11. Equality constrained minimization

- equality constrained minimization
- eliminating equality constraints
- Newton's method with equality constraints
- infeasible start Newton method
- implementation

Equality constrained minimization

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

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

Optimality conditions: $x^\star$ is optimal if and only if there exists a $\nu^\star$ such that

$$\nabla f(x^\star) + A^T \nu^\star = 0, \qquad Ax^\star = b$$

Equality constrained quadratic minimization

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

we assume $P \in \mathbf{S}_+^n$ (and $\mathbf{rank}(A) = p$)

Optimality condition

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

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

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T P x > 0 \tag{1}$$

  (see next page)

- equivalent condition for nonsingularity:

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

Proof

- $x^T P x = 0$ only if $Px = 0$ (because $P \succeq 0$); therefore (1) can be written

$$Ax = 0, \quad Px = 0 \qquad \Longrightarrow \qquad x = 0 \tag{2}$$

  in other words, $\mathbf{rank}(\begin{bmatrix} P \\ A \end{bmatrix}) = n$

- clearly, this is necessary for nonsingularity of the KKT matrix
- to show it is sufficient, suppose (1) holds, and consider x, y that satisfy

$$Px + A^T y = 0, \qquad Ax = 0$$

- inner product with x shows that $x^T P x = x^T(Px + A^T y) = 0$; hence

$$Px = 0, \qquad Ax = 0, \qquad A^T y = 0$$

- this implies $x = 0$ (from (2)) and $y = 0$ (from $\mathbf{rank}(A) = p$)

Eliminating equality constraints

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

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

- $\hat{x}$ is (any) particular solution
- range of $F \in \mathbf{R}^{n \times (n-p)}$ is nullspace of A ($\mathbf{rank}\, F = n - p$ and $AF = 0$)

Reduced or eliminated problem

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

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

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

Example

Optimal allocation with resource constraint

$$\begin{array}{ll}\text{minimize} & f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \cdots + x_n = b\end{array}$$

Reduced problem

- eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, *i.e.*, choose

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

- reduced problem with variables $x_1, \ldots, x_{n-1}$:

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

Newton step

Newton step Δx_{nt} of f at feasible x is given by 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

- Δx_{nt} solves second order approximation (with variable v)

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

- Δx_{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

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

Properties

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

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

- directional derivative in Newton direction:

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

- in general,

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

Newton's method with equality constraints

given: a starting point $x \in \mathrm{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$
repeat

1. *Newton step:* compute Newton step Δx_{nt} and Newton decrement $\lambda(x)$
2. *stopping criterion:* quit if $\lambda^2/2 \le \epsilon$
3. *line search:* choose step size t by backtracking line search
4. *update:* $x := x + t\Delta x_{\mathrm{nt}}$

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

Newton's method and elimination

Newton's method for reduced problem

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

- variables $z \in \mathbf{R}^{n-p}$
- $\hat{x}$ satisfies $A\hat{x} = b$; $\mathbf{rank}\, F = n - p$ and $AF = 0$
- Newton's method for $\tilde{f}$, started at $z^{(0)}$, generates iterates $z^{(k)}$

Newton's method with equality constraints

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

$$x^{(k+1)} = Fz^{(k)} + \hat{x}$$

hence, don't need separate convergence analysis

Newton step at infeasible points

2nd interpretation of page 11.7 extends to infeasible x (*i.e.*, $Ax \neq b$)

linearizing optimality conditions at infeasible x (with $x \in \mathbf{dom}\, f$) gives

$$\left[\begin{array}{cc} \nabla^2 f(x) & A^T \\ A & 0 \end{array}\right] \left[\begin{array}{c} \Delta x_{\mathrm{nt}} \\ w \end{array}\right] = -\left[\begin{array}{c} \nabla f(x) \\ Ax - b \end{array}\right] \qquad (3)$$

Primal-dual interpretation

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

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

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

$$\left[\begin{array}{cc} \nabla^2 f(x) & A^T \\ A & 0 \end{array}\right] \left[\begin{array}{c} \Delta x_{\mathrm{nt}} \\ \Delta \nu_{\mathrm{nt}} \end{array}\right] = -\left[\begin{array}{c} \nabla f(x) + A^T \nu \\ Ax - b \end{array}\right]$$

  same as (3) with $w = \nu + \Delta \nu_{\mathrm{nt}}$

Infeasible start Newton method

given: a starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$
repeat
  1. *Newton step:* compute primal and dual Newton steps $\Delta x_{\rm nt}$, $\Delta \nu_{\rm nt}$
  2. *backtracking line search on* $\|r\|_2$*:*
    $t := 1$
    while $\|r(x + t\Delta x_{\rm nt}, \nu + t\Delta \nu_{\rm nt})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$
      $t := \beta t$
  3. *update:* $x := x + t\Delta x_{\rm nt}$, $\nu := \nu + t\Delta \nu_{\rm nt}$
until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$

