. Department of Mathematics, Oakland University, MI. e-mail: sgkruk@acm.org

‡e-mail: franz.rendl@uni-klu.ac.at

§Tel.: 609-258-0876, Fax: 609-258-3796, e-mail: rvdb@princeton.edu, http://www.princeton.edu/~rvdb/

¶Research supported by the Natural Sciences and Engineering Research Council of Canada. e-mail: hwolkowi@orion.math.uwaterloo.ca, http://orion.math.uwaterloo.ca/~hwolkowi

search direction. We state some of the properties of this system that make it attractive for accurate solutions of semidefinite programs. We also offer some preliminary numerical results that highlight the conditioning of the system and the accuracy of the resulting solutions.

Keywords: Semidefinite programming; Newton direction; Symmetrization

1. INTRODUCTION

Primal-dual interior-point (p-d i-p) methods have proven to be successful for both linear programming (LP) and, more recently, for semidefinite programming (SDP) problems. Many of the techniques that have been used in linear programming have been extended to SDP. In fact, interior-point methods currently appear to be the only practical techniques for SDP.

The derivation of interior-point algorithms for SDP has followed closely the lessons learned from the applications in LP, *e.g.*, the primal-dual approach of [35, 31, 30]. This has led to specific choices for the form of the perturbed optimality conditions to which Newton's method is applied. In this paper, we present a modified paradigm for deriving these methods both for LP and for SDP.

1. Using the optimality conditions from the dual log-barrier problem, we obtain the system of three equations: primal feasibility; dual feasibility; and perturbed complementary slackness, see Section 1.2, program (1.2) and Eqs. (1.4a)–(1.4c).
2. The linear system for the Newton direction (N direction) for the log-barrier problem is ill-conditioned but can be solved in a stable way by using an augmented system, *i.e.*, this direction itself is *not ill-conditioned* and can be used successfully. However, the perturbed complementary slackness condition (1.4c) is nonlinear. We modify it to obtain a bilinear function and an equation on which Newton type methods converge faster.
3. We now find the Gauss-Newton search direction (GN direction) to the least squares problem for the resulting equations; equivalently we find the least squares solution of the linearized optimality conditions.

This paradigm is similar to what is currently done, *e.g.*, see [14]. In particular, item 1 is unchanged. However, the modifications of

the optimality conditions in item 2 is usually motivated by the ill-conditioning. Here we emphasize that this ill-conditioning can be avoided (see [34] Section 5) and that the true motivation should be to reduce the nonlinearity. The main novelty of our paradigm is item 3, where we use the least-squares solution to the overdetermined optimality conditions rather than a Newton direction on the symmetrized optimality conditions. This approach shows that the search direction for SDP can be found from the optimality conditions just as is done in LP. This is emphasized by the fact that the GN direction is always well defined, and is in contrast to current approaches that modify the optimality conditions to apply a Newton method to a square system.

1.1. Outline

In this paper we introduce a family of search directions for primal-dual interior-point methods for semidefinite programming. These directions arise from a Gauss-Newton method applied to minimizing the norm of the over-determined system of optimality conditions, rather than a Newton method applied to symmetrized optimality conditions. In addition, we compare several directions within the same interior-point algorithm to assess the accuracy of the resulting solutions and the effectiveness of the new approach. We conclude that the Gauss-Newton direction has interesting properties that warrant further study.

We complete this introduction with some background on SDP in Section 1.2. In Section 2 we define the GN direction and explore some of its properties in Section 2.1.1.

Solution techniques are discussed in Section 3 while Section 3.1 contains the details of the elimination approach used in our implementation of the GN direction.

Preliminary numerical tests appear in Section 4. These tests compare various new and known search directions. We see that the GN approach is competitive in terms of the iteration count and the accuracy of the solutions. In particular, it seems superior when applied to certain badly scaled problems.

Remark 1.1 In a detailed version of the present paper [34, Section 5], we discuss how ill-conditioning arises from log-barrier problems in NLP and how this can be avoided using an augmented system.

We conclude with several remarks and open questions in Section 5.

1.2. Optimality Conditions

We now include the notation and background information on SDP.[1] Let $\mathbb{S}^n$ denote the space of $n \times n$ symmetric matrices equipped with the trace inner product, $\langle A, B \rangle = \text{trace } AB$; and let $A \succeq 0$ (resp. $A \succ 0$) denote positive semidefiniteness (resp. positive definiteness), $A \succeq B$ denotes $A - B \succeq 0$, *i.e.*, $\mathbb{S}^n$ is equipped with the Löwner partial order. We let $\mathbb{S}^n_+$ denote the cone of symmetric positive semidefinite matrices; $\mathbb{M}^n$ denotes the set of $n \times n$ real matrices.

SDP is an extension of linear programming, denoted LP, where nonnegative variables are replaced by positive semidefinite matrices. The standard primal semidefinite program is

$$(PSDP) \qquad \begin{array}{lll} \mu^* := \min & & \langle C, X \rangle \\ \text{subject to} & & \mathcal{A}X = b \\ & & X \succeq 0, \end{array}$$

where the linear operator $\mathcal{A}: \mathbb{S}^n \to \mathbb{R}^m$ is defined as

$$[\mathcal{A}X]_i = \langle A_i, X \rangle,$$

and A_i, $i = 1, \ldots, m$, are given symmetric matrices. To avoid trivialities, we assume that μ^* is finite valued. Thus the problem has a bounded optimal value and a feasible solution. We also assume that the matrices $\{A_i: i = 1, \ldots, m\}$ are linearly independent.

Throughout this paper, calligraphy letters $(\mathcal{A}, \mathcal{Z})$ denote operators, uppercase Latin letters (X, F_d) denote matrices and lowercase Latin letters (b, f_p) denote vectors.

Problem PSDP has generated much interest. One reason is that there are many diverse applications: in discrete optimization see *e.g.* [3, 21, 49]; in engineering see *e.g.* [11, 63]; for matrix completions see *e.g.* [26, 1]. Another reason for the interest is that SDPs can be solved efficiently using interior-point methods. More applications and evidence of the current high level of activity can be found in the recent

[1] Comprehensive notation is available at URL: http://orion.uwaterloo.ca/~hwolkowi/henry/software/psd_tool.d/sdpnotation.d/notation.ps

theses: [2, 47, 23, 44, 27, 68, 33, 28, 56]; and in the recent books and notes [66, 10, 41, 65, 58, 64, 43].

We now follow the p-d i-p approach developed in [24]. We follow the usual derivation of p-d i-p methods in LP and first introduce the associated *log-barrier problem* for the dual SDP

$$\begin{array}{lll} & \max & b^t y \\ (DSDP) & \text{subject to} & \mathcal{A}^* y + Z = C \\ & & Z \succeq 0, \end{array} \tag{1.1}$$

where $\mathcal{A}^*$ denotes the *adjoint operator*, $\mathcal{A}^* y = \sum y_i A_i$. The *dual barrier problem* is:

$$\begin{array}{lll} & \max & b^t y + \mu \log\det Z \\ (DBP) & \text{subject to} & \mathcal{A}^* y + Z = C \\ & & Z \succ 0, \end{array} \tag{1.2}$$

where μ is a positive real number called the *barrier parameter*.

