The Trust-Region IPM

In this lecture, we will introduce the a new predictor step called the *trust-region step*, due to Lan, Monteiro and Tsuchiya [LMT09], that will be much more effective than affine scaling predictor (this is the name of the standard predictor step) at following long and straight segments of the central path. Enhanced with this new step, we will argue that TRUST-REGION IPM, defined below, is *locally optimal*. That is, the decrease in parameter at each iteration will be essentially as large as it could be given the local geometry of the central path. For this purpose, we will compare to what we call the "ideal step".

DEFINITION 1 (IDEAL STEP) *Let $z(\mu_1)$ be the central path point at $\mu_1 > 0$. For $\beta \in [0,1)$, let $\mu_2 \in [0, \mu_1]$ be the smallest parameter such that $[z(\mu_1), z(\mu_2)) \subset N_2(\beta)$. The β-ideal step at $z(\mu_1)$ is then defined as $\Delta z^{\mathrm{id}} := z(\mu_2) - z(\mu_1)$.*

In words, the ideal step from a central path point takes us as far down the central path as possible while remaining in the $N_2(\beta)$ neighborhood. One can easily modify the above definition to start from any iterate $z = (x, s, y) \in N_2(\beta)$, though the length of the step will not differ too dramatically when compared to the ideal step from the central path point with the same parameter. For the sake of simplicity of exposition, we will throughout the lecture restrict to analyzing steps that start from a point on the central path. With some additional effort, one can adapt the analyses to work from any starting point in the $N_2(\beta)$ neighborhood at the expense of slightly worse parameters.

1 Towards an Improved Predictor Step

A first immediate question is how much more powerful is the ideal step compared to the affine scaling in terms of reducing the gap. The simple answer is *infinitely*. As we will show below, it will in fact generically be the case that at the "end of the path", the ideal step takes us to the optimal solution in one step, while affine scaling provably requires an infinite number of steps. We expand on this in the following sections.

1.1 The Limitations of Affine Scaling

Let us assume that we start at $z(1)$ satisfying $(x(1), s(1)) = (1_n, 1_n)$, which is without loss of generality using path rescaling (recall homework 1 exercise 1). A "long-step" from $z(1)$ should correspond to $\mathbf{A}\Delta x = 0_n$, $\mathbf{A}^\top \Delta y + \Delta s = 0$ such that

$$(1_n + \Delta x)(1_n + \Delta s) \approx_{\beta\nu} \nu 1_n, \tag{1}$$

where $0 \leq \nu \ll 1$ (e.g., $\nu \leq 1/1000$). Here $\approx_{\beta\nu}$ requires the difference between both sides to have ℓ_2 norm at most $\beta\nu$. Let $\Delta z^p = (\Delta x^p, \Delta s^p, \Delta y^p)$ be the affine scaling direction at $z(1)$, which in this case satisfies

$$s(1)\Delta x^p + x(1)\Delta s^p = -x(1)s(1) \Leftrightarrow \Delta x^p + \Delta s^p = -1_n.$$

If $(\Delta x, \Delta s, \Delta y) = (1 - \nu)\Delta z^p$, as shown in the previous lecture, the error in (1) is precisely the *quadratic term* $\Delta x \Delta s$. Let us now analyze what this means for a single coordinate:

$$(1 + \Delta x_1)(1 + \Delta s_1) \approx_{\beta\nu} \nu.$$

Affine scaling forces $\Delta x_1 + \Delta s_1 = -(1-\nu)$ (recall $(\Delta x, \Delta s) := (1-\nu)(\Delta x^p, \Delta s^p)$), which means the error between the left and the right hand side is precisely $|\Delta x_1 \Delta s_1|$. Thus, the approximation we need can only hold if $|\Delta x_1 \Delta s_1| \leq \beta\nu$ (recall that we actually need an ℓ_2 error bound of $\beta\nu$ for all coordinates). We now propose a simple solution pattern on the first coordinate that breaks this requirement. Namely, letting $(\Delta x_1, \Delta s_1) = (-1+\nu, \beta)$ (or vice-versa), we have that $(1+\Delta x_1)(1+\Delta s_1) = \nu(1+\beta)$, so the error is $\nu\beta$. However, $|\Delta x_1 \Delta s_1| \geq (1-\nu)\beta \approx \beta \gg \nu\beta$. In particular, affine scaling can never generate a step that has this structure on any of its coordinates. As we will see in the next section, it may be the case that long steps are forced to have the above structure on each coordinate. This will then motivate the need for a new type of predictor step. At a high level, the heuristic hope that the quadratic term is tiny – which motivated the design of affine scaling – will typically not hold for long steps.

In the next subsection, we examine the precise pattern of steps that take us to the end of the central path, which will highlight this structure more clearly.

1.2 Stepping to the End of the Path

Let us now assume that we are "close" to the optimal solution $z^* := (x^*, s^*, y^*) := z(0)$. Let $B \cup N = [n]$ be the optimality partition satisfying $B = \text{support}(x^*)$ and $N = \text{support}(s^*)$ (in particular, $x^*_N = 0_N$ and $s^*_B = 0_B$). We quantify the closeness to z^* by requiring closeness only for coordinates in the support of (x^*, s^*):

$$\|(x^*_B - 1_B, s^*_N - 1_N)\|_2 \leq \beta. \tag{2}$$

Let $\Delta z^{\text{id}} := (\Delta x^{\text{id}}, \Delta s^{\text{id}}, \Delta y^{\text{id}}) = z^* - z(1)$, be the step to the end of the path. We will show that this is the β-ideal step, that is, that it stays inside the $N_2(\beta)$. The shape of this step is as follows:

$$(\underbrace{\Delta x^{\text{id}}_B}_{\|\cdot\|_2 \leq \beta}, \underbrace{\Delta x^{\text{id}}_N}_{=-1_N}) \approx (0_B, -1_N) \quad (\underbrace{\Delta s^{\text{id}}_B}_{=-1_B}, \underbrace{\Delta s^{\text{id}}_N}_{\|\cdot\|_2 \leq \beta}) \approx (-1_B, 0_N). \tag{Polarization Pattern}$$

In the above step, the primal step Δx^{id} uniformly cancels out the coordinates in N while not moving the coordinates in B by much (i.e., by at most β), and the dual step Δs^{id} uniformly cancels out the coordinates in B while not moving the coordinates in N by much (the role of B and N get switched when moving to the dual). We will say that a step satisfying this pattern is *polarized* with respect to the partition (B, N). We will not in general require perfect uniformity on the cancelling out side (N for the primal and B for the dual) as with the step above, however the *degree of uniformity* will essentially determine the length of the step (assuming the step doesn't leave the $N_2(\beta)$ neighborhood).

