SECTION C: CONTINUOUS OPTIMISATION
LECTURE 16: PRIMAL-DUAL PATH-FOLLOWING FOR LINEAR PROGRAMMING

HONOUR SCHOOL OF MATHEMATICS, OXFORD UNIVERSITY
HILARY TERM 2005, DR RAPHAEL HAUSER

1. Interior-Point Methods for Linear Programming. The barrier method we studied in Lecture 15 is an example of a so-called *interior-point method*: all intermediate solutions are in the interior of the admissible domain. Although originally devised for nonlinear programming, it was discovered in the mid 1980-ies that such algorithms can also solve linear programming problems very efficiently.

Crucially, these algorithms can be designed so that they run in *polynomial time*, that is, an upper bound on the number of bit operations performed by the algorithm until completion can be given as a polynomial in the bit-length of the input data (A, b, c) of a LP instance

$$\begin{aligned} \text{(P)} \qquad & \min_{x\in\mathbb{R}^n} c^{\mathrm{T}}x \\ & \text{s.t.} \quad Ax = b, \\ & \qquad x \geq 0. \end{aligned}$$

The simplex algorithm on the other hand is known to take a number of bit operations that is exponential in the bit-length of (A, b, c) on certain LP instances. The emergence of polynomial-time interior-point methods for LP has therefore been hailed as a great success, and it has sparked a huge research effort into improving these methods and extending them to other convex problem classes.

Today, interior-point methods are often the best approach to solving very large scale LP problems, and they are the method of choice for other important classes of convex programming. We cannot enter the discussion of methods that underlie industrial-strength software here, but we will use this lecture to outline the design and analysis of the simplest case of such an algorithm.

2. Perturbations of LP problems. Recall the primal and dual linear programming problems in standard form:

$$\begin{aligned} \text{(P)} \qquad & \min_{x\in\mathbb{R}^n} c^{\mathrm{T}}x \\ & \text{s.t.} \quad Ax = b, \\ & \qquad x \geq 0. \end{aligned}$$

$$\begin{aligned} \text{(D)} \qquad & \max_{y\in\mathbb{R}^m} b^{\mathrm{T}}y \\ & \text{s.t.} \quad A^{\mathrm{T}}y + s = c, \\ & \qquad s \geq 0. \end{aligned}$$

For the purposes of the developing a lean theory, we will make the following regularity assumptions:

Definition 2.1. *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, $\mathrm{rank}(A) = m$,
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 $A^{\mathrm{T}}y + s = c$ and $s > 0$ componentwise.

Note that these regularity assumptions are nothing else but the Slater constraint qualification both for (P) and (D). Whenever the standard LP regularity assumption does not hold, we can preprocess the input data A, b, c of the problems (P),(D) and obtain equivalent problems (P'),(D') that satisfy the assumption. Thus, there is no loss of generality in assuming that (P) and (D) are regular.

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

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

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

$$\begin{aligned} (\text{P})_\mu \qquad & \min_{x\in\mathbb{R}^n} c^{\mathrm{T}}x + \mu f(x) \\ & \text{s.t.} \quad Ax = b \\ & \qquad x > 0. \end{aligned}$$

$$\begin{aligned} (\text{D})_\mu \qquad & \max_{y\in\mathbb{R}^m} b^{\mathrm{T}}y - \mu f(s) \\ & \text{s.t.} \quad A^{\mathrm{T}}y + s = c, \\ & \qquad s > 0. \end{aligned}$$

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 $(\text{P})_\mu$ and $(\text{D})_\mu$ and found the following result:

Theorem 2.2. *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 $(P)_\mu$ and $(D)_\mu$ respectively if and only if the following system holds true:*

$$\begin{aligned} A^{\mathrm{T}}y + s &= c \\ Ax &= b \\ XSe &= \mu e \\ x, s &> 0, \end{aligned} \tag{2.1}$$

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

Notice how nicely this approximates the system of equations and inequalities that characterise optimal solutions for (P) and (D): since (P) and (D) satisfy the Slater conditions, Corollary 3.3 of Lecture 12 shows that (x, y, s) are primal-dual optimal if and only if the KKT conditions hold, that is,

$$\begin{aligned} A^{\mathrm{T}}y + s &= c \\ Ax &= b \\ XSe &= 0 \\ x, s &> 0. \end{aligned} \tag{2.2}$$

The only difference between (2.1) and (2.2) is that the complementarity equations $XSe = 0$ have been replaced by the approximate complementarity conditions $XSe = \mu e$.

