LECTURE 9. The Primal-Dual Method for SDP

Let us rephrase our previous analysis on the primal-dual method for LP, starting from the notion of *scaling*. Actually, the idea of scaling for solving linear programming was first considered by Dikin in 1967: it predates Karmarkar's work which initiated the burst of research interest on interior point methods. Dikin's idea is clear. Consider a linear program

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}} x \\ \text{subject to} & Ax = b \\ & x \in \Re^n_+. \end{array} \tag{1}$$

Suppose that $x_0 \in \Re^n_{++}$ is a given strongly feasible solution for (1). Let $X_0 := \operatorname{diag}(x_0)$, i.e., X^0 is a diagonal matrix whose ith diagonal component is simply the ith component of x_0. An interesting property of the convex cone $\Re^n_+$ is that it is invariant under the linear transformation: $X_0(\Re^n_+) = \Re^n_+$, or equivalently, $(X_0)^{-1}(\Re^n_+) = \Re^n_+$. Dikin's affine scaling algorithm is based on such a diagonal scaling transformation, which helps to relocate the iterate to a convenient position.

For instance,

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}} X_0 x \\ \text{subject to} & A X_0 x = b \\ & x \in \Re^n_+, \end{array}$$

is, on the one hand equivalent to (1), and on the other hand contains a feasible solution e — the all-one vector. In general, a linear program can be scaled by a positive diagonal matrix D as

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}} D x \\ \text{subject to} & A D x = b \\ & x \in \Re^n_+. \end{array} \tag{2}$$

This scaling can be understood as a variable transformation $x := D^{-1}x$. There is of course no reason to confine oneself to the primal problem only. It is natural to consider the dual of (1),

$$\begin{array}{ll} \text{maximize} & b^{\mathrm{T}}y \\ \text{subject to} & A^{\mathrm{T}}y + s = c \\ & s \in \Re^n_+. \end{array} \tag{3}$$

In fact, the dual of the scaled primal is a scaled dual problem itself:

$$\begin{array}{ll} \text{maximize} & b^{\mathrm{T}}Dy \\ \text{subject to} & DA^{\mathrm{T}}y + s = Dc \\ & s \in \Re^n_+, \end{array} \tag{4}$$

with the transformation $(y, s) := (y, Ds)$.

Interestingly, there is a particular scaling matrix D, that transforms a given primal-dual strong feasible solution to an identical location in the same cone. That is, for any given primal-dual pair $(x, s) \in \Re^n_{++} \times \Re^n_{++}$, there exists a scaling vector $d \in \Re^n_{++}$, such that $D = \text{diag}(d)$ and $D^{-1}x = Ds$. It is easy to see that such d vector is unique and can be computed as $d = (XS^{-1})^{0.5}e$. For ease of referencing, let us denote $D^{-1}x = Ds =: v \in \Re^n_{++}$. Clearly, $(d, v) = ((XS^{-1})^{0.5}e, (XS)^{0.5}e)$ is uniquely determined by the pair (x, s).

How about their relative positions to optimality (i.e. the duality gap between them)? We can easily compute that

$$c^{\mathrm{T}}x - b^{\mathrm{T}}y = x^{\mathrm{T}}s = (D^{-1}x)^{\mathrm{T}}Ds = v^{\mathrm{T}}v = \|v\|^2.$$

In other words, the duality gap is invariant under this transformation. Moreover, the duality gap can now be conveniently measured as the norm of vector v. Since we aim at reducing the duality gap, it is natural to update x and s so as to ensure that their corresponding v has small norm.

How about SDP? Surprisingly, a large part of the above analysis can be said almost literally the same for SDP. First of all, we note that the cone $\mathcal{S}_+^{n\times n}$ is invariant under the linear transformation $X \to LXL^{\mathrm{T}}$ where L is a nonsingular matrix. In this connection, the above transformation can be considered as *matrix scaling*. For a primal SDP problem

$$\begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i,\ i = 1, ..., m \\ & X \succeq 0, \end{array} \tag{5}$$

the scaled problem is

$$\begin{array}{ll} \text{minimize} & (L^{\mathrm{T}}CL) \bullet X \\ \text{subject to} & (L^{\mathrm{T}}A_iL) \bullet X = b_i,\ i = 1, ..., m \\ & X \succeq 0. \end{array} \tag{6}$$

The scaling is: $X \to LXL^{\mathrm{T}}$. With the same scaling matrix L, its dual SDP problem

$$\begin{array}{ll} \text{maximize} & b^{\mathrm{T}} y \\ \text{subject to} & \sum_{i=1}^{m} y_i A_i + Z = C,\ i = 1, ..., m \\ & Z \succeq 0, \end{array} \tag{7}$$

is scaled to

$$\begin{array}{ll} \text{maximize} & b^{\mathrm{T}} y \\ \text{subject to} & \sum_{i=1}^{m} y_i L^{\mathrm{T}} A_i L + Z = L^{\mathrm{T}} C L,\ i = 1, ..., m \\ & Z \succeq 0. \end{array} \tag{8}$$

The dual transformation is $Z \to L^{-\mathrm{T}} Z L^{-1}$.

