Network Optimization
Chapter 2: Mathematical Tools

Yiyin Wang
yiyinwang@sjtu.edu.cn
Department of Automation
Shanghai Jiao Tong University

Nov. 7, 2014

Acknowledgements

Slides borrow heavily from lectures by Stephen Boyd (Stanford), Lieven Vandenberghe (UCLA), and Mung Chiang (Princeton)

Part 3: Algorithms

- 2.4 Descent algorithm
- 2.5 Interior-point algorithm

Equality constrained minimization

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

Optimizer x^*. Optimized value $p^* = f(x^*)$

- $f : \mathbf{R}^n \to \mathbf{R}$ convex and twice differentiable
- $A \in \mathbf{R}^{p \times n}$ with rank $p < n$
- We assume p^* is finite and attained

Optimality conditions: KKT equations with $n + p$ equations in $n + p$ variables x^*, ν^*

$$\bigtriangledown f(x^*) + A^T \nu^* = 0, Ax^* = b$$

Equality constrained quadratic minimization

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

where $P \in \mathbf{S}^n_+$
Optimality condition

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

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

$$Ax = 0, \; x \neq 0, \; \to x^T P x > 0$$

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

Approach 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 ($\mathrm{rank}(F) = n - p$ and $AF = 0$)

The reduced or eliminated problem

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

- An unconstrained problem with variable $z \in \mathbf{R}^{n-p}$
- From solution z^*, obtain x^* and ν^* as

$$x^* = Fz^* + \hat{x}$$

Example

- Optimal allocation with resource constraint

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

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

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

- Reduced problem:

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

Approach 2: dual solution

- Primal problem

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

- Dual function:

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

- Dual problem:

$$\text{maximize} \ -b^T\nu - f^*(-A^T\nu)$$

Example

Solve

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

Dual problems:

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

Recover primal variable from dual variable:

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

Approach 3: direct derivation of Newton method

- Make sure initial point is feasible and $A\triangle x_{nt} = 0$
- Replace objective with second order Taylor approximation near x:

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

- Find Newton step $\triangle x_{nt}$ by solving:

$$
\begin{bmatrix} \triangledown^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \triangle x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\triangledown f(x) \\ 0 \end{bmatrix}
$$

where w is associated optimal dual variable of $Ax = b$

Approach 3: direct derivation of Newton method

Interpretations:

- $\triangle x_{nt}$ solves second order approximation problem (with variable v)
- $\triangle x_{nt}$ equations follow from linearizing optimality condintions

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

Approach 3: Newton decrement

$$\lambda(x) = (\triangle x_{nt}^T \bigtriangledown^2 f(x) \triangle x_{nt})^{1/2} = (- \bigtriangledown f(x)^T \triangle x_{nt})^{1/2}$$

Properties

- Gives an estimate of $f(x) - p^*$ using quadratic approximation $\hat{f}$:

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

- Directional derivative in Newton direction:

$$\frac{d}{dt} f(x + t\triangle x_{nt})|_{t=0} = -\lambda(x)^2$$

- In general, $\lambda(x) \neq (\bigtriangledown f(x)^T \bigtriangledown^2 f(x)^{-1} \bigtriangledown f(x))^{1/2}$

Inequality constrained minimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \dots, m \\ & Ax = b \end{array} \tag{1}$$

- f_i convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\operatorname{rank}(A) = p$
- We assume p^* is finite and attained
- We assume problem is strictly feasible: there exists $\tilde{x}$ with

$$\tilde{x} \in \operatorname{dom} f_0, \quad f_i(\tilde{x}) < 0, i = 1, \dots, m, \quad A\tilde{x} = b$$

Hence, strong duality holds and dual optimum is attained

Inequality constrained minimization

- Idea: reduce it to a sequence of linear equality constrained problem and apply Newton's method
- First, need to approximately formulate inequality constrained problem as an equality constrained problem

Barrier function

- Reformulation of (1) via indicator function:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

  where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise (indicator function of $\mathbf{R}_-$)

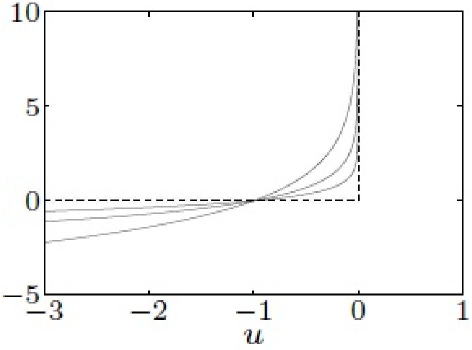
  - No inequality constraints, but objective function not differentiable

- Approximate the indicator function by a differentiable, closed, and convex function

$$\hat{I}_-(u) = -(1/t)\log(-u), \quad \mathrm{dom}\,\hat{I}_- = -\mathbf{R}_{++}$$

  where a larger parameter t gives more accurate approximation
  $\hat{I}_-$ increases to ∞ as u increases to 0

Barrier function



Use Newton method to solve approximation:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} \hat{I}_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

Logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \ \mathrm{dom}\,\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

- Convex (follows from composition rules)
- Twice continuously differentiable, with derivatives

$$\begin{aligned} \nabla\phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) \\ \nabla^2\phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x) \end{aligned}$$

Central path

- for $t > 0$, define $x^*(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

  for now, assume $x^*(t)$ exists and is unique for each $t > 0$
- Central path: solutions to above problem $x^*(t)$, characterized by
  1. Strict feasibility:

$$Ax^*(t) = b, f_i(x^*(t)) < 0, i = 1, \ldots, m$$

  2. Centrality condition: there exists $\hat{\nu} \in \mathbf{R}^p$ such that

$$t \bigtriangledown f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \bigtriangledown f_i(x^*(t)) + A^T \hat{\nu} = 0$$

