10. 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; hence, $\mathbf{dom}\, f$ is an open set
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume the optimal value $p^\star$ is finite and attained

Optimality conditions: x is optimal if and only if

$$x \in \mathbf{dom}\, f, \qquad Ax = b, \qquad \nabla f(x) + A^T \nu = 0 \tag{1}$$

for some $\nu \in \mathbf{R}^p$

Equality constrained quadratic minimization

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

where $P \in \mathbf{S}^n_+$ and $\mathbf{rank}(A) = p$

- optimality conditions from previous page:

$$Px + q + A^T \nu = 0, \qquad Ax = b$$

- this is a set of $n + p$ linear equations in $n + p$ variables

$$\left[\begin{array}{cc} P & A^T \\ A & 0 \end{array}\right] \left[\begin{array}{c} x \\ \nu \end{array}\right] = \left[\begin{array}{c} -q \\ b \end{array}\right]$$

coefficient matrix is called Karush–Kuhn–Tucker (KKT) matrix

Nonsingular KKT matrix

each of the following three conditions is equivalent to nonsingularity of

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$$

(assuming $P \succeq 0$ and $\mathbf{rank}(A) = p$)

1. the first block column has full column rank

$$\mathbf{rank}(\begin{bmatrix} P \\ A \end{bmatrix}) = n \tag{2}$$

2. P is positive definite on the nullspace of A

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T P x > 0 \tag{3}$$

3. the matrix $P + A^T A$ is positive definite

Exercise

show that the three conditions are equivalent to nonsingularity of the KKT matrix

Solution (condition 1)

- clearly, (2) is necessary (first n columns of KKT matrix must be independent)
- to show it is sufficient, we assume that (2) holds and show that

$$Px + A^T y = 0, \qquad Ax = 0 \tag{4}$$

  holds only if $x = 0$, $y = 0$

- inner product of x and first equation of (4) gives $x^T Px = x^T(Px + A^T y) = 0$
- $x^T Px = 0$ if and only if $Px = 0$ (for positive semidefinite P)
- hence (4) is equivalent to $Px = 0$, $Ax = 0$, $A^T y = 0$
- by the rank property (2) and $\mathbf{rank}(A) = p$, this holds only if $x = 0$, $y = 0$

Exercise

Condition 2

- (3) means the same as

$$x^T P x = 0, \quad Ax = 0 \qquad \Longrightarrow \qquad x = 0$$

- for $P \succeq 0$, can replace $x^T P x = 0$ with $Px = 0$, so (3) is the same as (2)

Condition 3

- the matrix $P + A^T A$ is positive definite if

$$x^T (P + A^T A) x = x^T P x + \|Ax\|_2^2 > 0 \quad \text{for all } x \neq 0$$

- for $P \succeq 0$, this is the same as condition 2 ($x^T P x > 0$ for nonzero x with $Ax = 0$)

Eliminating equality constraints

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

- the affine set defined by $Ax = b$ is represented as translate of range of F

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

- $\hat{x}$ is (any) particular solution of the linear equation $Ax = b$
- $F \in \mathbf{R}^{n \times (n-p)}$ is a full-rank matrix with range equal to nullspace of A
- the reformulated problem is unconstrained with variable $z \in \mathbf{R}^{n-p}$
- from solution $z^\star$, solution of optimality conditions (1) is

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

- elimination step can be expensive, obscure structure in A (*e.g.*, sparsity)

Newton step

we extend the definition of Newton step (p. 9.18) to equality-constrained problems

- assume x is feasible ($x \in \mathbf{dom}\, f$ and $Ax = b$)
- define Newton step Δx_{nt} at x as the solution v of

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

- $x + \Delta x_{\mathrm{nt}}$ solves problem if f is replaced with 2nd order approximation $\hat{f}$ at x

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

- $x + \Delta x_{\mathrm{nt}}$ solves optimality conditions (1), linearized at x:

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

Newton decrement

for the equality-constrained problem, we define the Newton decrement as

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

- $\lambda(x)^2$ is directional derivative of f at x in Newton direction $\Delta x_{\rm nt}$:

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

- $\lambda(x)$ is norm of Newton step in quadratic Hessian norm
- $\lambda(x)$ gives estimate of $f(x) - p^\star$, estimated using quadratic approximation $\hat{f}$:

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

- 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 $x \in \operatorname{dom} f$ 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 $\lambda^2/2 \leq \epsilon$
3. *line search:* choose step size t by backtracking line search
4. *update:* $x := x + t\Delta x_{\text{nt}}$

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

Newton's method and elimination

Newton's method for unconstrained optimization (after eliminating $Ax = b$)

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

suppose Newton method for g, started at $z^{(0)}$, generates iterates $z^{(k)}$

Newton's method with equality constraints (method on page 10.10)

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

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

hence, we don't need separate convergence analysis for the method on p. 10.10

Newton step at infeasible points

2nd interpretation of page 10.8 extends to infeasible x (*i.e.*, with $Ax \neq b$)

