COMP 9602: Convex Optimization

Algorithms for Equality Constrained Optimization

Dr. C. Wu

Department of Computer Science
The University of Hong Kong

Roadmap

Theory	convex set convex function standard forms of optimization problems, quasi-convex optimization linear program, integer linear program quadratic program geometric program semidefinite program vector optimization duality
Algorithm	unconstrained optimization equality constrained optimization interior-point method localization methods subgradient method decomposition methods

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 $\mathbf{rank}\, A = p < n$
- we assume $p^\star$ is finite and attained

optimality conditions: $x^\star$ is optimal iff there exists a $\nu^\star \in \mathbf{R}^p$ such that

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

(page 8, 4_Convex_Programs_I_C9602_Fall2018.pdf; also the KKT conditions)

solving the equality constrained optimization
<=> find the solution to the above KKT conditions (n+p equations)

Examples that are analytically solvable

- quadratic f(x) (with $P \in \mathbf{S}_+^n$)

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

Optimality conditions:

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

- coefficient matrix is called KKT matrix

if the KKT matrix is nonsingular, (x^*, ν^*) can be uniquely decided

Solving equality constrained optimization — Method 1

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 $AF = 0$)

(see also page 10, 4_Convex_Programs_I_C9602_Fall2018.pdf)

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

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

Solving equality constrained optimization —Method 2

- Solve the dual, then recover optimal primal variable x*
  - Lagrangian dual function:

$$\begin{aligned} g(\nu) &= -b^T\nu + \inf_x (f(x) + \nu^T Ax) \\ &= -b^T\nu - \sup_x ((-A^T\nu)^T x - f(x)) \\ &= -b^T\nu - f^*(-A^T\nu) \end{aligned}$$

  - The dual problem is

$$\max -b^T\nu - f^*(-A^T\nu)$$

- (possibly) unconstrained optimization: if $g(\nu)$ twice differentiable, descent methods can be applied
- strong duality holds
- reconstruct x*: x* minimizes $L(x, \nu^*)$

Example

$$\min f(x) = -\sum_{i=1}^{n} \log x_i$$

$$\text{s.t.} \quad Ax = b$$

Using

$$f^*(y) = \sum_{i=1}^{n}(-1 - \log(-y_i)) = -n - \sum_{i=1}^{n} \log(-y_i)$$

The dual problem is

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

After solving the dual, recover primal optimal point:

$$x_i = 1/(A^T\nu)_i$$

Solving equality constrained optimization —Method 3

- Newton's method with equality constraints

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

- General idea
  - start with a feasible $x^{(0)}$, s. t. $Ax^{(0)} = b$
  - find a feasible direction (Newton's direction or step):

$$\triangle x_{nt}, \text{ s.t. } x^{(k)} + t \triangle x_{nt} \text{ still satisfies } Ax = b$$

$$<=> A \triangle x_{nt} = 0$$

Newton's direction with equality constraints

- How to find $\triangle x_{nt}$

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

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

$$\Downarrow$$

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

$$\left[\begin{array}{cc} \nabla^2 f(x) & A^T \\ A & 0 \end{array} \right] \left[\begin{array}{c} v \\ w \end{array} \right] = \left[\begin{array}{c} -\nabla f(x) \\ 0 \end{array} \right]$$

KKT matrix (see also slide 4)

Newton's step is defined only at points
where KKT matrix is nonsingular

When f is nearly quadratic, $x + \triangle x_{nt}$ is a very good estimate of x*,
w is a good estimate of the optimal dual variable ν^*

Newton's decrement with equality constraints

- Newton's decrement

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

(The same $\triangle x_{nt}$ -based definition as for unconstrained optimization on page 10, 10_Alg_Unconstrained_Opt_C9602_Fall2018.pdf)

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

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 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.* $x := x + t\Delta x_{\mathrm{nt}}$.