Let us now analyze the centrality of $z(1) + \alpha\Delta z^{\text{id}} = (1_n, 1_n, y(1)) + \alpha\Delta z^{\text{id}}$ for $\alpha \in [0,1]$. Focusing on the coordinates in N and B separately:

$$\begin{aligned}
&(1_N + \alpha \underbrace{\Delta x^{\text{id}}_N}_{=-1_N})(1_N + \alpha\Delta s^{\text{id}}_N) - (1-\alpha)1_N = \alpha(1-\alpha)\Delta s^{\text{id}}_N. \\
&(1_B + \alpha\Delta x^{\text{id}}_B)(1_N + \alpha \underbrace{\Delta s^{\text{id}}_B}_{=-1_B}) - (1-\alpha)1_N = \alpha(1-\alpha)\Delta x^{\text{id}}_B.
\end{aligned} \tag{3}$$

From the above, the centrality distance satisfies:

$$\mathrm{dist}_{\mathrm{c}}(z(1)+\alpha\Delta z^{\mathrm{id}}) = \left\| \frac{(1_n+\alpha\Delta x^{\mathrm{id}})(1_n+\alpha\Delta s^{\mathrm{id}})}{1-\alpha} - 1_n \right\|_2 = \alpha \left\| (\Delta x_B^{\mathrm{id}}, \Delta s_N^{\mathrm{id}}) \right\|_2 \le \alpha\beta \le \beta.$$

Therefore, $z(1)+\alpha\Delta z^{\mathrm{id}} \in N_2(\beta)$, $\alpha \in [0,1]$. That is, the ideal step from $z(1)$ can follow the central path all the way until the end.

While we chose the end of the path as an illustrative example, long ideal steps can occur anywhere on the path. As an exercise in this direction, the reader can convince themselves that primal-dual central path for the pair $\min \sum_{i=1}^n \varepsilon^{i-1} x_i$, $\sum_{i=1}^n x_i = 1$, $x \ge 0_n$ and $\max y, s_i + y = \varepsilon^{i-1}$, $i \in [n]$, $s \ge 0_n$, has n long and straight segments for $\varepsilon > 0$ small enough, each requiring $O(1)$-trust region and affine scaling steps to traverse.

1.3 The Polarization Partition

Remaining with the above example, one may wonder how we can find the polarization partition (B,N) associated the ideal step? For this purpose, it turns out the affine scaling direction will reveal it to us. Recall that affine scaling satisfies $\Delta x^{\mathrm{id}} + \Delta s^{\mathrm{id}} = -1_n$ and the ideal step satisfies:

$$\Delta x^{\mathrm{id}} + \Delta s^{\mathrm{id}} = (-1_B, -1_N) + (\Delta x_B^{\mathrm{id}}, \Delta s_N^{\mathrm{id}}).$$

The orthogonality between the primal and dual movement will (perhaps surprisingly) imply that the affine scaling direction and the ideal step are not too far apart. Recalling that $\Delta x^{\mathrm{id}} - \Delta x^p \perp \Delta s^{\mathrm{id}} - \Delta x^p$, we have that

$$\begin{aligned} &(\Delta x^{\mathrm{id}} - \Delta x^p) + (\Delta s^{\mathrm{id}} - \Delta s^p) = (\Delta x_B^{\mathrm{id}}, \Delta s_N^{\mathrm{id}}) \Rightarrow \\ &\left\| (\Delta x^{\mathrm{id}} - \Delta x^p, \Delta s^{\mathrm{id}} - \Delta s^p) \right\|_2 = \left\| (\Delta x_B^{\mathrm{id}}, \Delta s_N^{\mathrm{id}}) \right\|_2 \le \beta. \end{aligned} \tag{4}$$

While affine scaling cannot do the job of the ideal step, the polarization pattern (3) together with the closeness (4) will imply that $(\Delta x^p, \Delta s^p)$ is also polarized according to (B,N) assuming β is small enough. The difference between the ideal and affine scaling step will be in the degree of uniformity on the cancellation side, i.e., perfect versus weak uniformity.

We now verify that $N = \{i \in [n] : \Delta x_i^p < \Delta s_i^p\}$ and $B = \{i \in [n] : \Delta s_i^p < \Delta x_i^p\}$ when $\beta < 1/3$. For $i \in N$, this follows since

$$\Delta x_i^p \overset{\text{Ineq. (4)}}{\le} \Delta x_i^{\mathrm{id}} + \beta = -1 + \beta \overset{\beta<1/3}{<} -2\beta \overset{\text{Ineq. (2)}}{\le} \Delta s_i^{\mathrm{id}} - \beta \overset{\text{Ineq. (4)}}{\le} \Delta s_i^p,$$

and a symmetric argument holds for $i \in B$ with the role of Δx^p and Δs^p reversed. This motivates the definition of the polarization partition with respect to affine scaling direction:

Definition 2 (Affine Scaling and Polarization Partition) *For $z = (x,s,y) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$, we define the affine scaling direction Δz^p at z by:*

$$s\Delta x^p + x\Delta s^p = -xs, \mathbf{A}\Delta x = 0_n, \mathbf{A}^\top \Delta y^p + \Delta s^p = 0_n. \qquad \text{(Affine Scaling Direction)}$$

