# Chapter 10 Equality constrained minimization

Last update on 2022-05-26 21:39

# Table of contents

# Equality constrained minimization

**equality constrained minimization problem**

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- $f$ convex and twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- assume optimal value $p^*$ is finite and attained

**optimality condition** (review)

$$x^* \text{ is optimal} \quad \Longleftrightarrow \quad \begin{array}{c} x^* \in \mathbf{dom}\, f, \quad Ax^* = b, \\ \text{there exists } \nu^* \text{ such that } \nabla f(x^*) + A^T \nu^* = 0 \end{array}$$

**equality constrained quadratic minimization** (with $P \in \mathbb{S}^n_+$)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

optimality condition

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T Px > 0$$

- equivalent condition for nonsingularity

$$P + A^T A \succ 0$$

# Eliminating equality constraints

represent solutions of $\{x \mid Ax = b\}$ as

$$\{x \mid Ax = b\} = \{Fz + \hat{x} \mid z \in \mathbb{R}^{n-p}\}$$

- $\hat{x}$ is any particular solution
- range of $F \in \mathbb{R}^{n \times (n-p)}$ is nullspace of $A$

**reduced or eliminated problem**

$$\text{minimize} \qquad f(Fz + \hat{x})$$

- unconstrained problem with variable $z \in \mathbb{R}^{n-p}$
- from solution $z^*$, obtain $x^*$ and $\nu^*$ as

$$x^* = Fz^* + \hat{x}, \qquad \nu^* = -\left(AA^T\right)^{-1} A \nabla f(x^*)$$

**example** optimal allocation with resource constraint

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + \cdots + x_n = b \end{array}$$

eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, namely, choose

$$\hat{x} = be_n, \qquad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)}$$

reduced problem

$$\text{minimize} \qquad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$$

# Newton step

Newton step $\Delta x_{\rm nt}$ of $f$ at feasible $x$ is given by the solution $v$ of

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

**interpretations**

- $\Delta x_{\rm nt}$ solves second order approximation (with variable $v$)

$$\begin{array}{ll} \text{minimize} & \widehat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

- $\Delta x_{\rm nt}$ equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

# Newton decrement

$$\lambda(x) = \left(\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}}\right)^{1/2} = \left(-\nabla f(x)^T \Delta x_{\mathrm{nt}}\right)^{1/2}$$

**interpretations**

- gives an estimate of $f(x) - p^*$ using quadratic approximation $\widehat{f}$

$$f(x) - \inf_{Ay=b} \widehat{f}(y) = (1/2)\lambda(x)^2$$

- directional derivative in Newton direction

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} f\left(x + t\Delta x_{\mathrm{nt}}\right)\right|_{t=0} = -\lambda(x)^2$$

- in general $\lambda(x) \neq \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$

# Newton's method with equality constraints

---

**given** starting point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$

**repeat**

1. Compute the Newton step and decrement $\Delta x_{\mathrm{nt}}$, $\lambda(x)$
2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$
3. *Line search.* Choose step size $t$ by backtracking line search
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$

---

- feasible descent method: $x^{(k)}$ feasible and $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$
- affine invariant

# Newton's method and elimination

**Newton's method for reduced problem**

$$\text{minimize } \tilde{f}(z) = f(Fz + \hat{x})$$

- $z \in \mathbb{R}^{n-p}$ are variables, $\hat{x}$ satisfies $A\hat{x} = b$, range of $F$ is the nullspace of $A$
- Newton's method for $\tilde{f}$ starts at $z^{(0)}$, generates iterates $z^{(k)}$

**relation to Newton's method with equality constraints**

when starting at $x^{(0)} = Fz^{(0)} + \hat{x}$, iterates are

$$x^{(k)} = Fz^{(k)} + \hat{x}$$

hence no separate convergence analysis is needed

# Newton step at infeasible points

Newton step $\Delta x_{\text{nt}}$ of $f$ at infeasible $x$ is given by the solution of

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{nt}} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$

**interpretation**

- $\Delta x_{\text{nt}}$ equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

## primal-dual interpretation

- write optimality condition as $r(y) = 0$ where

$$y = (x, \nu), \qquad r(y) = \left(\nabla f(x) + A^T\nu, Ax - b\right)$$

- linearizing $r(y) = 0$ gives

$$r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$$

which is equivalent to

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ \Delta \nu_{\rm nt} \end{bmatrix} = -\begin{bmatrix} \nabla f(x) + A^T\nu \\ Ax - b \end{bmatrix}$$

same as the above equation with $w = \nu + \Delta\nu_{\rm nt}$

# Infeasible start Newton method

---

**given** starting point $x \in \mathbf{dom}\, f$, $\nu$, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$

**repeat**

1. Compute primal and dual Newton steps $\Delta x_{\text{nt}}$, $\Delta \nu_{\text{nt}}$
2. *Backtracking line search on* $\|r\|_2$. $t := 1$.
   **while** $\|r(x + t\Delta x_{\text{nt}}, \nu + t\Delta \nu_{\text{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$, $\quad t := \beta t$
3. *Update*. $x := x + t\Delta x_{\text{nt}}, \nu := \nu + t\Delta \nu_{\text{nt}}$