A natural question thus arises: is it possible to find a scaling matrix L so that the pair X and (y, Z) is transformed to V and (y, V), just like linear programming? The answer is: Yes!

Lemma 1 *For any $X \succ 0$ and $Z \succ 0$ there is a nonsingular matrix L such that*

$$LXL^{\mathrm{T}} = L^{-\mathrm{T}} Z L^{-1} =: V. \tag{9}$$

Proof. $LXL^{\mathrm{T}} = L^{-\mathrm{T}} Z L^{-1} \iff L^{\mathrm{T}} L X L^{\mathrm{T}} L = Z \iff$
$X^{1/2} L^{\mathrm{T}} L X^{1/2} X^{1/2} L^{\mathrm{T}} L X^{1/2} = X^{1/2} Z X^{1/2} \iff$
$X^{1/2} L^{\mathrm{T}} L X^{1/2} = (X^{1/2} Z X^{1/2})^{1/2} \iff$
$L^{\mathrm{T}} L = X^{-1/2} (X^{1/2} Z X^{1/2})^{1/2} X^{-1/2}.$ □

1. In the SDP case, the V matrix is *not* unique! As a matter of fact, if the scaling matrix L transforms X and (y, Z) into V and (y, V), then for any orthonormal matrix Q, QL will transform X and (y, Z) into $Q^{\mathrm{T}} V Q$ and $(y, Q^{\mathrm{T}} V Q)$. In other words, what is uniquely determined is the set of the eigenvalues of V, but not the matrix itself.

2. To follow the previous point, it is clear that the duality gap, $\|V\|_F^2 = \mathrm{tr}\,(XZ)$, is invariant of the choice of the scaling matrix L, as long as $L^{\mathrm{T}} L = Z^{1/2}(Z^{1/2} X Z^{1/2})^{-1/2} Z^{1/2}$.

3. Since the choice of L is not unique, one choice is to make it symmetric, i.e.,

$$L = (Z^{1/2}(Z^{1/2} X Z^{1/2})^{-1/2} Z^{1/2})^{1/2}.$$

There is a simpler way to compute the primal-dual symmetric transformation.

Primal-Dual Symmetric Transformation

Compute a Cholesky factor of X:

$$X = GG^{\mathrm{T}}.$$

Compute a spectral decomposition:

$$G^{\mathrm{T}}ZG = Q^{\mathrm{T}}\Lambda Q.$$

Assemble L and V:

$$L = \Lambda^{1/4}QG^{-1}$$

and

$$V = \Lambda^{1/2}.$$

It is remarkable that the primal-dual transformation procedure is a two-way traffic, up to a certain degree. This fact is stated below and was first shown by Sturm and Zhang.

Theorem 1 *For any given* $W \in \mathcal{S}^{n\times n}_{++}$ *there exists a pair of feasible solutions,* $X \succ 0$ *feasible to (5) and* $Z \succ 0$ *feasible to (7), and a scaling matrix* L*, such that*

$$LXL^{\mathrm{T}} = L^{-\mathrm{T}} Z L^{-1} = W.$$

To prove this theorem we will need some preparations first. We will come back to this point later.

Since a pair of primal-dual strongly feasible solutions X and Z can be simultaneously transformed to a certain positive definite matrix V with $\|V\|_F^2 = \mathrm{tr}\,(XZ)$, it is natural to improve the solutions X and Z, in the sense that the corresponding V matrix is more desirable. The desirability can be roughly sated as: V needs to have a small matrix norm, and V preferably has a good condition number. This leads to the following generic primal-dual method for SDP.

Primal-dual method based on the V-space

Step 0. Start with a primal-dual interior solution pair (X, Z).

Step 1. Compute the corresponding V matrix.

Step 2. Change the pair (X, Z) so as to improve the corresponding V matrix.

Step 3. Return to **Step 1** if the improved solution is not good enough.

In **Step 2**, usually the Newton style linearization is applied. We will discuss this point in more details later.