For each $\mu > 0$, there is a corresponding Lagrangian:

$$L_\mu(X, y, Z) = b^t y + \mu \log\det Z + \langle X, (C - Z - \mathcal{A}^* y)\rangle. \tag{1.3}$$

With X_μ, $Z_\mu \succ 0$, the first-order optimality conditions for the saddle point of this Lagrangian are obtained by differentiation:

$$\nabla_X L_\mu = C - Z - \mathcal{A}^* y = 0 \tag{1.4a}$$

$$\nabla_y L_\mu = b - \mathcal{A}(X) = 0 \tag{1.4b}$$

$$\nabla_Z L_\mu = -X + \mu Z^{-1} = 0. \tag{1.4c}$$

There exists a unique solution (X_μ, y_μ, Z_μ) to these optimality conditions. The one-parameter family $\{(X_\mu, y_\mu, Z_\mu)\colon 0 < \mu\}$ is called the *central trajectory* or the central path. Given a point (X, y, Z) on the central trajectory it is easy to determine its associated μ value using (1.4c):

$$\mu = \frac{\langle Z, X\rangle}{n}. \tag{1.5}$$

We note that if the point is a feasible solution of the primal and the dual problem, then $\langle Z, X\rangle$ is the gap between the primal and the dual objective value.

For future reference, we express the central path equations as

$$F_\mu(s) = F_\mu(X, y, Z) := \begin{pmatrix} C - Z - \mathcal{A}^* y \\ b - \mathcal{A}(X) \\ -X + \mu Z^{-1} \end{pmatrix} = 0. \tag{1.6}$$

The equations correspond to: dual feasibility, primal feasibility, and complementary slackness, respectively. The solution s^* to $F_\mu(s) = 0$ satisfies the Karush-Kuhn-Tucker conditions (1.4) and is the optimal solution to the barrier problem. Path-following Interior-point methods approximately follow this trajectory for decreasing values of the parameter μ.

1.3. Bilinear Complementarity and Symmetrization

Since the objective of a path-following algorithm is to find an approximate solution of $F_\mu = 0$, a natural approach is to use a Newton step. But the complementarity equation is highly nonlinear, especially near the zero where Z will be rank-deficient. In LP (which can be viewed as a special case of SDP where all matrices are diagonal) the obvious solution has been to transform (1.4c) into the bilinear form

$$ZX - \mu I = 0. \tag{1.7}$$

The resulting function is less nonlinear since it is bilinear. Such modifications can improve the rate of convergence, see *e.g.* [8, 20, 46], and are the basis behind the success of algorithms for trust-region subproblems, see *e.g.* [46, 55, 51, 39].

Many of the algorithms from LP can be extended to SDP and retain many of their nice properties. (See *e.g.*, the books on interior-point methods: by Nesterov and Nemirovsky [41]; by Wright [65]; by Ye [66]; edited by Terlaky [58] and the handbook on SDP [64].)

However, there are several subtle differences with LP.

- One interesting difference is that duality gaps can exist for SDP in the absence of strictly feasible solutions (Slater's constraint qualification). This means that strong duality can fail, *i.e.*, the primal or dual may not be attained and/or there is a nonzero duality gap between the primal and dual optimal values. However, feasible interior-point methods usually assume strictly feasible solutions, so

we will not dwell on that here. (See *e.g.* [48] for a discussion on strong duality and [45] for the generic properties.)

- Another difference is that strict complementarity may fail at the optimum, *i.e.*, there may be no optimal primal-dual pair for which $X+Z$ is full rank. (See *e.g.* [5, 52] for a discussion on strict complementarity and [45] for the generic properties.)
- Finally, the difference we focus on is that the bilinear form of the complementarity Eq. (1.7), while it is equivalent in the LP case to the original Eq. (1.4c), prevents a Newton step in the SDP case.

A Newton step is not possible because replacing the third equation of $F_\mu = 0$ by (1.7) maps $\mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n$ into $\mathbb{S}^n \times \mathbb{R}^m \times \mathbb{M}^n$. The solution adopted by most interior-point algorithms is to neglect the skew-symmetric part of the product ZX by means of a symmetrization operator

$$H(M) := \frac{1}{2}(M + M^t), \tag{1.8}$$

to obtain

$$ZX + XZ - 2\mu I = 0. \tag{1.9}$$

As written above (1.9), the complementarity equation now maps $\mathbb{S}^n$ into $\mathbb{S}^n$ and a Newton step is possible. This formulation is precisely used by Alizadeh, Haeberly and Overton [6] but, in a very real sense, it is used by all popular interior-point algorithms. This is because all of the Monteiro-Zhang family of directions [38] can be expressed as the AHO direction obtained from the perturbed and symmetrized optimality conditions of an equivalent, scaled problem (see [59]).

$$\begin{array}{lll} & \min & \langle \widetilde{C}, \widetilde{X} \rangle \\ (Scaled - PSDP) & \text{subject to} & \widetilde{\mathcal{A}}\widetilde{X} = b \\ & & \widetilde{X} \succeq 0, \end{array} \tag{1.10}$$

where

$$\widetilde{C} := P^{-t} C P^{-1}, \quad \widetilde{A}_i := P^{-t} A_i P^{-1},$$

for various values of P, which is then scaled back *via*

$$(dX, dy, dZ) = (P^{-1}\widetilde{dX}P^{-t}, \widetilde{dy}, P^t\widetilde{dZ}P).$$

1. the H. K. M. direction, [24, 32, 36], corresponds to $P = Z^{1/2}$; this is equivalent to symmetrizing dX after solving the linear system with dX treated as a general matrix;
2. the dual H. K. M direction corresponds to $P = X^{-1/2}$.
3. the NT direction, [61], corresponds to $P = G^{-1}$, where G is the unique matrix such that $D := G^tZG = G^{-1}XG^{-t}$ is a diagonal matrix.

(See [6, 59, 64] for derivations of various Newton directions.)

The various scalings are not without effect on the resulting directions. Each guarantees certain properties that we will not discuss here. Our intent is to highlight the archetypical nature of the bilinear form of the complementarity equation and of the resulting symmetrization.

Of course, practical algorithms are not necessarily implemented *via* scalings (See *e.g.*, the software packages [60, 19, 12, 4].); scalings simply provide a unifying view which underlines the practical importance of the symmetrization in the current practice where a "square" system is used (see [50, 24, 36, 32, 38, 37, 6, 67, 61, 54]); it also serves to introduce what is truly novel in our approach.

2. OVER-DETERMINED OPTIMALITY CONDITIONS

We start from the bilinear form of the complementarity conditions,

$$ZX - \mu I = 0, \tag{2.11}$$

and recall that the symmetrization $ZX + XZ - 2\mu I = 0$ was introduced because of the insistence on a Newton step. Is there an alternative approach to the solution of (2.11)?

We now look at (2.11) as an operator mapping the symmetric matrix space $\mathbb{S}^n \times \mathbb{S}^n$ to the space of matrices $\mathbb{M}^n$. With this in mind, we see that the complete set of perturbed optimality conditions form an over-determined system of nonlinear equations. The standard approach,

e.g. [13], to an over-determined nonlinear system is to solve an equivalent least squares problem, *i.e.*, no symmetrization is needed. We now explore this possibility for possible advantages.

For reference, the complete system we are trying to solve is

$$\Phi_Z F_\mu(s) = \Phi_Z F_\mu(X, y, Z) = \begin{pmatrix} C - Z - \mathcal{A}^*(y) \\ b - \mathcal{A}(X) \\ ZX - \mu I \end{pmatrix} =: \begin{pmatrix} F_d \\ f_p \\ F_c \end{pmatrix} = 0, \tag{2.12}$$

where F_μ is defined in (1.6), Φ_Z: $\mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n \to \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{M}^n$, $\Phi_Z(A, b, C) = (A, b, Z \cdot C)$, (*i.e.*, the third argument is multiplied on the left by Z) and $(\cdot)^*$ denotes the adjoint operator. In contrast to (1.6), we are solving an over-determined but less nonlinear system.

