For SDP, we have from $\mathcal{A}U = (\langle A_i, U\rangle)_{i=1}^m$ and $\mathcal{A}^* w = \sum_j A_j w_j$ that

$$m_{ij} = \langle A_i, L_X^{-1} L_Y A_j\rangle, \quad 1 \le i, j \le m.$$

For general $A_1, \ldots, A_m$, we need to compute $L_X^{-1} L_Y A_j$ (which typically take $O(n^3)$ operations; see discussion later) for $j = 1, \ldots, m$, and then take its trace inner product with A_i (in $O(n^2)$ operations) for $i = 1, \ldots, m$ (or $i \le j$ if M is symmetric). Thus, the overall work is $O(mn^3 + m^2 n^2)$ operations. Notice that $m \le n(n+1)/2$ since $\mathrm{Null}(\mathcal{A}^*) = \{0\}$. Once M is computed, solving the reduced Newton equation (65) takes only $O(m^3)$ operations (using Gaussian elimination or QR factorization or, if $M \succ 0$, Cholesky factorization or conjugate gradient method). Thus, in contrast to LP, the main computational work lies in computing M rather than solving (65)!

For the MT direction (58) or the family of directions (59), it can be shown that $L_X^{-1} L_Y$ is positive definite (and also symmetric for (59)). Thus the corresponding directions exist and are unique. Moreover, M is symmetric and $M \succ 0$ for (59). For the AHO direction (57), M is generally not symmetric, so the work to solve (65) is greater.

If $A_1, \ldots, A_m$ have special structure, then we may be able to exploit the structure and compute M with less work. For the SDP relaxation of the MaxCut problem, we have $A_i = e^i (e^i)^T$ and $m = n$. Consider the family of directions (59), we have upon letting $V^j = L_Y A_j$ and $U^j = L_X^{-1} V^j$ that

$$m_{ij} = \mathrm{tr}[e^i (e^i)^T V^j] = (e^i)^T U^j e^i.$$

Now we show that L_X is invertible by solving $L_X U = V$ explicitly for U in terms of V: We have from (59) that

$$\frac{Z^\tau \bar{U} Z^{1-\tau} + Z^{1-\tau} \bar{U} Z^\tau}{2} = V,$$

where we let $Z = X \circ Y$ and $\bar{U} = X^{-1} \circ U$. Consider an eigen-decomposition of Z: $Z = Q \mathrm{Diag}(\zeta_1, \ldots, \zeta_n) Q^T$ with $Q^T Q = I$. Multiplying left and right by Q^T and Q, we can rewrite the above equation as

$$\mathrm{Diag}(\zeta_1^\tau, \ldots, \zeta_n^\tau) \hat{U} \mathrm{Diag}(\zeta_1^{1-\tau}, \ldots, \zeta_n^{1-\tau}) + \mathrm{Diag}(\zeta_1^{1-\tau}, \ldots, \zeta_n^{1-\tau}) \hat{U} \mathrm{Diag}(\zeta_1^\tau, \ldots, \zeta_n^\tau) = 2\hat{V},$$

where we let $\hat{U} = Q^T \bar{U} Q$ and $\hat{V} = Q^T V Q$. This yields

$$\zeta_k^\tau \hat{U}_{k\ell} \zeta_\ell^{1-\tau} + \zeta_k^{1-\tau} \hat{U}_{k\ell} \zeta_\ell^\tau = 2\hat{V}_{k\ell} \quad \forall k, \ell.$$

Thus we can solve for $\hat{U}$:

$$\hat{U}_{k\ell} = \frac{2\hat{V}_{k\ell}}{\zeta_k^\tau \zeta_\ell^{1-\tau} + \zeta_k^{1-\tau} \zeta_\ell^\tau} \quad \forall k, \ell.$$

Notice that $\hat{U}$ is symmetric since $\hat{V}$ is. Applying this to U^j and V^j yields

$$\hat{V}^j = Q^T V^j Q = Q^T (L_Y A_j) Q = Q^T (X^{1/2} e^j (e_j)^T X^{1/2}) Q = u^j (u^j)^T,$$

where we let $u^j = Q^T X^{1/2} e^j$, and

$$\begin{aligned} m_{ij} &= (e^i)^T X^{1/2} \bar{U}^j X^{1/2} e^i \\ &= (e^i)^T X^{1/2} Q \hat{U}^j Q^T X^{1/2} e^i \\ &= (u^i)^T \hat{U}^j u^i \\ &= \sum_{k,\ell} u^i_k \hat{U}^j_{k\ell} u^i_\ell \\ &= \sum_{k,\ell} u^i_k \frac{2\hat{V}^j_{k\ell}}{\zeta_k^\tau \zeta_\ell^{1-\tau} + \zeta_k^{1-\tau}\zeta_\ell^\tau} u^i_\ell \\ &= \sum_{k,\ell} u^i_k \frac{2u^j_k u^j_\ell}{\zeta_k^\tau \zeta_\ell^{1-\tau} + \zeta_k^{1-\tau}\zeta_\ell^\tau} u^i_\ell \end{aligned}$$