There is one particular V matrix that stands in the center of all attentions: positive multiple of the identity matrix. Indeed it is so central that it is called the *analytic central path* for SDP. That is, for any $\mu > 0$ there is a unique pair of primal-dual feasible solutions $X(\mu)$ and $Z(\mu)$ such that

$$X(\mu)Z(\mu) = \mu I.$$

In this case, the corresponding V is $\sqrt{\mu} I$. The primal central path is then defined as $\{X(\mu) \mid \mu > 0\}$, and the dual central path as $\{Z(\mu) \mid \mu > 0\}$.

The existence of the central path solutions for SDP can be shown without the V-space notion. Indeed one only need consider, for fixed $\mu > 0$, the following problem with a barrier term for (5):

$$\begin{array}{ll} \text{minimize} & C \bullet X - \mu \log \det X \\ \text{subject to} & A_i \bullet X = b_i,\ i = 1, ..., m. \end{array}$$

Under the condition that both (5) and (7) are strongly feasible, then the above problem is solvable with a unique optimal solution $X(\mu)$. Writing out the KKT optimality condition for the problem, where $y_i(\mu)$ is used as the Lagrangian multiplier for the ith constraint, yields

$$
\begin{aligned}
& C - \mu X(\mu)^{-1} - \textstyle\sum_{i=1}^m y_i(\mu) A_i = 0 \\
& X(\mu) \succ 0 \\
& A_i \bullet X(\mu) = b_i,\ i = 1, ..., m.
\end{aligned}
$$

Define $Z(\mu) := C - \sum_{i=1}^m y_i(\mu) A_i$ leads to a feasible solution for (7), with $X(\mu)Z(\mu) = \mu I$.

In this case, $X(\mu)$ and $Z(\mu)$ are commutable. Now we know that the central path solutions exist, and so it makes sense to follow the central path while driving μ to zero. Specifically, suppose that we scale the primal-dual solutions X and (y, Z) to V and (y, V). The resulting SDP problem pairs are (6) and (8). Let $\mu = \mathrm{tr}\,(XZ)/n = \|V\|_F^2/n$. Now we wish to make a move from V to $V + D_X$ for (6), and make a move from V to $V + D_Z$ for (8) simultaneously. Naturally, the new solutions will need to be feasible. We hope that after the update, the new pair is closer to the solution on the central path with μ being shifted to $\gamma\mu$ $(\gamma < 1)$.

This leads to the desired equation

$$(V + D_X)(V + D_Z) = \gamma\mu I.$$

By dropping the nonlinear term $D_X D_Z$, the equation is simplified to

$$V^2 + D_X V + V D_Z = \gamma\mu I,$$

and by symmetry,

$$V^2 + V D_X + D_Z V = \gamma\mu I.$$

Adding up these two terms,

$$\frac{1}{2}\left((D_X + D_Z)V + V(D_X + D_Z)\right) = \gamma\mu I - V^2. \tag{10}$$

Matrix equations of the type

$$XAX + XC + C^{\mathrm{T}}X = B$$

are called the algebraic Riccati equations as we know. Lemma 1 is concerned with the case $C = 0$, and (10) is related to the case $A = 0$. Due to its simplicity, the solution to (10) can be written explicitly as

$$D_X + D_Z = \gamma\mu V^{-1} - V. \tag{11}$$

In fact this is the unique solution, since the linear system $DV + VD = 0$ has only trivial solution $D = 0$.

It is called the *Nesterov-Todd direction*. We will see later the connection with the so-called Nesterov-Todd scaling. Let

$$\mathcal{L} = \{X \in \mathcal{S}^{n\times n} \mid A_i \bullet X = 0,\ i = 1, ..., m\},$$

and

$$\mathcal{L}^{\perp} = \left\{ \sum_{i=1}^{m} y_i A_i \,\middle|\, \exists y \in \Re^m \right\}.$$

The feasibility requires that $D_X \in \mathcal{L}$ and $D_Z \in \mathcal{L}^{\perp}$. Since $\mathcal{L}$ and $\mathcal{L}^{\perp}$ form an orthogonal complement for $\mathcal{S}^{n\times n}$, hence D_X and D_Z are uniquely determined: D_X is the projection of $\gamma\mu V^{-1} - V$ on $\mathcal{L}$ and D_Z is the projection on $\mathcal{L}^{\perp}$.

Is it possible to apply Newton method directly on without the notion of V-space? Yes, it is possible. Let X and (y, Z) be a pair of primal-dual solution. We wish to make a move along ΔX and $(\Delta y, \Delta Z)$ such that

$$(X + \Delta X)(Z + \Delta Z) = \gamma\mu I.$$

In the spirit of the primal-dual interior point method for linear programming by Kojima, Mizuno, and Yoshise, one ignores the nonlinear cross term $\Delta X \Delta Z$ and obtains the equation

$$X(\Delta Z) + (\Delta X)Z = \gamma\mu I - XZ \tag{12}$$

and $\Delta X \in \mathcal{L}$, $\Delta Z \in \mathcal{L}^{\perp}$.