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

Newton's method and elimination

- Newton's method with equality constraints is equivalent to Newton's method to solve the unconstrained reduced/eliminated 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 $x^{(k)} = Fz^{(k)} + \hat{x}$

- Therefore, convergence performance is exactly like the performance of Newton's method to solve unconstrained problems

Infeasible start Newton method

- A generalization that deals with infeasible initial points and iterates
  - let x be a point that we do not assume to be feasible
  - find $x + \triangle x_{nt}$ that solves the second-order approximation (see also slide 9)

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

  - The Newton method is based on primal-dual Newton step: both primal variable x and dual variable ν are updated

Infeasible start Newton method

- Primal-dual Newton step

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

where $\nu^+ = \nu + \triangle \nu_{nt}$ (equivalent to w on the previous slide)

$$\Updownarrow$$

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

primal Newton step   dual Newton step

- Residuals

$$r(x, \nu) = (\nabla f(x) + A^T \nu, Ax - b)^{\mathrm{T}}$$
$$= (r_{dual}(x, \nu), r_{pri}(x, \nu))^{\mathrm{T}}$$

optimality condition
<=> $r(x, \nu) = 0$

Infeasible start Newton method

- Primal-dual Newton step (an alternative way to derive)
  - write optimality condition as $r(y) = 0$, where

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

can be understood as $(r_{dual}(x,\nu), r_{pri}(x,\nu))$

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

Derivative of r evaluated at y

$$\Updownarrow$$

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

$$\Updownarrow$$

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

where $\nu^+ = \nu + \triangle \nu_{nt}$

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
- the norm of r decreases in the Newton's direction:

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

- if t=1, the next iterate will be feasible, and all the following iterates will be feasible

Example

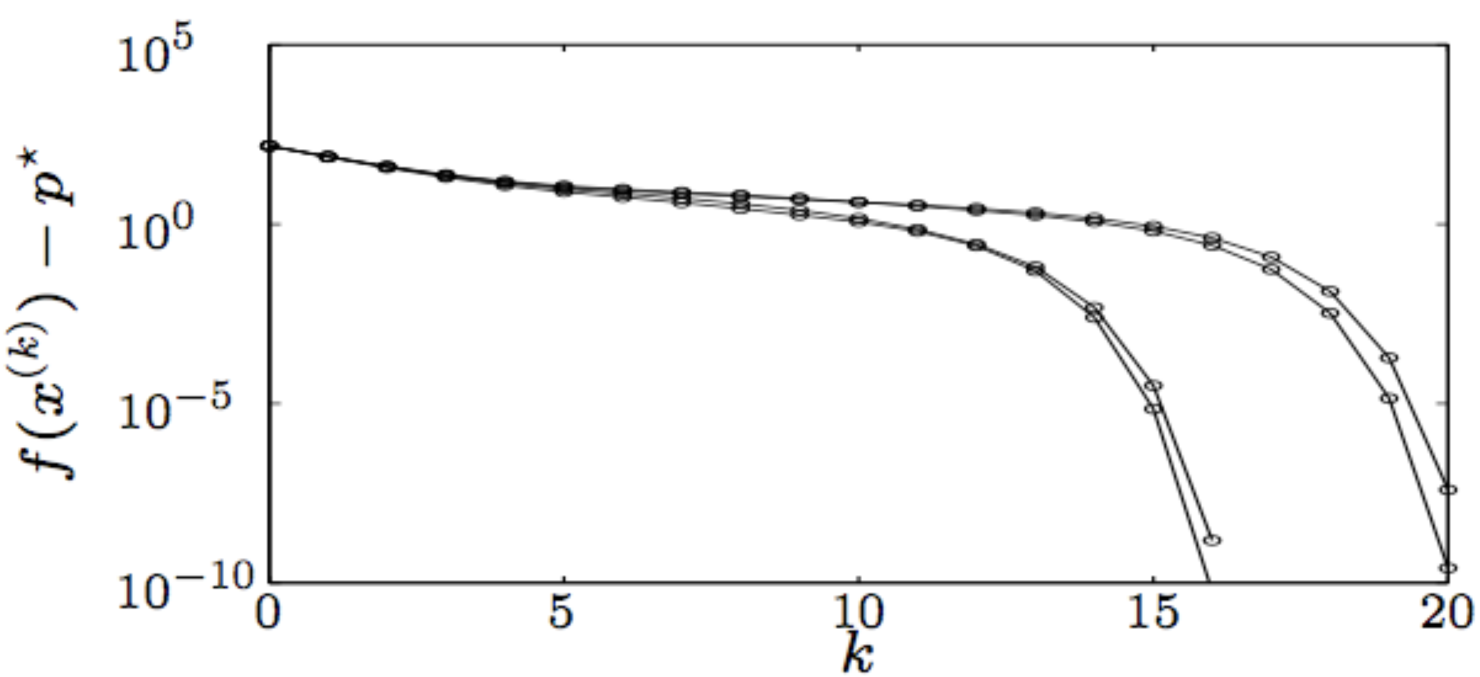
- Equality constraint analytic centering

