Instead of (49), we can use

$$\epsilon[\alpha] = \min\{\epsilon \mid (x[\alpha], y[\alpha], \mu[\alpha], \epsilon) \in \mathcal{N}\},$$

which may or may not be easy to compute, depending on the choice of $\mathcal{N}$.

As for the neighborhood $\mathcal{N}$, one choice is the same "narrow" neighborhood as (32) and (35):

$$\mathcal{N}_{\text{nar}}(\gamma) = \{(x, y, \mu, \epsilon) \mid \mathcal{A}x = b,\ y = \mathcal{A}^*\mu - c,\ x \in \text{int}K,\ \|x \circ y - \epsilon e\| \leq \gamma\epsilon\} \tag{52}$$

with $0 < \gamma < 1$. This is not as practical as other neighborhoods we will see later, but it yields the best worst-case iteration complexity. What about L_x and L_y?

- For LP, since $\frac{\partial}{\partial x_i}(x_i y_i) = y_i$ and $\frac{\partial}{\partial y_i}(x_i y_i) = x_i$, we see from (46) that

$$L_x u = y \circ u, \qquad L_y v = x \circ v. \tag{53}$$

  There has been much study of the primal-dual path-following method for LP, especially around 1990 to 2000. A good reference is the 1997 book by Steve Wright, Primal-Dual Interior-Point Methods.

- For SOCP, one possible choice of $\circ$ in (45) is the Jordan product associated with second-order cones:

$$x \circ y = [x_i \circ y_i]_{i=1}^N \qquad \text{with} \qquad x_i \circ y_i = \begin{bmatrix} (x_{j,i} y_{n,i} + x_{n,i} y_{j,i})_{j=1}^{n_i - 1} \\ x_i^T y_i \end{bmatrix}$$

  and $e = [e^{n_i}]_{i=1}^N$ ($e^{n_i} \in \mathbb{R}^{n_i}$ has 1 in the last entry and 0 else where). We have then

$$L_x u = y \circ u, \qquad L_y v = x \circ v.$$

  However, unlike LP, $x, y \in K_{\text{SOC}}$ does not ensure $x \circ y \in K_{\text{SOC}}$. Primal-dual path-following methods for SOCP were studied by Monteiro and Zhang around 1998. In general, interior-point methods for SOCP have not been too much studied.

- For SDP, the choice of L_X and L_Y is more complicated, as there are many choices! (These choices coincide if X, Y, μ are on the central path since X and Y commute there, but they can be quite different off the central path.) At issue is the linearization of the nonlinear equation in (45), which for SDP is

$$X^{1/2} Y X^{1/2} = \epsilon I, \tag{54}$$

  with $X, Y \succ 0$ and $\epsilon \geq 0$.

  0. A naive approach is to multiply (54) left and right by $X^{1/2}$ and $X^{-1/2}$, respectively, to obtain

$$XY = \epsilon I$$

  and then linearize (i.e., replace X, Y by $X + U$, $Y + V$, expand, and drop the higher-than-linear terms in U, V). This yields the (linear) Newton equation in U, V:

$$XY + UY + XV = \epsilon I, \tag{55}$$

What's wrong with this? Multiplying it on the right by Y^{-1}, it becomes $X+U+XVY^{-1} = \epsilon Y^{-1}$. While X and Y^{-1} are symmetric, XVY^{-1} is generally not symmetric even if V is symmetric, which means U is generally not symmetric even if it exists!

1. An early idea to overcome the difficulty of U asymmetry, proposed independently by Helmberg *et al.*, Kojima, Shindo, Hara, and Monteiro, is to multiply (55) left and right by $X^{-1/2}$ and $X^{1/2}$ (so it is symmetric in all except one term) and then symmetrize. This yields the (linear) Newton equation in U, V:

$$X \circ Y + \mathrm{Sym}[X^{-1/2}UYX^{1/2}] + X \circ V = \epsilon I,$$

where $\mathrm{Sym}[A] = (A + A^T)/2$ for any $A \in \mathbb{R}^{n\times n}$. Since $X^{-1/2}UYX^{1/2} = (X^{-1} \circ U)(X \circ Y)$, this is equivalent to the third equation in (48) with

$$L_X U = \mathrm{Sym}[(X^{-1} \circ U)(X \circ Y)], \qquad L_Y V = X \circ V. \tag{56}$$

The U, V satisfying (48) and (56), called "HRVW/KSH/M direction", can be shown to be symmetric, exist, and are unique.[10]

2. A second idea, proposed by Alizadeh, Haeberly, and Overton, is to directly symmetrize (55). This yields the (linear) Newton equation in U, V:

$$\mathrm{Sym}[XY + UY + XV] = \epsilon I. \tag{57}$$

This equation cannot be written in the form of the third equation in (48). Nevertheless, the (U, V) satisfying this equation and the first two equations of (48), called "AHO direction", can be shown to be symmetric, exist, and are unique if in addition $\mathrm{Sym}[XY] \succeq 0$. Thus, the AHO direction is usable under more restrictive condition. In practice it seems to yield more accurate solution.