2.1. The GN Direction

The classical approach to a system of nonlinear equations $F(s) = 0$ is the Gauss-Newton method: minimize the sum of squares of the equations and use an approximation to the Newton search direction, the GN direction. In more detail, replace $F(s) = 0$ by

$$\min \left\{ \frac{1}{2} \langle F(s), F(s) \rangle \right\} =: \min\{f(s)\}.$$

At some iterate s_k, express the quadratic approximation of f by

$$\hat{f}(s) := f(s_k) + \langle f'(s_k), d \rangle + \frac{1}{2} \langle f''(s_k) d, d \rangle,$$

where the second derivative $f''(s_k)$ is composed of first-order terms ($\langle F'(s_k), F'(s_k) \rangle$) and of second-order terms ($\langle F''(s_k), F(s_k) \rangle$). The Newton direction to minimize $\hat{f}$ is defined by $f''(s_k)d = -f'(s_k)$. The Gauss-Newton approach approximates the direction by ignoring the second-order terms of f''.

Under very mild assumptions, *e.g.*, full column rank of the derivative F', the Gauss-Newton method for solving nonlinear equations enjoys interesting properties, see *e.g.* [13, 15, 17, 16], some of which we

will show, as they apply to (2.12):

1. If the optimal value of the least squares problem is 0, then we get quadratic convergence asymptotically. In our case, if we assume that the duality gap is 0 and that both primal and dual SDP are attained, then there is a solution for the optimality conditions which means that the least squares problem does indeed have a (attained) minimum value of 0 for each $\mu > 0$.
2. The GN direction always exists. Thus there are no problems with existence as there are for some search directions used in the literature, see *e.g.* [37]. In fact, one can apply the Gauss-Newton approach to the scaled problems used to define other directions (1.10) and thus resolve the question of existence. This is already the object of some research [29]. In this sense we view the Gauss-Newton directions as a whole family even though we restrict our attention here to the unscaled problem.
3. The direction guarantees descent for the objective function of the least squares problem. We therefore have a merit function to monitor progress. (Though it is not a true potential function, see [40, page 5].) This aspect, as well as the implementation details related to step-length are the object of current research.
4. The direction is scale invariant under affine transformation of the space and also under orthogonal transformation of the problem data.
5. The search direction is found using the same matrix equations as for LP, with the exception that a least squares solution is used.

With this motivation, we now look at the nonlinear least squares problem

$$r_\mu^* = \min_s f_\mu(s) := \frac{1}{2}\langle \Phi_Z F_\mu(s), \Phi_Z F_\mu(s)\rangle, \tag{2.13}$$

where the Frobenius inner product is used for elements of $\mathbb{S}^n$ and the standard Euclidean inner product is used for elements of $\mathbb{R}^m$. The optimal value r_μ^* is called the residual.

Since the GN direction is formed by dropping the second order terms of the Hessian of f_μ from the Newton equation $f_\mu''(s)d_N = -f_\mu'(s)$, it is enlightening to consider the importance of the dropped terms in our application.

Recall that $F_c = F_c(X, Z) = ZX - \mu I$. The dropped second order terms therefore are $\langle F_c(X, Z), F_c''(X, Z)(\cdot, \cdot)\rangle$. If we denote the changes in Z, X as dZ, dX respectively, then

$$(Z + dZ)(X + dX) - \mu I = ZX - \mu I + ZdX + dXZ + dZdX.$$

The curvature in this direction is therefore

$$\langle (ZX - \mu I), dZdX \rangle.$$

The second order terms that are dropped from the Newton direction to get the GN direction are rather simple and similar to the LP case,

$$H := \frac{1}{2}\begin{bmatrix} 0 & (ZX - \mu I) \\ (ZX - \mu I) & 0 \end{bmatrix}.$$

This matrix is indefinite and can cause problems if included in the approximate Hessian. Note that the residual $r_\mu^* = 0$ in our case, which implies that the matrix of second derivatives is 0 at the optimum in (2.13). Thus the error is very small near the central path.

Finding the GN direction is therefore equivalent to solving the normal equations

$$((\Phi_Z F_\mu)')^*(\Phi_Z F_\mu)'(ds) = -((\Phi_Z F_\mu)')^*(\Phi_Z F_\mu), \tag{2.14}$$

which is equivalent to finding the least squares solution of the over-determined system

$$(\Phi_Z F_\mu)'(ds) = -(\Phi_Z F_\mu). \tag{2.15}$$

One popular method for solving the above system is to solve (2.15) using a sparse QR decomposition. (See [9, 57].) While the Newton direction for the least squares problem is solved from the system

$$(((\Phi_Z F_\mu)')^*(\Phi_Z F_\mu)' + H)(ds) = -((\Phi_Z F_\mu)')^*(\Phi_Z F_\mu) \tag{2.16}$$

which cannot exploit the normal equations. Moreover, the addition of the operator H introduces negative eigenvaluues and so the Newton direction may not point to the minimum.

With the above justification we now express the defining equations of the Gauss-Newton direction. We use operator notation because that viewpoint suggested the approach. The over-determined linear system

we are solving is:

$$\mathcal{A}^*(dy) + dZ = -(\mathcal{A}^*(y) + Z - C) \tag{2.17a}$$

$$\mathcal{A}(dX) = -(\mathcal{A}(X) - b) \tag{2.17b}$$

$$\mathcal{Z}(dX) + \mathcal{X}(dZ) = -(ZX - \mu I), \tag{2.17c}$$

where the operators $\mathcal{Z}, \mathcal{X}$ are defined as

$$\mathcal{Z}\colon \mathbb{S}^n \to \mathbb{M}^n, \quad \mathcal{Z}(M) := ZM; \quad \text{and} \quad \mathcal{X}\colon \mathbb{S}^n \to \mathbb{M}^n, \quad \mathcal{X}(M) := MX.$$

Sometimes an equivalent matrix formulation is useful, especially since it highlights that the iterates never leave the space of symmetric matrices. That is, we solve (2.17) for dX, dZ symmetric matrices. There is no symmetrization of the equations and neither is there a symmetrization of the resulting solution.

$$[J_{gn}]d = \begin{bmatrix} 0 & A^t & I \\ A & 0 & 0 \\ (Z \oslash I) & 0 & (I \oslash X) \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = -\begin{bmatrix} f_d \\ f_p \\ f_c \end{bmatrix}, \tag{2.18}$$

with the equivalences,

$$\begin{aligned} A &:= [\text{svec}(A_1), \ldots, \text{svec}(A_m)]^t, \\ dx &:= \text{svec}(dX), \\ dz &:= \text{svec}(dZ), \\ f_d &:= \text{svec}(F_d), \\ f_c &:= \text{avec}(F_c), \end{aligned}$$

where $\text{svec}(\cdot)\colon \mathbb{S}^n \to \mathbb{R}^{t(n)}$ stacks, column by column, the upper triangle of a symmetric matrix into a vector of size $t(n) := (n(n+1)/2)$ while multiplying the off-diagonal elements by $\sqrt{2}$ to maintain distances. The corresponding $\text{avec}(\cdot)\colon \mathbb{M}^n \to \mathbb{R}^{n^2}$ stacks the column of any matrix into a vector. The binary operator $\oslash\colon \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{M}^{n^2 \times t(n)}$ is what we will call the asymmetric Kronecker product, defined from the identity

$$\text{avec}(AXB) := (A \oslash B)\text{svec}(X).$$

The above system (2.18) is very sparse and most of the terms do not change at each iteration. We have to exploit this fact as is done in linear programming where block elimination is used.

2.1.1. Properties of the Gauss-Newton Direction

We assume that the operator $\mathcal{A}$ is onto, *i.e.*, the matrices A_k are linearly independent or, equivalently, $\mathcal{A}^*$ is one-one or full column rank. We first see that the Gauss-Newton direction exists and is well-defined without any other assumptions. In particular, we do not require a neighborhood around the central path.