The **normalized affine scaling** *direction is $\Delta\hat{z}^p := (\Delta\hat{x}^p, \Delta\hat{s}^p, \Delta\hat{y}^p) := (\frac{\Delta x^p}{x}, \frac{\Delta s^p}{s}, \frac{\Delta y^p}{\mu(z)})$, and the* **polarization partition** *at z is defined as $N = \{i \in [n] : \Delta\hat{x}_i < \Delta\hat{s}_i\}, B = [n] \setminus B$.*

Remark 1 The rescaling used to assume $(x(1), s(1)) = (1_n, 1_n)$ is the same as what is used to define the normalized affine scaling direction above. The normalized affine scaling direction $\Delta\hat{z} = (\Delta x^p, \Delta s^p, \Delta y^p)$ at $z \in \mathcal{P}_{++} \times \mathcal{D}_{++}$ in fact always satisfies

$$\Delta\hat{x}^p + \Delta\hat{s}^p = \frac{\Delta x^p}{x} + \frac{\Delta s^p}{s} = (XS)^{-1}(s\Delta x^p + x\Delta s^p) = (XS)^{-1}(-xs) = -1_n. \tag{5}$$

Furthermore, if $z = z(\mu) \in \mathrm{CP}$, the normalized movement subspaces $X(\mu)^{-1}\ker(\mathbf{A})$ and $S(\mu)^{-1}\operatorname{im}(\mathbf{A}^\top)$ are orthogonal complements since $X(\mu)S(\mu) = \mu\mathbf{I}_n$. In particular, $\Delta\hat{x}^p \perp \Delta\hat{s}^p$.

1.4 The Trust-Region Step

We are now in a position to define the new predictor step, the *trust-region step* due to Lan, Monteiro and Tsuchiya [LMT09], that will mimic the properties of the ideal step when one can take a long-step[1]. Beyond the end of the path analysis done above, it turns out that the affine scaling direction will in general reveal the polarization structure of the part of the path around the current iterate (if there is any) via the polarization partition (B, N) defined above. The trust-region step will then correspond to the step that is compatible with the polarization partition (B, N) and achieves the greatest "degree of cancellation". Specifically, for the primal, it will optimize the cancellation for the coordinates in N while being agnostic to the coordinates in B as long as they don't move much, and vice-versa for the dual.

Definition 3 (Trust-Region Step) *Let $z = (x, s, y) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$, $\beta \in (0, 1)$, and let (B, N) denote the polarization partition at z as defined in Definition 2. For $\beta > 0$, we define the β-*trust-region step *at z as the optimal solution $\Delta z^{tr} = (\Delta x^{tr}, \Delta s^{tr}, \Delta y^{tr})$, $\nu^{tr} \in [0, 1]$ to the* trust-region program *defined below*

$$\begin{aligned} &\min \quad \nu \\ &\left\|\frac{x_N + \Delta x_N}{x_N} - \nu 1_N\right\|_2 \le \beta\nu, \qquad \left\|\frac{\Delta x_B}{x_B}\right\|_2 \le \beta, \\ &\left\|\frac{s_B + \Delta s_B}{s_B} - \nu 1_B\right\|_2 \le \beta\nu, \qquad \left\|\frac{\Delta s_N}{s_N}\right\|_2 \le \beta, \\ &\mathbf{A}\Delta x = 0_n, \mathbf{A}^\top \Delta y + \Delta s = 0_n. \end{aligned} \tag{TR(β)}$$

If $N = \emptyset$, we set $\Delta x^{tr} = 0_n$, and similarly if $B = \emptyset$, we set $(\Delta s^{tr}, \Delta y^{tr}) = (0_n, 0_m)$.

Remark 2 The name trust-region step is derived from the trust-region constraints $\left\|\frac{\Delta x_B}{x_B}\right\|_2 \le \beta$, $\left\|\frac{\Delta s_N}{s_N}\right\|_2 \le \beta$, which isolate the region where we can "trust" the solution.

Remark 3 While the affine scaling direction can be computed by solving a linear system of equations, the trust-region step involves solving a non-trivial quadratic program. Due its special structure, it can in fact be solved up to $1 + \varepsilon$ relative accuracy using $\mathrm{poly}(n)\log(1/\varepsilon)$ arithmetic operations. We will not cover this here.

The following lemma justifies that the trust-region step is a good predictor step, in that it stays inside the relevant neighborhood. We defer the proof to the end of the notes as it is relatively similar to the end of path analysis.

[1]We note that [LMT09] had a related but different motivation, and did not directly compare to the ideal step as we do here. Our formulation of the trust-region step slightly differs from theirs to help simplify the presentation.

LEMMA 4 *Let* $z = (x, s, y) \in N(\beta/2)$, $\beta \in [0, 1/2]$, $\mu := \mu(z)$, *and let* $\Delta z^{tr}, \nu^{tr}$ *be the* $\bar{\beta} := \beta/8$ *trust-region step at* z. *Then, for all* $\alpha \in [0, 1)$, *we have*

$$\left\| \frac{(x + \alpha \Delta x^{tr})(s + \alpha \Delta s^{tr})}{((1 - \alpha) + \alpha \nu^{tr})\mu} - 1_n \right\|_2 \leq \beta. \tag{6}$$

If $\nu^{tr} = 0$, *then* $\mu(z + \Delta z^{tr}) = 0$ *and* $z + \Delta z^{tr} \in \overline{N}_2(\beta)$.

REMARK 4 A minor technical issue is that the optimality gap will not necessarily satisfy $\mu(z + z^{tr}) = \nu^{tr}\mu(z)$ when z^{tr} is the trust-region step (it will be multiplicatively very close to this however). The lemma therefore explicitly computes the distance to centrality with the appropriate choice of parameter. While this discrepancy will force us to keep track of the current parameter μ more explicitly in the forthcoming algorithm, it can for all intents and purposes be ignored.

