Lecture 13

Primal-dual interior point methods aim to solve the problem

$$\text{minimize} \quad \langle \boldsymbol{c}, \boldsymbol{x} \rangle \qquad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0} \tag{P}$$

for a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times d}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^d$, by applying Newton-type iterations to the optimality conditions of linear programming. More precisely, we have seen the following algorithm. Recall the feasible sets

$$\begin{aligned} \mathcal{F} &= \{(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{x}) : \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c},\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{s} \geq \boldsymbol{0}\} \\ \mathcal{F}^\circ &= \{(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{x}) : \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c},\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} > \boldsymbol{0},\ \boldsymbol{s} > \boldsymbol{0}\} \end{aligned}$$

The a simple primal-dual interior point method can be described as follows.

- Start with $(\boldsymbol{x}^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{s}^{(0)}) \in \mathcal{F}^\circ$;
- For each $k \geq 0$, compute the duality parameter

  $$\mu^{(k)} = \frac{1}{d} \sum_{i=1}^{d} x_i s_i$$

  and choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$. Solve

  $$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{A}^\top & \boldsymbol{I} \\ \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}^{(k)} & \boldsymbol{0} & \boldsymbol{X}^{(k)} \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{y} \\ \Delta \boldsymbol{s} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{X}^{(k)} \boldsymbol{S}^{(k)} \boldsymbol{e} + \sigma \mu^{(k)} \end{pmatrix}$$

  and compute

  $$\begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \\ \boldsymbol{s}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{k} \\ \boldsymbol{y}^{k} \\ \boldsymbol{s}^{k} \end{pmatrix} + \alpha_k \begin{pmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{y} \\ \Delta \boldsymbol{s} \end{pmatrix},$$

  for a small enough $\alpha_k > 0$ to ensure non-negativity.

In each iteration, a Newton step is taken in the direction of the *central path*. This is a curve in $\mathcal{F}^\circ$ defined as the set of solutions of

$$\begin{aligned} \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{s} - \boldsymbol{c} &= \boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} &= \boldsymbol{0} \\ \boldsymbol{X}\boldsymbol{S}\boldsymbol{e} &= \tau \boldsymbol{e} \\ \boldsymbol{x} &> \boldsymbol{0} \\ \boldsymbol{s} &> \boldsymbol{0}, \end{aligned} \tag{3.1}$$

where $\tau > 0$.

3.1 Path-following methods

A path-following method tries to ensure that each iterate is *close* to the central path. What it means to be close to the central path depends on the neighbourhood we choose. Here, we will look at the (one-sided) ∞-norm neighbourhood

$$\mathcal{N}_{-\infty}(\gamma) = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \in \mathcal{F}^\circ : x_i s_i \geq \gamma\mu,\ 1 \leq i \leq d\}$$

for some $\gamma \in (0, 1]$ (say, $\gamma = 10^{-3}$). In words, each $x_i s_i$ has to be at least some small multiple of their average value. To see what this has to do with the ∞-norm neighbourhood, consider the set of $\boldsymbol{x}$ such that

$$\|\boldsymbol{X}\boldsymbol{S}\boldsymbol{e} - \mu\boldsymbol{e}\|_\infty \leq (1-\gamma)\mu \iff \forall 1 \leq i \leq d,\ \gamma\mu \leq x_i s_i \leq 2 - \gamma,$$

and we are only interested in the lower inequality.

The so-called *long-step path-following* interior point method can then be described as follows.

- Start with $(\boldsymbol{x}^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{s}^{(0)}) \in \mathcal{N}_{-\infty}(\gamma)$;
- For each $k \geq 0$, compute the duality parameter

  $$\mu^{(k)} = \frac{1}{d}\sum_{i=1}^{d} x_i s_i$$

  and choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$. Solve

  $$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{A}^\top & \boldsymbol{I} \\ \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}^{(k)} & \boldsymbol{0} & \boldsymbol{X}^{(k)} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \\ \Delta\boldsymbol{s} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{X}^{(k)}\boldsymbol{S}^{(k)}\boldsymbol{e} + \sigma\mu^{(k)} \end{pmatrix}$$

  and compute

  $$\begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \\ \boldsymbol{s}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{k} \\ \boldsymbol{y}^{k} \\ \boldsymbol{s}^{k} \end{pmatrix} + \alpha_k \begin{pmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \\ \Delta\boldsymbol{s} \end{pmatrix},$$

  for a small enough $\alpha_k \in [0, 1]$ is the largest value such that $(\boldsymbol{x}^{(k+1)}, \boldsymbol{y}^{(k+1)}, \boldsymbol{s}^{(k+1)}) \in \mathcal{N}_{-\infty}(\gamma)$.

Remark 3.1. As noted at the end of Lecture 11, to find an initial point in $\mathcal{F}^\circ$ might not be trivial. In practice one can therefore also use the algorithm described above using *infeasible* points, though in this case we have to make sure that the residual norms $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$ and $\|\boldsymbol{c} - \boldsymbol{s} - \boldsymbol{A}^\top \boldsymbol{y}\|$ remain bounded.