THEOREM 2.1 *Suppose that both X and Z are positive definite. Then the linear operator of the overdetermined system* (2.17) *is full rank.*

Proof Suppose that (dX, dy, dZ) solves (2.17) with right hand side 0. Then $dZ = -\mathcal{A}^*(dy)$. Therefore, $\mathcal{A}(dX) = 0$ and

$$ZdXX^{-1} = \mathcal{A}^*(dy) \perp dX. \tag{2.19}$$

This means that the quadratic form

$$0 = \operatorname{trace}(-\mathcal{A}^*(dy)dX) = \operatorname{trace}(ZdXX^{-1}dX) = \|Z^{1/2}dXX^{-1/2}\|_F^2, \tag{2.20}$$

i.e., necessarily $dX = 0$. This means that both dy and dZ are 0 as well. ∎

However, to guarantee full rank of the Jacobian at the optimum, we need to assume that the primal and dual optima are both unique. This is equivalent to primal and dual non-degeneracy as defined and used in [5, 6] and shown to be generic, see also [53, 45].

THEOREM 2.2 *Suppose that (X, y, Z) are a unique optimal primal and dual solution (equivalently primal and dual non-degenerate) of PSDP. In addition, suppose that strict complementarity holds, i.e., $X + Z \succ 0$. Then 0 is the unique solution of the system* (2.17) *at $\mu = 0$.*

Proof Suppose that (dX, dy, dZ) solves (2.17). Then dX, dZ are symmetric since we solve this system in the symmetric matrix space and therefore (dX, dy, dZ) also solves the symmetrized system

$$ZdX + dZX + dXZ + XdZ = 0.$$

The result in [6, Theorem 1] implies that the Jacobian corresponding to the symmetrized system is nonsingular and so necessarily $(dX, dy, dZ) = 0$.

We can provide a self-contained proof using the approach above in Theorem 2.1, *i.e.*, if $W = dX$ and $S = \mathcal{A}^*(dy)$, then $ZW + SX = 0 = ZX$. Since Z and X commute, we can assume that they are diagonal. Strict complementarity and uniqueness imply that both S and W are 0. ∎

This, in combination with Theorem 2.1 above, suggest that we will not run into ill-conditioning difficulties. To pursue this idea, we can compare the singular values of the GN Jacobian to those of the AHO Jacobian, known to be the best-conditioned of the symmetric directions.

Consider this equivalent form of the over-determined system (2.17)

$$\mathcal{A}^*(dy) + dZ = -(\mathcal{A}^*(y) + Z - C), \tag{2.21a}$$

$$\mathcal{A}(dX) = -(\mathcal{A}(X) - b), \tag{2.21b}$$

$$H(ZdX + dZX) = -H(ZX - \mu I), \tag{2.21c}$$

$$K(ZdX + dZX) = -K(ZX - \mu I), \tag{2.21d}$$

where

$$H(M) := \frac{1}{2}[M + M^t], \quad \text{(symmetric part)}; \tag{2.22}$$

$$K(M) := \frac{1}{2}[M - M^t], \quad \text{(skew-symmetric part)}. \tag{2.23}$$

Note that the first three Eq. (2.21a)–(2.21c) correspond to the symmetrized (or AHO) system. Corresponding to the matrix formulation (2.18) we can write

$$[J_{gn}]d = P\begin{bmatrix} J_s \\ J_k \end{bmatrix} d = -\begin{bmatrix} F_s \\ F_k \end{bmatrix}, \tag{2.24}$$

for some permutation P, and where J_{gn} is the Jacobian of the Gauss-Newton system, $J_s = J_{aho}$ is the Jacobian of the symmetric (AHO)

system and J_k is the part of J_{gn} corresponding to the skew-symmetric component of the complementarity equation. To simplify the notation later, let

$$\bar{n} := t(n) + m + t(n), \tag{2.25}$$

$$\bar{m} := t(n) + m + n^2, \tag{2.26}$$

$$r := \bar{m} - \bar{n} = t(n) - t(n-1), \tag{2.27}$$

and note that J_{gn} is $(\bar{m} \times \bar{n})$, J_s is $(\bar{n} \times \bar{n})$, J_k is $(r \times \bar{n})$. Also let $\sigma(A)$ be the singular values of matrix A in non-increasing order. With this notation, from (2.24), we obtain the following interlacing relation for the singular values.

THEOREM 2.3 *The singular values of J_{gn} and of J_{aho} are related by $\sigma_k(J_{gn}) \geq \sigma_k(J_{aho}) \geq \sigma_{k+t(n-1)}(J_{gn})$, for $1 \leq k \leq \bar{n}$.*

Proof Follows directly from Corollary 3.1.3 of [25]. ∎

We can also bound the largest singular value.

THEOREM 2.4 *The largest singular value of the Gauss-Newton Jacobian is bounded by $\sigma_1(J_{gn}) \leq \sqrt{\sigma_1^2(J_{aho}) + \sigma_1^2(J_k)}$.*

Proof With the above definitions of J_{gn}, P, J_s, and of J_k,

$$\begin{aligned}
\sigma_1^2(J_{gn}) &= \lambda_1(J_{gn}^t J_{gn}) \\
&= \lambda_1\left([(PJ_s)^t (PJ_k)^t] \begin{bmatrix} PJ_s \\ PJ_k \end{bmatrix} \right) \\
&= \lambda_1(J_s^t J_s + J_k^t J_k) \\
&= \|J_s^t J_s + J_k^t J_k\|_2 \\
&\leq \|J_s^t J_s\|_2 + \|J_k^t J_k\|_2 \\
&= \lambda_1(J_s^t J_s) + \lambda_1(J_k^t J_k) \\
&= \sigma_1^2(J_s) + \sigma_1^2(J_k).
\end{aligned}$$

Therefore $\sigma_1(J_{gn}) \leq \sqrt{\sigma_1^2(J_{aho}) + \sigma_1^2(J_k)}$. ∎

These results have at least two interesting consequences. First, directly from (2.24) we can see that whenever Z and X commute, in particular on the central path, the Gauss-Newton and the AHO direction coincide. Second, it implies that the Gauss-Newton system is no closer to singularity than the AHO system is (strictly further if not on the central path). This is interesting in view of the numerical stability of the AHO direction compared to most members of the Monteiro-Zhang family. For a relation between the singularity of the system and the accuracy of the resulting direction, see the very interesting papers of Gu [22] and of Toh [62]. We believe these properties explain the good behavior of the GN step for the badly scaled problems described in Section 4.

The Gauss-Newton direction also enjoys two different invariance properties. First it is invariant under affine transformation of the variables. This is the classical meaning of scale-invariance of an algorithm. (See for example [18], Section 3.3.) Note that we do not imply that the function $\|\Phi_Z F_\mu(X, y, Z)\|^2$ is invariant, but only that the direction used to minimize that function is invariant.

In the following the notation $(\cdot)^\dagger$ indicates the Moore-Penrose inverse (See [7]).

THEOREM 2.5 *The GN direction, the least-square solution to* (2.15) *used to solve program* (2.13), *is invariant under affine transformation of the variables* $r = Hs + h$ *where* H *is non-singular.*

Proof The GN step in s-space, from current point s_c, is given by

$$s_+ = s_c - ((\Phi_Z F_\mu(s_c))')^\dagger (\Phi_Z F_\mu(s_c)).$$

Now, under the affine scaling $r = Hs + h$ we obtain that $\Phi_Z F_\mu(s) = \Phi_Z F_\mu(H^{-1}(r-h)) =: G(r)$. Using the full rank property in Theorem 2.2, we see that the GN step in the r-space, from the current point $r_c = Hs_c + h$, will be

$$\begin{aligned} r_+ &= r_c - (G(r_c)')^\dagger G(r_c) \\ &= Hs_c + h - ((\Phi_Z F_\mu(s_c))' H^{-1})^\dagger (\Phi_Z F_\mu(s_c)) \\ &= Hs_c + h - H((\Phi_Z F_\mu(s_c))')^\dagger (\Phi_Z F_\mu(s_c)) \\ &= Hs_+ + h. \end{aligned}$$

The GN step is therefore invariant under affine transformations of the space. ■

We should remark that this classical scale-invariance property of the Gauss-Newton step actually holds for any nonlinear function with full-rank Jacobian. In our specific case we can say more: the Gauss-Newton direction is invariant under orthogonal scaling of both the iterates and the problem data. This is called *Q-scale invariance* by Todd [59].

