Primal-Dual Path-Following for Linear Programming

Lecture 16, Continuous Optimisation
Oxford University Computing Laboratory, HT 2006
Notes by Dr Raphael Hauser (hauser@comlab.ox.ac.uk)

We consider LP instances in their standard primal and dual forms

$$\text{(P)} \quad \min_{x\in\mathbb{R}^n} c^{\mathsf{T}}x \quad \text{s.t.} \quad Ax = b, \quad x \geq 0 \qquad\qquad \text{(D)} \quad \max_{y\in\mathbb{R}^m} b^{\mathsf{T}}y \quad \text{s.t.} \quad A^{\mathsf{T}}y + s = c, \quad s \geq 0.$$

The barrier method we studied in Lecture 15 can also solve linear programming problems very efficiently. Crucially, these algorithms can be designed so that they run in *polynomial time*:

An upper bound on the number of bit operations can be given as a polynomial in the bit-length of the input data (A, b, c) of a LP instance

Definition 1: We say that (P) and (D) satisfy the standard LP regularity assumption if the following conditions are met:

i) A has linearly independent row vectors, that is, rank$(A) = m$,

ii) (P) is strictly feasible, that is, there exists a point $x \in \mathbb{R}^n$ such that $Ax = b$ and $x > 0$ componentwise,

iii) (D) is strictly feasible, that is, there exist points $(y, s) \in \mathbb{R}^m \times \mathbb{R}^n$ such that $A^{\mathsf{T}}y + s = c$ and $s > 0$ componentwise.

Note that these regularity assumptions are nothing else but the Slater constraint qualification both for (P) and (D).

The following notation will subsequently be used for the primal, dual and primal-dual feasible domains:

$$\begin{aligned}
&\mathcal{F}_P = \{x : Ax = b, x \geq 0\},\\
&\mathcal{F}_D = \{(y, s) : A^{\mathsf{T}}y + s = c, s \geq 0\},\\
&\mathcal{F}_P^\circ = \{x : Ax = b, x > 0\},\\
&\mathcal{F}_D^\circ = \{(y, s) : A^{\mathsf{T}}y + s = c, s > 0\},\\
&\mathcal{F}^\circ = \mathcal{F}_P^\circ \times \mathcal{F}_D^\circ.
\end{aligned}$$

Perturbations of LP problems: For $\mu > 0$ we consider the following perturbations of (P) and (D):

$$(\mathrm{P})_\mu \qquad \min_{x \in \mathbb{R}^n} c^\mathsf{T} x + \mu f(x) \qquad\qquad (\mathrm{D})_\mu \qquad \max_{y \in \mathbb{R}^m} b^\mathsf{T} y - \mu f(s)$$
$$\text{s.t.} \quad Ax = b \qquad\qquad\qquad\qquad \text{s.t.} \quad A^\mathsf{T} y + s = c,$$
$$x > 0 \qquad\qquad\qquad\qquad\qquad\qquad s > 0.$$

In both problems

$$\begin{aligned} f : \mathbb{R}^n_{++} &\to \mathbb{R} \\ x &\mapsto -\sum_{j=1}^n \log(x_j) \end{aligned}$$

is the logarithmic barrier function.

In Problem Set 6 we studied the duality/optimality theory of problems $(\mathrm{P})_\mu$ and $(\mathrm{D})_\mu$ and found the following result in which precisely the "perturbed" KKT conditions of Lecture 15 appear:

Theorem 1: Let (P),(D) satisfy the standard LP regularity assumption. Then $x(\mu) \in \mathbb{R}^n$ and $\big(y(\mu), s(\mu)\big) \in \mathbb{R}^m \times \mathbb{R}^n$ are optimal for $(\mathrm{P})_\mu$ and $(\mathrm{D})_\mu$ respectively if and only if the following system holds true:

$$\begin{aligned} A^\mathsf{T} y + s &= c \\ Ax &= b \\ XSe &= \mu e \\ x, s &> 0, \end{aligned} \tag{1}$$

where $X = \mathrm{Diag}(x)$, $S = \mathrm{Diag}(s)$ and $e = [\,1\ \dots\ 1\,]^\mathsf{T}$.