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

$$\left[\begin{array}{cc} \nabla^2 f(x) & A^T \\ A & 0 \end{array}\right]\left[\begin{array}{c} \Delta x_{\mathrm{nt}} \\ w \end{array}\right] = -\left[\begin{array}{c} \nabla f(x) \\ Ax - b \end{array}\right] \qquad (5)$$

Interpretation

- optimality condition (1) is nonlinear equation $r(x, \nu) = 0$, where

$$r(x, \nu) = \left[\begin{array}{c} \nabla f(x) + A^T \nu \\ Ax - b \end{array}\right]$$

- linearizing $r(y) = 0$ gives

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

  this is the same as (5) with $w = \nu + \Delta \nu_{\mathrm{nt}}$

Infeasible start Newton method

given: a starting point $x \in \operatorname{dom} f$, 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:*

   $t := 1$
   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
- therefore we use norm of r as merit function in line search
- directional derivative of norm of r, at $y = (x, \nu)$, in direction $\Delta y = (\Delta x_{\rm nt}, \Delta \nu_{\rm nt})$ is

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

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

- use matrix factorization for symmetric indefinite matrices (LDL^T factorization)
- if H is positive definite, solve by block elimination: two equations

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

- if H is not positive definite, first write KKT system 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 $Q \succeq 0$ chosen so that $H + A^TQA \succ 0$; then apply block elimination

Equality constrained analytic centering

Primal and dual problems

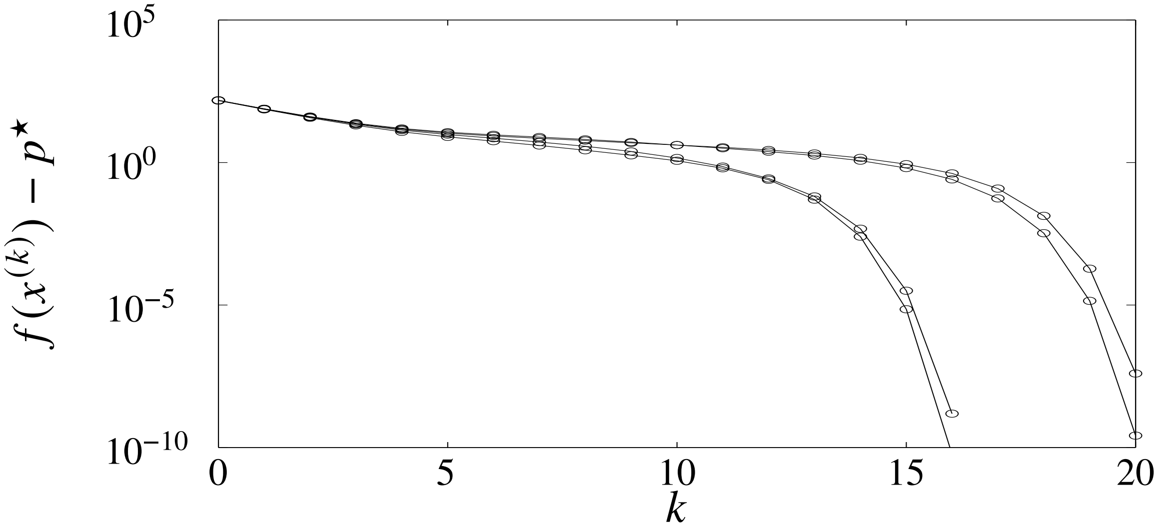
$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b \end{array} \qquad \text{maximize} \quad -b^T \nu + \sum_{i=1}^{n} \log(A^T \nu)_i + n$$

Algorithms

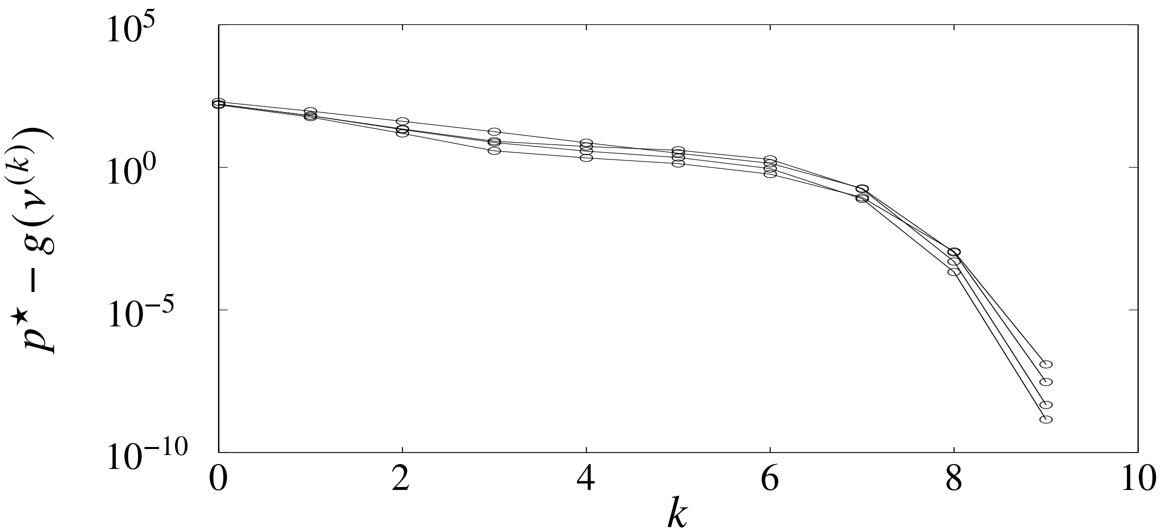
- we compare three versions of Newton's method
- $A \in \mathbf{R}^{100 \times 500}$
- starting points are different for the three methods