THEOREM 2.6 *Let (dX, dy, dZ) be the Gauss-Newton direction obtained at iterate (X, y, Z). Consider the scaled primal-dual pair*

$$min\{\langle \widetilde{C}, \widetilde{X}\rangle | \widetilde{\mathcal{A}}\widetilde{X} = b, \widetilde{X} \in \mathbb{S}^n_+\}, \tag{2.28}$$

$$max\{\langle b, \widetilde{y}\rangle | \widetilde{\mathcal{A}}^* \widetilde{y} + \widetilde{Z} = \widetilde{C}, \widetilde{Z} \in \mathbb{S}^n_+\}, \tag{2.29}$$

obtained from the original problem data with scaling

$$\widetilde{X} := PXP^t, \quad \widetilde{A}_i := P^{-t}A_iP^{-1}, \quad \widetilde{C} := P^{-t}CP^{-1}. \tag{2.30}$$

for some orthogonal P. Then the scaled vector $(PdXP^t, y, P^{-t}ZP^{-1})$ is the Gauss-Newton direction at the iterate $(\widetilde{X}, \widetilde{y}, \widetilde{Z})$ for the scaled problem.

Proof The Gauss-Newton direction can be obtained from the normal Eq. (2.14) since the Jacobian is full-rank. The normal equations for the scaled problem are

$$(\widetilde{\mathcal{A}}^*\widetilde{\mathcal{A}} + \widetilde{\mathcal{Z}}^*\widetilde{\mathcal{Z}})(dX) + (\widetilde{\mathcal{Z}}^*\widetilde{\mathcal{X}})(dZ) = -(\widetilde{\mathcal{A}}^*\widetilde{f}_p + \widetilde{\mathcal{Z}}^*\widetilde{F}_c), \tag{2.31a}$$

$$\widetilde{\mathcal{A}}\widetilde{\mathcal{A}}^*(dy) + \widetilde{\mathcal{A}}(dZ) = -(\widetilde{\mathcal{A}}\widetilde{F}_d), \tag{2.31b}$$

$$\widetilde{\mathcal{X}}^*\widetilde{\mathcal{Z}}(dX) + \widetilde{\mathcal{A}}^*(dy) + (I + \widetilde{\mathcal{X}}^*\widetilde{\mathcal{X}})(dZ) = -(\widetilde{F}_d + \widetilde{\mathcal{X}}^*\widetilde{F}_c). \tag{2.31c}$$

Substitute the scaled vector $(\widetilde{dX}, \widetilde{dy}, \widetilde{dZ})$ into the left-hand side of (2.31c) to get,

$$\begin{aligned}
&\widetilde{\mathcal{X}}^* \widetilde{Z}(dX) + \widetilde{\mathcal{A}}^*(dy) + (I + \widetilde{\mathcal{X}}^* \widetilde{\mathcal{X}})(dZ) \\
&\quad = \frac{1}{2}\{P^{-t} Z P^{-1} P dX P^t P X P^t \\
&\qquad + P X P^t P^{-t} Z P^{-1} P dX P^t\} \\
&\qquad + \sum P^{-t} A_i P^{-1} dy_i + P^{-t} dZ P^{-1} \\
&\qquad + \frac{1}{2}\{P^{-t} dZ P^{-1} P X P^t P X P^t \\
&\qquad\quad + P X P^t P X P^t P^{-t} dZ P^{-1}\} \\
&\quad = P^{-t}\left\{\frac{1}{2}\{Z dX X + X Z dX\} + \sum A_i dy_i \right. \\
&\qquad\quad \left. + dZ + \frac{1}{2}\{dZ X X + X X dZ\}\right\} P^{-1} \\
&\quad = -P^{-t}\{F_d + \mathcal{X}^* F_c\} P^{-1} \\
&\quad = -P^{-t}\left\{F_d + \frac{1}{2}\{F_c X + X F_c\}\right\} P^{-1} \\
&\quad = -\left\{P^{-t} F_d P^{-1} \right. \\
&\qquad\quad \left. + \frac{1}{2}\{P^{-t} F_c P^{-1} P X P^t + P X P^t P^{-t} F_c^t P^{-1}\}\right\} \\
&\quad = -\{\widetilde{F_d} + \widetilde{\mathcal{X}}^* \widetilde{F_c}\}.
\end{aligned}$$

Therefore $(\widetilde{dX}, \widetilde{dy}, \widetilde{dZ})$ satisfies (2.31c). Similar computations show that it also satisfies (2.31a)–(2.31b) and we can conclude that the Gauss-Newton direction is Q-scale invariant. ∎

3. CONSTRAINED LEAST SQUARES

The efficient calculation of our GN direction is still an open problem. This has to involve a clever implementation of a QR algorithm. Therefore, we present a constrained GN, denoted RGN, direction that is easier to calculate, see (3.34). The RGN direction still maintains most of the properties of the GN direction.

3.1. Basic Elimination

To reduce the execution time of the least-square solver, we can algebraically eliminate some variables using block elimination. The order of elimination is different from that used in LP or other SDP codes, where dX is eliminated first from the linearized complementarity condition. One advantage that we have is that we never invert Z.

It is simplest to first force dual feasibility and eliminate dZ from (2.17), which, we recall, is of size $(t(n)+m+n^2) \times (t(n)+m+t(n))$. We use

$$dZ = -\mathcal{A}^*(dy) - F_d,$$

to obtain the $(m+n^2) \times (t(n)+m)$ reduced *dual feasible* system

$$\begin{bmatrix} \mathcal{A} & 0 \\ \mathcal{Z} & -\mathcal{X}\mathcal{A}^* \end{bmatrix} \begin{bmatrix} dX \\ dy \end{bmatrix} \begin{bmatrix} -f_p \\ F_d X - F_c \end{bmatrix}. \tag{3.32}$$

The upper part of this equation is still sparse and can therefore be treated separately once for all iterations. Moreover, since the backsolve for dZ is a sum of matrices it is inexpensive and exploits any sparsity in $\mathcal{A}^*$.

A smaller system, everything else being equal, is faster to solve than a large one. In the absence of specific information about the structure of the problem, we pursue this idea further since it is possible to eliminate more variables and obtain an even smaller system. For a specific problem with known sparsity structure, maxcut for example, it is possible to eliminate variables while maintaining sparsity. In the general case we discuss here, the major bottleneck being the size of the system, we aim at a reduction in size with little concern for sparsity.

We now consider the primal constraint operator $\mathcal{A}$, which we write as

$$\mathcal{A} = [\mathcal{A}_B, \mathcal{A}_N],$$

where the $\mathcal{A}_B$ is a subset of size $m \times m$ of $\mathcal{A}$ that is easily invertible. We now divide $\mathcal{Z}, \mathcal{X}, dX$ and dZ in the same manner to obtain

$$dX_B = -\mathcal{A}_B^{-1} f_p - \mathcal{A}_B^{-1} \mathcal{A}_N dX_N, \tag{3.33}$$

which we substitute back into the dual feasible system (3.32), to obtain the $(n^2 \times t(n))$ system

$$[-\mathcal{Z}_B\mathcal{A}_B^{-1}\mathcal{A}_N + \mathcal{Z}_N, -\mathcal{X}\mathcal{A}^*]\begin{bmatrix} dX_N \\ dy \end{bmatrix} = \mathcal{X}F_d - F_c + \mathcal{Z}_B\mathcal{A}_B^{-1}f_p. \quad (3.34)$$

Albeit more costly to setup and backsolve than (3.32), this system, for large dense problems is faster to solve because the bottleneck is the least square solve. Moreover, the rank results established for the large system still hold.

