An Iterative Solver-Based Infeasible Primal-Dual Path-Following Algorithm for Convex QP

Zhaosong Lu

Renato D.C. Monteiro

Jerome O'Neal

(Georgia Tech)

In memory of Jos Sturm

University of Waterloo

Workshop on Large Scale Nonlinear and Semidefinite Programming

May 13, 2004

TALK OUTLINE

- **Convex QP Problem**
  - **Assumptions**
  - **Optimality conditions**
- **Primal-dual search directions**
  - **Usual normal equation versus**
  - **Augmented normal equation (ANE)**
- **Maximum Weight Basis preconditioner**
- **Inner-iteration complexity analysis**
- **Inexact P-D search direction**
- **The algorithm**
- **Outer-iteration complexity analysis**
- **Concluding remarks**

CONVEX QP PROBLEM

$$\min \left\{ \frac{1}{2}\|\mathbf{V}^T\mathbf{x}\|^2 + \mathbf{c}^T\mathbf{x} \;\;:\;\; \mathbf{A}\mathbf{x} = \mathbf{b},\;\; \mathbf{x} \geq \mathbf{0} \right\}$$

$$\max \left\{ -\frac{1}{2}\|\mathbf{V}^T\hat{\mathbf{x}}\|^2 + \mathbf{b}^T\mathbf{y} \;\;:\;\; \begin{array}{c} \mathbf{A}^T\mathbf{y} + \mathbf{s} - \mathbf{V}\mathbf{V}^T\hat{\mathbf{x}} = \mathbf{c}, \\ \mathbf{s} \geq \mathbf{0} \end{array} \right\}$$

where the data are $\mathbf{V} \in \Re^{n \times l}$, $\mathbf{A} \in \Re^{m \times n}$, $\mathbf{b} \in \Re^{m}$ and $\mathbf{c} \in \Re^{n}$, and the decision variables are $\mathbf{x} \in \Re^{n}$ and $(\hat{\mathbf{x}}, \mathbf{s}, \mathbf{y}) \in \Re^{n} \times \Re^{n} \times \Re^{m}$.

Remark: Hessian of O.F. is $\mathbf{Q} = \mathbf{V}\mathbf{V}^T$.

Assumptions:

1) **both problems have feasible solutions such that $\mathbf{x} > \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$.**

2) **$\mathbf{A}$ has full row rank.**

Optimality Conditions

$\mathbf{x}$ and $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ are optimal solutions of the primal and dual problems, respectively, iff, for some $\mathbf{z} \in \Re^{\mathbf{l}}$, the quadruple $\mathbf{w} = (\mathbf{x}, \mathbf{s}, \mathbf{y}, \mathbf{z})$ satisfies

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0} \\ \mathbf{A}^{\mathbf{T}}\mathbf{y} + \mathbf{s} + \mathbf{V}\mathbf{z} &= \mathbf{c}, \quad \mathbf{s} \geq \mathbf{0} \\ \mathbf{V}^{\mathbf{T}}\mathbf{x} + \mathbf{z} &= \mathbf{0} \\ \mathbf{X}\mathbf{s} &= \mathbf{0} \end{aligned}$$

where $\mathbf{X} = \mathbf{Diag}(\mathbf{x})$. We let $\mathcal{S}$ denote the set of all $\mathbf{w}$ satisfying the above equations.

P-D Search Directions

Given $\sigma \in [0,1]$, $\Delta w = (\Delta x, \Delta s, \Delta y, \Delta z)$ is determined by

$$\begin{aligned} A\Delta x &= -r_p \\ A^T\Delta y + \Delta s + V\Delta z &= -r_d \\ X\Delta s + S\Delta x &= -Xs + \sigma\mu e \\ V^T\Delta x + \Delta z &= -r_V \end{aligned}$$

where $X = \mathrm{Diag}(x)$, $S = \mathrm{Diag}(s)$, and

$$\mu = \mu(w) := x^T s/n, \quad (1)$$

$$r_p = r_p(w) := Ax - b, \quad (2)$$

$$r_d = r_d(w) := A^T y + s + Vz - c, \quad (3)$$

$$r_V = r_V(w) := V^T x + z, \quad (4)$$

One classical way to compute Δw leads to the usual normal equation

$$A(VV^T + X^{-1}S)^{-1}A^T\Delta y = g,$$

for some vector $g \in \Re^m$.