The central path: Equations (1) are called the *central path equations*.

Theorem 2: Let (P),(D) satisfy the standard LP regularity assumptions and let $\mu > 0$. Then the central path equations (1) have a unique solution $\big(x(\mu), y(\mu), s(\mu)\big)$.

Proof: See lecture notes. □

Definition 2: For $\mu > 0$ let us write $\big(x(\mu), y(\mu), s(\mu)\big)$ for the unique solution of the central path equations (1). Then the set $\{x(\mu) : \mu > 0\}$ is called the primal central path, $\{(y(\mu), s(\mu) : \mu > 0\}$ is the dual central path, and $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ is the primal-dual central path.

Theorem 3: The map

$$\mu \mapsto \big(x(\mu), y(\mu), s(\mu)\big)$$

is continuously differentiable. Furthermore, there exist x^* and (y^*, s^*) which are optimal solutions to (P) and (D) respectively such that

$$\lim_{\mu \downarrow 0} \big(x(\mu), y(\mu), s(\mu)\big) = (x^*, y^*, s^*).$$

Given an approximate solution (x, y, s) to the central path equations (1), we find a better approximation by applying Newton's method to find a zero of the map

$$\begin{bmatrix} x \\ y \\ s \end{bmatrix} \mapsto \begin{bmatrix} A^{\mathsf{T}} y + s - c \\ Ax - b \\ XSe - \mu e \end{bmatrix},$$

that is, we solve the system

$$\begin{bmatrix} 0 & A^{\mathsf{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} A^{\mathsf{T}} y + s - c \\ Ax - b \\ XSe - \mu e \end{bmatrix} \tag{2}$$

for $(\Delta x, \Delta y, \Delta s)$ and set

$$\begin{bmatrix} x^+ \\ y^+ \\ s^+ \end{bmatrix} := \begin{bmatrix} x \\ y \\ s \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix}.$$

Note that we have neglected the positivity constraints $x, s > 0$ of the central path equations.

We could enforce these by taking a damped Newton step $\alpha(\Delta x, \Delta y, \Delta s)^{\mathsf{T}}$ as in Lecture 15,

However, a nice feature of our algorithm will be that this issue is dealt with automatically through the notion of centrality developed below.

We will now describe and analyse an algorithm that iterates over points that satisfy the constraints

$$\begin{aligned} A^{\mathsf{T}} y + s &= c \\ Ax &= b \\ x, s &> 0 \end{aligned} \tag{3}$$

but not necessarily the equation $XSe = \mu e$.

This requires a starting point (x, y, s) that satisfies (3). This issue can be dealt with via a phase I type auxiliary problem. Thus, we may simply assume that such a point is available.

Centrality and the Duality Gap: In order to be able to assure that the iterates of our algorithm stay well inside the domain $x, s > 0$, we need a measure of centrality, or of "nearness" to the central path.

Definition 3: For all

$$\omega = (x, y, s) \in \mathcal{F}^\circ = \left\{(x, y, s) : A^{\mathsf{T}} y + s = c, Ax = b, x, s > 0\right\}$$

we define

$$\mu(\omega) := \frac{\sum_{j=1}^n x_j s_j}{n}.$$

Recall that LP duality showed that any feasible solution of (P) yields an upper bound on the optimal solution of (D), and any feasible solution of (D) yields a lower bound on the optimal solution of (P).

Definition 4: Let x and (y, s) be primal and dual feasible points. The duality gap associated with these solutions is defined as $c^\top x - b^\top y$. Strong LP duality shows that the duality gap becomes zero at a primal-dual optimal point $\omega^* = (x^*, y^*, s^*)$.