**until** $Ax = b$ and $\|r(x, \nu)\|_2 \leq \epsilon$

---

- not a descent method: $f\left(x^{(k+1)}\right) > f\left(x^{(k)}\right)$ is possible
- directional derivative of $\|r(y)\|_2$ in direction $\Delta y = (\Delta x_{\mathrm{nt}}, \Delta \nu_{\mathrm{nt}})$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\|r(y + t\Delta y)\|_2\bigg|_{t=0} = -\|r(y)\|_2$$

# Solving KKT systems

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

**solution methods**

- LDL$^\mathsf{T}$ factorization
- elimination with nonsingular $H$

$$AH^{-1}A^Tw = h - AH^{-1}g, \qquad Hv = -(g + A^Tw)$$

- elimination with singular $H$      first write as

$$\begin{bmatrix} H + A^TQA & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g + A^TQh \\ h \end{bmatrix}$$

with $Q \succeq 0$ for which $H + A^TQA \succ 0$, then apply elimination

# Equality constrained analytic centering

**primal problem**

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b \end{array}$$

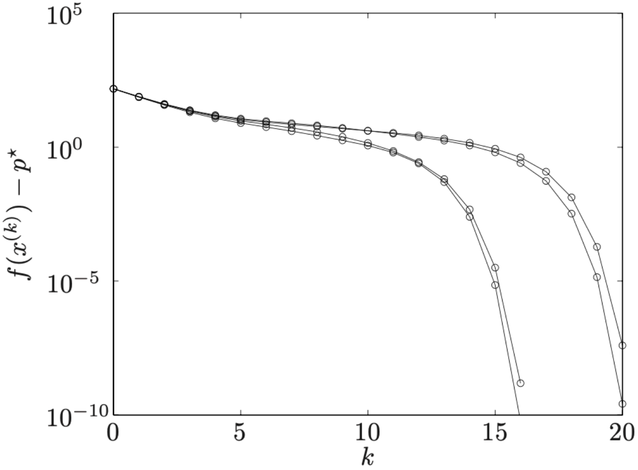
**dual problem**

$$\text{maximize} \quad -b^T\nu + \sum_{i=1}^{n} \log(A^T\nu)_i + n$$

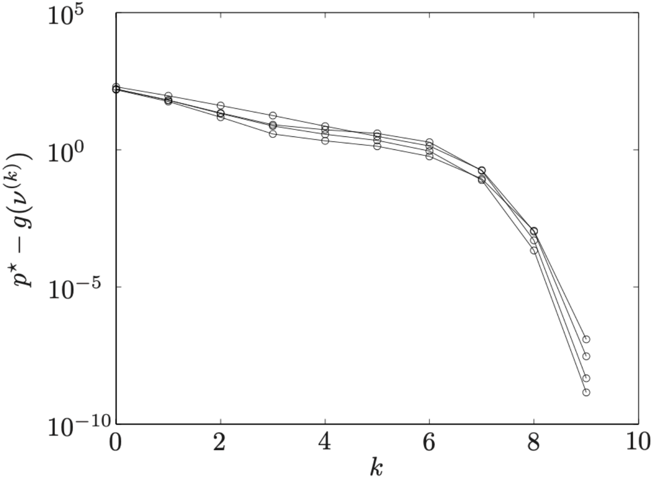
three methods for an example with $A \in \mathbb{R}^{100\times 500}$, different starting points

## Newton method with equality constraints

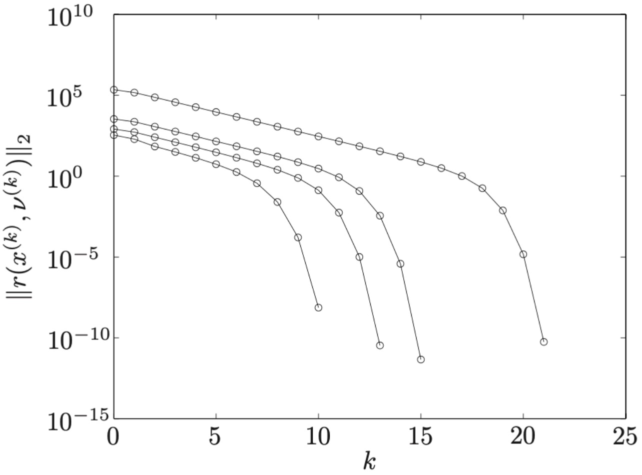
requires $x^{(0)} \succ 0$ and $Ax^{(0)} = b$



