MATH 4501558
FALL 2015

LECTURE 7

PRIMAL-DUAL INTERIOR POINT METHODS FOR LINEAR PROGRAMS

Linear program (primal problem)

$$\text{(LP)} \quad \begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^T x \\ &Ax = b \\ &x \geq 0 \end{aligned}$$

Given
$A \in \mathbb{R}^{m \times n}$
$c \in \mathbb{R}^n$
$b \in \mathbb{R}^m$

Karush-Kuhn-Tucker (KKT) conditions

(1) $A^T \lambda + s = c$

(2) $Ax = b$

(3) $x^T s = 0$ (that is $x_j s_j = 0 \ \forall j$)

(4) $x \geq 0$

(5) $s \geq 0$

THM

Suppose $x_* \in \mathbb{R}^n$ is such that (x_*, λ_*, s_*) satisfies the KKT conditions above for some $\lambda_* \in \mathbb{R}^m$ and $s_* \in \mathbb{R}^n$.

The point x_* is a global minimizer of (LP)

PROOF

For each $x \in \mathbb{R}^n$ s.t. $Ax = b$ and $x \geq 0$,

$$\begin{aligned} c^T x &\overset{\left(\substack{\text{due to} \\ c = A^T\lambda_* + s_*}\right)}{=} (A^T \lambda_* + s_*)^T x \\ &= \lambda_*^T A x + s_*^T x \\ &\overset{\left(\substack{\text{due to} \\ Ax = b}\right)}{=} \lambda_*^T b + s_*^T x \\ &\overset{\left(\substack{\text{due to} \\ s_*,\, x \geq 0}\right)}{\geq} \lambda_*^T b \end{aligned}$$

Specifically, $s_*^T x_* = 0$ which implies

$$c^T x_* = \lambda_*^T b.$$

Consequently, x_* is a global minimizer. $\square$

Dual Problem

In the KKT conditions, think

(λ, s) as variables

x as Lagrange ~~var~~ multiplier

Eq. (2) arising from a Lagrangian

Eq. (1) & (5) as constraints.

$$\text{(DLM1)} \quad \underset{\lambda \in \mathbb{R}^m,\ s \in \mathbb{R}^n}{\text{minimize}} \ -b^T\lambda \quad \left(\text{or } \begin{bmatrix} -b^T & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ s \end{bmatrix}\right)$$

$$A^T\lambda + s = c \quad \left(\text{equivalently } \begin{bmatrix} A^T & I \end{bmatrix}\begin{bmatrix} \lambda \\ s \end{bmatrix} = c\right)$$

$$s \geq 0 \quad \left(\text{or } \begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} \lambda \\ s \end{bmatrix} \geq 0\right)$$

KKT conditions

$$\begin{bmatrix} A \\ I \end{bmatrix} \underline{y} + \begin{bmatrix} 0 \\ I \end{bmatrix} x = \begin{bmatrix} -b \\ 0 \end{bmatrix}$$

$$x \geq 0, \quad x^T s = 0$$

$$A^T\lambda + s = c, \quad s \geq 0$$

$$\Longleftrightarrow$$

$$Ax = b$$

$$x \geq 0, \quad x^T s = 0$$

$$A^T\lambda + s = c, \quad s \geq 0$$

Equivalently,

$$\text{(DL)} \quad \underset{\lambda \in \mathbb{R}^m,\ s \in \mathbb{R}^n}{\text{maximize}} \ b^T\lambda$$

$$A^T\lambda + s = c$$

$$s \geq 0$$

<u>Note</u>

(λ_*, s_*) is a global max. of (DL)