Augmented Normal Equation (ANE)

The ANE is

$$\tilde{\mathbf{A}}\tilde{\mathbf{D}}^2\tilde{\mathbf{A}}^T \begin{pmatrix} \Delta \mathbf{y} \\ \Delta \mathbf{z} \end{pmatrix} = \mathbf{h}$$

where $\mathbf{D} := \mathbf{X}^{1/2}\mathbf{S}^{-1/2}$ **and**

$$\mathbf{h} := \tilde{\mathbf{A}} \begin{pmatrix} \mathbf{S}^{-1}\mathbf{r}_V - \mathbf{D}^2\mathbf{r}_d \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{r}_p \\ \mathbf{r}_V \end{pmatrix}$$

$$\tilde{\mathbf{A}} := \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{V}^T & \mathbf{I} \end{pmatrix}, \quad \tilde{\mathbf{D}} := \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

Next, we compute $\Delta \mathbf{x}$ **and** $\Delta \mathbf{s}$ **as**

$$\begin{aligned} \Delta \mathbf{s} &= -\mathbf{r}_d - \mathbf{A}^T\Delta \mathbf{y} - \mathbf{V}\Delta \mathbf{z}, \\ \Delta \mathbf{x} &= -\mathbf{D}^2\Delta \mathbf{s} - \mathbf{x} + \sigma\mu\,\mathbf{S}^{-1}\mathbf{e} \end{aligned}$$

Since $\tilde{\mathbf{D}}$ **is diagonal, standard methods for LP can be used to solve the ANE.**

Goal: Exploit the use of iterative (linear) solvers to obtain the solution of the ANE.

Difficulty: For degenerate CQP's, the coefficient matrix of the ANE becomes highly ill-conditioned as the iterates approach the solution set.

Remedy: Precondition the ANE to keep the condition number of $\tilde{A}\tilde{D}^2\tilde{A}^T$ under control.

No theoretically-good preconditioner is known for the usual normal equation. But a theoretically-good preconditioner is known for the ANE, namely the maximum weight basis preconditioner, due to Resende and Veiga 1993 (network flow) and Oliveira and Sorensen 1997 (general LP)

$$\tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T \begin{pmatrix} \tilde{\Delta y} \\ \tilde{\Delta z} \end{pmatrix} = \tilde{T}h$$

M.W.B. Algorithm

Start: Given $\tilde{\mathbf{A}} \in \Re^{\tilde{\mathbf{m}} \times \tilde{\mathbf{n}}}$ and $\tilde{\mathbf{d}} = \mathrm{diag}(\tilde{\mathbf{D}}) \in \Re^{\tilde{\mathbf{n}}}_{++}$,

1. Order the elements of $\tilde{\mathbf{d}}$ so that $\tilde{\mathbf{d}}_{\mathbf{1}} \geq \ldots \geq \tilde{\mathbf{d}}_{\tilde{\mathbf{n}}}$; order the columns of $\tilde{\mathbf{A}}$ accordingly.
2. Let $\mathcal{B} = \emptyset$, $\mathbf{j} = \mathbf{1}$.
3. **While** $|\mathcal{B}| < \tilde{\mathbf{m}}$ **do**
   - If $\tilde{\mathbf{A}}_{\mathbf{j}}$ is linearly indep. of $\{\tilde{\mathbf{A}}_{\mathbf{i}} : \mathbf{i} \in \mathcal{B}\}$, do $\mathcal{B} \leftarrow \mathcal{B} \cup \{\mathbf{j}\}$ and $\mathbf{j} \leftarrow \mathbf{j} + \mathbf{1}$.
4. Return to the original ordering of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{d}}$; determine the set $\mathcal{B}$ according to this ordering and set $\mathcal{N} := \{\mathbf{1}, \ldots, \tilde{\mathbf{n}}\} \backslash \mathcal{B}$.
5. Set $\tilde{\mathbf{B}} := \tilde{\mathbf{A}}_{\mathcal{B}}$ and $\tilde{\mathbf{D}}_{\mathcal{B}} := \mathrm{Diag}(\tilde{\mathbf{d}}_{\mathcal{B}})$.
6. Let $\tilde{\mathbf{T}} := \tilde{\mathbf{D}}_{\mathcal{B}}^{-\mathbf{1}} \tilde{\mathbf{B}}^{-\mathbf{1}}$.

end

Define

$$\varphi_{\tilde{A}} := \max\left\{\|\tilde{B}^{-1}\tilde{A}\|_F : \tilde{B} \text{ is a basis of } \tilde{A}\right\}$$

Theorem (Monteiro and O'Neal): Let $\tilde{T} = \tilde{T}(\tilde{A}, \tilde{d})$ be the preconditioner determined according to the M.W.B. Algorithm, and define $W := \tilde{T}\tilde{A}\tilde{D}^2\tilde{A}^T\tilde{T}^T$. Then, $\kappa(W) \le \varphi_{\tilde{A}}^2$.