This is exactly how we motivated the primal-dual barrier method in the nonlinear programming case, but here we see that there is a deeper reason for perturbing the complementarity equations in this fashion: the new equations correspond to the KKT equations of a primal-dual pair of perturbed problems $(\mathrm{P})_\mu$ and $(\mathrm{D})_\mu$.

3. The central path. Equations (2.1) are called the *central path equations*.

THEOREM 3.1. *Let (P),(D) satisfy the standard LP regularity assumptions and let* $\mu > 0$. *Then the central path equations* (2.1) *have a unique solution* $\big(x(\mu), y(\mu), s(\mu)\big)$.

Proof. Theorem 2.2 showed that (x, y, s) solves the central path equations if and only if x is optimal for $(\mathrm{P})_\mu$ and (y, s) is optimal for $(\mathrm{D})_\mu$. Since $(\mathrm{P})_\mu$ and $(\mathrm{D})_\mu$ are strictly convex problems, their optima are unique if they do exist. Thus, uniqueness is automatic once we have shown existence. Moreover, since $(\mathrm{P})_\mu$ and $(\mathrm{D})_\mu$ are of the same form, it suffices to establish this for $(\mathrm{P})_\mu$. We will rely on the well-known fact that any continuous function defined on a compact set has a minimiser in this set (this fact follows from the Bolzano-Weierstrass theorem).

Let $\hat{x}$ and $(\hat{y}, \hat{s})$ be strictly feasible primal and dual solutions, that is, the following system is satisfied:

$$\begin{aligned} A\hat{x} &= b \\ \hat{x} &> 0 \\ A^{\mathrm{T}}\hat{y} + \hat{s} &= c \\ \hat{s} &> 0. \end{aligned}$$

Note that $Ax = b$ implies

$$c^{\mathrm{T}}x = (A^{\mathrm{T}}\hat{y} + \hat{s})^{\mathrm{T}}x = b^{\mathrm{T}}\hat{y} + \hat{s}^{\mathrm{T}}x.$$

Therefore, $(\mathrm{P})_\mu$ is equivalent to

$$\begin{aligned} \min\ & \hat{s}^{\mathrm{T}}x + \mu f(x) \\ \text{s.t.}\ \ & Ax = b, \\ & \hat{s}^{\mathrm{T}}x + \mu f(x) \le \hat{s}^{\mathrm{T}}\hat{x} + \mu f(\hat{x}). \end{aligned}$$

Now note that since $\hat{x}, \hat{s} > 0$, there exist vectors $l, u > 0$ such that

$$\hat{s}^{\mathrm{T}}x + \mu f(x) \le \hat{s}^{\mathrm{T}}\hat{x} + \mu f(\hat{x}) \Rightarrow l \le x \le u.$$

Therefore, $(\mathrm{P})_\mu$ is equivalent to

$$\begin{aligned} (\mathrm{P'})_\mu \qquad \min\ & \hat{s}^{\mathrm{T}}x + \mu f(x) \\ \text{s.t.}\ \ & Ax = b, \\ & l \le x \le u. \end{aligned}$$

Since $\{x : Ax = b, l \le x \le u\}$ is a compact subset of $\mathbb{R}^n_{++}$ and $\hat{s}^{\mathrm{T}}x + \mu f(x)$ is continuous, the minimum of $(\mathrm{P'})_\mu$ is attained. □

DEFINITION 3.2. *For* $\mu > 0$ *let us write* $\big(x(\mu), y(\mu), s(\mu)\big)$ *for the unique solution of the central path equations* (2.1). *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.*

The central paths are smooth curves leading to a primal-dual optimal pair of solutions for the original problems (P),(D):

THEOREM 3.3. *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*

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

4. A primal-dual interior-point algorithm. The critical step of our primal-dual interior-point algorithm will be the following: given an approximate solution (x, y, s) to the central path equations (2.1), find a better approximation.