2 A Locally Optimal IPM

We are now ready to state our main local optimality result. In words, it states that taking the better of the affine scaling and the trust-region step (in terms of which reduces the gap the most), results in a step that is at least as long as the β-ideal step, under the provision that we increase the neighborhood size to $O(\beta)$. For simplicity of exposition, we state this only for a starting iterate on the central path. The proof is deferred to Section 3.

THEOREM 5 *Let* $z(\mu_1) \in \mathrm{CP}$, *let* Δz^{id} *denote the* β*-ideal step from* $z(\mu_1)$, $\beta \in (0, 1/50)$, *and let* $\mu_2 := \mu(z + \Delta z^{\mathrm{id}})$, $\nu := \mu_2/\mu_1$. *Then the following holds:*

1. *If* $\nu \geq 1/4$, *then* $z(\mu_1) + (1 - \nu)\Delta z^p \in N_2(25\beta)$, *where* Δz^p *is the affine scaling direction.*
2. *If* $\nu \leq 1/4$, *then* $(\Delta z^{\mathrm{id}}, \nu)$ *is a feasible solution to* $\mathrm{TR}(5\beta)$ *at* $z(\mu_1)$. *In particular, the value of* $\mathrm{TR}(5\beta)$ *is at most* ν.

2.1 The Trust Region IPM

We now present a full algorithm which at each iteration takes the best of the affine scaling and trust-region step. In contrast to the basic predictor-corrector algorithm, this IPM is guaranteed to terminate in a finite number of iterations. In the next lecture, we will give a combinatorial characterization of this number of iterations based on the notion of *straight-line complexity*.

REMARK 5 As mentioned previously, the trust-region step does not necessarily satisfy $\mu(z + \Delta z^{tr}) = \nu^{tr}\mu(z)$. We address the issue by explicitly keeping track of the parameter μ, and using this parameter to define the corrector step on line (4)). This will ensure that the corrected iterate z_c satisfies $\mu(z_c) = \mu\nu^{tr}$ and $z_c \in N_2(\beta/2)$.

3 Proof of Local Optimality

The main goal of this section is to prove Theorem 5, whose proof is given in Section 3.3.

Algorithm 1: TRUST-REGION IPM

Input : Constraint matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(\mathbf{A}) = m$, and initial iterate $z^0 := (x^0, s^0, y^0) \in N_2(\beta)$, $\beta \in (0, 1/2]$.

Output: $z^* = (x^*, s^*, y^*) \in \overline{N}_2(\beta)$, $\langle x^*, s^* \rangle = 0$.

$z = (x, s, y) \leftarrow (x^0, s^0, y^0)$;
$\mu \leftarrow \langle x, s \rangle / n$;
while $\mu > 0$ **do**
  `// Modified Corrector step.  Ensures` $\mu(z) = \mu$ `and` $z \in N(\beta/2)$.
  Compute Δz^c at z according $s\Delta x^c + x\Delta s^c = \mu 1_n - xs$, $\mathbf{A}\Delta x = 0_n$, $\mathbf{A}^\top \Delta y^c + \Delta s^c = 0_n$;
  $z \leftarrow z + \Delta z^c$;
  Compute Δz^p at z according to (Affine Scaling Direction);
  Choose $\alpha^p \in [0, 1]$ as large as possible so that $z + \alpha^p \Delta z^p \in \overline{N}_2(\beta)$;
  if $\alpha \geq 3/4$ **then**
    Compute $(\Delta z^{tr}, \nu^{tr})$ at z according to $\mathrm{TR}(\beta/8)$.
  else
    $\nu^{tr} = 1$;
  if $(1 - \alpha^p) < \nu^{tr}$ **then**
    $z \leftarrow z + \alpha^p \Delta z^p$;
    $\mu \leftarrow \mu(1 - \alpha^p)$;
  else
    $z \leftarrow z + \Delta z^{tr}$;
    $\mu \leftarrow \mu \nu^{tr}$;
return z;

3.1 Helper Propositions

We will need that the duality gap is behaves linearly on lines.

PROPOSITION 6 *For $z^i := (x^i, s^i, y^i) \in \mathcal{P} \times \mathcal{D}$, $i \in \{1, 2\}$, and $\alpha \in \mathbb{R}$, we have that*

$$\mu((1 - \alpha)z^1 + \alpha z^2) = (1 - \alpha)\mu(z^1) + \alpha\mu(z^2).$$

PROOF:

$$\begin{aligned} n\mu((1-\alpha)z^1 + \alpha z^2) &= \left\langle (1-\alpha)x^1 + \alpha x^2, (1-\alpha)s^1 + \alpha s^2 \right\rangle \\ &= \left\langle c, (1-\alpha)x^1 + \alpha x^2 \right\rangle - \left\langle b, (1-\alpha)y^1 + \alpha y^2 \right\rangle \\ &= (1-\alpha)\left(\left\langle c, x^1 \right\rangle - \left\langle b, y^1 \right\rangle \right) + \alpha \left(\langle c, x^2 \rangle - \langle b, y^2 \rangle \right) \\ &= (1-\alpha)\left\langle x^1, s^1 \right\rangle + \alpha \left\langle x^2, s^2 \right\rangle = n((1-\alpha)\mu(z^1) + \alpha\mu(z^2)), \end{aligned}$$

where the second and fourth equality follow from the gap formula. □

The next proposition gives simple estimates on the shape of solutions to the to bivariate quadratic equation $(1 + a)(1 + b) = \nu$ as a function of ν and the product $\gamma := ab$. In particular, this will establish that (a, b) "polarize" when γ and ν are suitably small.

PROPOSITION 7 *Let* $(1+a)(1+b) := \nu \in [0,1]$, $a \le b$ *and* $ab := \gamma$. *Then, the following holds:*

1. *If* $|\gamma| < 4$, $1+b \le \frac{1+\sqrt{|\gamma|}/2}{1-\sqrt{|\gamma|}/2}$.