What's the total work to find M? We can compute $X^{1/2}$, then Z, and then Q in $O(n^3)$ operations. We then compute each u^j in $O(n^2)$ operations. Computing each m_{ij} by the double sum takes $O(n^2)$ operations. (At least, I don't see a quicker way.) Thus the total work to compute M is $O(n^3 + mn^2 + m^2n^2) = O(n^4)$ operations (since $m = n$). However, for the NT direction, corresponding to $\tau = \frac{1}{2}$, the above expression simplifies to

$$m_{ij} = \sum_{k,\ell} u^i_k \frac{u^j_k u^j_\ell}{\zeta_k^{1/2}\zeta_\ell^{1/2}} u^i_\ell = \left(\sum_k \frac{u^i_k u^j_k}{\zeta_k^{1/2}} \right)^2 = \left((\hat{u}^i)^T \hat{u}^j \right)^2,$$

where $\hat{u}^i = (u^i_k/\zeta_k^{1/4})_{k=1}^n$. Thus in this case the total work to compute M is only $O(n^3)$ operations, of the same order as for the dual path-following method (though here the work is greater since it needs to find $X^{1/2}$, Q, and cannot exploit the sparsity of Y)!

4.4 More practical primal-dual path-following methods

The primal-dual method analyzed in Proposition 3 has the best worst-case iteration complexity, but in practice it is slower than other variants. We will look at two variants below. One replaces the narrow neighborhood by a wide neighborhood and another adds an extra predictor step to accelerate convergence. Also, the path-following methods we have seen so far require feasible interior starting points. For specially structured problems like the MaxCut SDP relaxation, such starting point can be found easily. But this is not true in general. We will see two approaches to handle this, both of which allow infeasible starting points.

4.4.1 A wide neighborhood

Primal-dual path-following methods can be significantly speeded up in practice by using a wider neighborhood (so that larger steps can be taken and still remain inside the neighborhood), even though the worst-case iteration complexity bound is worse. A popular choice is the "wide" neighborhood:

$$\mathcal{N}_{\text{wide}}(\gamma) = \{(x, y, \mu, \epsilon) \mid \mathcal{A}x = b,\ y = \mathcal{A}^*\mu - c,\ x \in \text{int}K,\ \min(x \circ y) \geq \gamma\epsilon\} \tag{66}$$

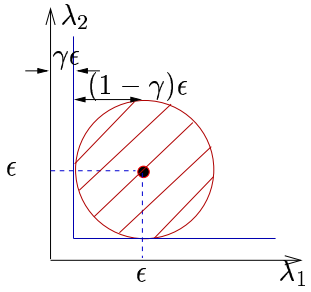
with $0 < \gamma < 1$. Here $\min(z) = \min_i z_i$ in the case of LP,

$$\min(z) = \min_i z_{n_i} - \sqrt{z_{1,i}^2 + \cdots + z_{n_i-1,i}^2}$$

in the case of SOCP, and

$$\min(Z) = \min_i \lambda_i(Z)$$

in the case of SDP. It's not hard to see that $\|z - \epsilon e\| \leq (1-\gamma)\epsilon$ implies $\min(z) \geq \gamma\epsilon$, and hence $\mathcal{N}_{\text{nar}}(1-\gamma) \subseteq \mathcal{N}_{\text{wide}}(\gamma)$. In fact, The latter is much larger than the former. For SDP, this is the difference between $\sqrt{\sum_i (\lambda_i(Z) - \epsilon)^2} \leq (1-\gamma)\epsilon$ and $\lambda_i(Z) \geq \gamma\epsilon$ for all i, where $Z = X \circ Y$.



The wide neighborhood was originally conceived for LP, for which it yields fast convergence in practice and has a worst-case iteration complexity that's only a factor of $\sqrt{n}$ worse than the best known. The above extensions to SOCP and SDP are very natural. However, the complexity analysis is not easy to extend. For SDP, a complexity analysis is known only for directions from the MZ family (see (60))

$$\mathcal{A}U = 0, \qquad V = \mathcal{A}^* w, \qquad H_P[XY + UY + XV] = \sigma\epsilon I, \tag{67}$$

where $0 < \sigma < 1$ and $P \in \mathbb{R}^{n \times n}$ is invertible, symmetric, and satisfies

$$PXYP^{-1} \text{ is symmetric.} \tag{68}$$

Monteiro and Zhang calls such P the "commutative" class of scaling matrices. This class includes the HRVW/KSH/M direction ($P = X^{-1/2}$), its "dual" ($P = Y^{-1/2}$), and the NT direction ($P = \left(X^{-1/2}(X \circ Y)^{1/2} X^{-1/2}\right)^{1/2}$), as can be checked.