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

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

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$^{\mathrm{T}}$ factorization
- elimination (if H nonsingular)

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

- elimination with singular H: 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 $H + A^TQA \succ 0$, and apply elimination

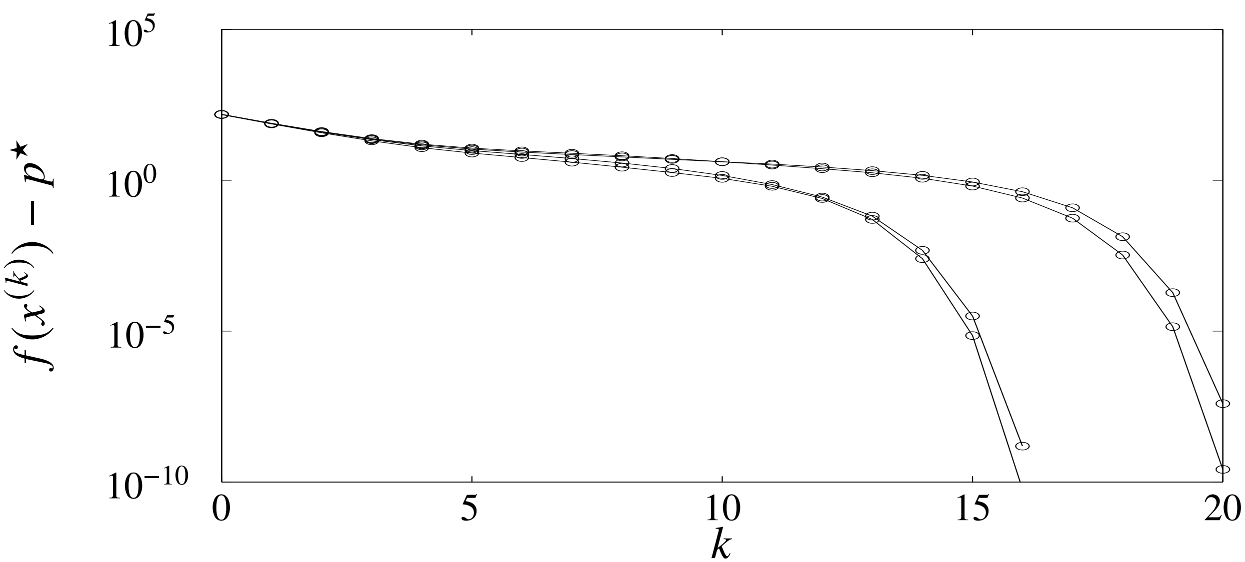
Equality constrained analytic centering

Primal problem: minimize $-\sum_{i=1}^n \log x_i$ subject to $Ax = b$

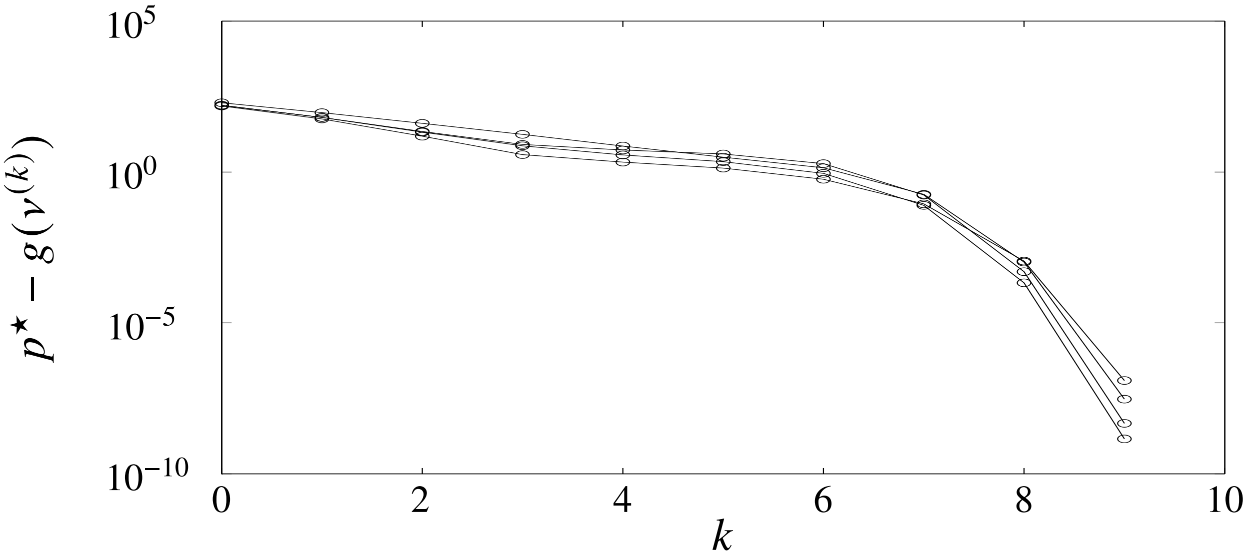
Dual problem: maximize $-b^T\nu + \sum_{i=1}^n \log(A^T\nu)_i + n$