We will assume that the iterative solver generates a sequence $\{u^j\}$ satisfying

$$\|v - Wu^j\| \le c(\kappa)\left[1 - \frac{1}{f(\kappa)}\right]^j \|v - Wu^0\|, \quad \forall j$$

where $v := \tilde{T}h$ and c and f are positive non-decreasing functions of $\kappa \equiv \kappa(W) > 0$.

Solver	$c(\kappa)$	$f(\kappa)$
SD	$\sqrt{\kappa}$	$(\kappa + 1)/2$
CG	$2\sqrt{\kappa}$	$(\sqrt{\kappa} + 1)/2$

Analysis of Inner Iterations

Proposition: **For any $\mathbf{u^0}$, the # of iterations to obtain $\mathbf{u^j}$ satisfying $\|\mathbf{v} - \mathbf{W}\mathbf{u^j}\| \le \xi\sqrt{\mu}$ is**

$$\mathcal{O}\left(f(\varphi^2_{\tilde{A}})\,\log\left(\frac{c(\varphi^2_{\tilde{A}})\,\|v - Wu^0\|}{\xi\sqrt{\mu}}\right)\right) \qquad (*)$$

It is possible to choose $\mathbf{u^0} = \mathbf{u^0}(\mathbf{w})$ and ξ so that

$$\begin{aligned} \|v - Wu^0\| &= \mathcal{O}(n\varphi_{\tilde{A}})\,\sqrt{\mu} \\ \xi^{-1} &= \mathcal{O}(\sqrt{n}\,) \end{aligned}$$

This choice of ξ is good to ensure that the number of outer iterations of the iterative solver-based IP method remains the same as its exact counterpart.

With the above choices, $(*)$ reduces to

$$\mathcal{O}\left(\,f(\varphi^2_{\tilde{A}})\;\left[\log c(\varphi^2_{\tilde{A}}) + \log(n\varphi_{\tilde{A}})\right]\,\right)$$

Computation of P-D Direction

Let $\mathbf{u^j}$ be s.t. $\|\mathbf{W u^j} - \mathbf{v}\| \leq \xi\sqrt{\mu}$ and define

$$\begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} := \tilde{\mathbf{T}}^{\mathbf{T}} \mathbf{u^j}$$

Next, we compute $\mathbf{\Delta s}$ and $\mathbf{\Delta x}$, respectively, from the equations:

$$\begin{aligned} \mathbf{A^T \Delta y + \Delta s + V \Delta z} &= -\mathbf{r_d} \\ \mathbf{X \Delta s + S \Delta x} &= -\mathbf{Xs} + \sigma\mu\mathbf{e} - \mathbf{p} \end{aligned}$$

where $\mathbf{p} \in \Re^{\mathbf{n}}$ is as explained below. Have:

$$\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{\mathbf{2}}\tilde{\mathbf{A}}^{\mathbf{T}} \begin{pmatrix} \Delta y \\ \Delta z \end{pmatrix} = \mathbf{h} + \tilde{\mathbf{T}}^{-\mathbf{1}}\tilde{\mathbf{f}}$$

for some $\tilde{\mathbf{f}}$ s.t. $\|\tilde{\mathbf{f}}\| \leq \xi\sqrt{\mu}$.

Using this fact, we easily see that

$$\begin{pmatrix} \mathbf{A\Delta x + r_p} \\ \mathbf{V^T \Delta x + \Delta z + r_V} \end{pmatrix} =$$

$$\mathbf{\tilde{A}} \begin{pmatrix} \mathbf{\Delta x} \\ \mathbf{\Delta z} \end{pmatrix} + \begin{pmatrix} \mathbf{r_p} \\ \mathbf{r_V} \end{pmatrix} = \mathbf{\tilde{T}^{-1}} \left[\mathbf{\tilde{f}} - \mathbf{\tilde{T}\tilde{A}} \begin{pmatrix} \mathbf{S^{-1}p} \\ \mathbf{0} \end{pmatrix} \right]$$