So far the situation looks quite similar to the V-space case. However, a crucial difference occurs at this point. Since XZ may not be symmetric, the number of equations arising from (12) is $n^2 + n(n+1)/2$ (including the requirement that $\Delta X \in \mathcal{L} \subseteq \mathcal{S}^{n\times n}$ and $\Delta Z \in \mathcal{L}^{\perp} \subseteq \mathcal{S}^{n\times n}$). However, since ΔX and ΔZ are symmetric matrices, in total they offer $n(n+1)$ free variables. Therefore, Equation (12) is in general unsolvable! Most natural remedy is of course to add more variables to make the system consistent and solvable. Here are some well known examples:

Kojima-Shindoh-Hara family search directions:

$$\begin{cases} X(\Delta Z + U_Z) + (\Delta X + U_X)Z = \gamma\mu I - XZ \\ U_X \text{ and } U_Z \text{ are skew symmetric matrices with } (U_X, U_Z) \in \mathcal{M} \\ \Delta X \in \mathcal{L},\ \Delta Z \in \mathcal{L}^{\perp}, \end{cases}$$

where $\mathcal{M}$ is a given subspace in the Cartesian product of $n \times n$ skew symmetric matrices with dimension $n(n-1)/2$. By adding U_X and U_Z as additional variables, the total number of variables has increased from $n(n+1)$ to $n(n+1) + 2\frac{n(n-1)}{2} = 2n^2$, while the number of equations has increased from $n^2 + n(n+1)/2$ to $n^2 + n(n+1)/2 + n(n-1)/2 = 2n^2$, hence a complete match. In this framework, the choice of $\mathcal{M}$ is still flexible, and some choices lead to the method of Helmberg et al. and Monteiro.

Alizadeh-Haeberly-Overton (AHO) direction:

$$\begin{cases} X(\Delta Z) + (\Delta X)Z + U = \gamma\mu I - XZ \\ \Delta X \in \mathcal{L},\ \Delta Z \in \mathcal{L}^\perp,\ U \text{ is skew symmetric.} \end{cases} \tag{13}$$

In this case, the number of variables has increased from $n(n+1)$ to $n(n+1) + n(n-1)/2 = n^2 + n(n+1)/2$, while the number of equations remains $n^2 + n(n+1)/2$: a complete match again. One may also eliminate the matrix U altogether. The system above is equivalent to:

$$\begin{cases} \left(X(\Delta Z) + (\Delta Z)X + (\Delta X)Z + Z(\Delta X)\right)/2 = \gamma\mu I - (XZ + ZX)/2 \\ \Delta X \in \mathcal{L},\ \Delta Z \in \mathcal{L}^\perp. \end{cases}$$

Having made the number of variables equal to the number of equations still does not guarantee the existence of a solution: the linear system may be degenerate. Let us elaborate on the AHO search direction a bit more. Monteiro and Zanjacomo show that if there exists $\nu > 0$ satisfying $\|Z^{1/2}XZ^{1/2} - \nu I\| \le \frac{\nu}{2}$ then the AHO direction is well-defined. In the V-matrix notion, this condition is equivalent to saying that the maximum eigenvalue of V^2 over the minimum eigenvalue of V^2 is no more than 3. This result can be slightly improved. First, using the primal-dual symmetric transformation, we can rewrite (13) as

$$VD_Z + D_XV + LUL^{-1} = \gamma\mu I - V^2.$$

We need only to show that the above system is nonsingular; i.e. $VD_Z + D_X V + L^{-1}UL = 0 \Longrightarrow (D_X, D_Z, U) = 0$. Note that for skew symmetric matrix U, $\mathrm{tr}\,(L^{-1}UL)^2 = -\|U\|_F^2$. For simplicity, we choose V to be diagonal. Denote the minimum eigenvalue of V to be m and the maximum eigenvalue of V to be M. On the other hand,

$$\mathrm{tr}\,(D_X V + V D_Z)^2 = \|V^{1/2} D_X V^{1/2}\|_F^2 + \|V^{1/2} D_Z V^{1/2}\|_F^2 + 2\mathrm{tr}\, D_X V^2 D_Z$$

and

$$\begin{aligned}
\mathrm{tr}\, D_X V^2 D_Z &= \sum_i \sum_j V_{ii}^2 (D_X)_{ij} (D_Z)_{ij} \\
&= \sum_i \sum_j V_{ii} V_{jj} (D_X)_{ij} (D_Z)_{ij} \left(\frac{V_{ii}}{V_{jj}} - \frac{mM}{V_{ii} V_{jj}} \right),
\end{aligned}$$

since D_X and D_Z are orthogonal.

