

ELE539A: Optimization of Communication Systems
Lecture 18: Interior Point Algorithm

Professor M. Chiang
Electrical Engineering Department, Princeton University

March 30, 2007

Lecture Outline

- Newton's method
- Equality constrained convex optimization
- Inequality constrained convex optimization
- Barrier function and central path
- Barrier method

- Summary of topics covered in the course
- Final projects
- Conclusions

Newton Method

Newton step:

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

Positive definiteness of $\nabla^2 f(x)$ implies that Δx_{nt} is a descent direction

Interpretation: linearize optimality condition $\nabla f(x^*) = 0$ near x ,

$$\nabla f(x + v) \approx \nabla f(x) + \nabla^2 f(x)v = 0$$

Solving this linear equation in v , obtain $v = \Delta x_{nt}$. Newton step is the addition needed to x to satisfy linearized optimality condition

Main Properties

- **Affine invariance:** given nonsingular $T \in \mathbf{R}^{n \times n}$ and let $\bar{f}(y) = f(Tx)$. Then Newton step for \bar{f} at y :

$$\Delta y_{nt} = T^{-1} \Delta x_{nt}$$

and

$$x + \Delta x_{nt} = T(y + \Delta y_{nt})$$

- **Newton decrement:**

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

Let \hat{f} be second order approximation of f at x . Then

$$f(x) - p^* \approx f(x) - \inf_y \hat{f}(y) = f(x) - \hat{f}(x + \Delta x_{nt}) = \frac{1}{2} \lambda(x)^2$$

Newton Method

GIVEN a starting point $x \in \text{dom } f$ and tolerance $\epsilon > 0$

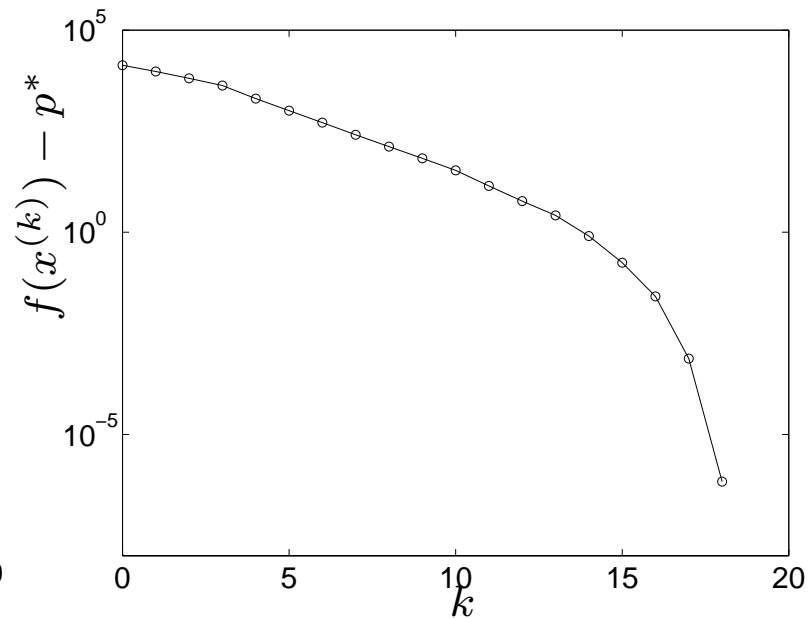
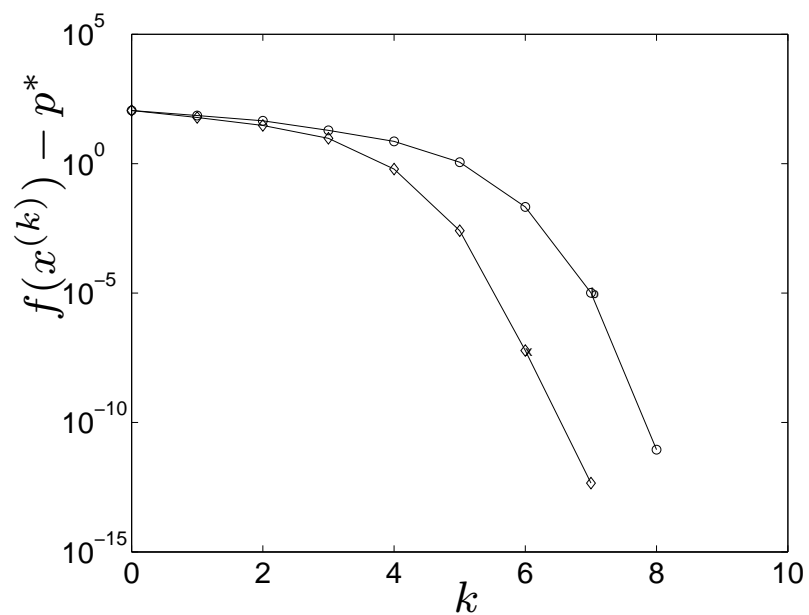
REPEAT