$\Longleftrightarrow$

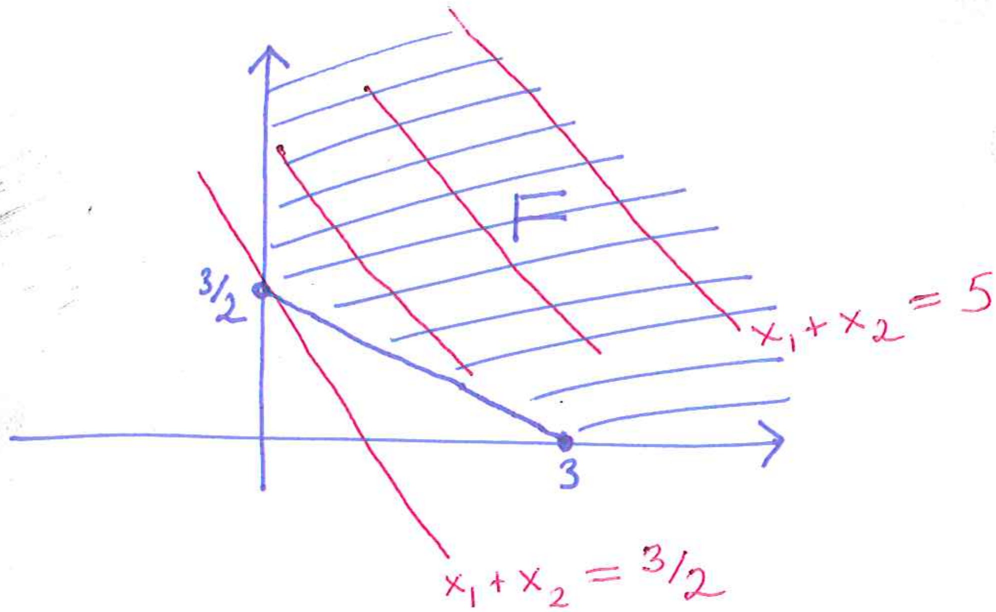
(λ_*, s_*) is a global min. of (DLM1)

Ex

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad x_1 + x_2$$

$$x_1 + 2x_2 \cancel{+} - 3 \geqslant 0$$

$$x_1, x_2 \geqslant 0$$



Equivalently,

$$\underset{x \in \mathbb{R}^3}{\text{minimize}} \quad x_1 + x_2$$

$$x_1 + 2x_2 \cancel{+} - 3 - x_3 = 0$$

$$x \geqslant 0$$

$$\boxed{\begin{aligned} A &= [1 \;\; 2 \;\; -1] \\ b &= +3 \\ c &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \end{aligned}}$$

Dual

$$\underset{\lambda \in \mathbb{R},\, s \in \mathbb{R}^3}{\text{maximize}} \quad +3\lambda$$

$$\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \lambda + s = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$s \geqslant 0$$

Remark

Letting $\lambda = \lambda_+ - \lambda_-$ where $\lambda_+, \lambda_- \geq 0$ are such that

$$[\lambda_+]_j := \max(\lambda_j, 0),$$

$$[\lambda_-]_j := \max(-\lambda_j, 0),$$

(DLM1) can be written as follows:

$$\text{(DLM2)} \qquad \begin{aligned} &\underset{\substack{\lambda_+, \lambda_- \in \mathbb{R}^m \\ s \in \mathbb{R}^n}}{\text{minimize}} \quad \begin{bmatrix} -b & b \end{bmatrix}^T \begin{bmatrix} \lambda_+ \\ \lambda_- \end{bmatrix} \\ &\begin{bmatrix} A^T & -A^T \end{bmatrix} \begin{bmatrix} \lambda_+ \\ \lambda_- \end{bmatrix} + s = c \\ &\lambda_+, \lambda_-, s \geq 0 \end{aligned}$$

(x_*, λ_*, s_*) satisfies KKT conditions (1)-(5).

$$\Longrightarrow$$

$(x_*, [\lambda_*]_+, [\lambda_*]_-, s_*)$ satisfies KKT conditions for (DLM2). (where $[\lambda_*]_+, [\lambda_*]_-$ are s.t. $\lambda_* = [\lambda_*]_+ - [\lambda_*]_-$)

$$\Longrightarrow$$

$(x_*, [\lambda_*]_+, [\lambda_*]_-, s_*)$ is a global minimizer of (DLM2).

$$\Longrightarrow$$

(x_*, λ_*, s_*) is a global minimizer of (DLMI).

$$\Longrightarrow$$

(x_*, λ_*, s_*) is a global maximizer of (DL).

THM

(1) The primal problem has a global minimizer x_* if and only if the dual problem (DL) has a global maximizer (λ_*, s_*).

(2) If the primal problem has a global minimizer (equivalently if the dual problem has a global maximizer),

$$\begin{array}{l} \underset{x \in \mathbb{R}^n}{\text{minimize}} \; c^T x \\ Ax = b \\ x \geq 0 \end{array} \quad = \quad \begin{array}{l} \underset{\lambda \in \mathbb{R}^m,\, s \in \mathbb{R}^n}{\text{maximize}} \; b^T \lambda \\ A^T \lambda + s = c \\ s \geq 0 \end{array}$$

PROOF

(1) (LP) has a global minimizer x_*

$$\Longleftrightarrow$$

KKT conditions holds for some (x_*, λ_*, s_*)

$$\Longleftrightarrow$$

(DL) has a global maximizer (λ_*, s_*).