Notice that if $M/m \le 2$ then

$$\left|\frac{V_{ii}}{V_{jj}} - \frac{mM}{V_{ii}V_{jj}}\right| \le 1.$$

Thus, if $VD_Z + D_X V + L^{-1}UL = 0$ and $M/m \le 2$, then on the one hand,

$$\begin{aligned}
& \operatorname{tr}(D_X V + V D_Z)^2 \\
= \; & \|V^{1/2} D_X V^{1/2}\|_F^2 + \|V^{1/2} D_Z V^{1/2}\|_F^2 + 2\operatorname{tr} D_X V^2 D_Z \\
\ge \; & \|V^{1/2} D_X V^{1/2}\|_F^2 + \|V^{1/2} D_Z V^{1/2}\|_F^2 - 2\|V^{1/2} D_X V^{1/2}\|_F \|V^{1/2} D_Z V^{1/2}\|_F \\
= \; & (\|V^{1/2} D_X V^{1/2}\|_F - \|V^{1/2} D_Z V^{1/2}\|_F)^2 \ge 0
\end{aligned}$$

and on the other hand,

$$\mathrm{tr}\,(D_X V + V D_Z)^2 = \mathrm{tr}\,(L^{-1} U L)^2 = -\|U\|_F^2.$$

Hence $U = 0$. Now $D_X V + V D_Z = 0$, and so $D_X = -V D_Z V^{-1}$, and consequently
$0 = \mathrm{tr}\, D_X D_Z = -\mathrm{tr}\, V D_Z V^{-1} D_Z = -\|V^{1/2} D_Z V^{-1/2}\|_F^2$, and
$D_Z = D_X = 0$.

Important observation: On the central path all search directions are identical. Recall that the central path is $\{(X, y, Z) \mid XZ = \mu I\}$, or $\{V \mid V = \sqrt{\mu} I\}$.

We shall now focus on the Nesterov-Todd direction (11). Let the neighborhood of the central path be

$$\mathcal{N}(\beta) := \{V \in \mathcal{S}^{n\times n}_{++} \mid \delta(V) \le \beta\},$$

where $0 < \beta < 1$ and

$$\delta(V) := \left\| I - \frac{1}{\mu} V^2 \right\|_F, \text{ and } \mu := \frac{I \bullet V^2}{n}, \text{ or } \mu = X \bullet Z/n.$$

Let the scaled primal affine subspace be

$$\mathcal{L}(L) := \{\Delta X \in \mathcal{S}^{n\times n} \mid (L^{\mathrm{T}} A_i L) \bullet \Delta X = 0, \text{ for } i = 1, 2, \ldots, m\}.$$

Clearly, $\mathcal{L} = \mathcal{L}(I)$.

Its dual (orthogonal) space is

$$\mathcal{L}^{\perp}(L) := \left\{ \sum_{i=1}^{m} \Delta y_i L^{\mathrm{T}} A_i L \in \mathcal{S}^{n\times n} \,\middle|\, \exists \Delta y \in \Re^m \right\}.$$

Let us denote the entire scaled primal-dual feasible set to be

$$\begin{aligned} \mathcal{F}(L) \quad := \quad & \{(X, y, Z) \mid (L^{\mathrm{T}} A_i L) \bullet X = b_i,\ i = 1, \ldots, m, \\ & Z = L^{\mathrm{T}} C L - \sum_{i=1}^{m} y_i L^{\mathrm{T}} A_i L,\ X, Z \in \mathcal{S}_{+}^{n\times n}\}. \end{aligned}$$

The V-space method requires to search for $(\Delta X, \Delta Z)$ with

$$\Delta X \in \mathcal{L}(I) \text{ and } \Delta Z \in \mathcal{L}^{\perp}(I).$$

The scaled directions are

$$D_X = L^{-1} \Delta X L^{-\mathrm{T}} \text{ and } D_Z = L^{\mathrm{T}} \Delta Z L$$

Let

$$X(t) = X + t\Delta X, \; Z(t) = Z + t\Delta Z,$$

and let $V(t)$ be the V-space image of $(X(t), Z(t))$.