The number $\mu(\omega)$ is useful in monitoring the progress of an algorithm because it is proportional to the duality gap: if $\omega = (x, y, s)$ is primal-dual feasible, then

$$c^\top x - b^\top y = x^\top (c - A^\top y) = x^\top s = n\mu(\omega).$$

It is thus reasonable to fix a number $\sigma \in (0, 1)$ and to set $\mu = \sigma\mu(\omega)$ in the system (2). That is to say, we are aiming to reduce the duality gap by a constant factor in each iteration.

Another interesting observation is that $\omega = (x, y, s) \in \mathcal{F}^\circ$ lies on the primal-dual central path if and only if $XSe = \mu(\omega)e$. This can be used to define a neighbourhood of the central path:

Definition 5: For $\theta \in (0, 1)$, let

$$\mathcal{N}_2(\theta) := \Big\{\omega = (x, y, s) \in \mathcal{F}^\circ : \|XSe - \mu(\omega)e\|_2 \le \theta\mu(\omega)\Big\}.$$

Note that this notion "distance θ" from the central path is homogenised by $\mu(\omega)$: for ω corresponding to a smaller duality gap, the distance must be proportionally smaller for ω to lie in $\mathcal{N}_2(\theta)$. That is to say, the neighbourhood narrows down as the central path approaches the optimal solution $\omega^* = (x^*, y^*, s^*)$ as guaranteed by Theorem 3. This feature is necessary to prevent the algorithm from going off-track.

The Main Motivation of the Algorithm: In each main iteration of our interior-point algorithm we aim at achieving two separate conflicting goals:

i) we want to reduce the duality gap by a constant factor,

ii) we want to stay near the central path, because we know that this will lead us to the optimal solution of the problem pair (P),(D).

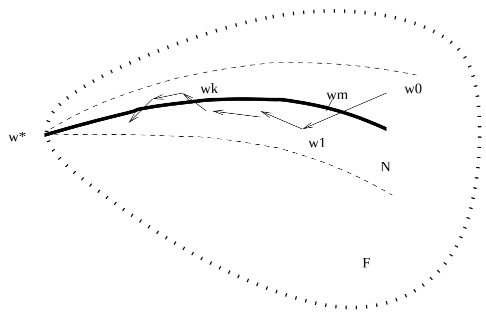
Starting from $\omega = (x, y, s)$, we aim for the point ω_μ that corresponds to the barrier parameter value $\mu = \sigma\mu(\omega)$, in order to reduce the duality gap by a constant factor.

If we start with a point $\omega_k \in \mathcal{N}_2(\theta)$, we want the update $\omega_{k+1} = \omega_k + \Delta\omega$ obtained from the solution of the system

$$\begin{bmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ X_k S_k e - \sigma\mu(\omega_k)e \end{bmatrix} \tag{4}$$

to end up in $\mathcal{N}_2(\theta)$ again, (see figure on next slide) so that we can apply the same analysis in each iteration.

Note that (4) was obtained from (2) by substituting $\omega_k = (x_k, y_k, s_k)$ and using the fact that ω_k is primal-dual feasible.

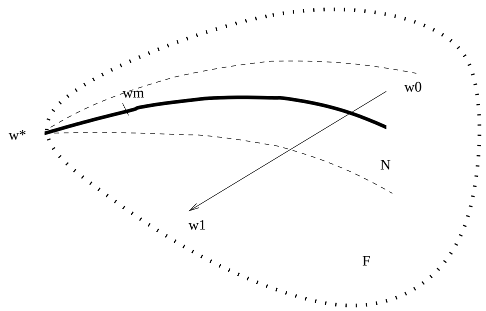


The choice of σ must therefore be sufficently large but still quantifiably low for the algorithm to be well-defined and efficient.

Thus, there must be a *functional dependence* between θ and σ.