2. *If* $\nu \in [0,1/4]$ *and* $|\gamma| \le 1/6$, *then*

$$a = -1 + \nu(1+c\gamma), \quad b = \frac{\gamma}{-1+\nu(1+c\gamma)}, \quad \text{for some } c \in [0,2].$$

Moreover, $|b| \le \frac{3}{2}|\gamma|$ *and* $b - a \ge 3/4 - 2|\gamma| \ge 5/12$.

PROOF:

Proof of (1). If $\gamma = ab \ge 0$, then $(1+a)(1+b) \in [0,1]$ implies that $a, b \le 0$, and hence $1+b \le 1$. Now assume $\gamma = ab \le 0$. Then, we can write $b = c\sqrt{|\gamma|}$ and $a = -\sqrt{|\gamma|}/c$ for $c > 0$, recalling that $a \le b$. Using that $1 \ge \nu = (1+a)(1+b) = 1 + (c - 1/c)\sqrt{|\gamma|} - |\gamma|$, one can verify that $c \le \frac{1}{1-\sqrt{|\gamma|}/2}$ (the assumption $|\gamma| < 4$ ensure the denominator is positive). In particular, $1+b \le 1 + \frac{\sqrt{|\gamma|}}{1-\sqrt{|\gamma|}/2} = \frac{1+\sqrt{|\gamma|}/2}{1-\sqrt{|\gamma|}/2}$.

Proof of (2). Since $\nu \in [0,1/4]$ and $|\gamma| \le 1/6$, assuming $c \in [0,2]$ we have

$$a = -1 + \nu(1+c\gamma) \le -1 + 1/4(1+2|\gamma|) = -3/4 + (1/2)|\gamma| \le -2/3, \tag{7}$$

$$|b| = \frac{|\gamma|}{|a|} \overset{\text{Eq. (7)}}{\le} (3/2)|\gamma| \le 1/4. \tag{8}$$

In particular, $a < b$ and $b - a \ge (3/4 - (1/2)|\gamma|) - (3/2)|\gamma| = (3/4) - 2|\gamma| \ge 5/12$.

We now prove that $c \in [0,2]$. If $\nu = 0$ or $\gamma = 0$, then since $a \le b$ we have $c = 0$ and the claim holds. If both ν, γ are non-zero, define

$$f(t) := (1 + (-1 + \nu(1+t\gamma)))(1 + \frac{\gamma}{-1+\nu(1+t\gamma)}) - \nu = \nu\gamma\left(\frac{t - (1+t\gamma)(1+t\nu)}{1-\nu(1+t\gamma)}\right).$$

$f(t)$ has precisely two roots corresponding to $-1+\nu(1+t\gamma) \in \{a, b\}$ (possibly a double root if $a = b$). If f has a root $r \in [0,2]$, then $r = c$ since $a \le b$ (given (7), (8)). Express $f(t) = \nu\gamma\frac{h(t)}{1-\nu(1+t\gamma)}$ with $h(t) = t - (1+t\gamma)(1+t\nu)$. By (7), the denominator $1 - \nu(1+t\gamma) \ge 2/3$ for $t \in [0,2]$, thus f is continuous on $[0,2]$ and has its sign determined by h. The existence of a root of f in $[0,2]$ now follows since $h(0) = -1$ and $h(2) = 2 - (1+2\gamma)(1+2\nu) \ge 2 - (1+2|\gamma|)(1+2\nu) = 2 - (1+2/6)(1+2/4) = 0$. □

3.2 Straight Line Segments in the ℓ_2 Neighborhood

The starting point of our analysis is the following characterization of when the straight-line segment between central path points lies in the ℓ_2-neighborhood. This is a generalization of the analysis done in Section 1.2 which works anywhere on the path.

Lemma 8 *Let* $\mu_1 > \mu_2 \geq 0$, $\nu = \frac{\mu_2}{\mu_1}$, $z(\mu_i) = (x(\mu_i), s(\mu_i), y(\mu_i)) \in \overline{\mathrm{CP}}$, $i \in \{1,2\}$, *and* $(\Delta\hat{x}, \Delta\hat{s}) := (\frac{x(\mu_2)-x(\mu_1)}{x(\mu_1)}, \frac{s(\mu_2)-s(\mu_1)}{s(\mu_1)})$. *Then* $[z(\mu_1), z(\mu_2)] \subset \overline{N}_2(\beta)$, $\beta \in (0,1)$, *if and only if*

$$\|\Delta\hat{x}\Delta\hat{s}\|_2 \leq (1+\sqrt{\nu})^2\beta. \tag{9}$$

Moreover, letting $\hat{z}^p := (\Delta\hat{x}^p, \Delta\hat{s}^p, \Delta\hat{y}^p)$ *denote the normalized affine scaling direction at* $z(\mu_1)$*, we have*

$$\|(\Delta\hat{x} - (1-\nu)\Delta\hat{x}^p, \Delta\hat{s} - (1-\nu)\Delta\hat{s}^p)\|_2 = \|\Delta\hat{x}\Delta\hat{s}\|_2 \,. \tag{10}$$

Proof: We will give a exact expression for the centrality error of $z^\alpha := (x^\alpha, s^\alpha, y^\alpha) := (1-\alpha)z(\mu_1) + \alpha z(\mu_2)$, $\alpha \in [0,1]$. By definition, note that

$$\nu 1_n = \frac{\mu_2}{\mu_1} 1_n = \frac{x(\mu_2)s(\mu_2)}{x(\mu_1)s(\mu_1)} = (1_n + \Delta\hat{x})(1_n + \Delta\hat{s}). \tag{11}$$