A $\mathbf{p} \in \Re^{\mathbf{n}}$ which makes the above r.h.s. $= \mathbf{0}$ may not exist. Instead, we introduce an extra variable $\mathbf{q} \in \Re^{\mathbf{l}}$ and consider

$$\mathbf{0} = \mathbf{\tilde{f}} - \mathbf{\tilde{T}\tilde{A}} \begin{pmatrix} \mathbf{S^{-1}p} \\ \mathbf{q} \end{pmatrix} = \mathbf{\tilde{f}} - \mathbf{\tilde{T}\tilde{A}\tilde{D}} \begin{pmatrix} \mathbf{(XS)^{-1/2}p} \\ \mathbf{q} \end{pmatrix}.$$

The above system has multiple solutions $(\mathbf{p}, \mathbf{q})$. Any such solution $(\mathbf{p}, \mathbf{q})$ satisfies

$$\begin{pmatrix} \mathbf{A\Delta x + r_p} \\ \mathbf{V^T \Delta x + \Delta z + r_V} \end{pmatrix} = \tilde{A} \begin{pmatrix} \mathbf{0} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{q} \end{pmatrix}$$

Proposition: There exists $(\mathbf{p}, \mathbf{q})$ such that

$$\begin{aligned} \|\mathbf{p}\| &\leq \|\mathbf{XS}\|^{1/2}\|\tilde{\mathbf{f}}\| \\ \|\mathbf{q}\| &\leq \|\tilde{\mathbf{f}}\| \end{aligned}$$

and the corresponding search direction $\Delta\mathbf{w} = (\Delta\mathbf{x}, \Delta\mathbf{s}, \Delta\mathbf{y}, \Delta\mathbf{z})$ satisfies

$$\begin{aligned} \mathbf{A}\Delta\mathbf{x} &= -\mathbf{r_p} \\ \mathbf{A^T}\Delta\mathbf{y} + \Delta\mathbf{s} + \mathbf{V}\Delta\mathbf{z} &= -\mathbf{r_d} \\ \mathbf{X}\Delta\mathbf{s} + \mathbf{S}\Delta\mathbf{x} &= -\mathbf{Xs} + \sigma\mu\mathbf{e} - \mathbf{p}, \\ \mathbf{V^T}\Delta\mathbf{x} + \Delta\mathbf{z} &= -\mathbf{r_V} + \mathbf{q}, \end{aligned}$$

Recall that $\tilde{\mathbf{f}}$ is the residual error for the preconditioned system. We will require it to satisfy $\|\tilde{\mathbf{f}}\| \leq \xi\sqrt{\mu} = \mathcal{O}(\sqrt{\mu})$. This clearly implies that $\mathbf{p}$ and $\mathbf{q}$ are $\mathcal{O}(\mu)$ and $\mathcal{O}(\sqrt{\mu})$, respectively.

The Neighborhood

Given an initial iterate $\mathbf{w^0} \in \Re^{\mathbf{2n}}_{++} \times \Re^{\mathbf{m+l}}$ and scalars $\theta > \mathbf{0}$ and $\eta, \gamma \in (\mathbf{0}, \mathbf{1})$, let $\mathcal{N}_{\mathbf{w^0}}(\eta, \gamma, \theta)$ denote the set of points $\mathbf{w} \in \Re^{\mathbf{2n}}_{++} \times \Re^{\mathbf{m+l}}$ satisfying

$$\mathbf{Xs} \geq (\mathbf{1} - \gamma)\mu\mathbf{e}, \qquad \eta \leq \mu/\mu_{\mathbf{0}},$$

$$(\mathbf{r_p}, \mathbf{r_d}) = \eta(\mathbf{r_p^0}, \mathbf{r_d^0}), \quad \|\mathbf{r_z} - \eta\mathbf{r_z^0}\| \leq \theta\sqrt{\mu}$$

All iterates of our algorithm lie in the following neighborhood:

$$\mathcal{N}_{\mathbf{w^0}}(\gamma, \theta) \;=\; \bigcup_{\eta \in [\mathbf{0}, \mathbf{1}]} \mathcal{N}_{\mathbf{w^0}}(\eta, \gamma, \theta).$$