COROLLARY 3.1 *Theorems* 2.1 *and* 2.2 *hold with system* (2.17) *replaced by system* (3.34).

Proof The proofs follow the same pattern as in the above mentioned theorems. For example, at each iteration we have

$$\mathcal{A}(dX) = 0, \quad \mathcal{A}^*(dy) + dZ = 0$$

and

$$ZdX + dZX = 0.$$

Therefore we can apply (2.19) and (2.20). ∎

Remark 3.1 Another approach is to do elimination in the same vein as is done in LP but preserving the least squares properties. This can be done using the explicit structure of the operators in the optimality conditions, see [34, Section 3.2].

4. NUMERICAL TESTS

4.1. Comparisons Between Directions

We first present some numerical results comparing the constrained GN directions to the well-known Alizadeh-Haeberly-Overton [6] (AHO), Helmberg-Rendl-Vanderbei-Wolkowicz [24]/Kojima-Shindoh-Hara [32]/Monteiro [36] (HKM) and Nesterov-Todd [42] (NT) directions. It was not our intention here to develop a completely new and efficient algorithm but rather to isolate and highlight the effects of the search directions. All tests were performed in Matlab using the SDPT3 code

[60] from Todd, Toh, Tutuncu. We modified the mainline of SDPT3 to allow further options, namely the GN and the constrained GN directions. The logic relating to step lengths and the updating of the centering parameters were untouched to allow a reasonable comparison of the relative value of the GN directions.

The first comparison (Fig. 1) illustrates the decrease in both the duality gap and in the infeasibility on random semidefinite programs. The problem has size ($n = 10$, $m = 15$) and the algorithm used a predictor-corrector approach on all directions. The restricted GN direction, seems numerically stable, and reduces the duality gap as much as the AHO direction in about as many iterations, typically one more. We should note here that the GN direction performs in exactly the same manner as the restricted GN on those problems, albeit more slowly.

By the constrained GN direction, we mean the elimination of dy and of part of dX to reduce the system size to $(n^2 \times t(n))$ by forcing the primal and dual feasibility equations. That is, we solved (3.34). On a random SDP, as is the case in this instance, with unknown and dense structure, this elimination will perform well.

The graph shows that the restricted GN direction performs very well on such problems in terms of the accuracy of the solutions. It does consistently as well or better than AHO.

The following (Figs. 2 and 3) are somewhat unfair tests meant to highlight the robustness of the GN directions with respect to the scaling of the problem. Again, the basic problem is randomly generated, but then the right-hand side is scaled *after the initial point is*

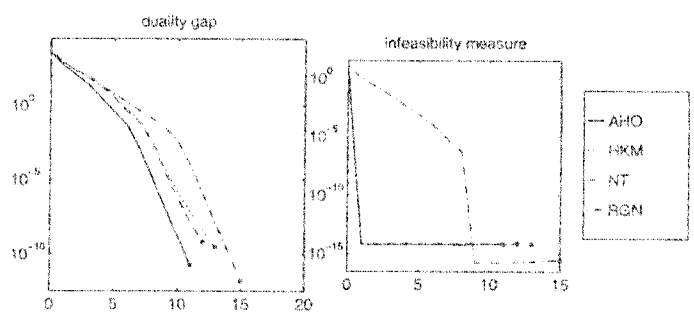

FIGURE 1 Random semidefinite program.

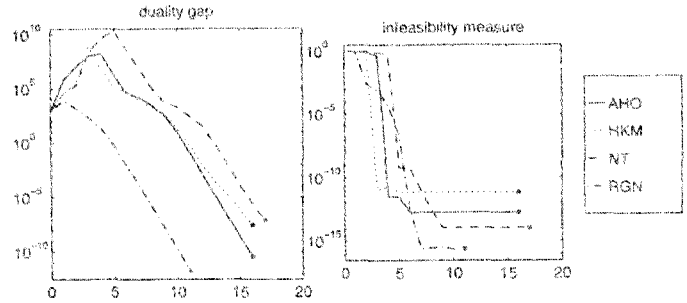

FIGURE 2 $b \leftarrow b*100$.

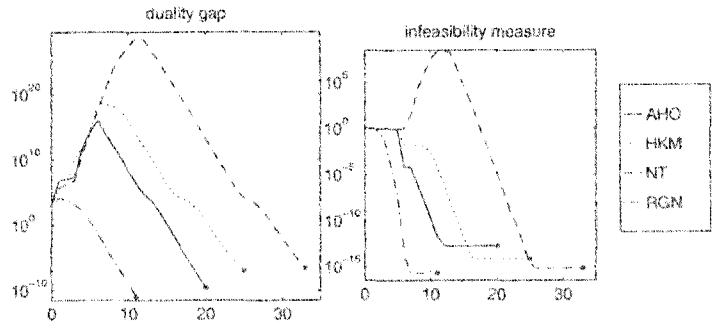

FIGURE 3 $b \leftarrow b*500$.

chosen. The problem therefore becomes difficult because the starting point is badly centered. The restricted GN direction is again used.

Table I compares the number of iterations and the accuracy of the solutions of the AHO and RGN directions on random problems. The columns represent, in order, the iteration count, primal feasibility

TABLE I Random problems with $n = 7$, $m = 28$

AHO				*RGN*			
Iter	*P feas*	*D feas*	*Gap*	*Iter*	*P feas*	*D feas*	*Gap*
9	1.2e−14	1.7e−15	4.4e−14	10	2.4e−16	1.8e−15	4.0e−15
9	8.0e−16	2.6e−16	5.8e−14	10	1.2e−16	2.6e−16	4.6e−15
9	1.8e−15	3.7e−16	1.1e−13	10	3.3e−16	5.3e−16	1.0e−14
9	1.6e−15	3.4e−16	7.4e−14	10	5.3e−16	3.9e−16	7.0e−15
9	2.8e−15	5.8e−16	4.5e−14	10	1.4e−16	4.8e−16	2.4e−15
9	1.0e−15	2.8e−16	8.0e−14	10	2.2e−16	1.9e−16	7.3e−15

TABLE II Equipartitioning problems with no slater point

	AHO			*GN*		
Size (n)	*P feas*	*D feas*	*Gap*	*P feas*	*D feas*	*Gap*
9	5.7e−04	2.1e−10	5.4e−07	1.8e−06	7.8e−12	6.7e−07
10	7.8e−04	2.8e−10	1.4e−06	4.9e−07	2.2e−12	2.1e−07
11	2.0e−04	1.4e−09	8.6e−07	3.1e−06	3.2e−11	8.9e−07
12	9.2e−04	1.6e−09	2.3e−06	1.5e−06	1.4e−11	5.4e−07
15	6.6e−04	6.0e−09	3.4e−06	3.7e−06	2.8e−11	8.9e−07
	AHO			RGN		
20	6.8e-04	1.2e-08	2.9e-05	2.0e-14	6.7e-11	9.7e-07

($\|b - \mathcal{A}(X)\|/\|b\|$), dual feasibility ($\|\mathcal{A}^*(y) + Z - C\|/\|C\|$), and the duality gap ($\langle X, Z\rangle$). The additional digit of accuracy for the RGN case is most likely due to the better conditioning of the system.

As far as work per iteration is concerned, the major cost of the restricted GN direction involves the solution of a linear least squares problem of size $n^2 \times t(n)$. This is in contrast to AHO, where the system to solve is of size m, and is then followed by the solution of a Lyapunov equation of order $t(n)$. Therefore, for problems with few constraints, the RGN iterations are more costly and they become competitive as m increases.