Letting $\hat{X} = PXP$, $\hat{U} = PUP$ and $\hat{Y} = P^{-1}YP^{-1}$, $\hat{V} = P^{-1}VP^{-1}$, we can rewrite the third equation in (67) as

$$\text{Sym}[\hat{X}\hat{Y} + \hat{U}\hat{Y} + \hat{X}\hat{V}] = \sigma\epsilon I. \tag{69}$$

Since P is symmetric and invertible, we have $\hat{X} \succ 0$, $\hat{Y} \succ 0$. Let $S_X U = \text{Sym}[XU]$. It's not hard to check that $S_{\hat{X}}$ and $S_{\hat{Y}}$ are self-adjoint and positive definite, so $S_{\hat{Y}}^{-1} S_{\hat{X}}$ is self-adjoint. Moreover, (68) is equivalent to $\hat{X}\hat{Y} = \hat{Y}\hat{X}$, so $\hat{X}$ and $\hat{Y}$ commute. Then it can be shown that $S_{\hat{X}}$ and $S_{\hat{Y}}$ commute, so that

$$G_{\hat{X},\hat{Y}} = S_{\hat{Y}}^{-1} S_{\hat{X}}$$

is positive definite. In fact, $\hat{X}$ and $\hat{Y}$ are simultaneously diagonalizable, i.e.,

$$\hat{X} = Q\text{Diag}(\xi)Q^T, \quad \hat{Y} = Q\text{Diag}(\psi)Q^T \quad \text{with} \quad Q^T Q = I, \tag{70}$$

where $Q \in \mathbb{R}^{n\times n}$ and $\xi, \psi \in (0,\infty)^n$ are the eigenvalues of $\hat{X}$ and $\hat{Y}$, respectively. For any $V \in \mathbb{S}^n$, letting $V = S_{\hat{Y}}U$ (so $U = S_{\hat{Y}}^{-1}V$), we have

$$Q^TVQ = \frac{\mathrm{Diag}(\psi)Q^TUQ + Q^TUQ\mathrm{Diag}(\psi)}{2} = \left[\frac{\psi_i+\psi_j}{2}(Q^TUQ)_{ij}\right]_{i,j}$$

and hence

$$Q^TUQ = \left[\frac{2}{\psi_i+\psi_j}(Q^TVQ)_{ij}\right]_{i,j}. \tag{71}$$

It readily follows that

$$G_{\hat{X},\hat{Y}}U = S_{\hat{Y}}^{-1}(S_{\hat{X}}U) = Q\left[\frac{\xi_i+\xi_j}{\psi_i+\psi_j}(Q^TUQ)_{ij}\right]_{i,j}Q^T.$$

We define the condition number of $G = G_{\hat{X},\hat{Y}}$ in the usual way: $\mathrm{cond}(G) = \lambda_{\max}(G)/\lambda_{\min}(G)$.

Below we estimate the improvement at each iteration of the path-following method using the wide neighborhood and MZ directions based on commutative scaling matrix. The proof, which streamlines that given in a 1998 paper of Monteiro and Zhang, differs from Propositions 3 because we work with $\lambda_{\min}(\cdot)$ instead of $\|\cdot - \epsilon I\|_F$. In particular, we can no longer use properties of the trace. Instead, we use the matrix facts (i) $A \in \mathbb{R}^{n\times n}$ and PAP^{-1} have the same eigenvalues for any invertible $P \in \mathbb{R}^{n\times n}$, (ii) $\min_i \mathrm{Re}\lambda_i(A) \geq \lambda_{\min}(\mathrm{Sym}[A])$, and (iii) $\lambda_{\min}(\cdot)$ is a concave function. (See the book Topics in Matrix Analysis by Horn and Johnson.) The estimate is for SDP, but it easily specializes to LP (i.e., A_i and X, Y are diagonal matrices) and possibly SOCP.

Proposition 4. *Consider SDP. If* $(X,Y,\mu,\epsilon) \in \mathcal{N}_{\mathrm{wide}}(\gamma)$ *with* $\epsilon = \frac{\langle X,Y\rangle}{n}$, *then* $(X_+,Y_+,\mu_+,\epsilon_+)$ *given by* (49), (50) *with* U,V *given by* (67) *and* (68),

$$0 < \alpha \leq \min\left\{1, \frac{2\sigma(1-\gamma)}{1-2\sigma+\frac{\sigma^2}{\gamma}}\cdot\frac{1}{n\sqrt{\mathrm{cond}(G_{\hat{X},\hat{Y}})}}\right\} \tag{72}$$

and $\theta = 1-\sigma$ *is also in* $\mathcal{N}_{\mathrm{wide}}(\gamma)$ *with* $\epsilon_+ = \frac{\langle X_+,Y_+\rangle}{n}$.