The Algorithm

1. Let $\epsilon > \mathbf{0}$, $\gamma \in (\mathbf{0}, \mathbf{1})$, $\theta > \mathbf{0}$, $\mathbf{w^0} \in \Re^{\mathbf{2n}}_{++} \times \Re^{\mathbf{m+l}}$, and $\mathbf{0} < \underline{\sigma} < \overline{\sigma} < \mathbf{4/5}$ be given. Set $\mathbf{k = 0}$.
2. If $\mu_{\mathbf{k}} := \mu(\mathbf{w^k}) \leq \epsilon$, stop;
3. Let $\mathbf{w} := \mathbf{w^k}$ and $\mu := \mu_{\mathbf{k}}$; choose $\sigma \in [\underline{\sigma}, \overline{\sigma}]$.
4. Build the precond. $\mathbf{\tilde{T}}$ using the M.W.B. Algorithm, and compute $\mathbf{W}$, $\mathbf{v}$, and $\mathbf{u^0}$.
5. Using $\mathbf{u^0}$ as start point for the iterative solver, find an approx. sol. $\mathbf{u}$ of $\mathbf{Wu = v}$ such that $\|\mathbf{Wu - v}\| \leq \xi\sqrt{\mu}$, where

$$\xi := \min\left\{\frac{\gamma\sigma}{\mathbf{4}\sqrt{\mathbf{n}}}, \left[\sqrt{\mathbf{1} + \left(\mathbf{1} - \frac{\gamma}{\mathbf{2}}\right)\sigma} - \mathbf{1}\right]\theta\right\}$$

6. Set $\binom{\mathbf{\Delta y}}{\mathbf{\Delta z}} = \mathbf{\tilde{T}^T u}$ and compute $(\mathbf{p}, \mathbf{q})$ and then $(\mathbf{\Delta s}, \mathbf{\Delta x})$ as explained above.
7. Compute $\tilde{\alpha} := \operatorname{argmax}\{\alpha \in [\mathbf{0}, \mathbf{1}] :$ $\mathbf{w} + \alpha'\mathbf{\Delta w} \in \mathcal{N}_{\mathbf{w^0}}(\gamma, \theta),\ \forall \alpha' \in [\mathbf{0}, \alpha]\}$.
8. Compute $\bar{\alpha} := \operatorname{argmin}\{(\mathbf{x} + \alpha\mathbf{\Delta x})^{\mathbf{T}}(\mathbf{s} + \alpha\mathbf{\Delta s}) :$ $\alpha \in [\mathbf{0}, \tilde{\alpha}]\}$.
9. Let $\mathbf{w^{k+1}} = \mathbf{w} + \bar{\alpha}\mathbf{\Delta w}$, set $\mathbf{k} \leftarrow \mathbf{k+1}$, and go to step 2.

Outer-iteration Analysis

Theorem: **Assume that γ, $\underline{\sigma}$, $\overline{\sigma}$ and θ are s.t.**

$$\max\left\{\gamma^{-1}, (1-\gamma)^{-1}, \underline{\sigma}^{-1}, \left(1-\tfrac{5}{4}\overline{\sigma}\right)^{-1}\right\} = \mathcal{O}(1)$$
$$\theta = \mathcal{O}(\sqrt{n})$$

and that the initial point $w^0 \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfies $(x^0, s^0) \geq (x^*, s^*)$ for some $w^* \in \mathcal{S}$. Then, an iterate $w^k \in \Re^{2n}_{++} \times \Re^{m+l}$ satisfying

$$\mu_k \leq \epsilon^2 \mu_0$$
$$\|(r_p^k, r_d^k)\| \leq \epsilon^2 \|(r_p^0, r_d^0)\|$$
$$\|r_V^k\| \leq \epsilon^2 \|r_V^0\| + \epsilon\theta\sqrt{\mu_0}$$

is generated within $\mathcal{O}\left(n^2 \log(1/\epsilon)\right)$ iterations.

Concluding Remarks

The dual residual is usually defined as

$$\mathbf{R_d} := \mathbf{A^T y} + \mathbf{s} - \mathbf{V V^T x} - \mathbf{c}$$

In terms of the residuals defined earlier, we have:

$$\mathbf{R_d} = \mathbf{r_d} - \mathbf{V}\, \mathbf{r_V}$$

Along the sequence of iterates of our algorithm, we have $\mathbf{r_d} = \mathcal{O}(\mu)$ **and** $\mathbf{r_V} = \mathcal{O}(\sqrt{\mu})$**, and hence**

$$\mathbf{R_d} = \mathcal{O}(\sqrt{\mu})$$

Conclusion: The primal and dual residuals converge to $\mathbf{0}$ at different rates, namely $\mathcal{O}(\mu)$ and $\mathcal{O}(\sqrt{\mu})$, respectively.