An obvious approach is to apply Newton's method to find a zero of the map

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

that is, we solve the system

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} A^{\mathrm{T}}y + s - c \\ Ax - b \\ XSe - \mu e \end{bmatrix} \tag{4.1}$$

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 $\alpha(\Delta x, \Delta y, \Delta s)^{\mathrm{T}}$ rather than the full step $(\Delta x, \Delta y, \Delta s)^{\mathrm{T}}$ when this takes $(\bar{x}_+, \bar{y}_+, \bar{s}_+)$ outside of the domain $x, s > 0$, just as we did in our barrier methods for nonlinear programming. 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

$$\begin{aligned} A^{\mathrm{T}}y + s &= c \\ Ax &= b \\ x, s &> 0 \end{aligned} \tag{4.2}$$

but not necessarily the equation $XSe = \mu e$. This requires a starting point (x, y, s) that satisfies (4.2). This issue can be dealt with via a phase I type auxiliary problem. Thus, we may simply assume that such a point is available.

4.1. 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 4.1. *For all* $\omega = (x, y, s) \in \mathcal{F}^\circ = \{(x, y, s) : A^{\mathrm{T}}y + s = c, Ax = b, x, s > 0\}$ *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 4.2. *Let* x *and* (y, s) *be primal and dual feasible points. The duality gap associated with these solutions is defined as* $c^{\mathrm{T}}x - b^{\mathrm{T}}y$.

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

$$c^{\mathrm{T}}x - b^{\mathrm{T}}y = x^{\mathrm{T}}(c - A^{\mathrm{T}}y) = x^{\mathrm{T}}s = n\mu(\omega).$$

It is thus reasonable to fix a number $\sigma \in (0, 1)$ and to set $\mu = \sigma\mu(\omega)$ in the system (4.1). 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 4.3. *For* $\theta \in (0, 1)$, *let*

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

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 3.3. This feature is necessary to prevent the algorithm from going off-track.

4.2. 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).

We decided to 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

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ X_k S_k e - \sigma\mu(\omega_k)e \end{bmatrix} \tag{4.3}$$

to end up in $\mathcal{N}_2(\theta)$ again, see Figure 4.1, so that we can apply the same analysis in each iteration. Note that (4.3) was obtained from (4.1) by substituting $\omega_k = (x_k, y_k, s_k)$ and using the fact that ω_k is primal-dual feasible.

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 $\mathcal{N}_2(\theta)$, see Figure 4.2. 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 in Section 5 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:

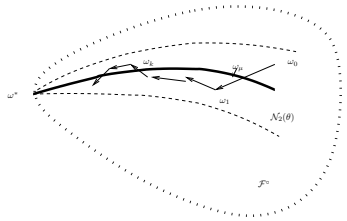

FIG. 4.1. *The SPF algorithm follows the central path without leaving the neighbourhood* $\mathcal{N}_2(\theta)$

4.3. The Algorithm. We are ready to formulate the primal-dual short-step path-following algorithm for LP:

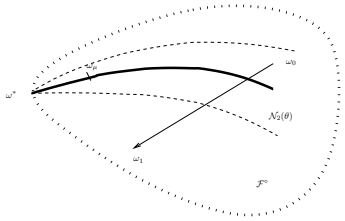

FIG. 4.2. *Choosing σ too small and aiming at a point ω_μ too far away makes the update leave the narrow neighbourhood $\mathcal{N}_2(\theta)$.*

ALGORITHM 4.4 (SPF).
S0 Choose $\theta, \delta \in (0,1)$ be such that

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

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 (4.3) with $\omega = \omega_k$ for $\Delta\omega := (\Delta x, \Delta y, \Delta s)$
compute $\omega_{k+1} = \omega_k + \Delta\omega$
end

