CS 2601 Linear and Convex Optimization

12. Newton's method for equality constrained problems

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

- Equality constrained convex QP
- Newton's method

Equality constrained convex QP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{QP}$$

where $\boldsymbol{Q} \in \mathbb{R}^n$, $\boldsymbol{Q} \succeq \boldsymbol{O}$, $\boldsymbol{A} \in \mathbb{R}^{k\times n}$, $\operatorname{rank}\boldsymbol{A} = k$.

- The Lagrange/KKT conditions $\nabla\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}$ gives the KKT system of the problem,

$$\begin{cases} \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}\boldsymbol{x}^* + \boldsymbol{g} + \boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} = \boldsymbol{0} \end{cases} \quad \text{or} \quad \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} -\boldsymbol{g} \\ \boldsymbol{b} \end{bmatrix}$$

  The coefficient matrix $\boldsymbol{K} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix}$ is called the KKT matrix.

- Any solution to the KKT system gives an optimal solution $\boldsymbol{x}^*$ with corresponding Lagrange multiplier $\boldsymbol{\lambda}^*$.

Question. When is the KKT system solvable?

Null space of KKT matrix

通过证明相互属于

Lemma.

$$\text{Null}(\boldsymbol{K}) = \left\{ \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{bmatrix} : \boldsymbol{d} \in \text{Null}(\boldsymbol{A}) \cap \text{Null}(\boldsymbol{Q}) \right\}$$

Proof. Denote the RHS by S. It is trivial that $S \subset \text{Null}(\boldsymbol{K})$. To show $\text{Null}(\boldsymbol{K}) \subset S$. Let $\begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} \in \text{Null}(\boldsymbol{K})$, i.e. $\boldsymbol{Q}\boldsymbol{d} + \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0}$, $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$. Then

$$\boldsymbol{d}^T\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{d}^T(-\boldsymbol{A}^T\boldsymbol{\lambda}) = -(\boldsymbol{A}\boldsymbol{d})^T\boldsymbol{\lambda} = -\boldsymbol{0}^T\boldsymbol{\lambda} = 0$$

Since $\boldsymbol{Q} \succeq \boldsymbol{O}$, $\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{0}$ [1], so $\boldsymbol{d} \in \text{Null}(\boldsymbol{A}) \cap \text{Null}(\boldsymbol{Q})$ and $\boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0}$. Since $\boldsymbol{A}^T$ has full column rank, $\boldsymbol{\lambda} = \boldsymbol{0}$. Thus $\text{Null}(\boldsymbol{K}) \subset S$.

---

[1] Proof. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be an orthonormal eigenbasis of $\boldsymbol{Q}$ and $\boldsymbol{Q}\boldsymbol{v}_i = \alpha_i\boldsymbol{v}_i$. Note $\alpha_i \geq 0$, since $\boldsymbol{Q} \succeq \boldsymbol{O}$. Expand $\boldsymbol{d}$ in the eigenbasis, $\boldsymbol{d} = \sum_{i=1}^{n} \beta_i\boldsymbol{v}_i$. Then $\boldsymbol{Q}\boldsymbol{d} = \sum_{i=1}^{n} \beta_i\alpha_i\boldsymbol{v}_i$.

$$0 = \boldsymbol{d}^T\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{d}^T\left(\sum_{i=1}^{n} \alpha_i\beta_i\boldsymbol{v}_i\right) = \sum_{i=1}^{n} \alpha_i\beta_i\boldsymbol{d}^T\boldsymbol{v}_i = \sum_{i=1}^{n} \alpha_i\beta_i^2 \implies \alpha_i\beta_i^2 = 0, \forall i$$

so $\alpha_i\beta_i = 0$ for all i, and $\boldsymbol{Q}\boldsymbol{d} = \sum_{i=1}^{n} \alpha_i\beta_i\boldsymbol{v}_i = \boldsymbol{0}$.