1. Compute Newton step and decrement: $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$ and $\lambda = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$
2. Stopping criterion: QUIT if $\frac{\lambda^2}{2} \leq \epsilon$
3. Line search: choose a step size $t > 0$
4. Update: $x := x + t\Delta x$

Advantages of Newton method: Fast, Robust, Scalable

Error vs. Iteration for Problems in \mathbf{R}^{100} and \mathbf{R}^{10000}

$$\text{minimize } c^T x - \sum_{i=1}^M \log(b_i - a_i^T x), \quad x \in \mathbf{R}^{100} \text{ and } \in \mathbf{R}^{10000}$$



Equality Constrained Problems

Solve a convex optimization with **equality** constraints:

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

$f : \mathbf{R}^n \rightarrow \mathbf{R}$ is twice differentiable

$A \in \mathbf{R}^{p \times n}$ with $\text{rank } p < n$

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

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

Approach 1: Can be turned into an **unconstrained** optimization, after eliminating the equality constraints

Approach 2: Dual Solution

Dual function:

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

Dual problem:

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

Example: solve

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

Dual problem:

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

Recover primal variable from dual variable:

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

Example With Analytic Solution

Convex **quadratic minimization** over equality constraints:

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

Optimality condition:

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

If KKT matrix is nonsingular, there is a unique optimal primal-dual pair x^*, ν^*

If KKT matrix is singular but solvable, any solution gives optimal x^*, ν^*

If KKT matrix has no solution, primal problem is unbounded below

Approach 3: Direct Derivation of Newton Method

Make sure initial point is feasible and $A\Delta x_{nt} = 0$

Replace objective with second order Taylor approximation near x :

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

Find **Newton step** Δx_{nt} by solving:

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

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

Newton's method (Newton decrement, affine invariance, and stopping criterion) stay the **same**

Inequality Constrained Minimization

Let $f_0, \dots, f_m : \mathbf{R}^n \rightarrow \mathbf{R}$ be convex and twice continuously differentiable and $A \in \mathbf{R}^{p \times n}$ with rank $p < n$:

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

Assume the problem is strictly feasible and an optimal x^* exists

Idea: reduce it to a sequence of linear **equality** constrained problems and apply Newton's method

First, need to **approximately** formulate inequality constrained problem as an equality constrained problem

Barrier Function

Make inequality constraints implicit in the objective:

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

where I_- is **indicator function**:

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

No inequality constraints, but objective function not differentiable

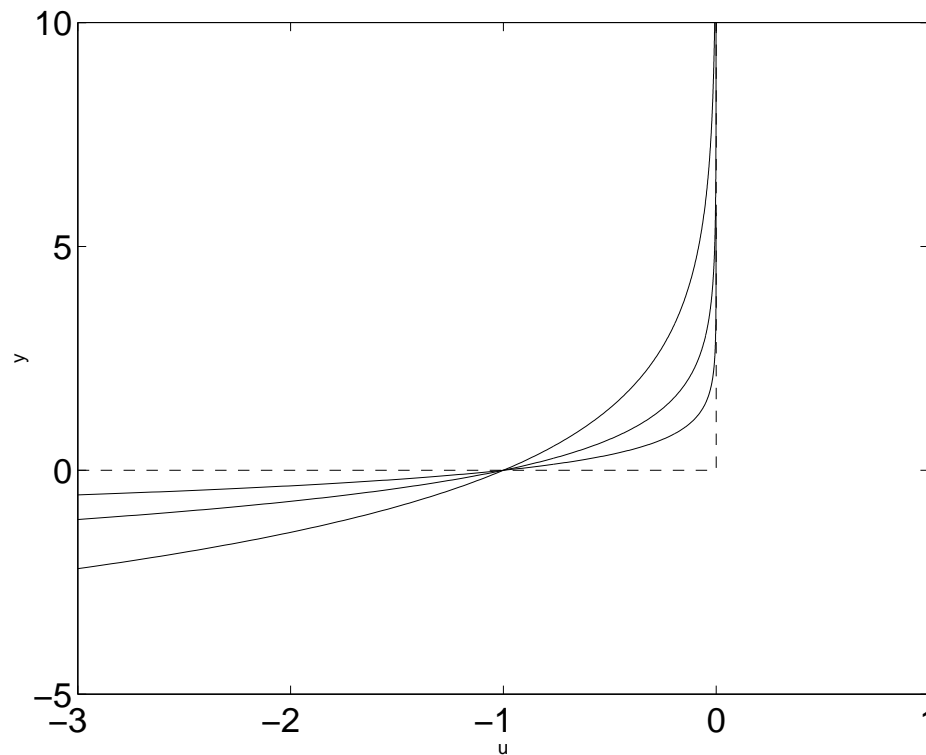
Approximate indicator function by a **differentiable**, closed, and convex function:

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

where a larger parameter t gives more accurate approximation

\hat{I}_- increases to ∞ as u increases to 0

Log Barrier



Use Newton method to solve approximation:

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

Log Barrier

Log barrier function:

$$\phi(x) = - \sum_{i=1}^m \log(-f_i(x)), \quad \text{dom } \phi = \{x \in \mathbf{R}^n \mid f_i(x) < 0, i = 1, \dots, m\}$$

Approximation better if t is large, but then Hessian of $f_0 + (1/t)\phi$ varies rapidly near boundary of feasible set. **Accuracy Stability tradeoff**

Solve a **sequence of approximation** with larger t , using **Newton method** for each step of the sequence

Gradient and Hessian of log barrier function:

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x)$$

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

Central Path

Consider the family of optimization problems parameterized by $t > 0$:

$$\begin{aligned} &\text{minimize} && t f_0(x) + \phi(x) \\ &\text{subject to} && Ax = b \end{aligned}$$

Central path: solutions to above problem $x^*(t)$, characterized by:

1. **Strict feasibility**:

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

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

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

Every central point gives a **dual feasible point**. Let

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

Central Path

Dual function

$$g(\lambda^*(t), \nu^*(t)) = f_0(x^*(t)) - m/t$$

which implies **duality gap** is m/t . Therefore, **suboptimality gap**

$$f_0(x^*(t)) - p^* \leq m/t$$

Interpretation as **modified KKT condition**:

x is a central point $x^*(t)$ iff there exists λ, ν such that

$$Ax = b, \quad f_i(x) \leq 0, \quad i = 1, \dots, m$$

$$\lambda \succeq 0$$

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

$$-\lambda_i f_i(x) = 1/t, \quad i = 1, \dots, m$$

Complementary slackness is **relaxed** from 0 to $1/t$

Example

Inequality form LP:

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && Ax \preceq b \end{aligned}$$