5. Convergence Analysis of Algorithm SPF. Let us now analyse the convergence of the primal-dual short-step path-following method SPF. Our main theorem shows that θ and σ were chosen so that the iterates never leave the narrow neighbourhood $\mathcal{N}_2(\theta)$ of the central path, and that the duality gap shrinks at a geometric rate:

THEOREM 5.1. *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 5.1 is that it takes only logarithmically many iterations to reduce the duality gap below a desired threshold $\epsilon > 0$:

COROLLARY 5.2. *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*

$$c^{\mathrm{T}}x_k - b^{\mathrm{T}}y_k \le \epsilon.$$

Theorem 5.1 readily follows from the following result:

LEMMA 5.3. *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*

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ XSe - \sigma\mu(\omega)e \end{bmatrix} \tag{5.1}$$

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:

$$\mu_+ = \frac{e^{\mathrm{T}}X_+S_+e}{n}, \tag{5.2}$$

$$\|X_+S_+e - \mu_+e\| \le \theta\mu_+, \tag{5.3}$$

$$x_+, s_+ > 0. \tag{5.4}$$

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

We will establish the validity of Claims (5.2), (5.3), and (5.4) separately after proving the following two technical lemmas:

LEMMA 5.4. $\Delta x^{\mathrm{T}}\Delta s = 0$.

Proof. This follows readily from the first two blocks of equations in (5.1). □

LEMMA 5.5. *Let $u, v \in \mathbb{R}^n$ be such that $u^{\mathrm{T}}v \ge 0$ and let $U = \mathrm{Diag}(u)$, $V = \mathrm{Diag}(v)$. Then*

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

Proof. First note that for all $\alpha, \beta \in \mathbb{R}$ such that $\alpha\beta \ge 0$ it is the case that

$$|\alpha\beta| \le \frac{(\alpha+\beta)^2}{4}. \tag{5.5}$$

Consider the index sets $I := \{j : u_jv_j \ge 0\}$ and $J := \{j : u_jv_j < 0\}$. Using this notation we can write $u^{\mathrm{T}}v = \sum_{j\in I}|u_jv_j| - \sum_{j\in J}|u_jv_j| \ge 0$. Therefore,

$$\begin{aligned}\|UVe\|^2 &= \sum_{j\in I}(u_jv_j)^2 + \sum_{j\in J}(u_jv_j)^2 = \|uv_I\|_2^2 + \|uv_J\|_2^2 \\ &\le \|uv_I\|_1^2 + \|uv_J\|_1^2 \le 2\|uv_I\|_1^2,\end{aligned}$$

where uv_I is a vector of length $|I|$ with components $\{u_jv_j : j \in I\}$ and similarly for uv_J, and where we used the well-known fact that $\|\cdot\|_2 \le \|\cdot\|_1$. Finally, this proves that

$$\|UVe\| \le \sqrt{2}\|uv_I\|_1 \overset{(5.5)}{\le} \frac{\sqrt{2}}{4}\sum_{j\in I}(u_j+v_j)^2 \le \frac{\|u+v\|^2}{2^{3/2}}.$$

□

LEMMA 5.6. *Claim* (5.2) *is true.*

Proof. First note that (5.1) implies

$$\begin{aligned} X_+S_+e &= (X+\Delta X)(S+\Delta S)e = XSe + X\Delta s + S\Delta x + \Delta X\Delta Se \\ &= \mu_+ e + \Delta X\Delta Se. \end{aligned} \tag{5.6}$$

Using Lemma 5.4 in conjunction with (5.6), we obtain

$$e^{\mathrm{T}}X_+S_+e = n\mu_+ + e^{\mathrm{T}}\Delta X\Delta Se = n\mu_+ + \Delta x^{\mathrm{T}}\Delta s = n\mu_+.$$