Solution of KKT system

Let $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$ and $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{d}$. The problem QP is equivalent to

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & f(\boldsymbol{x}_0 + \boldsymbol{d}) = f(\boldsymbol{x}_0) + \boldsymbol{g}^T\boldsymbol{d} + \boldsymbol{x}_0^T\boldsymbol{Q}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T\boldsymbol{Q}\boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{d} = \boldsymbol{0} \end{aligned}$$

If $\boldsymbol{d}_0 \in \mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q})$, then $\boldsymbol{d}_0$ is a feasible direction along which f reduces to an affine function

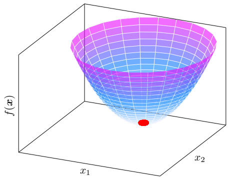
$$f(\boldsymbol{x}_0 + t\boldsymbol{d}_0) = f(\boldsymbol{x}_0) + t\boldsymbol{g}^T\boldsymbol{d}_0$$

1. $\mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q}) = \{\boldsymbol{0}\} \implies \boldsymbol{K}$ nonsingular & unique solution.

2. $\mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q}) \neq \{\boldsymbol{0}\} \implies \boldsymbol{K}$ singular.
   2.1 $\boldsymbol{g} \perp \mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q}) \implies$ infinitely many solutions.
   2.2 $\boldsymbol{g} \not\perp \mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q}) \implies$ no solution and $f^* = -\infty$.

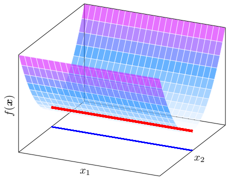
Note. When $\boldsymbol{A} = \boldsymbol{O}$, this reduces to the unconstrained case (cf. slide 10 of §5 part 2).

Unconstrained vs constrained problems



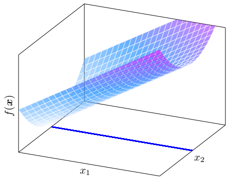
$\boldsymbol{Q} \succ \boldsymbol{O}$

unique solution



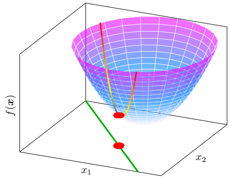
$\boldsymbol{g} \perp \mathrm{Null}(\boldsymbol{Q}) \neq \{\mathbf{0}\}$

infinitely many solutions



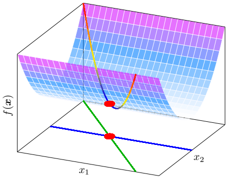
$\boldsymbol{g} \not\perp \mathrm{Null}(\boldsymbol{Q}) \neq \{\mathbf{0}\}$

no solution



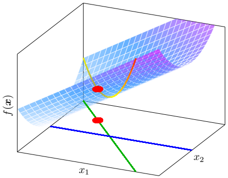
$\boldsymbol{Q} \succ \boldsymbol{O}$

unique solution



$\mathrm{Null}(\boldsymbol{Q}) \cap \mathrm{Null}(\boldsymbol{A}) = \{\mathbf{0}\}$

unique solution

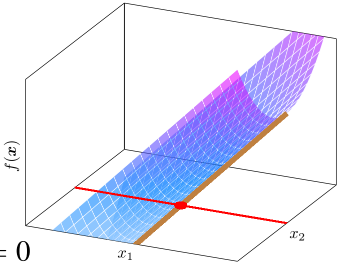


$\mathrm{Null}(\boldsymbol{Q}) \cap \mathrm{Null}(\boldsymbol{A}) = \{\mathbf{0}\}$

unique solution

Unsolvable KKT system (case 2.2)

Example.



$$\min_{x_1,x_2} \quad f(x_1,x_2) = \frac{1}{2}x_2^2 + x_1$$
$$\text{s.t.} \quad x_2 = 0$$

This is a convex QP with

$\boldsymbol{Q} =$ diag$\{0,1\}$, $\boldsymbol{g} = (1,0)^T$, $\boldsymbol{A} = (0,1)$, $b = 0$

The KKT system is

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

which has no solution, since $0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot \lambda \neq -1$.

Note $f^* = -\infty$, and $\boldsymbol{e}_1 = (1,0)^T \in \mathrm{Null}(\boldsymbol{A}) \cap \mathrm{Null}(\boldsymbol{Q})$ and $\boldsymbol{g}^T\boldsymbol{e}_1 \neq 0$, i.e. there is a feasible direction along which the linear term dominates.