Proof. Since $(X,Y,\mu,\epsilon) \in \mathcal{N}_{\mathrm{wide}}(\gamma)$ we have $\mathcal{A}X = b$, $Y = \mathcal{A}^*\mu - C$, $X \succ 0$, and $\lambda_{\min}(X\circ Y) \geq \gamma\epsilon$. Since $\mathcal{A}U = 0$ so that $V = \mathcal{A}^*w$, we have from (49) that

$$\mathcal{A}X[\alpha] = b, \qquad Y[\alpha] = \mathcal{A}^*\mu[\alpha] - C \qquad \forall 0 \leq \alpha \leq 1.$$

We also have from (49) that

$$X[\alpha]Y[\alpha] = (X+\alpha U)(Y+\alpha V) = (1-\alpha)XY + \alpha(XY+UY+XV) + \alpha^2UV. \tag{73}$$

Assuming $X[\alpha] \succeq 0$ (which will be verified later), so $X[\alpha]Y[\alpha]$ has the same eigenvalues as

$X[\alpha] \circ Y[\alpha]$ and hence

$$\begin{aligned}
\lambda_{\min}(X[\alpha] \circ Y[\alpha]) &= \lambda_{\min}(X[\alpha]Y[\alpha]) \\
&= \lambda_{\min}(PX[\alpha]Y[\alpha]P^{-1}) \\
&\geq \lambda_{\min}\left(\mathrm{Sym}[PX[\alpha]Y[\alpha]P^{-1}]\right) \\
&= \lambda_{\min}\left(H_P[(1-\alpha)XY + \alpha(XY + UY + XV) + \alpha^2 UV]\right) \\
&= \lambda_{\min}\left((1-\alpha)H_P[XY] + \alpha H_P[XY + UY + XV] + \alpha^2 H_P[UV]\right) \\
&= \lambda_{\min}\left((1-\alpha)H_P[XY] + \alpha(\sigma\epsilon I + \alpha H_P[UV])\right) \\
&\geq (1-\alpha)\lambda_{\min}\left(H_P[XY]\right) + \alpha\lambda_{\min}\left(\sigma\epsilon I + \alpha H_P[UV]\right) \\
&= (1-\alpha)\lambda_{\min}\left(PXYP^{-1}\right) + \alpha\left(\sigma\epsilon + \alpha\lambda_{\min}(H_P[UV])\right) \\
&= (1-\alpha)\lambda_{\min}\left(X \circ Y\right) + \alpha\sigma\epsilon + \alpha^2\lambda_{\min}(H_P[UV]) \\
&\geq (1-\alpha)\gamma\epsilon + \alpha\sigma\epsilon + \alpha^2\lambda_{\min}(H_P[UV]), \qquad (74)
\end{aligned}$$

where the second inequality uses $0 \leq \alpha \leq 1$ and the concavity of $\lambda_{\min}(\cdot)$; the sixth equality uses the symmetry of $PXYP^{-1}$.

For any $0 \leq \alpha \leq 1$, we have from (73) and $\langle U, V \rangle = 0$ that

$$\begin{aligned}
\frac{\langle X[\alpha], Y[\alpha] \rangle}{n} &= (1-\alpha)\frac{\langle X, Y \rangle}{n} + \alpha\frac{\langle X, Y \rangle + \langle U, Y \rangle + \langle X, V \rangle}{n} + \alpha^2\frac{\langle U, V \rangle}{n} \\
&= (1-\alpha)\epsilon + \alpha\frac{\mathrm{tr}[XY + UY + XV]}{n} \\
&= (1-\alpha)\epsilon + \alpha\frac{\mathrm{tr}[\sigma\epsilon I]}{n} \\
&= (1-\alpha)\epsilon + \alpha\sigma\epsilon \\
&= (1-\alpha\theta)\epsilon \\
&= \epsilon[\alpha], \qquad (75)
\end{aligned}$$

where the third equality uses (67); the fifth equality uses $\theta = 1 - \sigma$; and the last equality uses (49).

It follows from (74) and (75) that $\lambda_{\min}(X[\alpha] \circ Y[\alpha]) \geq \epsilon[\alpha]$ whenever

$$(1-\alpha)\gamma\epsilon + \alpha\sigma\epsilon + \alpha^2\lambda_{\min}(H_P[UV]) \geq \gamma(1 - \alpha + \alpha\sigma)\epsilon$$

which simplifies to

$$\sigma(1-\gamma)\epsilon + \alpha\lambda_{\min}(H_P[UV]) \geq 0. \qquad (76)$$

Then (76) is satisfied by a constant α provided $-\lambda_{\min}(H_P[UV])$ is uniformly bound from above.

We have upon letting $\hat{U} = PUP$, $\hat{V} = P^{-1}VP^{-1}$, and $G = G_{\hat{X},\hat{Y}}$ that