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

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

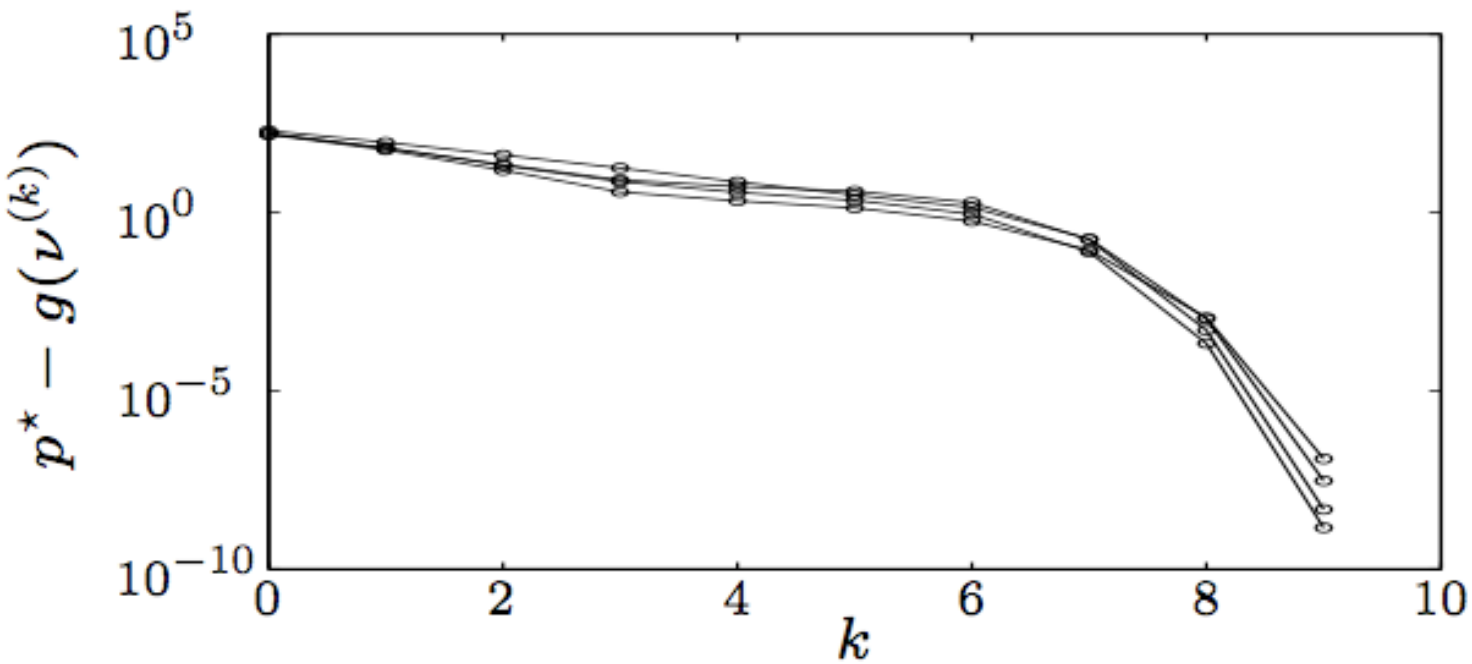
three methods for an example with $A \in \mathbf{R}^{100\times 500}$, different starting points