(2) Letting x_* be the global minimizer of (LP), and (λ_*, s_*) be the corresponding Lagrange multipliers, we have

$$\begin{aligned}
\underset{\substack{x \in \mathbb{R}^n \\ Ax = b \\ x \geq 0}}{\text{minimize}} \; c^T x &= c^T x_* \\
&= (A^T \lambda_* + s_*)^T x_* \\
&= \lambda_*^T (A x_*) + s_*^T x_* \\
&= \lambda_*^T b
\end{aligned}$$

Furthermore, since (x_*, λ_*, s_*) satisfies the KKT conditions

$$\underset{\substack{\lambda \in \mathbb{R}^m, s \in \mathbb{R}^n \\ A^T \lambda + s = c \\ s \geq 0}}{\text{maximize}} \; b^T \lambda = b^T \lambda_* .$$

$\square$

Newton's method (for nonlinear systems)

Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable.

Finds $x_* \in \mathbb{R}^n$ such that

$$F(x_*) = 0.$$

Linear approximation about $x^{(k)} \in \mathbb{R}^n$

$$L^{(k)}(x) = F(x^{(k)}) + \underbrace{F'(x^{(k)})}_{\substack{n \times n \\ \text{Jacobian}}} (x - x^{(k)})$$

<u>Ex</u>

$$F(x) = \begin{bmatrix} x_1^2 + x_2^2 \\ x_1^2 - x_2^2 \end{bmatrix}$$

linear approximation about $(1,1)$

$$L(x) = F(1,1) + F'(1,1)\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 + 2x_2 - 2 \\ 2x_1 - 2x_2 \end{bmatrix}$$

Newton's method generates a sequence $\{x^{(k)}\}$ such that

$$L^{(k)}(x^{(k+1)}) = 0$$

$$F(x^{(k)}) + F'(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

that is

$$x^{(k+1)} = x^{(k)} + p_k$$

where p_k is the solution of

$$F'(x^{(k)})\, p_k = -F(x^{(k)}).$$

Remark

Pure Newton's method for unconstrained optimization is the Newton's method above with $F(x) = \nabla f(x)$.

Primal-Dual Interior Point Methods

Based on application of Newton's method to find (x, λ, s) such that

$$A^T \lambda + s = c$$

$$Ax = b$$

$$XSe = 0 \quad \left(\begin{array}{c}\text{that is} \\ x^T s = 0\end{array}\right)$$

$$x, s \geq 0$$

where

$$X := \mathrm{diag}(x) := \begin{bmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & \ddots & \\ 0 & & & x_n \end{bmatrix}$$

$$S := \mathrm{diag}(s) := \begin{bmatrix} s_1 & & & 0 \\ & s_2 & & \\ & & \ddots & \\ 0 & & & s_n \end{bmatrix}$$

$$e := [1 \ 1 \ \dots \ 1]^T.$$

Newton's method can be employed to solice the first three equations. But a solution by Newton's method would typically not satisfy $x, s \geq 0$.

Instead solve

$$\text{(C1)} \quad A^T\lambda + s = c$$
$$\text{(C2)} \quad Ax = b$$
$$\text{(C3)} \quad [x_j s_j] = \tau$$

$$\left(\begin{array}{l} \text{Equivalently} \\ F(x,\lambda,s) := \\ \begin{bmatrix} A^T\lambda + s - c \\ Ax - b \\ XSe - \tau e \end{bmatrix} = 0 \end{array} \right.$$

by Newton's method, impose

$$x, s \geqslant 0$$

by mean's of a line search.

Pseudocode (Overview)

Given $(x^{(0)}, \lambda^{(0)}, s^{(0)})$ and τ, with $x^{(0)} > 0$, $s^{(0)} > 0$

$r_c \leftarrow A^T \lambda^{(0)} + s^{(0)} - c$

$r_b \leftarrow A x^{(0)} - b$, $k \leftarrow 0$

repeat until convergence

(1) Solve

$$\underbrace{\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^{(k)} & 0 & X^{(k)} \end{bmatrix}}_{F'(x^{(k)}, \lambda^{(k)}, s^{(k)})} \begin{bmatrix} \Delta x^{(k)} \\ \Delta \lambda^{(k)} \\ \Delta s^{(k)} \end{bmatrix} = \underbrace{\begin{bmatrix} -r_c \\ -r_b \\ -X^{(k)}S^{(k)}e + \tau e \end{bmatrix}}_{-F(x^{(k)}, \lambda^{(k)}, s^{(k)})}$$

for $(\Delta x^{(k)}, \Delta\lambda^{(k)}, \Delta s^{(k)})$.