Log barrier function:

$$\phi(x) = - \sum_{i=1}^m \log(b_i - a_i^T x)$$

with gradient and Hessian:

$$\begin{aligned} \nabla \phi(x) &= \sum_{i=1}^m \frac{a_i}{b_i - a_i^T x} = A^T d \\ \nabla^2 \phi(x) &= \sum_{i=1}^m \frac{a_i a_i^T}{(b_i - a_i^T x)^2} = A^T \mathbf{diag}(d^2) A \end{aligned}$$

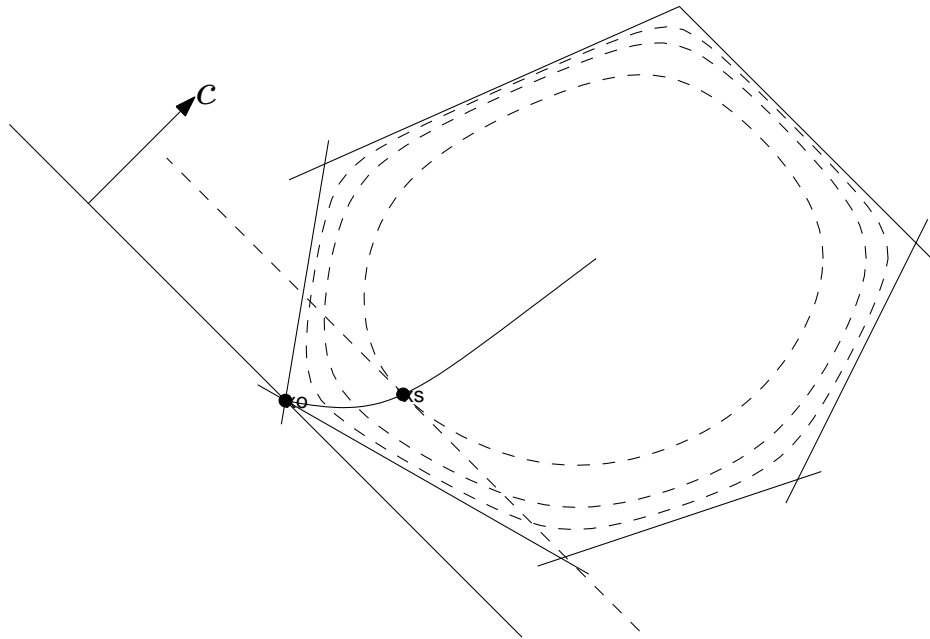
where $d_i = 1/(b_i - a_i^T x)$

Centrality condition becomes:

$$tc + A^T d = 0$$

c is parallel to $\nabla\phi(x)$.

Therefore, hyperplane $c^T x^*(t)$ is tangent to level set of ϕ



Barrier Method

GIVEN a strictly feasible point $x \in \mathbf{dom} f$, $t := t^{(0)} > 0$, $\mu > 1$ and tolerance $\epsilon > 0$

REPEAT

1. Centering step: compute $x^*(t)$ by minimizing $tf_0 + \phi$ subject to $Ax = b$, starting at x
2. Update: $x := x^*(t)$
3. Stopping criterion: QUIT if $\frac{m}{t} \leq \epsilon$
4. Increase t : $t := \mu t$

Other names: Sequential Unconstrained Minimization Technique (SUMT) or path-following method