$$\begin{aligned}
-\lambda_{\min}(H_P[UV]) &= -\lambda_{\min}(\mathrm{Sym}[PUVP^{-1}]) \\
&= -\lambda_{\min}(\mathrm{Sym}[\hat{U}\hat{V}]) \\
&= \lambda_{\max}(-\mathrm{Sym}[\hat{U}\hat{V}]) \\
&\leq \|-\mathrm{Sym}[\hat{U}\hat{V}]\|_F \\
&= \frac{\|\hat{U}\hat{V} + \hat{V}\hat{U}\|_F}{2} \\
&\leq \frac{\|\hat{U}\hat{V}\|_F + \|\hat{V}\hat{U}\|_F}{2} \\
&= \|\hat{U}\hat{V}\|_F \\
&\leq \lambda_{\max}(\hat{U})\|\hat{V}\|_F \\
&\leq \|\hat{U}\|_F\|\hat{V}\|_F \\
&= \|G^{1/2}G^{-1/2}\hat{U}\|_F\|G^{-1/2}G^{1/2}\hat{V}\|_F \\
&\leq \lambda_{\max}(G^{1/2})\|G^{-1/2}\hat{U}\|_F\lambda_{\max}(G^{-1/2})\|G^{1/2}\hat{V}\|_F \\
&= \mathrm{cond}\,(G)^{1/2}\,\|G^{-1/2}\hat{U}\|_F\|G^{1/2}\hat{V}\|_F \\
&\leq \mathrm{cond}\,(G)^{1/2}\,\frac{\|G^{-1/2}\hat{U}\|_F^2 + \|G^{1/2}\hat{V}\|_F^2}{2}, \qquad (77)
\end{aligned}$$

where the last inequality uses $ab \leq (a^2+b^2)/2$. We have from (69) and symmetry of $\hat{X}\hat{Y}$ that

$$S_{\hat{Y}}\hat{U} + S_{\hat{X}}\hat{V} = \sigma\epsilon I - \hat{X}\hat{Y}.$$

Applying $S_{\hat{Y}}^{-1}$ and then $G^{-1/2}$ to both sides yields

$$G^{-1/2}\hat{U} + G^{1/2}\hat{V} = G^{-1/2}S_{\hat{Y}}^{-1}(\sigma\epsilon I - \hat{X}\hat{Y}).$$

Since $S_{\hat{X}}$ and $S_{\hat{Y}}$ are positive definite and commute, we have $G^{-1/2} = S_{\hat{X}}^{-1/2}S_{\hat{Y}}^{1/2}$. Also, $\langle G^{-1/2}\hat{U}, G^{1/2}\hat{V}\rangle = \langle \hat{U}, \hat{V}\rangle = \langle U, V\rangle = 0$. Thus

$$\begin{aligned}
\|G^{-1/2}\hat{U}\|_F^2 + \|G^{1/2}\hat{V}\|_F^2 &= \|G^{-1/2}\hat{U} + G^{1/2}\hat{V}\|_F^2 \\
&= \|S_{\hat{X}}^{-1/2}S_{\hat{Y}}^{-1/2}(\sigma\epsilon I - \hat{X}\hat{Y})\|_F^2 \\
&= \left\| Q\left[\sqrt{\frac{2}{\xi_i+\xi_j}}\sqrt{\frac{2}{\psi_i+\psi_j}}\tilde{Z}_{ij}\right]_{i,j} Q^T \right\|_F^2 \\
&= \left\| \left[\sqrt{\frac{2}{\xi_i+\xi_j}}\sqrt{\frac{2}{\psi_i+\psi_j}}\tilde{Z}_{ij}\right]_{i,j} \right\|_F^2 \\
&= \sum_{i=1}^n \frac{1}{\xi_i\psi_i}(\sigma\epsilon - \xi_i\psi_i)^2 \\
&= \sum_{i=1}^n \frac{\sigma^2\epsilon^2}{\xi_i\psi_i} - 2\sigma\epsilon n + \sum_{i=1}^n \xi_i\psi_i \\
&\leq \frac{\sigma^2\epsilon}{\gamma}n - 2\sigma\epsilon n + \epsilon n,
\end{aligned}$$

where the third equality uses the formula (70), (71) for $S_{\hat{Y}}^{-1}$ (and the observation that $S_{\hat{Y}}^{-1/2}$ is obtained by taking square root of each $\frac{2}{\psi_i+\psi_j}$ term in (71)) and we let $\tilde{Z} = Q^T(\sigma\epsilon I - \hat{X}\hat{Y})Q = \mathrm{Diag}(\sigma\epsilon - \xi_1\psi_1, \ldots, \sigma\epsilon - \xi_n\psi_n)$; the inequality uses

$$\gamma\epsilon \leq \lambda_i(X \circ Y) = \lambda_i(PXYP^{-1}) = \lambda_i(\hat{X}\hat{Y}) = \xi_i\psi_i \quad \forall i$$

and $\epsilon n = \langle X, Y\rangle = \langle \hat{X}, \hat{Y}\rangle = \sum_i \xi_i\psi_i$. This combined with (77) shows that (76) holds whenever α satisfies (72).