  Every central point gives a dual feasible point. Let

$$\lambda_i^*(t) = -\frac{1}{tf_i(x^*(t))}, i = 1, \ldots, m, \nu^*(t) = \frac{\hat{\nu}}{t}$$

Central path

- Thus, $x^*(t)$ minimizes the Lagrangian

$$L(x, \lambda^*(t), \nu^*(t)) = f_0(x) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x) + \nu^*(t)^T (Ax - b)$$

  where $\lambda_i^*(t)$ and $\nu^*(t)$ are defined as before

- This confirms the intuitive idea that $f_0(x^*(t)) \to p^*$ if $t \to \infty$

$$\begin{aligned} p^* &\geq g(\lambda^*(t), \nu^*(t)) \\ &= L(x^*(t), \lambda^*(t), \nu^*(t)) \\ &= f_0(x^*(t)) - m/t \end{aligned}$$

Interpretation via KKT conditions

$x = x^*(t), \lambda = \lambda^*(t), \nu = \nu^*(t)$ satisfy

1. Primal constraints: $f_i(x) \leq 0, i = 1, \ldots, m, Ax = b$
2. Dual constraints: $\lambda \succeq 0$
3. Approximate complementary slackness: $-\lambda_i f_i(x) = 1/t, i = 1, \ldots, m$
4. Gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T \nu = 0$$

Difference with KKT is that condition 3 (complementary slackness) is relaxed from 0 to $1/t$

Barrier method

Given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$
Repeat

1. Centering step. Compute $x^*(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$
2. Update: $x := x^*(t)$
3. Stopping criterion: quit if $m/t < \epsilon$
4. Increase t. $t := \mu t$

- Terminates with $f_0(x) - p^* \leq \epsilon$ (stopping criterion follows from $f_0(x^*(t)) - p^* \leq m/t$)
- Centering usually done using Newton's method, starting at current x

Remarks

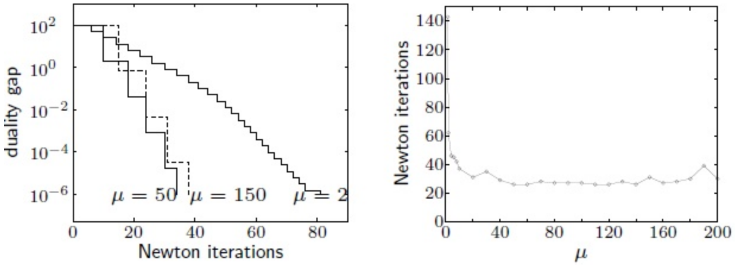
- Choice of μ: tradeoff number of inner iterations with number of outer iterations. Large μ means fewer outer iterations, more inner iterations; typical values: $\mu = 10 - 20$
- Choice of $t^{(0)}$: tradeoff number of inner iterations within the first outer iteration with number of outer iterations
- Number of centering steps (outer iterations) required:

$$\left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

plus the initial centering step (to compute $x^*(t^{(0)})$), where m is the number of inequality constraints and ϵ is desired accuracy

Examples

Inequality form LP ($m = 100$ inequalities, $n = 50$ variables)

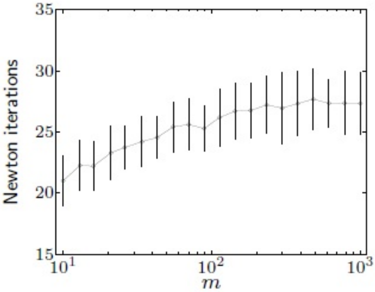


- Starts with x on central path ($t^{(0)} = 1$, duality gap 100)
- Terminates when $t = 10^8$ (gap 10^{-6})
- Centering uses Newton's method with backtracking
- Total number of Newton iterations not very sensitive for $\mu \geq 10$

Family of standard LPs ($A \in \mathbf{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m, solve 100 randomly generated instances



Number of iterations grows very slowly as m ranges over a 100 : 1 ratio

Feasibility and phase I method

How to compute a strictly feasible point to start barrier method?

Feasible problem: find x such that

$$f_i(x) \leq 0, i = 1, \ldots, m, \quad Ax = b \tag{2}$$

Consider a phase I optimization problem in variables $x \in \mathbf{R}^n, s \in R$

$$\begin{array}{rl} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, i = 1, \ldots, m \\ & Ax = b, x \succeq 0 \end{array} \tag{3}$$

- If x, s feasible, with $s < 0$, then x is strictly feasible for (2)
- If optimal value $\bar{p}^*$ of (3) is positive, then problem (2) is infeasible
- If $\bar{p}^* = 0$ and attained, then problem (2) is feasible (but not strictly); If $\bar{p}^* = 0$ and not attained, then problem (2) is infeasible

Phase I method

1. Apply barrier method to solve phase I problem (stop when $s < 0$)
2. Use the resulted strictly feasible point for the original problem to start barrier method for the original problem

Summary

- Convert equality constrained optimization into unconstrained optimization
- Solve a general convex optimization with interior point methods
- Turn inequality constrained problem into a sequence of equality constrained problems that increasingly accurate approximation of the original problem
- Polynomial time (in theory) and much faster (in practice): about $25 - 50$ least-squares effort for a wide range of problem sizes

Topics covered by this chapter

- Fundamentals of convex sets, convex functions, convex problems, duality
- Applications of linear programming (LP) and semidefinite programming (SDP)
- Algorithms for unconstrained, equality and inequality constrained minimization: descent and interior-point algorithm