3. A third idea, proposed by Monteiro and Tsuchiya, is to linearize $X^{1/2}$. In particular, expanding $X+U = (X^{1/2}+W)^2$ yields $U = X^{1/2}W+WX^{1/2}+o(\|W\|)$. Similarly expanding $(X^{1/2} + W)Y(X^{1/2} + W) = \epsilon I$ and dropping the $o(\|W\|)$ terms yields the (linear) Newton equation in U, V, W:

$$U = X^{1/2}W + WX^{1/2}, \qquad X \circ Y + WYX^{1/2} + X^{1/2}YW + X \circ V = \epsilon I.$$

Rewriting the first equation as $X^{-1} \circ U = \mathrm{Sym}[WX^{-1/2}]$ (which can be seen to have a unique solution $W \in \mathbb{S}^n$ for each $U \in \mathbb{S}^n$) and using $WYX^{1/2} = WX^{-1/2}(X \circ Y)$, we see that this is equivalent to the third equation in (48) with

$$L_X U = \mathrm{Sym}[WX^{-1/2}(X \circ Y)] \text{ with } \mathrm{Sym}[WX^{-1/2}] = X^{-1} \circ U, \qquad L_Y V = X \circ V. \tag{58}$$

The U, V satisfying (48) and (58), called "MT direction", can be shown to be symmetric, exist, and are unique. (This is not easy to show.) Notice that (58) reduces to (56) when $WX^{-1/2}$ is symmetric.

[10]The symmetry of U is due to the fact that $\mathrm{Sym}[WZ] = R$ has a unique solution $W \in \mathbb{S}^n$ for any $Z \succ 0$ and any $R \in \mathbb{S}^n$. In fact, by using the eigen-decomposition of Z: $Z = Q\mathrm{Diag}(\zeta_1, \ldots, \zeta_n)Q^T$ with $Q^TQ = I$, we can rewrite this as $Q^TWQ\mathrm{Diag}(\zeta_1, \ldots, \zeta_n) + \mathrm{Diag}(\zeta_1, \ldots, \zeta_n)Q^TWQ = 2Q^TRQ$ and hence $W = Q\left[\frac{2(Q^TRQ)_{ij}}{\zeta_i+\zeta_j}\right]_{i,j} Q^T$.

4. Can the previous directions be generalized? One generalization, parameterized by $0 \le \tau \le \frac{1}{2}$, is

$$L_X U = \mathrm{Sym}[(X \circ Y)^\tau (X^{-1} \circ U)(X \circ Y)^{1-\tau}], \qquad L_Y V = X \circ V. \tag{59}$$

   Here $\tau = 0$ corresponds to the HRVW/KSH/M direction, and $\tau = \frac{1}{2}$ corresponds to a search direction proposed by Nesterov and Todd (NT direction). Although the NT direction looks complicated, it can be easier to compute, as we shall see for the SDP relaxation of MaxCut. The U, V satisfying (48) and (59) can be shown to be symmetric, exist, and are unique.

5. Another generalization, proposed by Monteiro and Zhang, is to first multiply (55) left and right by P and P^{-1}, where $P \in \mathbb{R}^{n\times n}$ is invertible (possibly depending on X, Y), and then symmetrize. The resulting scaled Newton equation is

$$H_P[XY + UY + XV] = \epsilon I. \tag{60}$$

   where for simplicity we denote $H_P[A] = \mathrm{Sym}[PAP^{-1}]$. This yields the AHO direction when $P = I$, the HRVW/KSH/M direction when $P = X^{-1/2}$, and (with some algebra) the NT direction when $P^T P = X^{-1/2}(X \circ Y)^{1/2} X^{-1/2}$. The (U, V) satisfying (60) and the first two equations of (48), called the "MZ family", can be shown to be symmetric, exist, and are unique under suitable assumptions on P and X, Y.