The analysis is based on a number of technical estimations of several useful quantities. We define

$$U(t) := \frac{1}{2}(V + tD_X)(V + tD_Z) - \frac{1}{2}(V + tD_Z)(V + tD_X)$$

and

$$W(t) := \frac{1}{2}(V + tD_X)(V + tD_Z) + \frac{1}{2}(V + tD_Z)(V + tD_X).$$

Notice that $U(t)$ is skew-symmetric, whereas $W(t)$ is symmetric. This implies that

$$U(t) \perp W(t). \tag{14}$$

Lemma 2 *Suppose that $\delta(V) < 1$ and $0 \le t < t^*$. There holds*

$$\delta(V(t))^2 = \left\| \frac{1}{\mu(t)} W(t) - I \right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}.$$

Proof. Notice that

$$\mu(t)^2\delta(V(t))^2 = \left\|V(t)^2\right\|_F^2 - n\mu(t)^2, \tag{15}$$

and

$$\left\|V(t)^2\right\|_F^2 = \operatorname{tr}(V(t)^2 \cdot V(t)^2) = \operatorname{tr}((V+tD_X)(V+tD_Z))^2 = \operatorname{tr}(W(t)+U(t))^2.$$

Now using the skew-symmetricity of $U(t)$ and using (14),

$$\left\|V(t)^2\right\|_F^2 = (W(t) - U(t)) \bullet (W(t) + U(t)) = \|W(t)\|_F^2 - \|U(t)\|_F^2 . \tag{16}$$

As $\operatorname{tr} W(t) = I \bullet W(t) = n\mu(t)$, it follows from (15) and (16) that

$$\begin{aligned} \delta(V(t))^2 &= \frac{\|W(t)\|_F^2 - n\mu(t)^2}{\mu(t)^2} - \frac{\|U(t)\|_F^2}{\mu(t)^2} \\ &= \left\|\frac{1}{\mu(t)}W(t) - I\right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}. \end{aligned}$$

The lemma is proven. $\square$

We also see that

$$\begin{aligned} n\mu(t) &= \operatorname{tr} W(t) = \operatorname{tr} V^2 + \frac{t}{2}\left(V(D_X + D_Z) + (D_X + D_Z)V\right) \\ &= n\mu + \frac{t}{2}\operatorname{tr}\left(V(\gamma\mu V^{-1} - V) + (\gamma\mu V^{-1} - V)V\right) \\ &= n\mu + t\operatorname{tr}(\gamma\mu I - V^2) \\ &= (1 - t + \gamma t)n\mu. \end{aligned}$$

Letting $G := \frac{1}{2\mu}D_X D_Z + \frac{1}{2\mu}D_Z D_X$, we have

$$\begin{aligned} W(t) &= V^2 + \frac{t}{2}V(D_X + D_Z) + \frac{t}{2}(D_X + D_Z)V + t^2\mu G \\ &= (1-t)V^2 + t\gamma\mu I + t^2\mu G. \end{aligned}$$

Using Lemma 2, it follows that

$$\begin{aligned}\mu(t)^2\delta(V(t))^2 &= \left\|(1-t)(V^2-\mu I)+t^2\mu G\right\|_F^2 - \|U(t)\|_F^2\\ &\le \left\|(1-t)(V^2-\mu I)+t^2\mu G\right\|_F^2 .\end{aligned}$$

Applying the triangle inequality to the above relation, we obtain

$$\mu(t)\delta(V(t)) \le (1-t)\mu\delta(V) + t^2 \|\mu G\|_F . \tag{17}$$

Lemma 3 *There holds*

$$\|G\|_F \le \frac{1}{2\mu}\|D_X + D_Z\|_F^2 .$$

Proof. Using the triangle inequality and the properties of the matrix norm, we have

$$
\begin{aligned}
\|D_X D_Z + D_Z D_X\|_F &\le \|D_X D_Z\|_F + \|D_Z D_X\|_F \\
&\le 2\,\|D_X\|_F\,\|D_Z\|_F \\
&\le \|D_X\|_F^2 + \|D_Z\|_F^2 \\
&= \|D_X + D_Z\|_F^2 ,
\end{aligned}
$$

where the last step is due to the Pythagoras theorem. The lemma follows from the definition of G. $\square$

Lemma 4 *There holds*

$$\frac{1}{\mu}\|D_X + D_Z\|_F^2 \le \frac{\gamma^2\delta(V)^2}{1-\delta(V)} + n(1-\gamma)^2.$$

Proof. By (11) we have

$$\|D_X + D_Z\|_F^2 = \gamma^2 \left\|\mu V^{-1} - V\right\|_F^2 + (1-\gamma)^2 n\mu.$$