Unsolvable KKT system (cont'd)

If the KKT system has no solution, then the problem (QP) is either infeasible (impossible when $\operatorname{rank} \boldsymbol{A} = k$) or unbounded below.

- KKT system has no solution iff
$$\begin{bmatrix} -\boldsymbol{g} \\ \boldsymbol{b} \end{bmatrix} \notin \operatorname{Range}(\boldsymbol{K}) = \operatorname{Range}(\boldsymbol{K}^T) = \operatorname{Null}(\boldsymbol{K})^\perp$$
- There exists $\begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\lambda} \end{bmatrix} \in \operatorname{Null}(\boldsymbol{K})$ s.t. $\begin{bmatrix} -\boldsymbol{g} \\ \boldsymbol{b} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\lambda} \end{bmatrix} \neq 0$, i.e. $\boldsymbol{g}^T\boldsymbol{v} \neq \boldsymbol{b}^T\boldsymbol{\lambda}$.
- By the previous lemma, $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$, $\boldsymbol{Q}\boldsymbol{v} = \boldsymbol{0}$, $\boldsymbol{\lambda} = \boldsymbol{0}$. So $\boldsymbol{g}^T\boldsymbol{v} \neq 0$.
- If $\boldsymbol{x}_0$ is feasible, then $\boldsymbol{x}_0 + t\boldsymbol{v}$ is feasible for any $t \in \mathbb{R}$, since $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$. Since $\boldsymbol{Q}\boldsymbol{v} = \boldsymbol{0}$,
$$f(\boldsymbol{x}_0 + t\boldsymbol{v}) = f(\boldsymbol{x}_0) + t(\boldsymbol{Q}\boldsymbol{x}_0 + \boldsymbol{g})^T\boldsymbol{v} + \frac{1}{2}t^2\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} = f(\boldsymbol{x}_0) + t(\boldsymbol{g}^T\boldsymbol{v})$$
which goes to $-\infty$, as $t \to -\operatorname{sgn}(\boldsymbol{g}^T\boldsymbol{v}) \cdot \infty$.

Note. Similar to slide 6, $\boldsymbol{v} \in \operatorname{Null}(\boldsymbol{A}) \cap \operatorname{Null}(\boldsymbol{Q})$ and $\boldsymbol{g}^T\boldsymbol{v} \neq 0$.

Nonsingularity of KKT matrix (case 1)

If the KKT matrix $\boldsymbol{K}$ is nonsingular, then the KKT system has a unique solution, which is optimal.

Recall $\boldsymbol{Q} \succeq \boldsymbol{O}$ and $\operatorname{rank} \boldsymbol{A} = k$. The following conditions are equivalent

1. $\boldsymbol{K}$ is nonsingular
2. $\mathrm{Null}(\boldsymbol{Q}) \cap \mathrm{Null}(\boldsymbol{A}) = \{\boldsymbol{0}\}$, i.e. $\boldsymbol{Q}$ and $\boldsymbol{A}$ have no nontrivial common nullspace, i.e. $\boldsymbol{Ax} = \boldsymbol{0}$, $\boldsymbol{Qx} = \boldsymbol{0}$ only have the trivial solution $\boldsymbol{x} = \boldsymbol{0}$.
3. $\boldsymbol{Ax} = \boldsymbol{0}, \boldsymbol{x} \neq \boldsymbol{0} \implies \boldsymbol{x}^T \boldsymbol{Qx} > 0$, i.e. $\boldsymbol{Q}$ is positive definite on the nullspace of $\boldsymbol{A}$.
4. $\boldsymbol{F}^T \boldsymbol{QF} \succ \boldsymbol{O}$ for any $\boldsymbol{F} \in \mathbb{R}^{n \times (n-k)}$ s.t. $\mathrm{Range}(\boldsymbol{F}) = \mathrm{Null}(\boldsymbol{A})$, i.e. the columns of $\boldsymbol{F}$ are linearly independent solutions of $\boldsymbol{Ax} = \boldsymbol{0}$.

In particular, if $\boldsymbol{Q} \succ \boldsymbol{O}$, then $\boldsymbol{K}$ is nonsingular (by 3).