(2) Find α_k such that $x^{(k)} + \alpha_k \Delta x^{(k)} > 0$ and $s^{(k)} + \alpha_k \Delta s^{(k)} > 0$

(3) $(x^{(k+1)}, \lambda^{(k+1)}, s^{(k+1)}) \leftarrow (x^{(k)}, \lambda^{(k)}, s^{(k)}) + \alpha_k (\Delta x^{(k)}, \Delta\lambda^{(k)}, \Delta s^{(k)})$

(4) $r_c \leftarrow A^T \lambda^{(k+1)} + s^{(k+1)} - c$

$r_b \leftarrow A x^{(k+1)} - b$

$k \leftarrow k+1$

end

Decrease τ gradually.

THM

Suppose (LP) has a solution, and $A \in \mathbb{R}^{m \times n}$ is full rank. Following are equivalent:

(1) There exists (x, λ, s) such that

$$Ax = b, \quad A^T \lambda + s = c$$
$$x > 0 \quad \text{and} \quad s > 0;$$

(2) For each $\tau > 0$, there exists a unique $(x(\tau), y(\tau), s(\tau))$ such that

$$Ax(\tau) = b, \quad A^T \lambda(\tau) + s(\tau) = c$$
$$[x(\tau)]_j \, [s(\tau)]_j = \tau \qquad j = 1, \dots, n$$
$$x(\tau) > 0 \quad \text{and} \quad s(\tau) > 0.$$

Proof follows by applying implicit function thm to

$$F : \mathbb{R}^{2n+m} \times \mathbb{R} \longrightarrow \mathbb{R}^{2n+m}$$

$$F(x, \lambda, s, \tau) := \begin{bmatrix} A^T\lambda + s - c \\ Ax - b \\ XSe - \tau e \end{bmatrix}$$

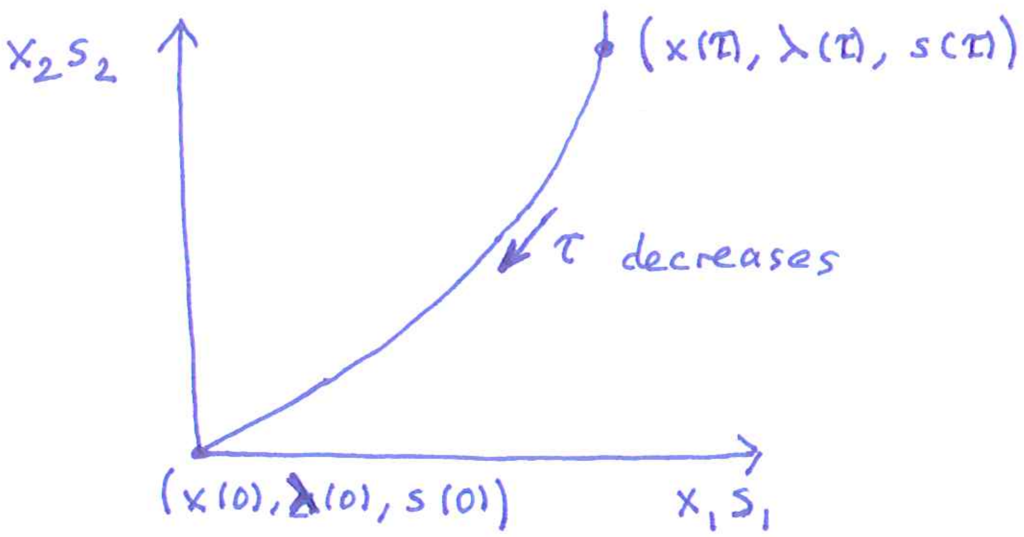
exploiting the invertibility of

$$F'_{(x,\lambda,s)}(x, \lambda, s, \tau) = \begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix}$$

whenever $x, s > 0$.

Central Path

$$\left\{ (x(\tau), \lambda(\tau), s(\tau)) \;\middle|\; \begin{array}{l} Ax(\tau) = b \\ A^T\lambda(\tau) + s(\tau) = c \\ [x(\tau)]_j\,[s(\tau)]_j = \tau \;\; \forall j \\ x(\tau), s(\tau) > 0 \end{array} \right\}$$



Long-step Path-following Algorithm

Given $(x^{(k)}, \lambda^{(k)}, s^{(k)})$ (estimate for a solution for KKT cond.)