We will see that a good choice of parameters is obtained as in the initialisation step S0 of the following primal-dual "short-step" path-following (SPF) algorithm:

Unfortunately, if σ is chosen too small and we aim for too radical a reduction of the duality gap in each iteration, then ω_{k+1} will lie outside of $\mathcal{N}_2(\theta)$.



Algorithm (SPF):

S0 Choose $\theta, \delta \in (0,1)$ be such that

$$\frac{\theta^2 + \delta^2}{2^{3/2}(1-\theta)} \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta.$$

Set $\sigma := 1 - \frac{\delta}{\sqrt{n}}$ and choose $\omega_0 = (x_0, y_0, s_0) \in \mathcal{N}_2(\theta)$.

S1 For $k = 0, 1, \ldots$ repeat

solve (4) with $\omega = \omega_k$ for $\Delta\omega := (\Delta x, \Delta y, \Delta s)$

compute $\omega_{k+1} = \omega_k + \Delta\omega$

end

Theorem 4: The sequence $(\omega_k)_{\mathbb{N}}$ generated by Algorithm SPF satisfies $\omega_k \in \mathcal{N}_2(\theta)$ for all $k \in \mathbb{N}$, and

$$\mu(\omega_k) = \left(1 - \frac{\delta}{\sqrt{n}}\right)^k \mu(\omega_0).$$

An immediate consequence of Theorem 4 is that it takes only logarithmically many iterations to reduce the duality gap below a desired threshold $\epsilon > 0$:

Corollary 1: After at most $k = O\Big(\sqrt{n} \log \frac{n \times \mu(\omega_0)}{\epsilon}\Big)$ iterations Algorithm SPF produces a point $\omega_k = (x_k, y_k, s_k) \in \mathcal{F}^\circ$ such that

$$c^\top x_k - b^\top y_k \leq \epsilon.$$

Theorem 4 readily follows from the following result:

Lemma 1: Let $\omega = (x, y, s) \in \mathcal{N}_2(\theta)$ and let $\omega_+ = (x_+, y_+, s_+) = \omega + \Delta\omega$, where $\Delta\omega = (\Delta x, \Delta y, \Delta s)$ solves the system

$$\begin{bmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ XSe - \sigma\mu(\omega)e \end{bmatrix} \tag{5}$$

with $\sigma = 1 - \frac{\delta}{\sqrt{n}}$ and θ, δ chosen as in the initialisation step of Algorithm SPF. Then $\omega_+ \in \mathcal{N}_2(\theta)$ and $\mu(\omega_+) = \sigma\mu(\omega)$.

Proof: Let $\mu_+ = \Big(1 - \frac{\delta}{\sqrt{n}}\Big)\mu(\omega)$. We claim that the following three relations hold true:

$$\mu_+ = \frac{e^\top X_+ S_+ e}{n}, \tag{6}$$

$$\|X_+ S_+ e - \mu_+ e\| \leq \theta\mu_+, \tag{7}$$

$$x_+, s_+ > 0. \tag{8}$$

Clearly, these relations imply that the lemma holds true. □

We will establish the validity of Claims (6), (7), and (8) separately after introducing the following two technical lemmas:

Lemma 2: $\Delta x^\top \Delta s = 0$.

Proof: This follows readily from the first two blocks of equations in (5). □

Lemma 3: Let $u, v \in \mathbb{R}^n$ be such that $u^\top v \geq 0$ and let $U = \mathrm{Diag}(u)$, $V = \mathrm{Diag}(v)$. Then

$$\|UVe\| \leq \frac{\|u + v\|^2}{2^{3/2}}.$$

Proof: See notes. □

Lemma 4: Claim (6) is true.