Proof

We show $1 \iff 2 \iff 3 \iff 4$.

- $(1 \Leftrightarrow 2)$. By the lemma on slide 4,

$$\boldsymbol{K} \text{ nonsingular } \iff \operatorname{Null}(\boldsymbol{K}) = \{\mathbf{0}\} \iff \operatorname{Null}(\boldsymbol{A}) \cap \operatorname{Null}(\boldsymbol{Q}) = \{\mathbf{0}\}$$

- $(2 \Leftrightarrow 3.)$ Since $\boldsymbol{Q} \succeq \boldsymbol{O}$, $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = 0$ iff $\boldsymbol{Q}\boldsymbol{x} = \mathbf{0}$ (footnote on slide 4).

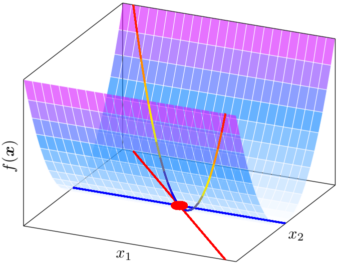
$$\begin{aligned} 2 &\iff \boldsymbol{A}\boldsymbol{x} = \mathbf{0} \text{ and } \boldsymbol{Q}\boldsymbol{x} = \mathbf{0} \text{ implies } \boldsymbol{x} = \mathbf{0} \\ &\iff \boldsymbol{A}\boldsymbol{x} = \mathbf{0} \text{ and } \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = 0 \text{ implies } \boldsymbol{x} = \mathbf{0} \iff 3 \end{aligned}$$

- $(3 \Leftrightarrow 4.)$ Note $\boldsymbol{x} \in \operatorname{Null}(\boldsymbol{A})$ iff $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{z}$, and $\boldsymbol{F}\boldsymbol{z} \neq \mathbf{0}$ iff $\boldsymbol{z} \neq \mathbf{0}$,

$$\begin{aligned} 3 &\iff \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} > 0 \text{ if } \boldsymbol{A}\boldsymbol{x} = \mathbf{0}, \boldsymbol{x} \neq \mathbf{0} \\ &\iff \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} > 0 \text{ if } \boldsymbol{x} = \boldsymbol{F}\boldsymbol{z}, \boldsymbol{z} \neq \mathbf{0} \\ &\iff \boldsymbol{z}^T\boldsymbol{F}^T\boldsymbol{Q}\boldsymbol{F}\boldsymbol{z} > 0 \text{ if } \boldsymbol{z} \neq 0 \\ &\iff \boldsymbol{F}^T\boldsymbol{Q}\boldsymbol{F} \succ \boldsymbol{O} \end{aligned}$$

Example

$$\min_{x_1,x_2} \quad f(x_1,x_2) = \frac{1}{2}x_2^2$$
$$\text{s.t.} \quad x_1 + 2x_2 = b$$



Trivial with solution $x_1^* = b, x_2^* = 0$.

But let's check the condition on slide 8.

$$\boldsymbol{Q} = \text{diag}\{0,1\}, \quad \boldsymbol{A} = (1,2)$$

Let $\boldsymbol{F} = (2,-1)^T$. Then $\mathrm{Range}(\boldsymbol{F}) = \mathrm{Null}(\boldsymbol{A})$, and

$$\boldsymbol{F}^T\boldsymbol{Q}\boldsymbol{F} = [1] \succ \boldsymbol{O}$$

By 4 of slide 8, the KKT matrix in nonsingular, so $\exists$ a unique solution.

Note. The unconstrained problem $\min_{\boldsymbol{x}} f(\boldsymbol{x})$ has infinitely many solutions. But this does not prevent the constrained problem from having a unique solution, as $\boldsymbol{Q} \succ \boldsymbol{O}$ on $\mathrm{Null}(\boldsymbol{A})$ (see 3 on slide 8).