① σ_k μ_k will replace τ

σ_k → centrality parameter

μ_k → duality measure

$$\sigma_k \in (0,1)$$

$$\mu_k := \frac{1}{n} \sum_{j=1}^{n} x_j^{(k)} s_j^{(k)}$$

② $(x^{(k+1)}, \lambda^{(k+1)}, s^{(k+1)}) \in N_{-\infty}(\gamma)$

will be enforced in a line search

$$N_{-\infty}(\gamma) := \{ (x,\lambda,s) \in F^{\circ} \mid x_j s_j \geq \gamma \mu \quad j = 1, \ldots, n \}$$

$$\mu \to \frac{1}{n} \sum_{j=1}^{n} x_j s_j$$

$\gamma \in (0,1)$ → fixed parameter

$$F^{\circ} := \{ (x,\lambda,s) \mid Ax = b,\ A^T\lambda + s = c,\ x, s > 0 \}$$

Pseudocode (Parameters $\gamma \in (0,1)$ and $\sigma_{min}, \sigma_{max}$ s.t. $0 < \sigma_{min} < \sigma_{max} < 1$)

Choose $(x^{(0)}, \lambda^{(0)}, s^{(0)})$ such that $x^{(0)}, s^{(0)} > 0$.

$r_b^{(0)} \longleftarrow A x^{(0)} - b$

$r_c^{(0)} \longleftarrow A^T \lambda^{(0)} + s^{(0)} - c$

$k \leftarrow 0$

repeat the following

(0) Choose $\sigma_k \in [\sigma_{min}, \sigma_{max}]$.

(1) Solve

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^{(k)} & 0 & X^{(k)} \end{bmatrix} \begin{bmatrix} \Delta x^{(k)} \\ \Delta \lambda^{(k)} \\ \Delta s^{(k)} \end{bmatrix} = \begin{bmatrix} -r_c^{(k)} \\ -r_b^{(k)} \\ -X^{(k)} S^{(k)} e + \sigma_k \mu^{(k)} e \end{bmatrix}$$

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

(2) Line search: Choose $\alpha_k \in (0, 1]$ such that

$$\left(x^{(k)} + \alpha_k \Delta x^{(k)}, \lambda^{(k)} + \alpha_k \Delta \lambda^{(k)}, s^{(k)} + \alpha_k \Delta s^{(k)}\right) \in N_{-\infty}(\gamma) \cancel{\mu^{(k)}(\alpha)}$$

and as large as possible.

(3) $(x^{(k+1)}, \lambda^{(k+1)}, s^{(k+1)}) \longleftarrow$

$(x^{(k)}, \lambda^{(k)}, s^{(k)}) + \alpha_k (\Delta x^{(k)}, \Delta \lambda^{(k)}, \Delta s^{(k)})$

$k \leftarrow k+1$

Line search - step (2) above

$\alpha_k \leftarrow 1$

While $\left[(x^{(k)}, \lambda^{(k)}, s^{(k)}) + \alpha_k (\Delta x^{(k)}, \Delta\lambda^{(k)}, \Delta s^{(k)})\right] \notin N_{-\infty}(\gamma)$

$\qquad \alpha_k \leftarrow \alpha_k / 2$

end

Convergence

For simplicity suppose

$$A x^{(0)} = b$$

$$A^T \lambda^{(0)} + s^{(0)} = c.$$

This implies

$$A x^{(k)} = b$$

$$A^T \lambda^{(k)} + s^{(k)} = c.$$

~~can be verified~~ by ~~induction~~

for each k. Can be verified by induction, i.e., assume

$$r_c^{(k)} = r_b^{(k)} = 0$$

$$\Longrightarrow$$

$$A^T \Delta\lambda^{(k)} + \Delta s^{(k)} = 0 \quad \text{and} \quad A \Delta x^{(k)} = 0$$

$$\Longrightarrow$$

$$A^T \lambda^{(k+1)} + s^{(k+1)} = 0 \quad \text{and} \quad A x^{(k+1)} = 0$$

LEMMA

Letting

$$\left(x^{(k)}(\alpha), \lambda^{(k)}(\alpha), s^{(k)}(\alpha)\right) :=$$

$$\left(x^{(k)}, \lambda^{(k)}, s^{(k)}\right) + \alpha\left(\Delta x^{(k)}, \Delta\lambda^{(k)}, \Delta s^{(k)}\right),$$

we have

$$\left(x^{(k)}(\alpha), \lambda^{(k)}(\alpha), s^{(k)}(\alpha)\right) \in N_{-\infty}(\gamma)$$

for all $\alpha \in \left[0, 2\sqrt{2}\,\gamma \frac{1-\gamma}{1+\gamma} \frac{\sigma_k}{n}\right]$.