Usually use Newton method for Centering Step

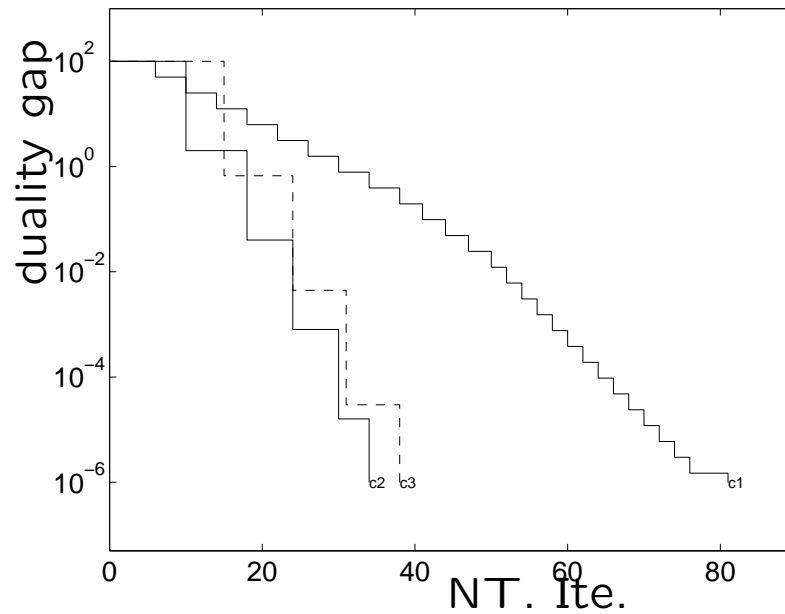
Remarks

- Each centering step does not need to be exact
- Choice of μ : tradeoff number of **inner iterations** with number of **outer iterations**
- Choice of $t^{(0)}$: tradeoff number of **inner iterations within the first outer iteration** with number of **outer iterations**
- Number of centering steps required:

$$\frac{\log(m/(\epsilon t^{(0)}))}{\log \mu}$$

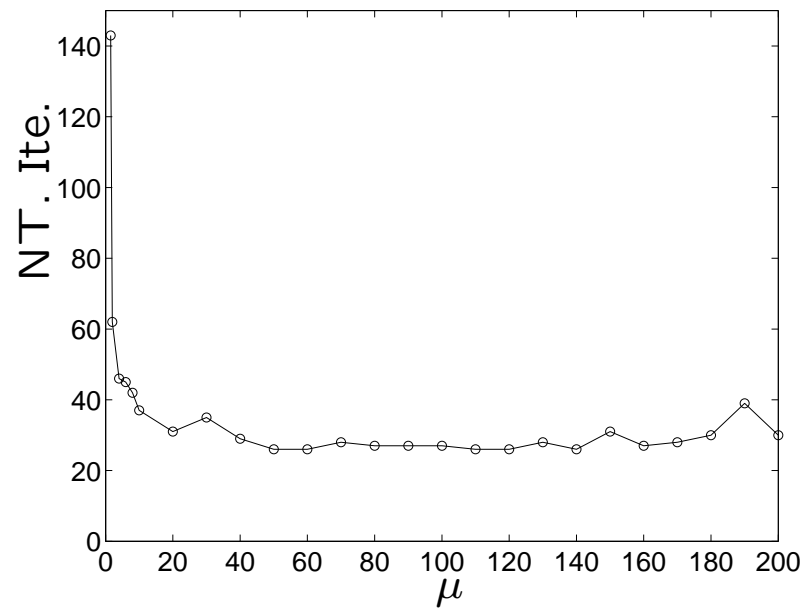
where m is the number of inequality constraints and ϵ is desired accuracy