However,

$$\left\|\mu V^{-1} - V\right\|_F^2 \le \left\|\mu V^{-1}\right\|^2 \left\|I - \frac{1}{\mu}V^2\right\|_F^2 = \delta(V)^2 \left\|\mu V^{-1}\right\|^2$$

As $V^2 \succeq (1-\delta(V))\mu I$, it follows that $V^{-2} \preceq \frac{1}{(1-\delta(V))\mu} I$ and therefore,

$$\left\|\mu V^{-1}\right\|^2 \le \mu/(1-\delta(V)).$$

Combining the above relations, the lemma follows. $\square$

Summarizing, we have

Lemma 5 *Suppose* $\delta(V) < 1$*. For* $0 \le t < t^*$ *there holds*

$$(1 - t + \gamma t)\delta(V(t)) \le (1 - t)\delta(V) + \frac{t^2}{2}\left(\frac{\gamma^2\delta(V)^2}{1 - \delta(V)} + n(1 - \gamma)^2\right).$$

In the sequel, we investigate how we can choose the parameter γ to guarantee feasibility of the full Newton step. Our basic observation is that if

$$\delta(V(t)) < 1 \text{ for } 0 \le t \le 1$$

then, by continuity, $V(1) \succ 0$, implying that the full Newton step is feasible.

Suppose $V \in \mathcal{N}(\beta)$. From Lemma 5, it follows for $\gamma \in (0,1]$ that

$$(1-t+\gamma t)\delta(V(t)) \leq (1-t)\beta + \frac{1}{2}(\gamma t)^2 \left(\frac{\beta^2}{1-\beta} + n\left(\frac{1-\gamma}{\gamma}\right)^2 \right) \quad (18)$$

yielding the following lemma.

Lemma 6 *Let $\gamma \in (0,1]$ and $\beta \in (0,1)$. If $V \in \mathcal{N}(\beta)$ and*

$$\frac{\beta^2}{1-\beta} + n\left(\frac{1-\gamma}{\gamma}\right)^2 \leq 1 \quad (19)$$

then $t^* > 1$ and $V(1) \succ 0$.

Proof. Let $t \in [0, t^*)$. Based on (18) and (19) we obtain

$$(1 - t + \gamma t)\delta(V(t)) \leq (1 - t)\beta + \frac{1}{2}(\gamma t)^2$$

so that

$$\delta(V(t)) < 1 \text{ for } 0 \leq t \leq 1.$$

By continuity this implies that

$$X(t) \succ 0 \text{ and } Z(t) \succ 0 \text{ for } 0 \leq t \leq 1.$$

The lemma is proven. □

Primal-dual method based on the V-space

Data: $\epsilon > 0$, $(X^{(0)}, Z^{(0)})$ with $V^{(0)} \in \mathcal{N}(\beta)$.

Step 0 Set $k = 0$.

Step 1 If $X^{(k)} \bullet Z^{(k)} < \epsilon$ then stop.

Step 2 Choose $\gamma^{(k)} \in [0,1]$ and solve $(\Delta X^{(k)}, \Delta Z^{(k)})$ from

$$\begin{aligned} & \Delta X^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z^{(k)} D(X^{(k)}, Z^{(k)}) \\ = \quad & \gamma^{(k)} \mu^{(k)} (Z^{(k)})^{-1} - X^{(k)}, \\ & \Delta X^{(k)} \in \mathcal{L}(I) \text{ and } \Delta Z^{(k)} \in \mathcal{L}^{\perp}(I), \end{aligned}$$

with $\mu^{(k)} = X^{(k)} \bullet Z^{(k)} / n$.

Step 3 Choose $t^{(k)}$ and let

$$X^{(k+1)} = X^{(k)} + t^{(k)} \Delta X^{(k)}, \; Z^{(k+1)} = Z^{(k)} + t^{(k)} \Delta Z^{(k)}.$$

Step 4 Set $k = k + 1$ and return to Step 1.

In the above notation, $D(X^{(k)}, Z^{(k)})$ is $L^{\mathrm{T}}L$ where L is the scaling matrix for $X^{(k)}$ and $Z^{(k)}$. We will choose the parameters such that $V^{(k)} \in \mathcal{N}(\beta)$ for all k.

The short step algorithm

Choose

$$\beta = \frac{1}{2},$$

$$\gamma^{(k)} = \frac{1}{1 + 1/\sqrt{2n}} \text{ for } k = 0, 1, 2, \ldots$$

and

$$t^{(k)} = 1 \text{ for } k = 0, 1, 2, \ldots.$$

We let

$$\gamma := \frac{1}{1 + 1/\sqrt{2n}}$$

so that

$$\gamma^{(k)} = \gamma \text{ for } k = 0, 1, 2, \ldots .$$

Based on the technical results of the previous section, we obtain the following result for the short step algorithm.