Finally, since $X[0] = X \succ 0$, $Y[0] = Y \succ 0$, $X[\alpha]$, $Y[\alpha]$ are continuous in α and $\lambda_{\min}(\cdot)$ is continuous on $\mathbb{S}^n$, we also have that $X[\alpha] \succ 0$, $X[\alpha] \succ 0$ for all α in the interval (72) (otherwise $\lambda_{\min}X[\alpha] \circ Y[\alpha] = 0$ for some α in this interval, contradicting $\lambda_{\min}(X[\alpha] \circ Y[\alpha]) \geq \epsilon[\alpha]$, as is implied by (76)). ■

We need $0 < \sigma < 1$ to ensure that $\theta > 0$ and the interval (72) is nonempty, so that $\epsilon_+ < \epsilon$. It's easily verified that, for the NT direction ($P = \left(X^{-1/2}(X \circ Y)^{1/2}X^{-1/2}\right)^{1/2}$), $\hat{X} = \hat{Y}$ and hence

$$\mathrm{cond}(G_{\hat{X},\hat{Y}}) = 1.$$

For the HRVW/KSH/M direction ($P = X^{-1/2}$) or its dual ($P = Y^{-1/2}$), it can be shown that

$$\mathrm{cond}(G_{\hat{X},\hat{Y}}) \leq \frac{n}{\gamma}$$

whenever $\frac{\langle X,Y\rangle}{n} = \epsilon$ and $\lambda_{\min}(X \circ Y) \geq \gamma\epsilon$. It follows from Proposition 4 that, when the NT direction is used, we have $\alpha = \Theta(\frac{1}{n})$ so that $\epsilon_+ = (1 - \Theta(\frac{1}{n}))\epsilon$. This in turn implies $O\left(n \log\left(\frac{\epsilon^0}{\epsilon^{\text{final}}}\right)\right)$ iterations until termination. This is worse by a $\sqrt{n}$ factor compared to using a narrow neighborhood (see Proposition 3). For the HRVW/KSH/M direction, the iteration complexity is worse by another $\sqrt{n}$ factor. However, in practice the wide neighborhood yields faster convergence (by taking larger stepsize α, as long as the new iterate stays inside the neighborhood), even though its worst-case complexity is worse. Can Proposition 4 be extended to other primal-dual directions?

4.4.2 Predictor-corrector methods

Primal-dual path-following methods stay near the central path by moving (x, y, μ) along a direction (u, v, w) towards the central path (sometimes called the "centering" direction) and then decreasing ϵ. Propositions 3 and 4 show that solving the Newton equation (48) with $0 < \sigma \leq 1$ yields such a centering direction. However, the solution we are seeking corresponds to $\epsilon = 0$, so intuitively we may be able to speed up convergence by also using directions corresponding to $\sigma = 0$ in (48) (sometimes called the "affine scaling" direction), with stepsize chosen to keep (x, y, μ) near the central path. This is in the spirit of predictor-corrector methods for equation solving.

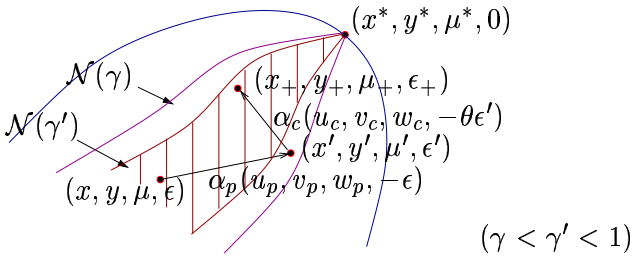
For example, at each iteration, we might move the current $(x, y, \mu, \epsilon) \in \mathcal{N}(\gamma)$ (the neighborhood can be narrow or wide) along an affine scaling direction $(u_p, v_p, w_p, -\epsilon)$ to obtain

$$(x', y', \mu', \epsilon') = (x, y, \mu, \epsilon) + \alpha_p(u_p, v_p, w_p, -\epsilon),$$

with $\alpha_p > 0$ chosen as large as possible such that $(x', y', \mu', \epsilon') \in \mathcal{N}(\gamma')$ for some $\gamma < \gamma' < 1$. This is the predictor step. Then move $(x', y', \mu', \epsilon')$ along a centering direction $(u_c, v_c, w_c, -\theta\epsilon')$ to obtain

$$(x_+, y_+, \mu_+, \epsilon_+) = (x', y', \mu', \epsilon') + \alpha_c(u_c, v_c, w_c, -\theta\epsilon'),$$

with $\alpha_c > 0$ and $0 < \theta < 1$ chosen so that $(x_+, y_+, \mu_+, \epsilon_+) \in \mathcal{N}(\gamma)$. This is the corrector step (check?). Thus the predictor step aims to speed up convergence while the corrector step ensures robustness of the method.



For LP, the directions (u_p, v_p, w_p) and (u_c, v_c, w_c) are typically computed by solving (48) with $\sigma = 0$ and $\sigma = 1$, respectively. This has the key advantage that the left-hand matrix M in the reduced Newton (Schur) equation (65) is the same for both directions, so that a Cholesky factorization of M can be reused instead of computed twice. For this version of the predictor-corrector method, local superlinear convergence can be shown (in addition to polynomial time complexity) for suitable choices of the parameters γ, γ', θ. (For example, γ and $\gamma' - \gamma$ need to be sufficiently small.) However, for some large-scale problems (such as signal denoising by Wavelets), Cholesky factorization is impractical for solving the Schur equation and iterative methods like conjugate gradient method are used instead. In such cases, there is no advantage in having the same M, and faster local convergence (quadratic rate) can be achieved when the centering direction (u_c, v_c, w_c) is computed by solving (48) with $\sigma = 1$ and (x, y, μ, ϵ) replaced by $(x', y', \mu', \epsilon')$.