Progress of Barrier Method for an LP Example

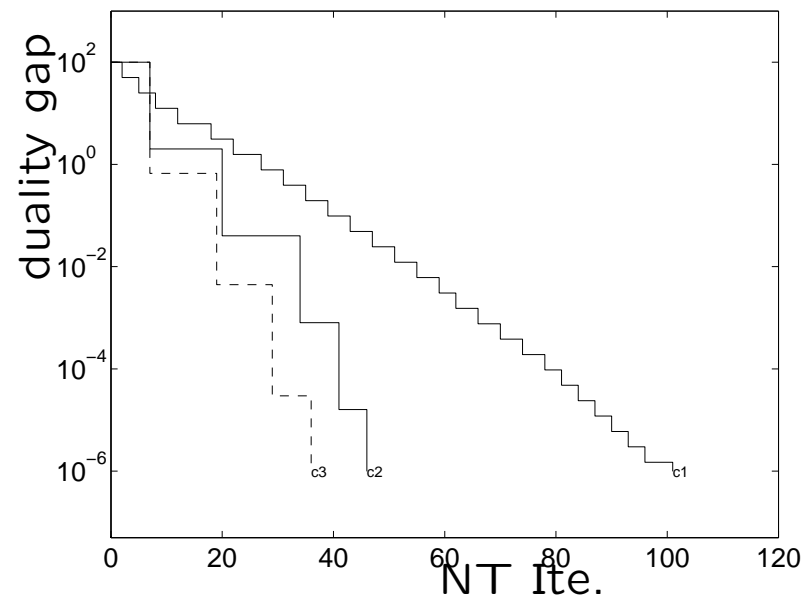


Three curves for $\mu = 2, 50, 150$

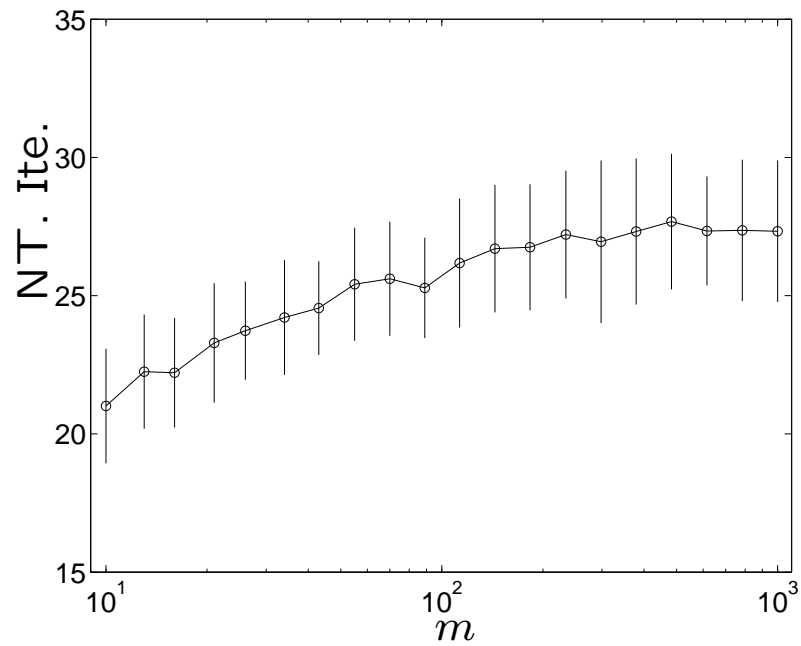
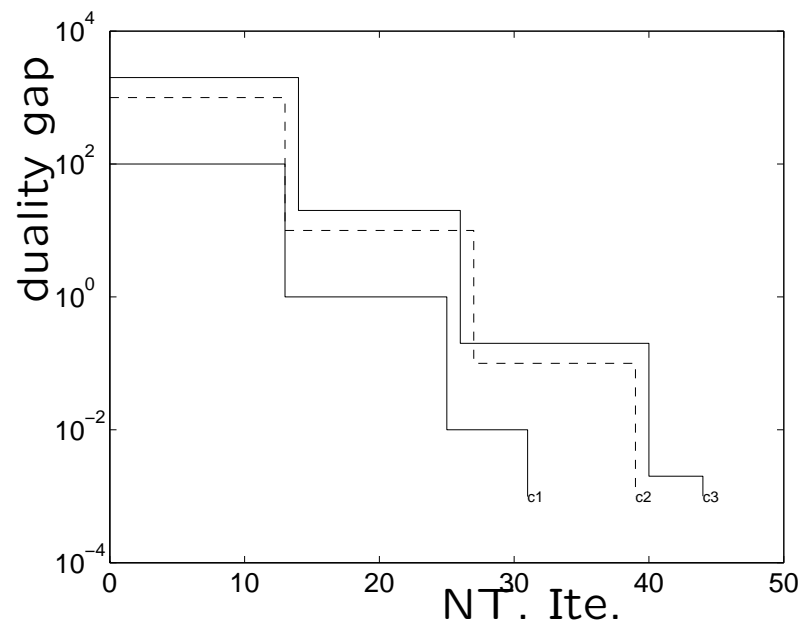
Tradeoff of μ Parameter for a Small LP



Progress of Barrier Method for a GP Example



Insensitive to Problem Size



Three curves for $m = 50, 500, 1000$, $n = 2m$.