Visualising the algorithm

The feasible set $\mathcal{F}^\circ$ usually lives in a space that can't be easily visualised, but if the dual version is two-dimensional,

$$\{\boldsymbol{y} \in \mathbb{R}^2 : \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c},\ \boldsymbol{s} \geq \boldsymbol{0}\},$$

then we have a chance to see how the trajectories of the iterates in $\boldsymbol{y}$ look like. Consider, for example, the linear programming problem whose dual is given by

$$\begin{aligned} &\text{maximize} && y_1 + y_2 \\ &\text{subject to} && 0.2py_1 + y_2 + s_p = 1 + 0.01p^2,\ 0 \leq p \leq 10. \end{aligned}$$

The points $\boldsymbol{y} = (0,0)^\top$, $\boldsymbol{x} = (1,\dots,1)^\top/11$ and $\boldsymbol{s} = \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{y}$ are strictly feasible starting points. The trajectory in the $\boldsymbol{y}$-plane of the algorithm (red path) and the constraint equations (blue lines) are given in the following diagram, where the parameter $\sigma = 0.5$ was used. It is instructive to play around with the parameter σ and to try to

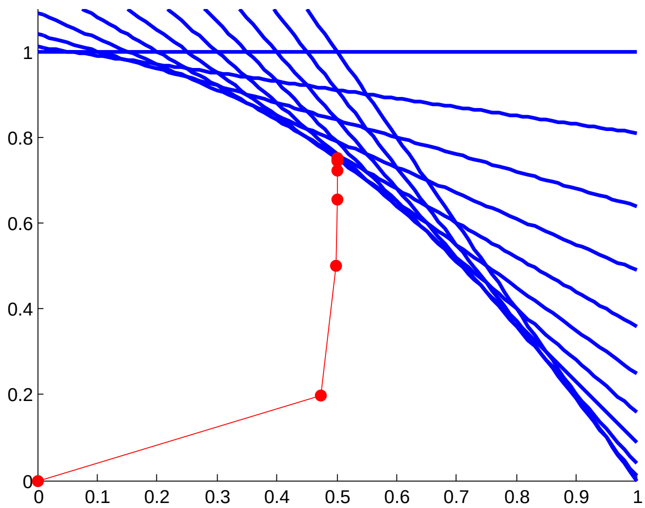

Figure 3.1: Trajectory of long-step path-following in the $y_1 - y_2$ plane.

determine the form of the central path in this example.

Another way to visualise the trajectory is to plot the pairs $x_i s_i$ and $x_j y_j$ against each other. Figure **??** shows the trajectory of the above example in the $x_2 s_2 - x_5 s_5$ plane. Note that the central path, plotted in blue, is trivial in these coordinates, as it is defined by the property of the $x_i s_i = \tau$ being equal.

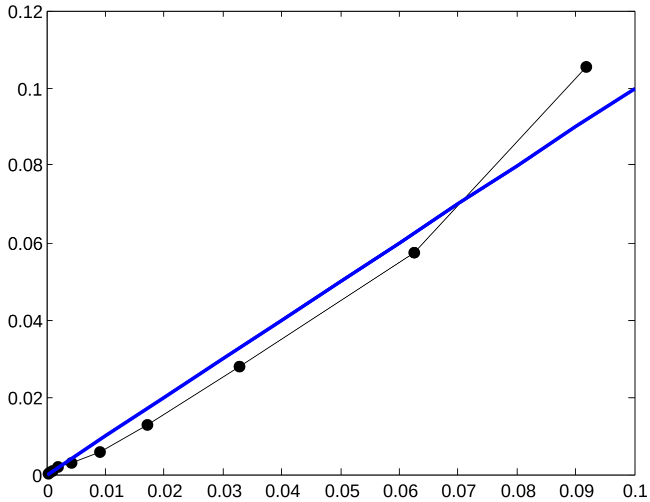

Figure 3.2: Trajectory and central path in $x_2s_2 - x_5s_5$ coordinates.

3.2 Analysis of Path-following

In the analysis of the long-step path-following algorithm, it is enough to establish that the duality measure $\mu^{(k)}$ converges to 0 as $k \to \infty$. The reason is that $\mu = 0$ forces all the products $x_is_i = 0$, and since by design the other constraints are satisfied, this means that the sequence of points converges to a solution. The first theorem tells us that the μ_k decrease as k increases. An elementary proof is given in Theorem 14.3 in Nocedal and Wright. It depends crucially on the assumption that the iterates remain inside the neighbourhood $\mathcal{N}_{-\infty}(\gamma)$ of the central path.

Theorem 3.2. *Given parameters γ, $\sigma_{\min}$ and $\sigma_{\max}$, there is a constant $\delta > 0$, independent of d, such that*

$$\mu_{k+1} \leq \left(1 - \frac{\delta}{d}\right) \mu_k. \tag{3.1}$$

The next theorem gives a bound on the number of iterations needed to reduce the duality measure beyond any given ε.

Theorem 3.3. *Let $\varepsilon > 0$ and $\gamma \in (0,1)$. Let $(\boldsymbol{x}^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{s}^{(0)}) \in \mathcal{N}_{-\infty}(\gamma)$ be a starting point such that the duality measure satisfies $\mu^{(0)} \leq \varepsilon^{-\kappa}$ for some constant κ. Then there is an index $K = O(d \log(1/\varepsilon))$ such that for all $k > K$,*

$$\mu_k \leq \varepsilon.$$

In particular, the long-step path-following algorithm converges.

Proof. Repeatedly applying (**??**), we get

$$\mu_k \leq \left(1 - \frac{\delta}{d}\right)^k \mu_0.$$

Taking logarithms on both sides,

$$\begin{aligned}\log \mu_k \le k \log\left(1 - \frac{\delta}{d}\right) + \log \mu_0 &\le k \log\left(1 - \frac{\delta}{d}\right) + \kappa \log\left(\frac{1}{\varepsilon}\right) \\ &\le k\frac{-\delta}{d} + \kappa \log\left(\frac{1}{\varepsilon}\right).\end{aligned}$$

We have $\mu_k < \varepsilon$ if

$$-k\frac{\delta}{d} + \kappa\left(\frac{1}{\varepsilon}\right) \le \log \varepsilon,$$

of equivalently, if

$$k \ge (1+\kappa)\frac{d}{\delta}\log\left(\frac{1}{\varepsilon}\right) = K.$$

This was to be shown.

□