6. Notice that (60) corresponds to replacing X, Y, U, V in (57) by, respectively, PXP^T, $P^{-T}YP^{-1}$, $PUP^T, P^{-T}VP^{-1}$. Making the same replacement in (58) yields another generalization, which can be shown to yield the NT direction when $P = Y^{1/2}$ and the HRVW/KSH/M direction when $P = (YXY)^{1/2}$. (Check?) The (U, V) satisfying (48) and this generalization of (58), called the "X-MT family", can be shown to be symmetric, exist, and are unique under suitable assumptions on P and X, Y.

For SDP, is there a unified way to generate all the different directions? There have been some studies of this by, e.g., Kojima et al. using the notion of inexact Newton directions, but I think we still don't have a full understanding yet. Are some directions "better" than others? We will compare the work of computing them later. What about the number of iterations? In practice, greater solution accuracy is reported using the AHO direction, but it's theoretical property is not so nice. We will see that a wide neighborhood can be used in conjunction with NT and HRVW/KSH/M directions and still achieve "polynomial" iteration complexity. Overall, the AHO, NT, and HRVW/KSH/M directions seem to be more popular. The 1999 paper by Todd, A study of search directions in primal-dual interior-point methods for semidefinite programming, compares 20 directions using many different measures.

Another open question: When specialized to SOCP (for starter, consider a single SOC), are the above directions still different? Or are some of them the same? I am not aware that this question has been studied.

Like Proposition 2 for the dual path-following method, we can prove a similar iteration complexity for the primal-dual path-following method using the family of directions (59) and narrow neighborhood (52). The complexity analysis with the MZ and X-MT families are more complicated.

Proposition 3. *Consider* $\circ$ *and* L_x, L_y *given by* (46), (53) *for LP or by* (47), (59) *for SDP. If* $(x, y, \mu, \epsilon) \in \mathcal{N}_{\rm nar}(\gamma)$ *and* $\gamma < 2-\sqrt{3}$, *then* $(x_+, y_+, \mu_+, \epsilon_+)$ *given by* (48), (49), (50) *with* $\sigma = 1$, $\alpha = 1$, $\theta = \frac{\gamma(1-4\gamma+\gamma^2)}{(1-\gamma)^2(\gamma+\sqrt{n})}$ *is also in* $\mathcal{N}_{\rm nar}(\gamma)$.

Proof. We prove this for SDP only. LP is recovered as a special case when C and $A_1, \ldots, A_m$ are diagonal matrices.

We have from (48), (59), and $\sigma = 1$ that

$$\mathcal{A}U = 0, \qquad V = \mathcal{A}^* w, \qquad \frac{1}{2}(Z^\tau \bar{U} Z^{1-\tau} + Z^{1-\tau} \bar{U} Z^\tau) + \bar{V} = \epsilon I - Z. \tag{61}$$

where for simplicity we let $Z = X \circ Y$, $\bar{U} = X^{-1} \circ U$, and $\bar{V} = X \circ V$. Since $(x, y, \mu, \epsilon) \in \mathcal{N}_{\rm nar}(\gamma)$ we have $\mathcal{A}X = b$, $Y = \mathcal{A}^*\mu - C$, $X \succ 0$, and $\|R\|_F \le \gamma\epsilon$, where we let $R = \epsilon I - Z$. Since $\alpha = 1$ so that $X_+ = X + U$, $Y_+ = Y + V$, and $\mu_+ = \mu + w$, the first two equations in (61) imply

$$\mathcal{A}X_+ = b, \qquad Y_+ = \mathcal{A}^*\mu_+ - C.$$

Also, $|\lambda_i(R)| \le \|R\|_F \le \gamma\epsilon$ for all i, so that (since $\lambda_i(R) = \epsilon - \lambda_i(Z)$)

$$(1-\gamma)\epsilon \le \lambda_i(Z) \le (1+\gamma)\epsilon, \quad \forall i. \tag{62}$$