The centrality vector for z^α is given by:

$$\begin{aligned}
\frac{x^\alpha s^\alpha}{\mu(z^\alpha)} &\overset{\text{Proposition 6}}{=} \frac{x^\alpha s^\alpha}{(1-\alpha)\mu_1 + \alpha\mu_2} \overset{x(\mu_1)s(\mu_1)=\mu_1 1_n}{=} \frac{1}{(1-\alpha)+\alpha\nu}\frac{x^\alpha}{x(\mu_1)}\frac{s^\alpha}{s(\mu_1)} \\
&= \frac{(1_n + \alpha\Delta\hat{x})(1_n + \alpha\Delta\hat{s})}{(1-\alpha)+\alpha\nu} = \frac{(1-\alpha)1_n + \alpha(1_n+\Delta\hat{x})(1_n+\Delta\hat{s}) - \alpha(1-\alpha)\Delta\hat{x}\Delta\hat{s}}{(1-\alpha)+\alpha\nu} \\
&\overset{\text{Eq. (11)}}{=} \frac{(1-\alpha)1_n + \alpha\nu 1_n - \alpha(1-\alpha)\Delta\hat{x}\Delta\hat{s}}{(1-\alpha)+\alpha\nu} = 1_n - \frac{\alpha(1-\alpha)}{(1-\alpha)+\alpha\nu}\Delta\hat{x}\Delta\hat{s}.
\end{aligned}$$

Using the above, the centrality error $\mathrm{dist_c}(z^\alpha)$ is:

$$\mathrm{dist_c}(z^\alpha) := \left\| \frac{x^\alpha s^\alpha}{\mu(z^\alpha)} - 1_n \right\|_2 = \frac{\alpha(1-\alpha)}{(1-\alpha)+\alpha\nu} \|\Delta\hat{x}\Delta\hat{s}\|_2 \,.$$

Therefore $\mathrm{dist_c}(z^\alpha) \leq \beta, \forall \alpha \in [0,1)$ iff $\max_{\alpha\in[0,1]} \frac{\alpha(1-\alpha)}{(1-\alpha)+\alpha\nu} \|\Delta\hat{x}\Delta\hat{s}\|_2 \leq \beta$. The expression $\frac{\alpha(1-\alpha)}{(1-\alpha)+\alpha\nu}$ is uniquely maximized in $[0,1]$ at $\alpha = (1+\sqrt{\nu})^{-1}$ where it achieves the value $\left(1+\sqrt{\nu}\right)^{-2}$. In particular, $[z(\mu_1), z(\mu_2)] \subset N_2(\beta)$ iff $\|\Delta\hat{x}\Delta\hat{s}\|_2 \leq (1+\sqrt{\nu})^2\beta$, as needed.

For the moreover, note by construction that $\Delta\hat{x} - (1-\nu)\Delta\hat{x}^p \in X(\mu_1)^{-1}\ker(\mathbf{A})$ and $\Delta\hat{s} - (1-\nu)\Delta\hat{s}^p \in S(\mu_1)^{-1}\operatorname{im}(\mathbf{A}^\top)$, which are orthogonal by Remark 1. Therefore

$$\begin{aligned}
\|(\Delta\hat{x} - (1-\nu)\Delta\hat{x}^p, \Delta\hat{s} - (1-\nu)\Delta\hat{s}^p)\|_2^2 &:= \|\Delta\hat{x} - (1-\nu)\Delta\hat{x}^p\|_2^2 + \|\Delta\hat{s} - (1-\nu)\Delta\hat{s}^p\|_2^2 \\
&\overset{\text{orthogonality}}{=} \|\Delta\hat{x} + \Delta\hat{s} - (1-\nu)(\Delta\hat{x}^p + \Delta\hat{s}^p)\|_2^2 \\
&\overset{\text{Eq. (5)}}{=} \|\Delta\hat{x} + \Delta\hat{s} + (1-\nu)1_n\|_2^2 \\
&\overset{\text{Eq. (11)}}{=} \|1_n + \Delta\hat{x} + \Delta\hat{s} - (1_n+\Delta\hat{x})(1+\Delta\hat{s})\|_2^2 = \|\Delta\hat{x}\Delta\hat{s}\|_2^2 \,.
\end{aligned}$$

□

3.3 Proof of Theorem 5

Proof: Let $\Delta\hat{z}^{\mathrm{id}} := (\Delta\hat{x}^{\mathrm{id}}, \Delta\hat{s}^{\mathrm{id}}, \Delta\hat{y}^{\mathrm{id}}) := (\frac{\Delta x^{\mathrm{id}}}{x(\mu_1)}, \frac{\Delta s^{\mathrm{id}}}{s(\mu_1)}, \frac{\Delta\hat{y}^{\mathrm{id}}}{\mu_1})$ denote the normalized ideal step. Similarly, let $\Delta z^p := (\Delta x^p, \Delta s^p, \Delta y^p)$, $\Delta\hat{z}^p := (\Delta\hat{x}^p, \Delta\hat{s}^p, \Delta y^p)$ denote the predictor and normalized affine scaling direction at $z(\mu_1)$. By definition of Δz^{id}, recall that $[z(\mu_1), z(\mu_2)] \subset \overline{N}_2(\beta)$, where $z(\mu_2) = z + \Delta z^{\mathrm{id}}$.

Combining (9), (10), (11) from Lemma 8, letting $(\Gamma_x, \Gamma_s) := ((1-\nu)\Delta\hat{x}^p - \Delta\hat{x}^{\mathrm{id}}, (1-\nu)\Delta\hat{s}^p - \Delta\hat{s}^{\mathrm{id}})$, using that $\nu \in [0,1]$, we have that

$$\left\|\Delta\hat{x}^{\mathrm{id}}\Delta\hat{s}^{\mathrm{id}}\right\|_\infty \le \left\|\Delta\hat{x}^{\mathrm{id}}\Delta\hat{s}^{\mathrm{id}}\right\|_2 \le (1+\sqrt{\nu})^2\beta \le 4\beta \tag{12}$$

$$\|(\Gamma_x, \Gamma_s))\|_2 \le (1+\sqrt{\nu})^2\beta \le 4\beta, \tag{13}$$

$$\left(\frac{x(\mu_2)}{x(\mu_1)}\right)\left(\frac{s(\mu_2)}{s(\mu_1)}\right) = \underbrace{(1_n + \Delta\hat{x}^{\mathrm{id}})}_{\ge 0_n}\underbrace{(1_n + \Delta\hat{s}^{\mathrm{id}})}_{\ge 0_n} = \nu 1_n. \tag{14}$$