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

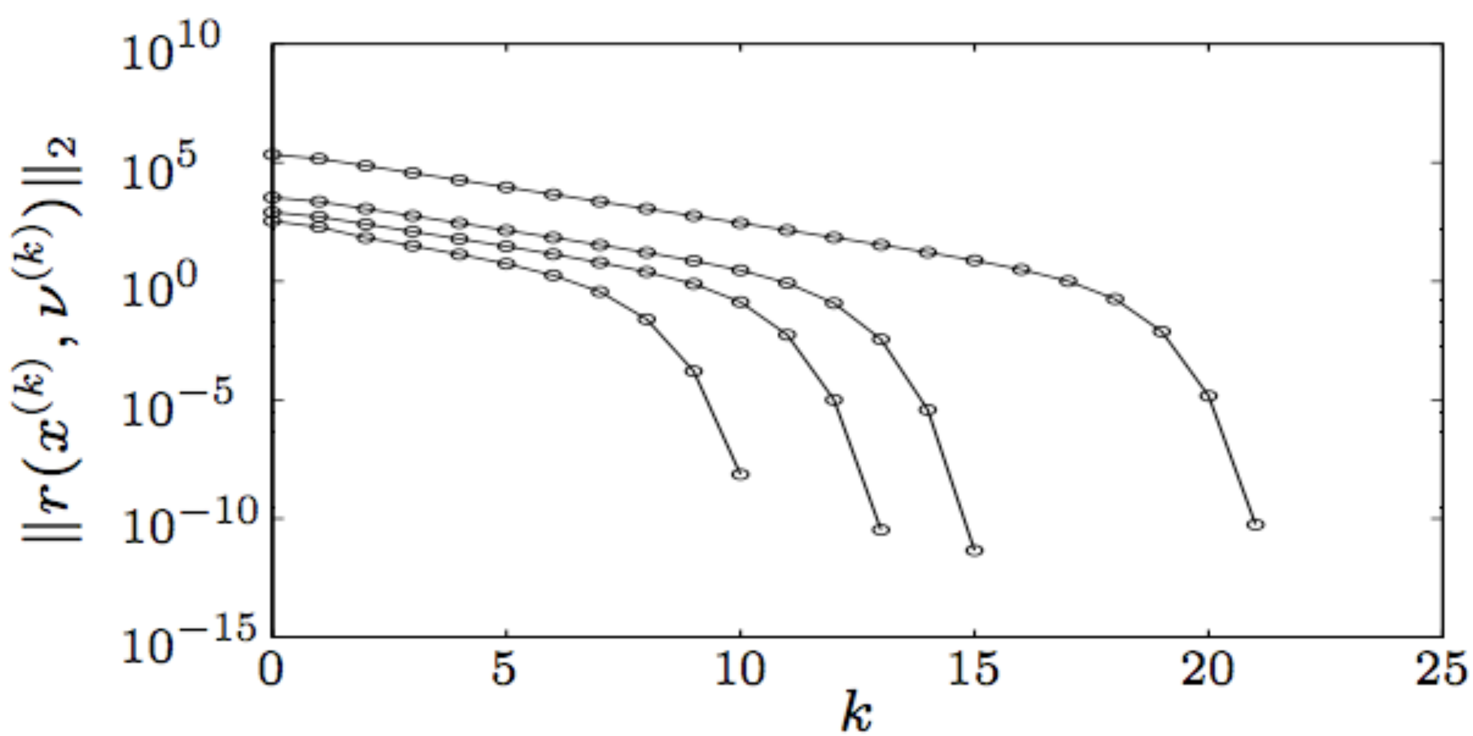


Example

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



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



- Reference
  - Chapter 10, Convex Optimization.
- Acknowledgement
  - Some materials are extracted from the slides created by Prof. Stephen Boyd for the textbook