Lemma 7 *For the short step algorithm, we have*

$$V^{(k)} \in \mathcal{N}(\beta)$$

for $k = 0, 1, \ldots$.

Proof. Let $k \in \{0, 1, 2, \ldots\}$. For the given choice of parameters, we have

$$\frac{\beta^2}{1-\beta} + n\left(\frac{1-\gamma}{\gamma}\right)^2 = \frac{1}{2} + \frac{1}{2} = 1.$$

Hence, we have from Lemma 6 that the maximum step length t^* is greater than 1. Using (18) with $t = 1$ it thus follows that

$$\gamma\delta(V^{(k+1)}) \leq \frac{\gamma^2}{2}$$

which implies

$$\delta(V^{(k+1)}) < \frac{1}{2}.$$

As $V^{(0)} \in \mathcal{N}(\beta)$ by hypothesis, the lemma follows by induction. $\square$

Theorem 2 *The short step algorithm computes an ϵ-optimal solution in*

$$O\left(\sqrt{n}\log\frac{X^{(0)}\bullet Z^{(0)}}{\epsilon}\right)$$

iterations.

Proof. We know from Lemma 7 that all iterates of the short step algorithm are contained in $\mathcal{N}(\beta)$. For given $k \in \{0, 1, \ldots\}$,

$$X^{(k+1)} \bullet Z^{(k+1)} = n\mu^{(k)}(1) = \gamma X^{(k)} \bullet Z^{(k)},$$

and so

$$\begin{aligned} \log X^{(k+1)} \bullet Z^{(k+1)} &= \log\gamma + \log X^{(k)} \bullet Z^{(k)} = (k+1)\log\gamma + \log X^{(0)} \bullet Z^{(0)} \\ &= \log X^{(0)} \bullet Z^{(0)} - (k+1)\log\left(1 + \frac{1}{\sqrt{2n}}\right) \\ &\leq \log X^{(0)} \bullet Z^{(0)} - \frac{k+1}{2\sqrt{2n}}. \end{aligned}$$

The theorem is proven. $\square$

The predictor-corrector algorithm

Let $\beta = \frac{1}{2}$. For $k = 0, 1, 2, \ldots$ let

(even iterations: corrector)

$$\gamma^{(2k)} = 1, t^{(2k)} = 1$$

(odd iterations: predictor)

$$\gamma^{(2k+1)} = 0,$$

and $t^{(2k+1)}$ is maximal with respect to

$$V^{(2k+1)}(t) \in \mathcal{N}(\beta) \text{ for } 0 \le t \le t^{(2k+1)}.$$

At the start of a corrector iteration, we have $V \in \mathcal{N}(\frac{1}{2})$. From (18) we have $\delta(V(1)) \le \frac{1}{2}\frac{\beta^2}{1-\beta} = \frac{1}{4}$. Predictor iterations therefore always start with $V \in \mathcal{N}(\frac{1}{4})$. Using Lemma 5, it follows with $\gamma = 0$ that

$$\delta(V(t)) \le \frac{1}{4} + \frac{nt^2}{2(1-t)}.$$

This means that if $0 \le t \le \frac{2}{1+\sqrt{1+8n}}$ then $\delta(V(t)) \le \frac{1}{2}$, implying that the step lengths in the predictor iterations are never shorter than $\frac{2}{1+\sqrt{1+8n}}$. Similar to Theorem 2, this yields:

Theorem 3 *The predictor-corrector algorithm computes an ϵ-optimal solution in*

$$O\left(\sqrt{n}\log\frac{X^{(0)} \bullet Z^{(0)}}{\epsilon}\right)$$

iterations.

Key References:

- J.F. Sturm and S. Zhang, *Symmetric Primal-Dual Path Following Algorithms for Semidefinite Programming*, Applied Numerical Mathematics, 29, 301 – 315, 1999.
- Yu. Nesterov and M.J. Todd, *Self–Scaled Barriers and Interior–Point Methods for Convex Programming*, Mathematics of Operations Research 22, 1 – 42, 1997.
- J.F. Sturm and S. Zhang, *On Weighted Centers for Semidefinite Programming*, European Journal of Operational Research, 126, 391 – 407, 2000.
- Z.Q. Luo, J.F. Sturm and S. Zhang, *Superlinear Convergence of a Symmetric Primal-Dual Path Following Algorithm for Semidefinite Programming*, SIAM Journal on Optimization, 8, 59 – 81, 1998.