For SDP, the primal-dual direction is not unique, and the feasible sets $\mathcal{F}(P)$ and $\mathcal{F}(D)$ are not polyhedral, which complicate matters. One version of the predictor-corrector method chooses (U_p, V_p, w_p) to be the AHO affine scaling direction, i.e.,

$$\mathcal{A}U_p = 0, \quad V_p = \mathcal{A}^* w_p, \quad \mathrm{Sym}[XY + U_p Y + X V_p] = 0,$$

and (U_c, V_c, w_c) to solve (48) with $0 < \sigma \leq 1$ and (X, Y, μ, ϵ) replaced by $(X', Y', \mu', \epsilon')$. For this version, local superlinear convergence, i.e.,

$$\|(X_+, Y_+, \mu_+) - (X^*, Y^*, \mu^*)\|_F = o(\|(X, Y, \mu) - (X^*, Y^*, \mu^*)\|_F)$$

can be shown (in addition to polynomial time complexity) for suitable choices of the parameters γ, γ', θ, and assuming also that the optimal primal and dual solutions X^*, (Y^*, μ^*) satisfy "strict complementarity" and "nondegeneracy", namely,

$$X^* + Y^* \succ 0, \qquad \mathcal{A}U = 0,\ V = \mathcal{A}^* w,\ UY^* + X^* U = 0 \quad \Longrightarrow \quad U = V = 0,\ w = 0.$$

Why the AHO affine scaling direction? This direction has the key property that it's well defined and unique at (X^*, Y^*, μ^*) under the above assumption, which existing convergence analysis needs. This contrasts with NT and HRVW/KSH/M directions which seem not well defined at (X^*, Y^*, μ^*). This does not mean superlinear convergence is not possible with other directions, but a proof would need new ideas.

4.4.3 Infeasible methods

The path-following methods of previous sections start with an (x, y, μ, ϵ) in some neighborhood $\mathcal{N}_{\text{nar}}(\gamma)$ or $\mathcal{N}_{\text{wide}}(\gamma)$. For specially structured problems such as MaxCut SDP relaxation, such (x, y, μ, ϵ) can be found easily. What if it cannot be found easily?

One approach is to allow the equality constraints to be violated, i.e., $\mathcal{A}x \neq b$ and/or $y \neq \mathcal{A}^*\mu - c$. Accordingly, we modfiy $\mathcal{A}u = 0$ and $v = \mathcal{A}^*w$ in (48) to

$$\mathcal{A}u = b - \mathcal{A}x, \quad v = \mathcal{A}^*w + \mathcal{A}^*\mu - c - y.$$

However, in general $\langle u, v\rangle \neq 0$, which complicates the complexity analysis. Nonetheless, by suitably estimating $\langle u, v\rangle$, a complexity analysis is possible under the primal and dual feasible interior assumption (21). As the analysis is fairly involved, we will not go into details here. What if (21) fails? Can this be detected by the method (possibly by monitoring the divergence of x or (y, μ))? This seems not easy, though some study has been made. An alternative approach will be considered in the next subsection.

4.4.4 Self-dual homogeneous embedding

A second approach to finding a starting point for path-following methods, originally proposed by Erling Andersen and Yinyu Ye for LP, is to reformulate the basic primal-dual optimality condition (Karush-Kuhn-Tucker condition) as a self-dual conic optimization problem with (essentially) the identity element e (with respect to the symmetric multiplication $\circ$) being a feasible point. Then e would be feasible for the dual problem, and since $e \circ e = e$, we have a point on the central path to start the path-following methods from previous subsections.

Specifically, it's readily verified that the following conic optimization problem

$$\begin{array}{ll} \max_{u,v} & \langle d, u\rangle \\ \text{s.t.} & \mathcal{B}u + v = d, \\ & u \in K,\ v \in K^\circ, \end{array} \tag{78}$$

where K is a nonempty closed convex cone in $\mathbb{H}$, $d \in \mathbb{H}$, and $\mathcal{B}$ is a linear mapping from $\mathbb{H}$ to $\mathbb{H}$ having the skew-symmetry property $\mathcal{B}^* = -\mathcal{B}$, is self-dual, i.e., its dual equals itself.

Now, consider the primal problem (18) and its dual (20). For simplicity, consider the case of SDP. (The same idea specializes to LP and SOCP.) A basic condition for X and (Z, μ) to be optimal solutions is primal and dual feasibility plus no duality gap:

$$\begin{array}{lll} \mathcal{A}X = b, & C - \mathcal{A}^*\mu = Z, & \\ X \succeq 0, & Z \preceq 0, & \langle C, X\rangle \geq b^T\mu. \end{array} \tag{79}$$

This is a sufficient condition for primal-dual optimality by weak duality, which in fact implies the last inequality must hold with equality (if it has a solution at all). However, the inequality is needed to obtain a self-dual problem. The sufficient condition (79) is also necessary for LP (when $A_1, \ldots, A_m$, C are diagonal matrices) and more generally when the cone in (18) is polyhedral.