The final comparison highlights one the main strengths of the GN directions: the numerical accuracy of the solution they provide. The problems under consideration fail to have strictly feasible points in the primal space. That is, the Slater condition fails. The primal constraint force the diagonal elements of the solution to be all ones and the sum of the elements to be zero, making these problems numerically fairly difficult. The primal objective matrix is randomly chosen. Table II displays feasibility and the gap of the AHO and of GN directions on some of these problems. All stopping criteria were discarded (except for a large maximum iteration count) and the algorithms were left to run to obtain the best possible solution given their respective search directions.

5. FUTURE DIRECTIONS

All the search directions of the Monteiro–Zhang family used in interior-point algorithms for semidefinite programming can be derived

as the AHO direction of a scaled problem. In this sense, the symmetrization and the resulting AHO system is the basic direction finding paradigm in currently popular algorithms.

In this paper we have introduced an alternative approach. Instead of projecting onto the symmetric space, we view the modified, perturbed optimality conditions as a nonlinear operator to which we apply the Gauss–Newton method. The result is an over-determined system of equations and a new family of directions based on the least-squares solution of this system.

We have investigated some of the properties of the resulting unscaled direction, which we call GN, as it compares to AHO and have found that the GN direction is well-defined, coincides with the standard direction in the LP case, coincides with AHO on the central path in the SDP case, and is invariant under orthogonal scaling of the data and affine scaling of the space. Moreover, the conditioning of the system of equations allows a numerically robust implementation, as exhibited by our preliminary experiments.

The current research aims at establishing polytime convergence of algorithms based on the GN direction, investigating the effect of various scalings and reducing the computational cost while exploiting sparsity. We are also considering different but related families of directions. First, we are exploring the use of more Hessian information, possibly within an indefinite trust-region framework and second, to reduce the size of the GN system, we have tried to eliminates redundant equations from the optimality conditions.

For, surprisingly, there are redundant nonlinear equations in (2.12). We can remove the lower triangular part from the nonlinear block to get the equivalent system

$$\mathcal{T}(ZX - \mu I) = 0, \tag{5.35}$$

where the linear operator $\mathcal{T}: \mathbb{M}^n \to \mathbb{S}^n$, ignores the strictly lower triangular part of the matrix, *i.e.*, the i,j components are

$$(\mathcal{T}(W))_{ij} = \begin{cases} W_{ij} & \text{if } i \leq j \\ W_{ji} & \text{if } i > j \end{cases}.$$

Note $\mathcal{T}(M) = PMP^{-1}$ implies $P^{-1}\mathcal{T}(M)P = M$ which shows that this does not fall under the general symmetrization framework (1.8) of the

Monteiro-Zhang family yet the resulting system is now square; the domain and range space are the same. Also note that $\mathcal{T}$ is the composition of two linear maps: first apply the simple coordinate (orthogonal) linear projection onto the subspace of upper triangular matrices in the space of $n \times n$ matrices; then apply the one–one mapping to the space of symmetric matrices formed by duplicating the upper triangular part in the lower triangular part. We have $\mathcal{T}(\mathcal{T}(W)) = \mathcal{T}(W)$, *i.e.*, $\mathcal{T}$ is idempotent (a projection). In fact, if we identify the upper triangular matrices with $\mathbb{S}^n$, then we can consider $\mathcal{T}$ as orthogonal as well.

The modified optimality conditions become

$$\begin{pmatrix} C - Z - \mathcal{A}^* y \\ b - \mathcal{A}(X) \\ \mathcal{T}(ZX - \mu I) \end{pmatrix} = 0. \tag{5.36}$$

This transformation does not introduce new nonlinearities but does guarantee that we map between the same spaces. In fact, and this is the surprise, we can prove (see [34] Lemma 2.1) that we do not lose information in the optimality conditions when we only consider the upper triangular parts.

References

[1] Alfakih, A., Khandani, A. and Wolkowicz, H. (1999). Solving Euclidean distance matrix completion problems *via* semidefinite programming. *Comput. Optim. Appl.*, 12(1–3), 13–30. Computational optimization – a tribute to Olvi Mangasarian, Part I.

[2] Alizadeh, F., Combinatorial optimization with interior point methods and semidefinite matrices. *Ph.D. Thesis*, University of Minnesota, 1991.

[3] Alizadeh, F. (1995). Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5, 13–51.

[4] Alizadeh, F., Haeberly, J.-P., Nayakkankuppam, M. V., Overton, M. L. and Schmieta, S., SDPpack user's guide – version 0.9 Beta. *Technical Report TR 1997–737*, Courant Institute of Mathematical Sciences, NYU, New York, NY, June, 1997.

[5] Alizadeh, F., Haeberly, J.-P. A. and Overton, M. L. (1997). Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, 77, 111–128.

[6] Alizadeh, F., Haeberly, J.-P. A. and Overton, M. L. (1998). Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM J. Optim.*, 8, 746–768.

[7] Ben-israel, A. and Greville, T. N. E., *Generalized Inverses: Theory and Applications*. Wiley-Interscience, 1974.

[8] Adi Ben-israel, Newton's method with modified functions. In: *Recent Developments in Optimization Theory and Nonlinear Analysis* (Jerusalem, 1995), Volume 204 of *Contemp. Math.*, pp. 39–50. Amer. Math. Soc., Providence, RI, 1997.

[9] Bjorck, A., *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 1996.
[10] Boyd, S., Balakrishnan, V., Feron, E. and El Ghaoui, L. (1993). Control system analysis and synthesis *via* linear matrix inequalities. *Proc. ACC*, pp. 2147–2154.
[11] Boyd, S., El Ghaoui, L., Feron, E. and Balakrishnan, V., *Linear Matrix Inequalities in System and Control Theory*, Volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June, 1994.
[12] Brixius, N., Potra, F. A. and Sheng, R., SDPHA: a matlab implementation of homogeneous interior-point algorithms for semidefinite programming. *Technical Report*, University of Iowa, Iowa City, IA, 1998.
[13] Dennis, J. E. Jr. and Robert B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.
[14] El-bakry, A. S., Tapia, R. A., Tsuchiya, T. and Zhang, Y. (1996). On the formulation and theory of the Newton interior-point method for nonlinear programming. *J. Optim. Theory Applic.*, **89**, 507–541.
[15] Elhallabi, M., An inexact hybrid algorithm for nonlinear systems of equations. *Technical Report CRPC-TR94372*, Department of Mathematical Sciences, Rice University, Houston, TX, 1994.
[16] Elhallabi, M. and Tapia, R. A., An inexact trust-region feasible-point algorithm: globalization of Gauss-Newton method. *Technical Report TR93-43*, Department of Mathematical Sciences, Rice University, Houston, TX, 1993.
[17] Elhallabi, M. and Tapia, R. A., An inexact trust-region feasible-point algorithm for nonlinear systems of equalities and inequalities. *Technical Report CRPC-TR95543*, Department of Mathematical Sciences, Rice University, Houston, TX, 1995.
[18] Fletcher, R., Practical Methods of Optimization, Vol. 1, *Unconstrained Optimization*. John Wiley & Sons, New York, NY, 1980.
[19] Fujisawa, K., Kojima, M. and Nakata, K., SDPA semidefinite programming algorithm. *Technical Report*, Dept. of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1995.
[20] Gerlach, J. (1994). Accelerated convergence in Newton's method. *SIAM Review*, **36**, 272–276.
[21] Goemans, M. X. (1997). Semidefinite programming in combinatorial optimization. *Math. Programming*, **79**, 143–162.
[22] Gu, M., On primal-dual interior point methods for semidefinite programming. *Technical Report CAM 97-12*, Department of Mathematics, University of California, Los Angeles, CA 90095-1555, 1997.
[23] Helmberg, C., An interior-point method for semidefinite programming and max-cut bounds. *Ph.D. Thesis*, Graz University of Technology, Austria, 1994.
[24] Helmberg, C., Rendl, F., Vanderbei, R. J. and Wolkowicz, H. (1996). An interior-point method for semidefinite programming. *SIAM J. Optim.*, **6**(2), 342–361.
[25] Horn, R. A. and Johnson, C. R., *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.
[26] Johnson, C. R., Kroschel, B. and Wolkowicz, H. (1998). An interior-point method for approximate positive semidefinite completions. *Comput. Optim. Appl.*, **9**(2), 175–190.
[27] Karisch, S., Nonlinear Approaches for Quadratic Assignment and Graph Partition Problems. *Ph.D. Thesis*, University of Graz, Graz, Austria, 1995.
[28] De Klerk, E., Interior point methods for semidefinite programming. *Ph.D. Thesis*, Delft University, 1997.
[29] De Klerk, E., Peng, J., Roos, C. and Terlaky, T., A scaled Gauss-Newton primal-dual search direction for semidefinite programming. *Technical Report*, Delft University of Technology, Faculty of Technical Mathematics and Informatics, Delft, The Netherlands, 1999.