Proof of (1). By applying (14) together with Proposition 7 part (1) to the coordinates of $(\Delta\hat{s}, \Delta\hat{x})$, we have that

$$\max\left\{\left\|1_n + \Delta\hat{x}^{\mathrm{id}}\right\|_\infty, \left\|1_n + \Delta\hat{s}^{\mathrm{id}}\right\|_\infty\right\} \le \frac{1+\sqrt{\beta}}{1-\sqrt{\beta}}.$$

The centrality error can now be expressed as follows:

$$\begin{aligned}
\mathrm{dist_c}(z(\mu_1) + (1-\nu)\Delta z^p) &= \left\|\frac{(x(\mu_1) + (1-\nu)\Delta x^p)(s(\mu_1) + (1-\nu)\Delta s^p)}{\mu_2} - 1_n\right\|_2 \\
&= \left\|\frac{(x(\mu_1) + (1-\nu)\Delta x^p)(s(\mu_1) + (1-\nu)\Delta s^p)}{\nu x(\mu_1)s(\mu_1)} - 1_n\right\|_2 \\
&= \left\|\frac{1}{\nu}(1_n + \Delta\hat{x}^{\mathrm{id}} + \Gamma_x)(1_n + \Delta\hat{s}^{\mathrm{id}} + \Gamma_s) - 1_n\right\|_2 \\
&= \frac{1}{\nu}\left\|(1_n + \Delta\hat{s}^{\mathrm{id}})\Gamma_x + (1_n + \Delta\hat{x}^{\mathrm{id}})\Gamma_s + \Gamma_x\Gamma_s\right\|_2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{dist_c}(z(\mu_1) + (1-\nu)\Delta z^p) &\overset{\nu\ge 1/4}{\le} 4\left\|(1_n + \Delta\hat{s}^{\mathrm{id}})\Gamma_x + (1_n + \Delta\hat{x}^{\mathrm{id}})\Gamma_s + \Gamma_x\Gamma_s\right\|_2 \\
&\le 4\left(\left\|(1_n + \Delta\hat{s}^{\mathrm{id}})\Gamma_x + (1_n + \Delta\hat{x}^{\mathrm{id}})\Gamma_s\right\|_2 + 4\|\Gamma_x\Gamma_s\|_2\right) \\
&\overset{\text{orthogonality}}{\le} 4\sqrt{\|(1_n + \Delta\hat{s}^{\mathrm{id}})\Gamma_x\|_2^2 + \|(1_n + \Delta\hat{x}^{\mathrm{id}})\Gamma_s\|_2^2} + 4\|\Gamma_x\Gamma_s\|_2 \\
&\le 4\max\left\{\left\|1_n + \Delta\hat{s}^{\mathrm{id}}\right\|_\infty, \left\|1_n + \Delta\hat{x}^{\mathrm{id}}\right\|_\infty\right\}\|(\Gamma_x, \Gamma_s)\|_2 + \frac{4}{2}(\|\Gamma_x\|_2^2 + \|\Gamma_s\|_2^2) \\
&\le 4\left(\frac{1+\sqrt{\beta}}{1-\sqrt{\beta}}\right)(4\beta) + 2(4\beta)^2 \overset{\beta\le 1/50}{\le} 4(\frac{4}{3})(4\beta) + \beta \le 25\beta.
\end{aligned}$$

Proof of (2). Define the partition (B, N) given by $N = \{i \in [n] : \Delta\hat{x}_i^{\mathrm{id}} < \Delta\hat{s}_i^{\mathrm{id}}\}$ and $B = [n] \setminus B$. Given (13), $4\beta \le 1/6$ and $\nu \le 1/4$, by Proposition 7 part (2) applied to the coordinates $(\Delta\hat{x}^{\mathrm{id}}, \Delta\hat{s}^{\mathrm{id}})$, we have that

$$\begin{aligned}
&(1-\nu)\min\{\min_{i\in N}(\Delta\hat{s}_i^p - \Delta\hat{x}_i^p), \min_{i\in B}(\Delta\hat{x}_i^p - \Delta\hat{s}_i^p)\} \ge \\
&\min\{\min_{i\in N}\Delta\hat{s}_i^{\mathrm{id}} - \Delta\hat{x}_i^{\mathrm{id}}, \min_{i\in B}\Delta\hat{x}_i^{\mathrm{id}} - \Delta\hat{s}_i^{\mathrm{id}}\} - 2\|(\Gamma_x, \Gamma_s)\|_2 \ge \frac{5}{12} - 8\beta \overset{\beta\le 1/50}{\ge} 1/4.
\end{aligned}$$

Therefore the polarization partition induced by the normalized affine scaling direction $\Delta\hat{z}^p$ at $z(\mu_1)$ agrees with the partition (B, N) defined above.

Using $\nu \le 1/4$, we strengthen the bound (12) to $\left\|\Delta\hat{x}^{\mathrm{id}}\Delta\hat{s}^{\mathrm{id}}\right\|_2 \le (1+\sqrt{\nu})^2\beta \le 9/4\beta$. By applying Proposition 7 part (2) again, we certify that $(\Delta z^{\mathrm{id}}, \nu)$ is a feasible solution for TR(5β):