**Newton method applied to dual problem** requires $A^T\nu^{(0)} \succ 0$



**infeasible start Newton method** requires $x^{(0)} \succ 0$



## dominant steps of three methods

1. use block elimination to solve KKT system

$$\begin{bmatrix} \mathbf{diag}(x)^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = \begin{bmatrix} \mathbf{diag}(x)^{-1}\mathbf{1} \\ 0 \end{bmatrix}$$

   reduces to solving

$$A\,\mathbf{diag}(x)^2 A^T w = b$$

2. solve Newton system

$$A\,\mathbf{diag}\left(A^T\nu\right)^{-2} A^T \Delta\nu = -b + A\,\mathbf{diag}\left(A^T\nu\right)^{-1}\mathbf{1}$$

3. use block elimination to solve KKT system

$$\begin{bmatrix} \mathbf{diag}(x)^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\nu \end{bmatrix} = \begin{bmatrix} \mathbf{diag}(x)^{-1}\mathbf{1} - A^T\nu \\ b - Ax \end{bmatrix}$$

   reduces to solving

$$A\,\mathbf{diag}(x)^2 A^T w = 2Ax - b$$

**comparison of complexity per iteration**

- in each case, solve
$$ADA^T w = h$$
with $D$ positive diagonal
- complexity per iteration of three methods is identical

# Network flow optimization

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} \phi_i(x_i) \\ \text{subject to} & Ax = b \end{array}$$

- directed (connected) graph with $n$ arcs and $p + 1$ nodes
- $x_i$ is flow through arc $i$
- $\phi_i$ is cost flow function for arc $i$ (with $\phi_i''(x) > 0$)
- $A$ is (reduced) node-arc incidence matrix
- $b \in \mathbb{R}^p$ is (reduced) source vector

## KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

- $H = \mathbf{diag}(\phi_1''(x_1), \ldots, \phi_n''(x_n))$ with positive diagonal
- solve via elimination

$$AH^{-1}A^T w = h - AH^{-1}g, \qquad Hv = -(g + A^T w)$$

sparsity pattern of coefficient matrix is given by graph connectivity

$$\begin{aligned} (AH^{-1}A^T)_{ij} \neq 0 \qquad & \Longleftrightarrow \qquad (AA^T)_{ij} \neq 0 \\ & \Longleftrightarrow \qquad \text{nodes } i \text{ and } j \text{ are connected by an arc} \end{aligned}$$

# Analytic center of linear matrix inequality

$$
\begin{array}{ll}
\text{minimize} & -\log\det X \\
\text{subject to} & \mathbf{tr}(A_i X) = b_i, \qquad i = 1, \ldots, p
\end{array}
$$

where $X \in \mathbb{S}^n$ is the variable, $A_i \in \mathbb{S}^n$, $b_i \in \mathbb{R}$

**optimality conditions**

$$
X^* \succ 0, \qquad -(X^*)^{-1} + \sum_{j=1}^{p} \nu_j^* A_i = 0, \qquad \mathbf{tr}(A_i X^*) = b_i, \qquad i = 1, \ldots, p
$$

## Newton equation at feasible $X$

$$X^{-1}\Delta X X^{-1} + \sum_{j=1}^{p} w_j A_i = X^{-1}, \qquad \mathbf{tr}(A_i \Delta X) = 0, \qquad i = 1, \ldots, p$$

- follows from linear approximation

$$(X + \Delta X)^{-1} \approx X^{-1} - X^{-1}\Delta X X^{-1}$$

- $n(n+1)/2 + p$ variables in $\Delta X$ and $w$

**solution by block elimination**

- compute $\Delta X$ from first equation

$$\Delta X = X - \sum_{j=1}^{p} w_j X A_j X$$

- substitute $\Delta X$ in second equation

$$\sum_{j=1}^{p} \mathbf{tr}(A_i X A_j X) w_j = b_i, \qquad i = 1, \ldots, p$$

a (dense) positive definite set of linear equations with variable $w \in \mathbb{R}^p$

**flop count** (dominant terms) using Cholesky factorization $X = LL^T$

- form $p$ products $L^T A_j L$: $(3/2)pn^3$
- form $p(p+1)/2$ inner products $\mathbf{tr}((L^T A_i L)(L^T A_j L))$: $(1/2)p^2n^2$
- solve for $w_j$ via Cholesky factorization: $(1/3)p^3$