THM (Convergence)

Suppose that the long-step path-following algorithm chooses step-lengths α_k such that

$$\alpha_k \geqslant 2\sqrt{2}\,\gamma \frac{1-\gamma}{1+\gamma} \frac{\sigma_k}{n}.$$

We have

$$\mu^{(k+1)} \leqslant \left(1 - \frac{\delta}{n}\right) \mu^{(k)}$$

where $\delta := 2\sqrt{2}\,\gamma \frac{1-\gamma}{1+\gamma} \min\left\{\sigma_{min}(1-\sigma_{min}), \sigma_{max}(1-\sigma_{max})\right\}$.

PROOF

$$\mu^{(k+1)} = \frac{1}{n} \sum_{j=1}^{n} x_j^{(k+1)} s_j^{(k+1)}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left(x_j^{(k)} + \alpha_k \Delta x_j^{(k)}\right)\left(s_j^{(k)} + \alpha_k \Delta s_j^{(k)}\right)$$

$$= \underbrace{\frac{1}{n} \sum_{j=1}^{n} x_j^{(k)} s_j^{(k)}}_{\mu^{(k)}} +$$

$$\frac{\alpha_k}{n} \sum_{j=1}^{n} \Delta x_j^{(k)} s_j^{(k)} +$$

$$\frac{\alpha_k}{n} \sum_{j=1}^{n} x_j^{(k)} \Delta s_j^{(k)} +$$

$$\frac{\alpha_k^2}{n} \underbrace{\sum_{j=1}^{n} \Delta x_j^{(k)} \Delta s_j^{(k)}}_{0 \text{ (see note below)}}$$

$$= \mu^{(k)} + \frac{\alpha_k}{n} \sum_{j=1}^{n} s_j^{(k)} \Delta x_j^{(k)} + x_j^{(k)} \Delta s_j^{(k)}$$

Furthermore, notice

$$S^{(k)} \Delta x^{(k)} + X^{(k)} \Delta s^{(k)} = -X^{(k)} S^{(k)} e + \sigma_k \mu^{(k)} e$$

$$\Longrightarrow$$

$$s_j^{(k)} \Delta x_j^{(k)} + x_j^{(k)} \Delta s_j^{(k)} = -x_j^{(k)} s_j^{(k)} + \sigma_k \mu^{(k)}.$$

Consequently,

$$\mu^{(k+1)} = \mu^{(k)} + \frac{\alpha_k}{n} \sum_{j=1}^{n} \left(-x_j^{(k)} s_j^{(k)} + \sigma_k \mu^{(k)}\right)$$

$$= \mu^{(k)} + \alpha_k \sigma_k \mu^{(k)} - \alpha_k \mu^{(k)}$$

$$= \mu^{(k)} \left[1 - (1 - \sigma_k) \alpha_k\right].$$

Exploiting $\alpha_k \geq 2\sqrt{2}\,\gamma \,\frac{1-\gamma}{1+\gamma}\,\frac{\sigma_k}{n}$, we deduce

$$\mu^{(k+1)} \leq \mu^{(k)}\left[1 - \underbrace{(1-\sigma_k)\,\sigma_k}_{\geq \min\{\sigma_{min}(1-\sigma_{min}),\,\sigma_{max}(1-\sigma_{max})\}}\,\frac{2\sqrt{2}}{n}\,\gamma\,\frac{1-\gamma}{1+\gamma}\right]$$

$$\leq \mu^{(k)}\left[1 - \frac{\delta}{n}\right].$$

$\square$

Note

$$\left[\Delta x^{(k)}\right]^T \Delta s^{(k)} = 0$$

since

$$A^T \Delta\lambda^{(k)} + \Delta s^{(k)} = 0$$

$$A\,\Delta x^{(k)} = 0$$

$$\Longrightarrow$$

$$0 = \underbrace{\left[\Delta x^{(k)}\right]^T A^T}_{0} \Delta\lambda^{(k)} + \left[\Delta x^{(k)}\right]^T \Delta s^{(k)} = \left[\Delta x^{(k)}\right]^T \Delta s^{(k)}$$