$$\left\|\frac{x_N(\mu_1)+\Delta x_N^{\mathrm{id}}}{x_N(\mu_1)} - \nu 1_N\right\|_2 = \left\|\Delta\hat{x}_N^{\mathrm{id}} + (1-\nu)1_N\right\|_2 \le 2\nu\left\|\Delta\hat{x}_N^{\mathrm{id}}\Delta\hat{s}_N^{\mathrm{id}}\right\|_2 \le 5\nu\beta,$$
$$\left\|\frac{\Delta x_B^{\mathrm{id}}}{x_B(\mu_1)}\right\|_2 = \left\|\Delta\hat{x}_B^{\mathrm{id}}\right\|_2 \le \frac{3}{2}\left\|\Delta\hat{x}_B^{\mathrm{id}}\Delta\hat{s}_B^{\mathrm{id}}\right\|_2 \le 5\beta,$$
$$\left\|\frac{s_B(\mu_1)+\Delta s_B^{\mathrm{id}}}{s_B(\mu_1)} - \nu 1_B\right\|_2 = \left\|\Delta\hat{s}_B^{\mathrm{id}} + (1-\nu)1_B\right\|_2 \le 2\nu\left\|\Delta\hat{x}_B^{\mathrm{id}}\Delta\hat{s}_B^{\mathrm{id}}\right\|_2 \le 5\nu\beta,$$
$$\left\|\frac{\Delta s_N^{\mathrm{id}}}{s_N(\mu_1)}\right\|_2 = \left\|\Delta\hat{s}_N^{\mathrm{id}}\right\|_2 \le \frac{3}{2}\left\|\Delta\hat{x}_N^{\mathrm{id}}\Delta\hat{s}_N^{\mathrm{id}}\right\|_2 \le 5\beta.$$

□

4 Proof of Validity of the Trust-Region Step (Lemma 4)

PROOF: For $\alpha \in [0,1]$, we claim that $(\alpha\Delta z^{tr}, (1-\alpha)+\alpha\nu^{tr})$ is also TR$(\bar\beta)$ solution:

$$\left\|\frac{x_N+\alpha\Delta x_N^{tr}}{x_N} - ((1-\alpha)+\alpha\nu^{tr})1_N\right\|_2 = \alpha\left\|\frac{x_N+\Delta x_N^{tr}}{x_N} - \nu^{tr}1_N\right\|_2 \le \alpha\bar\beta\nu^{tr} \le \bar\beta\nu^{tr},$$
$$\left\|\frac{\alpha\Delta x_B^{tr}}{x_B}\right\|_2 \le \alpha\bar\beta \le \bar\beta,$$

The reasoning for $\alpha(\Delta s^{tr}, \Delta y^{tr})$ is analogous. Given the above, it suffices to prove the statement for $\alpha = 1$ and $\nu^{tr} > 0$ (only needed to avoid division by 0). From the TR$(\bar\beta)$ guarantees, we have:

$$\begin{aligned}\left\|\frac{(x_N+\Delta x_N^{tr})(s_N+\Delta s_N^{tr})}{x_N s_N} - \nu^{tr}1_N\right\|_2 &\le \left\|\frac{x_N+\Delta x_N^{tr}}{x_N} - \nu^{tr}1_N\right\|_2 + \left\|\frac{(x_N+\Delta x_N^{tr})(\Delta s_N^{tr})}{x_N s_N}\right\|_2 \\ &\le \bar\beta\nu^{tr} + \left\|\frac{x_N+\Delta x_N^{tr}}{x_N}\right\|_\infty \left\|\frac{\Delta s_N^{tr}}{s_N}\right\|_2 \\ &\le \bar\beta\nu^{tr} + ((1+\bar\beta)\nu^{tr})(\bar\beta) = (2\bar\beta+\bar\beta^2)\nu^{tr} \overset{\bar\beta\le 1/5}{\le} 2.2\bar\beta.\end{aligned}$$

By a symmetric argument $\left\|\frac{(x_B+\Delta x_B^{tr})(s_B+\Delta s_B^{tr})}{x_B s_B} - \nu^{tr}1_B\right\|_2 \le 2.2\bar\beta\nu^{tr}$. Combining both bounds:

$$\left\|\frac{(x+\Delta x^{tr})(s+\Delta s^{tr})}{xs} - \nu^{tr}1_n\right\|_2 \le \sqrt{2}(2.2\bar\beta\nu^{tr}). \tag{15}$$

We now derive (6):

$$
\begin{aligned}
\left\| \frac{(x+\Delta x^{tr})(s+\Delta s^{tr})}{\nu^{tr}\mu} - 1_n \right\|_2 &\leq \left\| \frac{xs}{\mu} - 1_n \right\|_2 + \left\| \frac{xs}{\mu\nu^{tr}} \left(\frac{(x+\Delta x^{tr})(s+\Delta s^{tr})}{xs} - \nu^{tr} 1_n \right) \right\|_2 \\
&\leq \left\| \frac{xs}{\mu} - 1_n \right\|_2 + \left\| \frac{xs}{\mu\nu^{tr}} \right\|_\infty \left\| \frac{(x+\Delta x^{tr})(s+\Delta s^{tr})}{xs} - \nu^{tr} 1_n \right\|_2 \\
&\overset{z\in N_2(\beta/2)}{\leq} \frac{\beta}{2} + \frac{1+\beta/2}{\nu^{tr}} \left\| \frac{(x+\Delta x^{tr})(s+\Delta s^{tr})}{xs} - \nu^{tr} 1_n \right\|_2 \\
&\overset{\text{Eq. (15)},\ \beta\leq 1/2}{\leq} \frac{\beta}{2} + (5/4)(\sqrt{2}(2.2\bar{\beta})) \leq \frac{\beta}{2} + 4\bar{\beta} \overset{\bar{\beta}=\frac{\beta}{8}}{\leq} \beta.
\end{aligned}
$$

If $\nu^{tr} = 0$, the TR($\bar{\beta}$) guarantees imply that $(x+\Delta x^{tr})(s+\Delta s^{tr}) = 0_n$ and hence $\mu(z+\Delta z^{tr}) = 0$. By Proposition 6, we have $\mu(z+\alpha\Delta z^{tr}) = (1-\alpha)\mu(z)$, and thus (6) implies that $z+\alpha\Delta z^{tr} \in N_2(\beta)$ for $\alpha \in [0,1)$. In particular, $z+\Delta z^{tr} \in \overline{N}_2(\beta)$. □

References

[LMT09] Guanghui Lan, Renato DC Monteiro, and Takashi Tsuchiya. A polynomial predictor-corrector trust-region algorithm for linear programming. *SIAM Journal on Optimization*, 19(4):1918–1946, 2009.