We have, as in the proof of Proposition 2, that

$$\begin{aligned}
& \|X_+ \circ Y_+ - \epsilon I\|_F^2 \\
= \;& \mathrm{tr}\left[(X_+ Y_+ - \epsilon I)^2\right] \\
= \;& \mathrm{tr}\left[((X+U)(Y+V) - \epsilon I)^2\right] \\
= \;& \mathrm{tr}\left[\left((I+\bar{U})(Z+\bar{V}) - \epsilon I\right)^2\right] \\
= \;& \mathrm{tr}\left[\left((I+\bar{U})\left(\epsilon I - \frac{1}{2}(Z^\tau \bar{U} Z^{1-\tau} + Z^{1-\tau}\bar{U}Z^\tau)\right) - \epsilon I\right)^2\right] \\
= \;& \mathrm{tr}\left[\left(\epsilon\bar{U} - \frac{1}{2}(Z^\tau \bar{U} Z^{1-\tau} + Z^{1-\tau}\bar{U}Z^\tau) - \frac{1}{2}(\bar{U}Z^\tau \bar{U} Z^{1-\tau} + \bar{U}Z^{1-\tau}\bar{U}Z^\tau)\right)^2\right] \\
\le \;& \left(\|\epsilon\bar{U} - \frac{1}{2}(Z^\tau \bar{U} Z^{1-\tau} + Z^{1-\tau}\bar{U}Z^\tau)\|_F + \frac{1}{2}\|\bar{U}Z^\tau \bar{U} Z^{1-\tau}\|_F + \frac{1}{2}\|\bar{U}Z^{1-\tau}\bar{U}Z^\tau\|_F\right)^2 \\
\le \;& \left(\|\bar{U}\|_F\gamma\epsilon + \|\bar{U}\|_F^2\lambda_{\max}(Z)\right)^2,
\end{aligned} \tag{63}$$

where the third and fifth equalities use $X = X^{1/2}X^{1/2}$ and $\mathrm{tr}[AB] = \mathrm{tr}[BA]$ $(A, B \in \mathbb{R}^{n\times n})$; the fourth equality uses (61); the first inequality uses $\mathrm{tr}[A^2] \le \|A\|_F^2$ $(A \in \mathbb{R}^{n\times n})$ and the triangle inequality; the second inequality uses $\|\bar{U}Z^\omega\bar{U}Z^{1-\omega}\|_F \le \|\bar{U}\|_F\lambda_{\max}(Z)$ $(0 \le \omega \le 1)$ and $\|\epsilon\bar{U} - \frac{1}{2}(Z^\tau\bar{U}Z^{1-\tau} + Z^{1-\tau}\bar{U}Z^\tau)\|_F \le \|\bar{U}\|_F\gamma\epsilon$ (see Eq. (25) and Lemma 4.1(d) in the paper: Tseng, P., Search directions and convergence analysis..., Optimization Methods and Software, Vol. 9, 1998, 245-268).

Now we bound $\|\bar{U}\|_F$. We have from (62) that

$$\begin{aligned}
(1-\gamma)\epsilon\|\bar{U}\|_F^2 &\leq \lambda_{\min}(Z)\|\bar{U}\|_F^2 \\
&\leq \|Z^{\tau/2}\bar{U}Z^{(1-\tau)/2}\|_F^2 \\
&= \operatorname{tr}\left[\bar{U}Z^{\tau}\bar{U}Z^{1-\tau}\right] \\
&= \langle \bar{U}, \frac{1}{2}(Z^{\tau}\bar{U}Z^{1-\tau} + Z^{1-\tau}\bar{U}Z^{\tau})\rangle \\
&= \langle \bar{U}, R - \bar{V}\rangle \\
&= \langle \bar{U}, R\rangle \\
&\leq \|\bar{U}\|_F\|R\|_F,
\end{aligned}$$

where the second inequality uses $\|Z^{\omega}\bar{U}Z^{1-\omega}\|_F \geq \|\bar{U}\|_F\lambda_{\min}(Z)$ (see Eq. (24) in the above paper); the third equality uses (61) and $R = \epsilon I - Z$; and the fourth equality uses $\langle \bar{U}, \bar{V}\rangle = 0$ (which can be shown similarly as in the proof of Proposition 2). Thus

$$\|\bar{U}\|_F \leq \frac{\|R\|_F}{(1-\gamma)\epsilon} \leq \frac{\gamma}{1-\gamma}. \tag{64}$$

Using (49), (50), $\alpha = 1$, and the triangle inequality yields

$$\begin{aligned}
\|Y_+^{1/2}X_+Y_+^{1/2} - \epsilon_+ I\|_F &= \|Y_+^{1/2}X_+Y_+^{1/2} - \epsilon(1-\theta)I\|_F \\
&\leq \|Y_+^{1/2}X_+Y_+^{1/2} - \epsilon I\|_F + \epsilon\theta\|I\|_F \\
&\leq \frac{\gamma^2}{1-\gamma}\epsilon + \frac{\gamma^2}{(1-\gamma)^2}(1+\gamma)\epsilon + \epsilon\theta\sqrt{n} \\
&= \frac{2\gamma^2}{(1-\gamma)^2}\epsilon + \epsilon\theta\sqrt{n} \\
&= \gamma\epsilon(1-\theta) \\
&= \gamma\epsilon_+,
\end{aligned}$$

where the second inequality uses (62), (63), and (64); the third equality uses the definition of θ. Here $\theta > 0$ because $\gamma < 2 - \sqrt{3}$. Finally, (43) implies $|\lambda_i(\bar{U})| \leq \|\bar{U}\|_F \leq \gamma < 1$ for all i, so that $\lambda_i(I + \bar{U}) = 1 + \lambda_i(\bar{U}) > 0$ for all i, implying $I + \bar{U} \succ 0$ and hence $X_+ = X^{1/2}(I + \bar{U})X^{1/2} \succ 0$. ■

