Convex Optimization

Stephen Boyd   Lieven Vandenberghe

Revised slides by Stephen Boyd, Lieven Vandenberghe, and Parth Nobel

10. Equality constrained minimization

Outline

Equality constrained minimization

Newton's method with equality constraints

Infeasible start Newton method

Implementation

Equality constrained minimization

- equality constrained smooth minimization problem:

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

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

- $f(x) = (1/2)x^T Px + q^T x + r$, $P \in \mathbf{S}^n_+$
- $\nabla f(x) = Px + q$
- optimality conditions are a **system of linear equations**

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

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

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

- equivalent condition for nonsingularity: $P + A^T A \succ 0$

Eliminating equality constraints

- represent feasible set $\{x \mid Ax = b\}$ as $\{Fz + \hat{x} \mid z \in \mathbf{R}^{n-p}\}$
  - $\hat{x}$ is (any) **particular solution** of $Ax = b$
  - 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**: minimize $f(Fz + \hat{x})$
- 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 resource allocation

- allocate resource amount $x_i \in \mathbf{R}$ to agent i
- agent i cost if $f_i(x_i)$
- resource budget is b, so $x_1 + \cdots + x_n = b$
- resource allocation problem is

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

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

$$\hat{x} = be_n, \qquad F = \left[\begin{array}{c} I \\ -\mathbf{1}^T \end{array}\right] \in \mathbf{R}^{n \times (n-1)}$$

- reduced problem: minimize $f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$

Outline

Equality constrained minimization

Newton's method with equality constraints

Infeasible start Newton method

Implementation

Newton step

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

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

- 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

- Newton decrement for equality constrained minimization is

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

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

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

- directional derivative in Newton direction:

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

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

Newton's method with equality constraints

---

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

repeat

1. Compute the Newton step and decrement Δx_{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.* $x := x + t\Delta x_{\text{nt}}$.

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

Newton's method and elimination

- reduced problem: minimize $\tilde{f}(z) = f(Fz + \hat{x})$
  - variables $z \in \mathbf{R}^{n-p}$
  - $\hat{x}$ satisfies $A\hat{x} = b$; $\mathbf{rank}\, F = n - p$ and $AF = 0$
- (unconstrained) Newton's method for $\tilde{f}$, started at $z^{(0)}$, generates iterates $z^{(k)}$
- iterates of Newton's method with equality constraints, started at $x^{(0)} = Fz^{(0)} + \hat{x}$, are

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

- hence, don't need separate convergence analysis

Outline

Equality constrained minimization

Newton's method with equality constraints

Infeasible start Newton method

Implementation

Newton step at infeasible points

- with $y = (x, \nu)$, write optimality condition as $r(y) = 0$, where

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

  is **primal-dual residual**

- consider $x \in \mathbf{dom} f$, $Ax \neq b$, *i.e.*, x is infeasible
- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

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

- $(\Delta x_{\mathrm{nt}}, \Delta \nu_{\mathrm{nt}})$ is called **infeasible** or **primal-dual** Newton step at x

Infeasible start Newton method

---

given starting point $x \in \textbf{dom} f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.

repeat

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

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

Outline

Solving KKT systems

- feasible and infeasible Newton methods require solving KKT system

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

- in general, can use $\mathrm{LDL}^{\mathrm{T}}$ factorization

- or elimination (if H nonsingular and easily inverted):
  - solve $AH^{-1}A^Tw = h - AH^{-1}g$ for w
  - $v = -H^{-1}(g + A^Tw)$

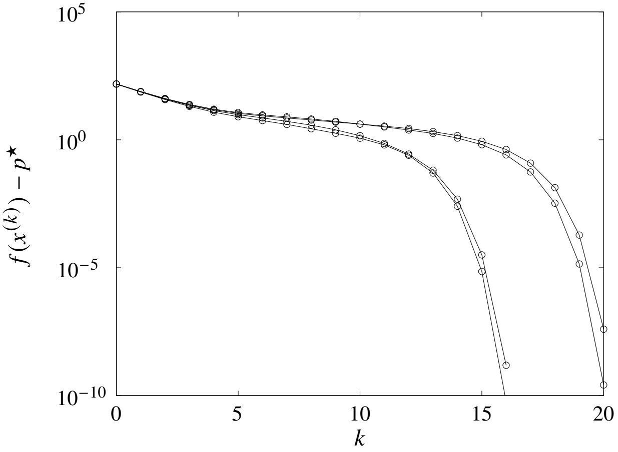
Example: Equality constrained analytic centering