Proof:

- First note that (5) implies

$$\begin{aligned} X_+ S_+ e &= (X+\Delta X)(S+\Delta S)e \\ &= XSe + X\Delta s + S\Delta x + \Delta X \Delta S e \\ &= \mu_+ e + \Delta X \Delta S e. \end{aligned} \tag{9}$$

- Using Lemma 2 in conjunction with (9), we obtain

$$e^\mathsf{T} X_+ S_+ e = n\mu_+ + e^\mathsf{T} \Delta X \Delta S e = n\mu_+ + \Delta x^\mathsf{T} \Delta s = n\mu_+. \quad \square$$

Moreover, Lemma 2 shows that $(D^{-1}\Delta x)^\mathsf{T}(D\Delta s) = 0$, which makes it possible to apply Lemma 3 to find

$$\begin{aligned} \|\Delta X \Delta S e\| &= \|(D^{-1}\Delta X)(D\Delta S)e\| \\ &\leq 2^{-3/2}\|D^{-1}\Delta x + D\Delta s\|^2 \\ &\overset{(11)}{\leq} \frac{\|XSe - \mu_+ e\|^2}{2^{3/2} \times \min\{x_j s_j\}}. \end{aligned} \tag{12}$$

- Because of the assumption $\omega \in \mathcal{N}_2(\theta)$, we have

$$\|XSe - \mu(\omega)e\| \leq \theta\mu(\omega) \tag{13}$$

and hence, $x_j s_j \geq (1-\theta)\mu(\omega)$ for all j.

Lemma 5: Claim (7) is true.

Proof:

- (9) shows that

$$\|X_+ S_+ e - \mu_+ e\| = \|\Delta X \Delta S e\|. \tag{10}$$

- To bound the right hand side of this equation, consider the matrix $D = X^{\frac{1}{2}} S^{-\frac{1}{2}}$. The last block of equations of (5) multiplied by $X^{-\frac{1}{2}} S^{-\frac{1}{2}}$ can then be written as

$$D^{-1}\Delta x + D\Delta s = (XS)^{-\frac{1}{2}}(\mu_+ e - XSe). \tag{11}$$

- Substituting this in (12), we find

$$\|\Delta X \Delta S e\| \leq \tag{14}$$

$$\leq \frac{\|XSe - \mu_+ e\|^2}{2^{3/2}(1-\theta)\mu(\omega)}$$

$$= \left(2^{3/2}(1-\theta)\mu(\omega)\right)^{-1} \times \left(\|XSe - \mu(\omega)e\|^2 + \|(\mu(\omega) - \mu_+)e\|^2\right) \tag{15}$$

$$\overset{(13)}{\leq} \frac{\theta^2\mu(\omega)^2 + \delta^2\mu(\omega)^2}{2^{3/2}(1-\theta)\mu(\omega)} = \frac{\theta^2+\delta^2}{2^{3/2}(1-\theta)}\mu(\omega) \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta\mu(\omega)$$

$$= \theta\mu_+,$$

where (15) holds because

$$e^\mathsf{T} XSe - \mu(\omega)e^\mathsf{T} e = \sum_{j=1}^n x_j s_j - n\mu(\omega) = 0$$

shows that $XSe - \mu e \perp e$. $\square$

Lemma 6: Claim (8) is true.

Proof: We only treat the case $\theta \leq 1/2$ which is easier to understand. For the case $\theta \leq 1/2$ see lecture notes.

- (7) shows that $(x_+)_j(s_+)_j \geq (1-\theta)\mu_+ > 0$ for all j. So, if $(x_+)_j < 0$ for some j then $(s_+)_j$ is negative too, and then

$$\Delta x_j \Delta s_j \geq (x_+)_j(s_+)_j > (1-\theta)\mu_+. \tag{16}$$

On the other hand,

$$\Delta x_j \Delta s_j \leq \|\Delta X \Delta S e\| \overset{(7),(10)}{\leq} \theta\mu_+. \tag{17}$$

- The combination of (16) and (17) yields

$$(1-\theta)\mu_+ \leq \Delta x_j \Delta s_j < \theta\mu_+$$

which implies the contradiction $2\theta > 1$ and proves the claim. $\square$

I hope you enjoyed the course,

thanks for listening!

Reading Assignment: Lecture-Note 16.