[30] Kojima, M., Megiddo, N. and Mizuno, S. (1993). A primal-dual infeasible-interior-point algorithm for linear programming. *Math. Programming*, **61**, 263–280.

[31] Kojima, M., Mizuno, S. and Yoshise, A., A primal-dual interior point algorithm for linear programming. In: Megiddo, N. Ed., *Progress in Mathematical Programming: Interior Point and Related Methods*, pp. 29–47. Springer Verlag, New York, 1989.

[32] Kojima, M., Shindoh, S. and Hara, S. (1997). Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7(1), 86–125.

[33] Kruk, S., Semidefinite programming applied to nonlinear programming. *Master's Thesis*, University of Waterloo, 1996.

[34] Kruk, S., Muramatsu, M., Rendl, F., Vanderbei, R. J. and Wolkowicz, H., The Gauss-Newton direction in linear and semidefinite programming. *Technical Report CORR 98-16*, University of Waterloo, Waterloo, Canada, 1998.

[35] Megiddo, N., Pathways to the optimal set in linear programming. In: Megiddo, N. Ed., *Progress in Mathematical Programming: Interior Point and Related Methods*, pp. 131–158. Springer Verlag, New York, 1989. Identical version In: *Proceedings of the 6th Mathematical Programming Symposium of Japan*, Nagoya, Japan, pp. 1–35, 1986.

[36] Monteiro, R. D. C. (1997). Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7(3), 663–678.

[37] Monteiro, R. D. C. and Zanjácomo, P. R. (1997). A note on the existence of the Alizadeh-Haeberly-Overton direction for semidefinite programming. *Math. Programming*, 78(3, Ser. A), 393–396.

[38] Monteiro, R. D. C. and Zhang, Y. (1998). A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Programming*, **81**(3, Ser. A), 281–299.

[39] Moré, J. J. and Sorensen, D. C. (1983). Computing a trust region step. *SIAM J. Sci. Statist. Comput.*, 4, 553–572.

[40] Nazareth, L., Computer solution of linear programs: non-simplex algorithms II. *Technical Report*, Dept. of Pure and Applied Mathematics, Washington State University, Pullman, 1996.

[41] Nesterov, Y. E. and Nemirovski, A. S., *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Publications. SIAM, Philadelphia, USA, 1994.

[42] Nesterov, Y. E. and Todd, M. J. (1997). Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, **22**(1), 1–42.

[43] Pardalos, P. and Wolkowicz, H. Eds., *Topics in Semidefinite and Interior-Point Methods*, The Fields Institute for Research in Mathematical Sciences, Communications Series, Providence, RI, 1998. American Mathematical Society.

[44] Pataki, G., Cone Programming and Eigenvalue Optimization: Geometry and Algorithms. *Ph.D. Thesis*, Carnegie Mellon University, Pittsburgh, PA, 1996.

[45] Pataki, G. and Tunçel, L., On the generic properties of convex optimization problems in conic form. *Technical Report CORR 97-16*, Department of Combinatorics and Optimization, Waterloo, Ont, 1997.

[46] Powell, M. J. D., The use of band matrices for second derivative approximations in trust region algorithms. In: *Advances in Nonlinear Programming* (Beijing, 1996), pp. 3–28. Kluwer Acad. Publ., Dordrecht, 1998.

[47] Ramana, M. V., An Algorithmic Analysis of Multiquadratic and Semidefinite Programming Problems. *Ph.D. Thesis*, Johns Hopkins University, Baltimore, Md, 1993.

[48] Ramana, M. V., Tunçel, L. and Wolkowicz, H. (1997). Strong duality for semidefinite programming. *SIAM J. Optim.*, 7(3), 641–662.

[49] Rendl, F. (1999). Semidefinite programming and combinatorial optimization. *Appl. Numer. Math.*, **29**, 255–281.

[50] Rendl, F., Vanderbei, R. J. and Wolkowicz, H. (1995). Max–min eigenvalue problems, primal-dual interior point algorithms, and trust region subproblems. *Optim. Methods Softw.*, **5**, 1–16.
[51] Rendl, F. and Wolkowicz, H. (1997). A semidefinite framework for trust region subproblems with applications to large scale minimization. *Math. Programming*, **77**(2, Ser. B), 273–299.
[52] Shapiro, A. (1997). First and second order analysis of nonlinear semidefinite programs. *Math. Programming*, **77**, 301–320.
[53] Shapiro, A. and Fan, M. K. H. (1995). On eigenvalue optimization. *SIAM J. Optim.*, **5**, 552–569.
[54] Sheng, R. and Potra, F. A., Nonsymmetric search directions for semidefinite programming. *Technical Report*, University of Iowa, Iowa City, IA, 1997.
[55] Sorensen, D. C. (1982). Newton's method with a model trust region modification. *SIAM J. Numer. Anal.*, **19**, 409–426.
[56] Sturm, J. F., Primal-Dual Interior Point Approach to Semidefinite Programming. *Ph.D. Thesis*, Erasmus University Rotterdam, 1997.
[57] Sun, C., Dealing with dense rows in the solution of sparse linear least squares problems. *Technical Report TR95-227*, Department of Computer Science, Cornell University, Ithaca, NY, 1995.
[58] Terlaky, T. Ed., *Interior point methods of mathematical programming*. Kluwer Academic Publishers, Dordrecht, 1996.
[59] Todd, M. J. (1999). A study of search directions in primal-dual interior-point methods for semidefinite programming. *Optim. Methods Softw.*, **11** and **12**, 1–46.
[60] Todd, M. J., Toh, K. C. and Tutuncu, R. H., A MATLAB software package for semidefinite programming. *Technical Report*. School of OR and IE, Cornell University, Ithaca, NY, 1996.
[61] Todd, M. J., Toh, K. C. and Tutuncu, R. H. (1998). On the Nesterov-Todd direction in semidefinite programming. *SIAM J. Optim.*, **8**(3), 769–796.
[62] Toh, K. C., Search directions for primal-dual interior point methods in semidefinite programming. *Technical Report*, Department of Mathematics, National University of Singapore, Singapore, 1997.
[63] Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Rev.*, **38**(1), 49–95.
[64] Wolkowicz, H., Saigal, R. and Vandenberghe, L. Eds., *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, MA, 2000. xxvi + 654 pages.
[65] Wright, S., *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, Pa. 1996.
[66] Ye, Y., *Interior Point Algorithms: Theory and Analysis*. Wiley-Interscience series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 1997.
[67] Zhang, Y., Basic equalities and inequalities in primal-dual interior-point methods for semidefinite programming. *Technical Report*, Department of Mathematics and Statistics, University of Maryland Baltimore County, 1995.
[68] Zhao, Q., Semidefinite Programming for Assignment and Partitioning Problems. *Ph.D. Thesis*, University of Waterloo, 1996.