How to reformulate (79) into the form (78)? First, we put (79) into a homogeneous form by scaling the constants b and C by a new variable $\tau \geq 0$:

$$\begin{array}{lllll} -\mathcal{A}X & +b\tau & +\nu & = 0, & \mu \in \mathbb{R}^m, \quad \nu \in \{0\}, \\ \mathcal{A}^*\mu & -C\tau & +Z & = 0, & X \succeq 0, \quad Z \preceq 0, \\ -b^T\mu & +\langle C, X\rangle & +\rho & = 0, & \tau \geq 0, \quad \rho \leq 0. \end{array} \tag{80}$$

Here we also introduced an artificial variable ν (fixed at 0) and a nonpositive slack variable ρ to put the equations in the self-dual form (78). Then any solution of (79) is a solution of (80) with $\tau = 1$, $\nu = 0$, and ρ appropriately chosen. Conversely, any solution of (80) with $\tau > 0$ yields $(X, Z, \mu)/\tau$ as a solution of (79). What if $\tau = 0$? We will address this later.

Second, we borrow a trick from LP for starting the simplex method. Let $F(\mu, X, \tau, \nu, Z, \rho)$ denote the left-hand side of the set of three linear equations in (80). Let

$$r = F(0, I, 1, 0, -I, -1)$$

and augment these equations by a new column $-r\theta$ with θ a new nonnegative variable. Then $(0, I, 1, 0, -I, -1)$ together with $\theta = 1$ is a solution of the augmented equations. We then minimize θ so to drive θ to zero. To maintain skew-symmetric form of the equations, we introduce a new row $\langle r, \cdot\rangle$ augmented by a zero. The reformulated problem is

$$\begin{array}{ll} \max\limits_{\substack{\mu, X, \tau, \theta \\ \nu, Z, \rho, \pi}} & \beta\theta \\ \text{s.t.} & \left[\begin{array}{cccc} 0 & -\mathcal{A} & b & | \\ \mathcal{A}^* & 0 & -C & -r \\ -b^T & \langle C, \cdot\rangle & 0 & | \\ — & \langle r, \cdot\rangle & — & 0 \end{array}\right] \left[\begin{array}{c} \mu \\ X \\ \tau \\ \theta \end{array}\right] + \left[\begin{array}{c} \nu \\ Z \\ \rho \\ \pi \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \beta \end{array}\right] \\ & y \in \mathbb{R}^m, \ X \succeq 0, \ \tau \geq 0, \ \theta \geq 0, \\ & \nu \in \{0\}, \ Z \preceq 0, \ \rho \leq 0, \ \pi \leq 0, \end{array} \tag{81}$$

where we let $\beta = \langle r, (0, I, 1)\rangle - 1$. Some calculation yields $\beta = -(n+2)$, so maximizing $\beta\theta$ is equivalent to minimizing θ. It is readily seen that (81) has the self-dual form (78). Moreover, our choice of r and θ implies that $(\mu, X, \tau, \theta, \nu, Z, \rho, \pi) = (0, I, 1, 1, 0, -I, -1, -1)$ is feasible. This can be used to start a primal-dual path-following method to solve (81). By exploiting the special structure of (81), the work per iteration can be made comparable to the primal-dual methods we saw before.

Notice that $(0, 0, 0, 0, 0, 0, 0, \beta)$ is feasible for (81), so it's in fact optimal (since $\theta \geq 0$ on the feasible set) and the optimal objective value is 0 (i.e., $\theta = 0$). So what good is solving (81)? When we solve it using the path-following method, the optimal solution found, say $(\mu^*, X^*, \tau^*, 0, 0, Z^*, \rho^*, \pi^*)$, can be shown to be "maximally complementary" in the sense that $X^* - Z^*$ (which is positive semidefinite) has maximum rank among all optimal solutions. (Caution: This may not be true if another method is used to solve (81).) Then, it can be shown the following:

- If $\tau^* > 0$, then $\frac{X^*}{\tau^*}$ and $\frac{(Z^*, \mu^*)}{\tau^*}$ solve (79) and hence are optimal solutions of (18) and its dual (20).

- If $\tau^* = 0$ and $\rho^* > 0$, then a primal or dual improving direction is found (i.e., either $\mathcal{A}X^* = 0$, $\langle C, X^* \rangle > 0$ or $Z^* = -\mathcal{A}^*\mu^*$, $b^T\mu^* < 0$). This implies the primal or the dual problem is infeasible or unbounded.
- If $\tau^* = 0$ and $\rho^* = 0$, then (79) has no solution and no primal or dual improving direction exists.

More discussions and references on this can be found in Section 3.3 and Chapter 4 of the 2002 book *Aspects of Semidefinite Programming* by de Klerk.

The self-dual homogeneous embedding (81) for SDP was proposed by de Klerk et al. and was implemented in the popular SDP solver SeDuMi by Jos Sturm (who died tragically young). In practice, (81) is solved only inexactly within some tolerance.