Equality constrained analytic centering

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

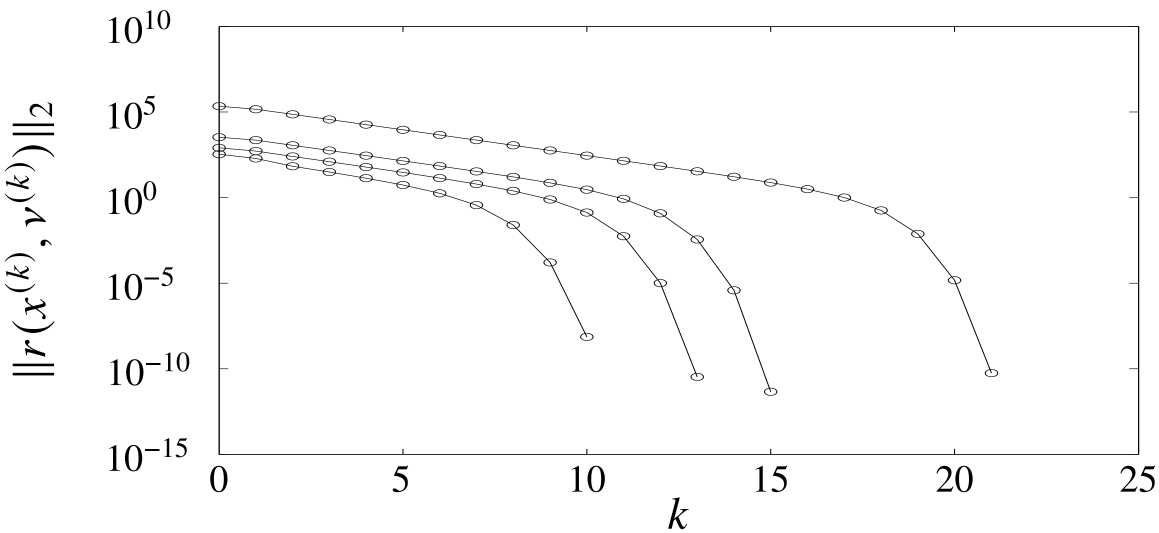


2. (unconstrained) Newton method applied to dual problem (requires $A^T\nu^{(0)} \succ 0$)



Equality constrained analytic centering

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



Equality constrained analytic centering

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 is flow through arc i; ϕ_i is 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
- $\textbf{rank}\, A = p$ if graph is connected

KKT system for network flow problem

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

- $H = \mathbf{diag}(\phi_1''(x_1), \ldots, \phi_n''(x_n))$, positive diagonal
- can be solved by block elimination:

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

sparsity pattern of coefficient matrix $AH^{-1}A^T$ is given by graph connectivity

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

Analytic center of linear matrix inequality

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

- variable is $X \in \mathbf{S}^n$
- includes an implicit constraint $X \succ 0$

Optimality conditions

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

Newton equation at feasible X

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

- follows from linear approximation $(X + \Delta X)^{-1} \approx X^{-1} - X^{-1}\Delta X X^{-1}$
- a set of $\frac{1}{2}n(n+1) + p$ linear equations in $\frac{1}{2}n(n+1) + p$ variables ΔX, w

Solution by block elimination

- eliminate ΔX from first equation:

$$\Delta X = X - \sum_{j=1}^{p} w_j X A_j X \tag{6}$$

- substitute expression for ΔX in second equation

$$\sum_{j=1}^{p} \operatorname{tr}(A_i X A_j X) w_j = b_i, \quad i = 1, \ldots, p \tag{7}$$

  this is a dense positive definite set of linear equations $Gw = b$ with

$$G_{ij} = \operatorname{tr}(A_i X A_j X), \quad i, j = 1, \ldots, p$$

- first solve (7) by Cholesky factorization of G; then substitute w in (6) to get ΔX

Complexity of block elimination method

the dominant terms in a flop count:

- Cholesky factorization $X = LL^T$ ($(1/3)n^3$ flops)
- form p products $L^T A_j L$ ($(3/2)pn^3$ flops)
- compute $p(p+1)/2$ elements of G

$$G_{ij} = \mathrm{tr}((L^T A_i L)(L^T A_j L))$$

  ($(1/2)p^2n^2$ flops)
- solve (7) via Cholesky factorization ($(1/3)p^3$ flops)

complexity is *cubic* in n (although KKT system has $\frac{1}{2}n(n+1) + p$ variables)