□

LEMMA 5.7. *Claim* (5.3) *is true.*

Proof. Equation (5.6) shows that

$$\|X_+S_+e - \mu_+ e\| = \|\Delta X\Delta Se\|. \tag{5.7}$$

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 (5.1) multiplied by $X^{-\frac{1}{2}}S^{-\frac{1}{2}}$ can then be written as

$$D^{-1}\Delta x + D\Delta s = (XS)^{-\frac{1}{2}}(\mu_+ e - XSe). \tag{5.8}$$

Moreover, Lemma 5.4 shows that $(D^{-1}\Delta x)^{\mathrm{T}}(D\Delta s) = 0$, which makes it possible to apply Lemma 5.5 to find

$$\begin{aligned} \|\Delta X\Delta Se\| &= \|(D^{-1}\Delta X)(D\Delta S)e\| \\ &\le 2^{-3/2}\|D^{-1}\Delta x + D\Delta s\|^2 \\ &\overset{(5.8)}{\le} \frac{\|XSe - \mu_+ e\|^2}{2^{3/2}\times\min\{x_j s_j\}}. \end{aligned} \tag{5.9}$$

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

$$\|XSe - \mu(\omega)e\| \le \theta\mu(\omega) \tag{5.10}$$

and hence, $x_j s_j \ge (1-\theta)\mu(\omega)$ for all j. Substituting this in (5.9), we find

$$\begin{aligned} \|\Delta X\Delta Se\| &\le \frac{\|XSe - \mu_+ e\|^2}{2^{3/2}(1-\theta)\mu(\omega)} \\ &= \big(2^{3/2}(1-\theta)\mu(\omega)\big)^{-1} \times \big(\|XSe - \mu(\omega)e\|^2 + \|(\mu(\omega)-\mu_+)e\|^2\big) \qquad (5.11) \\ &\overset{(5.10)}{\le} \frac{\theta^2\mu(\omega)^2 + \delta^2\mu(\omega)^2}{2^{3/2}(1-\theta)\mu(\omega)} = \frac{\theta^2+\delta^2}{2^{3/2}(1-\theta)}\mu(\omega) \le \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta\mu(\omega) \\ &= \theta\mu_+, \end{aligned}$$

where (5.11) holds because

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

shows that $XSe - \mu e \perp e$. □

LEMMA 5.8. *Claim* (5.4) *is true.*

Proof. We begin with the case $\theta \le 1/2$ which is easier to understand. Relation (5.3) shows that $(x_+)_j(s_+)_j \ge (1-\theta)\mu_+ > 0$ for all j. So, if $(x_+)_j < 0$ for some j then $(s_+)_j$ is negative too, and then

$$\Delta x_j \Delta s_j \ge (x_+)_j(s_+)_j > (1-\theta)\mu_+. \tag{5.12}$$

On the other hand,

$$\Delta x_j \Delta s_j \le \|\Delta X\Delta Se\| \overset{(5.3),(5.7)}{\le} \theta\mu_+. \tag{5.13}$$

The combination of (5.12) and (5.13) yields

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

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

If we do not assume $\theta \le 1/2$, we can proceed as follows: let

$$\omega(\alpha) := \big(x(\alpha), y(\alpha), s(\alpha)\big) := (x,y,s) + \alpha(\Delta x, \Delta y, \Delta s).$$

Proceeding as in the proof of (5.2) and (5.3), one can establish that for all $\alpha \in [0,1]$,

$$\mu\big(\omega(\alpha)\big) = \frac{e^{\mathrm{T}}X(\alpha)S(\alpha)e}{n} = \left(1 - \frac{\alpha\delta}{\sqrt{n}}\right)\mu(\omega),$$

and $\|X(\alpha)S(\alpha)e - \mu(\omega(\alpha))e\| \le \theta\mu(\omega(\alpha))$. Thus, $x_j(\alpha)s_j(\alpha) \ge (1-\theta)\mu(\omega(\alpha))$ for all $\alpha \in [0,1]$. This implies $x_j(\alpha), s_j(\alpha) > 0$ for all $\alpha \in [0,1]$ by continuity of $x(\alpha), s(\alpha)$ and by the fact that $(1-\theta)\mu(\omega(\alpha)) > 0$. In particular, $x_+ = x(1) > 0$ and $s_+ = s(1) > 0$, which proves the claim. □