Outline

- Equality constrained convex QP
- Newton's method

Newton direction for equality constrained problem

Consider the second-order Taylor approximation for f at a feasible $\boldsymbol{x}_k$,

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & h(\boldsymbol{d}) \triangleq \hat{f}(\boldsymbol{x}_k + \boldsymbol{d}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{b} \end{aligned}$$

Using $\boldsymbol{A}\boldsymbol{x}_k = \boldsymbol{b}$,

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & h(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{d} = \boldsymbol{0} \end{aligned}$$

KKT system for this quadratic problem is

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}_k) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}_k) \\ \boldsymbol{0} \end{bmatrix}$$

The Newton direction $\boldsymbol{d}_k$ is given by the solution to the KKT system.
We will assume the KKT matrix is always nonsingular (cf. slides 4 & 8).

Newton's method for equality constrained problem

1: initialization $\boldsymbol{x} \leftarrow \boldsymbol{x}_0 \in X$ ▷ $\boldsymbol{x}_0$ is feasible, i.e. $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$
2: **repeat**
3: Compute Newton's direction $\boldsymbol{d}$ by solving

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ \boldsymbol{0} \end{bmatrix}$$

4: $t \leftarrow 1$ ▷ backtracking line search on lines 4-7
5: **while** $f(\boldsymbol{x} + t\boldsymbol{d}) > f(\boldsymbol{x}) + \alpha t \nabla f(\boldsymbol{x})^T \boldsymbol{d}$ **do**
6: $t \leftarrow \beta t$
7: **end while**
8: $\boldsymbol{x} \leftarrow \boldsymbol{x} + t\boldsymbol{d}$
9: **until** $\|\boldsymbol{d}\| \leq \delta$
10: **return** $\boldsymbol{x}$

Note. We cannot use $\|\nabla f(\boldsymbol{x})\| \leq \delta$ as stopping criterion now, as $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ no longer holds in general. [BV] uses $\sqrt{\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d}} \leq \delta$.

Note. This is called a feasible descent method, since all $\boldsymbol{x}_k$ are feasible and $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$ unless $\boldsymbol{x}_k$ is optimal.

Newton's method and constraint elimination

Let $\boldsymbol{F} \in \mathbb{R}^{n\times(n-k)}$ be a matrix whose columns are linearly independent solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Fix a feasible $\tilde{\boldsymbol{x}} \in \boldsymbol{X}$. Every $\boldsymbol{x} \in X$ has a unique representation of the form $\boldsymbol{x} = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}$,

$$\boldsymbol{X} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} = \{\tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z} : \boldsymbol{z} \in \mathbb{R}^{n-k}\}$$

Constrained problem reduces to unconstrained problem by $\boldsymbol{x} = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}$,

$$(C) : \begin{cases} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{cases} \qquad \Longleftrightarrow \qquad (U) : \min_{\boldsymbol{z}} \; g(\boldsymbol{z}) = f(\tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z})$$

Applying Newton's method to (C) with initial point $\boldsymbol{x}_0 = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}_0$ is equivalent to applying Newton's method to (U) with initial point $\boldsymbol{z}_0$:

If $\{\boldsymbol{x}_k\}$ and $\{\boldsymbol{z}_k\}$ are the iterates for (C) and (U), respectively, then

$$\boldsymbol{x}_k = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}_k, \quad \forall k$$

Newton's method has same convergence properties for (C) and (U).

Proof

We only need to show $\boldsymbol{x}_1 = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}_1$ and then use induction.

Let $\Delta\boldsymbol{x}_0$ and $\Delta\boldsymbol{z}_0$ denote the Newton directions for (C) and (U), i.e. $\Delta\boldsymbol{x}_0$ satisfies

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}_0) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x}_0 \\ \boldsymbol{\lambda}_0 \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}_0) \\ \boldsymbol{0} \end{bmatrix}$$

and $\Delta\boldsymbol{z}_0$ satisfies

$$\nabla^2 g(\boldsymbol{z}_0)\Delta\boldsymbol{z}_0 = -\nabla g(\boldsymbol{z}_0)$$

1. Both $\Delta\boldsymbol{x}_0$ and $\Delta\boldsymbol{z}_0$ are well-defined
2. $\Delta\boldsymbol{x}_0 = \boldsymbol{F}\Delta\boldsymbol{z}_0$. This also shows $\Delta\boldsymbol{x}_0 = \boldsymbol{0}$ iff $\Delta\boldsymbol{z}_0 = \boldsymbol{0}$.
3. Backtracking line search gives the same step size t_0 in both cases
4. By 2 and 3,

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + t_0\Delta\boldsymbol{x}_0 = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}_0 + t_0\boldsymbol{F}\Delta\boldsymbol{z}_0 = \tilde{\boldsymbol{x}} + \boldsymbol{F}(\boldsymbol{z}_0 + t_0\Delta\boldsymbol{z}_0) = \tilde{\boldsymbol{x}} + \boldsymbol{F}\boldsymbol{z}_1$$

Proof (cont'd)

More details.