- **primal problem:** minimize $-\sum_{i=1}^{n} \log x_i$ subject to $Ax = b$
- **dual problem:** maximize $-b^T\nu + \sum_{i=1}^{n} \log(A^T\nu)_i + n$
  - recover $x^\star$ as $x_i^\star = 1/(A^T\nu)_i$
- three methods to solve:
  - Newton method with equality constraints
  - Newton method applied to dual problem
  - infeasible start Newton method

  these have **different requirements for initialization**
- we'll look at an example with $A \in \mathbf{R}^{100\times 500}$, different starting points

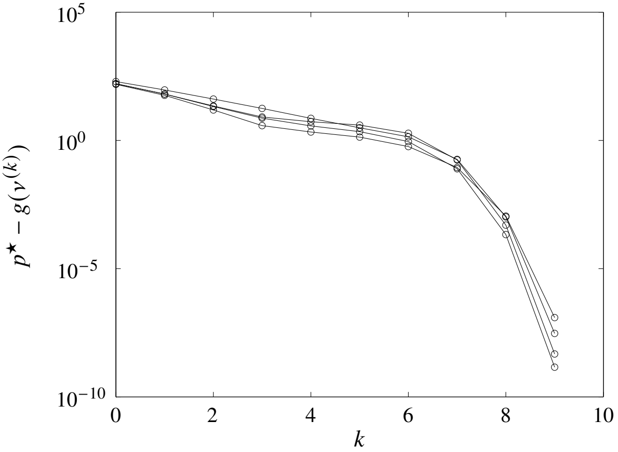
Newton's method with equality constraints

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



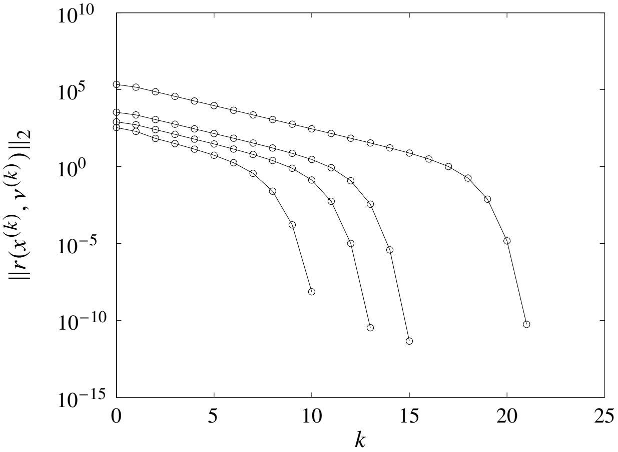
Newton method applied to dual problem

- requires $A^T \nu^{(0)} \succ 0$



Infeasible start Newton method

- requires $x^{(0)} \succ 0$



Complexity per iteration of three methods is identical

- for feasible Newton method, 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 $A\,\mathbf{diag}(x)^2 A^T w = b$
- for Newton system applied to dual, solve $A\,\mathbf{diag}(A^T\nu)^{-2}A^T\Delta\nu = -b + A\,\mathbf{diag}(A^T\nu)^{-1}\mathbf{1}$
- for infeasible start Newton method, 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 $A\,\mathbf{diag}(x)^2 A^T w = 2Ax - b$

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

Example: Network flow optimization

- directed graph with n arcs, $p+1$ nodes
- x_i: flow through arc i; ϕ_i: strictly convex flow cost function for arc i
- **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 incidence matrix** $A \in \mathbf{R}^{p\times n}$ is $\tilde{A}$ with last row removed
- $\mathbf{rank}\, A = p$ if graph is connected
- flow conservation is $Ax = b$, $b \in \mathbf{R}^p$ is (reduced) source vector

- **network flow optimization problem**: minimize $\sum_{i=1}^n \phi_i(x_i)$ subject to $Ax = b$

KKT system

- KKT system is

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

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

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

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

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

Analytic center of linear matrix inequality

- minimize $-\log\det X$ subject to $\mathbf{tr}(A_i X) = b_i, \ \ i = 1, \ldots, p$
- optimality conditions

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

- Newton step ΔX at feasible X is defined by

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

- 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 to get $\Delta X = X - \sum_{j=1}^{p} w_j X A_j X$
- substitute ΔX in second equation to get

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

- a dense positive definite set of linear equations with variable $w \in \mathbf{R}^p$
- form and solve this set of equations to get w, then get ΔX from equation above

Flop count

- find Cholesky factor L of X $\quad (1/3)n^3$
- form p products $L^T A_j L$ $\quad (3/2)pn^3$
- form $p(p+1)/2$ inner products $\mathbf{tr}((L^T A_i L)(L^T A_j L))$ to get coefficent matrix $\quad (1/2)p^2 n^2$
- solve $p \times p$ system of equations via Cholesky factorization $\quad (1/3)p^3$
- flop count dominated by $pn^3 + p^2 n^2$
- cf. naïve method, $(n^2 + p)^3$