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

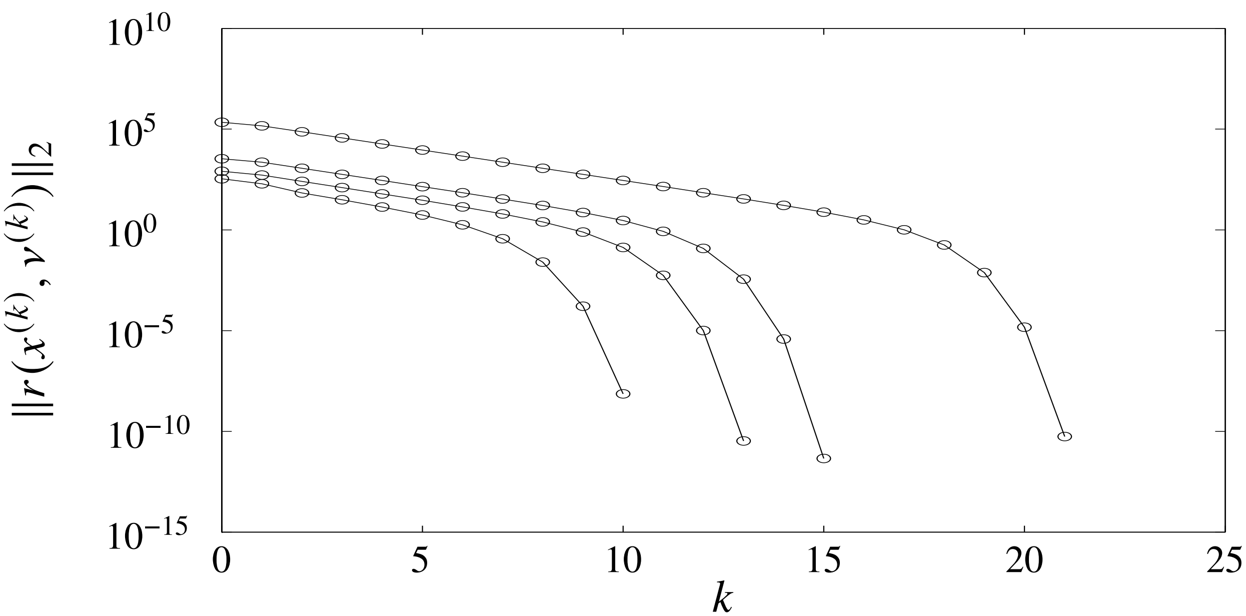
1. Newton method with equality constraints (requires $x^{(0)} \succ 0$, $Ax^{(0)} = b$)



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



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



Complexity per iteration of the three methods is identical

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

3. use block elimination to solve KKT system

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

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

conclusion: in each case, solve $ADA^T w = h$ with D positive diagonal

Network flow optimization

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

- directed graph with n arcs, $p+1$ nodes
- x_i: flow through arc i; ϕ_i: cost flow function for arc i (with $\phi_i''(x) > 0$)
- node-incidence matrix $\tilde{A} \in \mathbf{R}^{(p+1)\times n}$ defined as

$$
\tilde{A}_{ij} = \begin{cases} 1 & \text{arc } j \text{ leaves node } i \\ -1 & \text{arc } j \text{ enters node } i \\ 0 & \text{otherwise} \end{cases}
$$

- reduced node-incidence matrix $A \in \mathbf{R}^{p\times n}$ is $\tilde{A}$ with last row removed
- $b \in \mathbf{R}^p$ is (reduced) source vector
- $\mathbf{rank}\, A = p$ if graph is connected

KKT system

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

- $H = \mathbf{diag}(\phi_1''(x_1), \ldots, \phi_n''(x_n))$, 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

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

Analytic center of linear matrix inequality

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

variable $X \in \mathbf{S}^n$

Optimality conditions

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

Newton equation at feasible X:

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

- follows from linear approximation $(X + \Delta X)^{-1} \approx X^{-1} - X^{-1} \Delta X X^{-1}$
- $n(n+1)/2 + p$ variables ΔX, w

Solution by block elimination

- eliminate ΔX from first equation: $\Delta X = X - \sum_{j=1}^{p} w_j X A_j X$
- substitute ΔX in second equation

$$\sum_{j=1}^{p} \mathrm{tr}(A_i X A_j X) w_j = b_i, \quad i = 1, \ldots, p \tag{4}$$

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

Complexity: 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 $\mathrm{tr}((L^T A_i L)(L^T A_j L))$: $(1/2)p^2 n^2$
- solve (4) via Cholesky factorization: $(1/3)p^3$