Phase I Method

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

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

$$\begin{aligned} &\text{minimize} && s \\ &\text{subject to} && f_i(x) \leq s, \quad i = 1, \dots, m \\ &&& Ax = b \end{aligned}$$

Strictly feasible point: for any $x^{(0)}$, let $s = \max f_i(x^{(0)})$

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

Not Covered

- Self-concordance analysis and complexity analysis (polynomial time)
- Numerical linear algebra (large scale implementation)
- Generalized inequalities (SDP)

Other (sometimes more efficient) algorithms:

- Primal-dual interior point method
- Ellipsoid methods
- Analytic center cutting plane methods

Lecture Summary

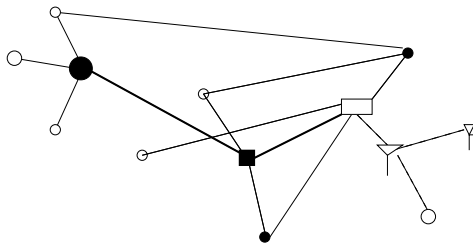
- Convert equality constrained optimization into unconstrained optimization
- Solve a general convex optimization with interior point methods
- Turn inequality constrained problem into a sequence of equality constrained problems that are increasingly accurate approximation of the original problem
- Polynomial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

Readings: Chapters 9.1-9.3, 9.5, 10.1-10.2, 11.1-11.4 in Boyd and Vandenberghe

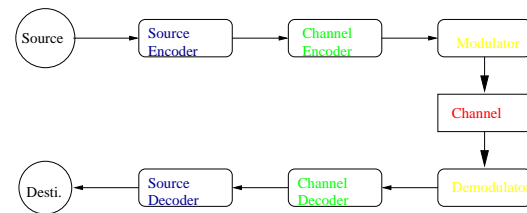
Why Take This Course

- Learn the **tools and mentality** of optimization (surprisingly useful for other study you may engage in later on)
- Learn **classic and recent results on optimization of communication systems** (over a surprisingly wide range of problems)
- Enhance the ability to do original research in academia or industry
- Have a fun and productive time on the final project

Where We Started



Application
Presentation
Session
Transport
Network
Link
Physical



$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

Topics Covered: Methodologies

- Nonlinear optimization (linear, convex, nonconvex), Pareto optimization, Dynamic programming, Integer programming
- Convex optimization, convex sets and convex functions
- Lagrange duality and KKT optimality condition
- LP, QP, QCQP, SOCP, SDP, GP, SOS
- Gradient method, Newton's method, interior point method
- Distributed algorithms and decomposition methods
- Nonconvex optimization and relaxations
- Stochastic network optimization and Lyapunov function
- Robust LP and GP
- Network flow problems
- NUM and generalized NUM

Topics Covered: Applications

- **Information theory problems:** channel capacity and rate distortion for discrete memoryless models
- **Detection and estimation problems:** MAP and ML detector, multi-user detection
- **Physical layer:** DSL spectrum management
- **Wireless networks:** resource allocation, power control, joint power and rate allocation
- **Network algorithms and protocols:** multi-commodity flow problems, max flow, shortest path routing, network rate allocation, TCP congestion control, IP routing
- **Reverse engineering,** heterogeneous protocols
- **Congestion, collision, interference** management
- **Layering as optimization decomposition**
- Eigenvalue and norm optimization problems

Some of the Topics Not Covered

- Circuit switching and optimal routing
- Packet switch architecture and algorithms
- Multi-user information theory problems
- More digital signal processing like blind equalization
- Large deviations theory and queuing theory applications
- System stability and control theoretic improvement of TCP
- Inverse optimization, robust optimization
- Self concordance analysis of interior point method's complexity
- Ellipsoid method, cutting plane method, simplex method ...