We see from Proposition 3 that $\theta^{-1} = O(\sqrt{n})$. Thus, like the dual path-following method using narrow neighborhood, the primal-dual method terminates after $O\left(\sqrt{n}\log\left(\frac{\epsilon^0}{\epsilon_{\text{final}}}\right)\right)$ iterations. Proposition 3 can be extended to allow for $\sigma < 1$, provided γ is sufficiently small (depending on σ). The algebra becomes more complicated, however. Proposition 3 treats LP and SDP. SOCP can be treated by reformulating it as an SDP, and taking care to exploit the structure when computing the Newton direction. Can SOCP be treated directly?

What about computing the Newton direction? In general, we need to solve a linear equation of the form

$$\mathcal{A}u = 0, \quad v = \mathcal{A}^*w, \quad L_x u + L_y v = r$$

for some linear mappings L_x and L_y and $r \in \mathbb{H}$ (see (48) and (60)). Does this have a solution? Is it unique? Assuming L_x is invertible, then we can apply L_x^{-1} to the third equation to obtain $u + L_x^{-1} L_y v = r$ and then using the first two equations to reduce this to

$$Mw = \mathcal{A} L_x^{-1} r \qquad \text{with} \qquad M = \mathcal{A} L_x^{-1} L_y \mathcal{A}^*. \tag{65}$$

(This is sometimes called the Schur equation.) Since $\text{Null}(\mathcal{A}^*) = \{0\}$, it follows that M is positive definite (and hence invertible and (65) has unique solution), provided that $L_x^{-1} L_y$ is positive definite.

For LP, we have $x \in \mathbb{R}^n$ and $\mathcal{A}x = Ax$, where $A \in \mathbb{R}^{m \times n}$. Also, the linear mapping L_x can be represented as matrix multiplication by a diagonal matrix with diagonal entries $y_1, \ldots, y_n$ and similarly for L_y. Thus

$$M = A\text{Diag}(x_1/y_1, \ldots, x_n/y_n)A^T.$$

Then M is symmetric, positive definite (since $\text{Null}(A^T) = \{0\}$), and its entries are computable in $O(m^2 n)$ operations (less if A is sparse). Notice that $m \leq n$ since $\text{Null}(A^T) = \{0\}$. Moreover, M is often sparse when A is sparse, and the sparsity pattern is independent of D. Then the equation (65) can be efficiently solved using Cholesky factorization of M in $O(n^3)$ operations (and the permutation matrix needs to be computed only once). If A is sparse except for a few columns, then we can use a product-form Cholesky factorization proposed by Goldfarb and Scheinberg. Alternatively, we can drop the dense columns and find Cholesky factorization for the reduced A. The solution of the original equation can be recovered using Sherman-Morrison-Woodbury inverse updates. If M is dense, then preconditioned conjugate gradient (CG) method is a possible alternative way to solve (65). However, M becomes increasingly ill-conditioned as x, y approaches optimality (since $x_i \to 0$ or $y_i \to 0$ for each i by complementarity), so preconditioned CG can take many iterations to solve (65) accurately (unless a good preconditioner can be found).

For the LP reformulation of the convex approximation of the compressed sensing problem that we saw in week 1 (see (3)), we have $\mathcal{A}x = Au - Av$, where $x = \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^{2n}$. Correspondingly

$$\begin{aligned} M &= \begin{pmatrix} A & -A \end{pmatrix} \text{Diag}\left(\frac{u_1}{w_1}, \ldots, \frac{u_n}{w_n}, \frac{v_1}{z_1}, \ldots, \frac{v_n}{z_n}\right) \begin{pmatrix} A^T \\ -A^T \end{pmatrix} \\ &= A\text{Diag}\left(\frac{u_1}{w_1} + \frac{v_1}{z_1}, \ldots, \frac{u_n}{w_n} + \frac{v_n}{z_n}\right) A^T, \end{aligned}$$

where $y = \begin{bmatrix} w \\ z \end{bmatrix}$. Thus, by exploiting its structure, the LP can be solved as if it has only n variables instead of $2n$.