1. Both $\Delta \boldsymbol{x}_0$ and Δz_0 are well-defined. By the chain rule,

$$\nabla^2 g(z_0) = \boldsymbol{F}^T \nabla^2 f(\boldsymbol{x}_0) \boldsymbol{F}, \quad \nabla g(z_0) = \boldsymbol{F}^T \nabla f(\boldsymbol{x}_0)$$

   Since we assume the KKT matrix of (C) is nonsingular, by 4 on slide 8, $\nabla^2 g(z_0) \succ \boldsymbol{O}$. Hence both $\Delta \boldsymbol{x}_0$ and Δz_0 are well-defined.

2. $\Delta \boldsymbol{x}_0 = \boldsymbol{F} \Delta z_0$.
   - Pre-multiplying the first KKT equation by $\boldsymbol{F}^T$,

$$\boldsymbol{F}^T \nabla^2 f(\boldsymbol{x}_0) \Delta \boldsymbol{x}_0 + (\boldsymbol{A}\boldsymbol{F})^T \boldsymbol{\lambda}_0 = -\boldsymbol{F}^T \nabla f(\boldsymbol{x}_0)$$

   - Since the columns of $\boldsymbol{F}$ are solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, $\boldsymbol{A}\boldsymbol{F} = \boldsymbol{O}$. Thus

$$\boldsymbol{F}^T \nabla^2 f(\boldsymbol{x}_0) \Delta \boldsymbol{x}_0 = -\boldsymbol{F}^T \nabla f(\boldsymbol{x}_0)$$

   - $\boldsymbol{A} \Delta \boldsymbol{x}_0 = \boldsymbol{0}$ by the second KKT equation, so $\Delta \boldsymbol{x}_0 = \boldsymbol{F}\boldsymbol{u}$ for some $\boldsymbol{u}$.

$$\boldsymbol{F}^T \nabla^2 f(\boldsymbol{x}_0) \boldsymbol{F}\boldsymbol{u} = -\boldsymbol{F}^T \nabla f(\boldsymbol{x}_0) \iff \nabla^2 g(z_0) \boldsymbol{u} = -\nabla g(z_0)$$

     Thus $\boldsymbol{u} = \Delta z_0$ and $\Delta \boldsymbol{x}_0 = \boldsymbol{F}\boldsymbol{u} = \boldsymbol{F} \Delta z_0$

Proof (cont'd)

3. Backtracking line search gives the same step size t_0.

Note

$$f(\boldsymbol{x}_0) = f(\tilde{\boldsymbol{x}} + \boldsymbol{F}z_0) = g(z_0)$$

$$f(\boldsymbol{x}_0 + t\Delta\boldsymbol{x}_0) = f(\tilde{\boldsymbol{x}} + \boldsymbol{F}(z_0 + t\Delta z_0)) = g(z_0 + t\Delta z_0)$$

$$\nabla f(\boldsymbol{x}_0)^T \Delta\boldsymbol{x}_0 = \nabla f(\boldsymbol{x}_0)^T \boldsymbol{F}\Delta z_0 = \nabla g(z_0)^T \Delta z_0$$

Thus the test condition in backtracking line search for f

$$f(\boldsymbol{x}_0 + t\Delta\boldsymbol{x}_0) > f(\boldsymbol{x}_0) + \alpha t \nabla f(\boldsymbol{x}_0)^T \Delta\boldsymbol{x}_0$$

is exactly the same as that for g,

$$g(z_0 + t\Delta z_0) > g(z_0) + \alpha t \nabla g(\boldsymbol{x}_0)^T \Delta z_0$$