Final Projects: Timeline

March 30 – May 24:

- Talk to me as often as needed
- Start early

April 27: Interim project report due

Mid May: Project Presentation Day

May 24: Final project report due

Many student continue to extend the project

Final Projects: Topics

Wenjie Jiang: Content distribution networking

Suhas Mathur: Wireless network security

Hamed Mohsenian: Multi-channel MAC

Martin Suchara: Stochastic simulation for TCP/IP

Guanchun Wang: Google with feedback

Yihong Wu: Practical stepsize

Yiyue Wu: Coding

Yongxin Xi: Image analysis

Minlan Yu: Network embedding

Dan Zhang: To be decided

The Optimization Mentality

- What can be varied and what cannot?
- What're the objectives?
- What're the constraints?
- What's the dual?

Build models, analyze models, prove theorems, compute solutions, design systems, verify with data, refine models ...

Warnings:

- Remember the restrictive assumptions for application needs
- Understand the limitations: discrete, nonconvex, complexity, robustness ... (still a lot can be done for these difficult issues)

Optimization of Communication Systems

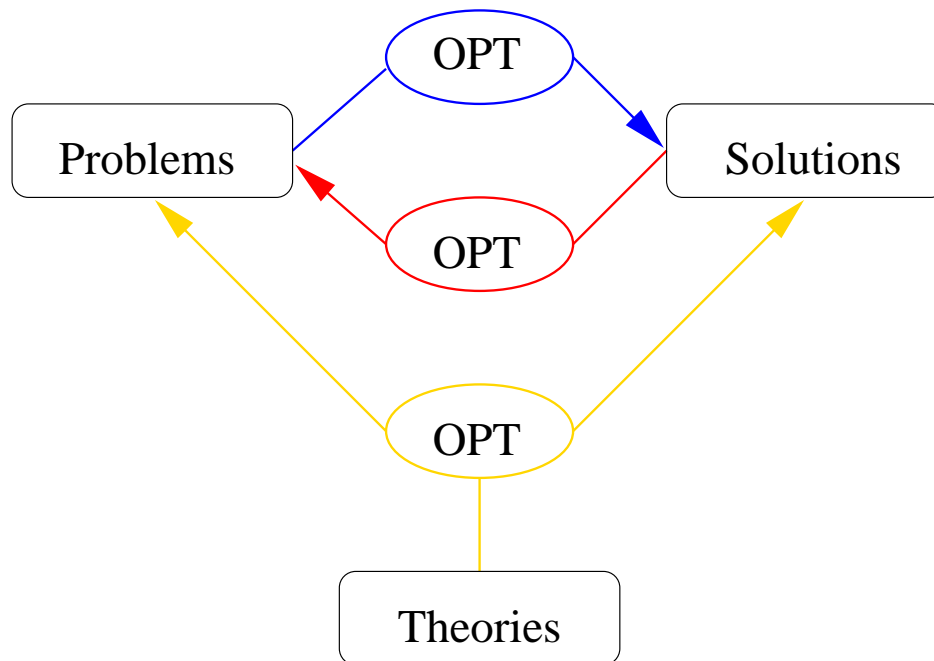
Three ways to apply optimization theory to communication systems:

- Formulate the problem as an optimization problem
- Interpret a given solution as an optimizer/algorithm for an optimization problem
- Extend the underlying theory by optimization theoretic techniques

A remarkably powerful, versatile, widely applicable and not yet fully recognized framework

Applications in communication systems also stimulate new developments in optimization theory and algorithms

Optimization of Communication Systems



Optimization Beyond Optimality

- Modeling (reverse engineering, fairness, etc.)
- Architecture
- Robustness
- Design for Optimizability

This Is The Beginning

Quote from Thomas Kailath:

A course is not about what you've covered, but what you've **uncovered**

**To Everyone In ELE539A
You've Been Amazing**