

A Tutorial on Convex Optimization II: Duality & Interior Point Methods

Haitham Hindi

Palo Alto Research Center (PARC)

using lecture slides courtesy of
Stephen Boyd, *EE Dept, Stanford*
Lieven Vandenberghe, *EE Dept, UCLA*

Outline

- Convexity review
- Duality from practical view point
- Interior point methods overview

1. Convexity

- motivation
- convex sets and functions
- new standard convex problem classes

Motivation: why optimize?

1. consensus on exactly what the problem is
2. enables automation of design (CAD) via optimization
3. discover limits of performance
4. if can solve exact optimization problem we're done
5. if can't solve/implement exact solution, have benchmark

General optimization problem: standard form

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

- $x = (x_1, \dots, x_n)$ is *optimization variable*
- f_0 is *objective* or *cost function*;
 $f_i(x) \leq 0$ are the *inequality constraints*;
 $h_i(x) = 0$ are the *equality constraints*
- x is *feasible* if it satisfies the constraints;
the *feasible set* C is the set of all feasible points;
- *optimal value* is $f^* = \inf_{x \in C} f_0(x)$,
optimal point: any $x^* \in C$ s.t. $f(x^*) = f^*$;

Geometric interpretation

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

- $f_i(x) \leq 0$ means x lies in the 0-sublevel sets of $f_i(x)$, ie:

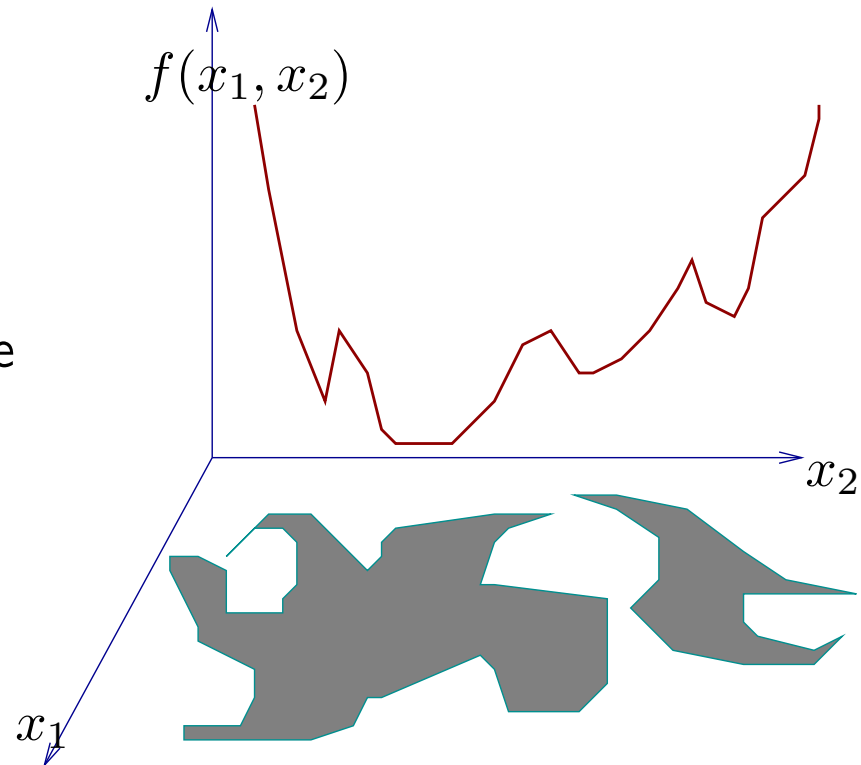
$$x \text{ feasible} \quad \Rightarrow \quad x \in \bigcap_i \{x \mid f_i(x) \leq 0\}$$

- $h_i(x) = 0$ means that x lies in the 0-solution sets of $h_i(x)$, ie:

$$x \text{ feasible} \quad \Rightarrow \quad x \in \bigcap_i \{x \mid h_i(x) = 0\}$$

Practical problems with optimization

- local minima
- slow convergence
- stopping criterion (how do we know we're done?)
- numerical issues
- finding initial feasible point



Usually solved using gradient descent, genetic algorithms, simulated annealing, etc...

All these problems go away if problem is **convex**!

What makes optimization hard or easy: Convexity

emerging (and correct) view:

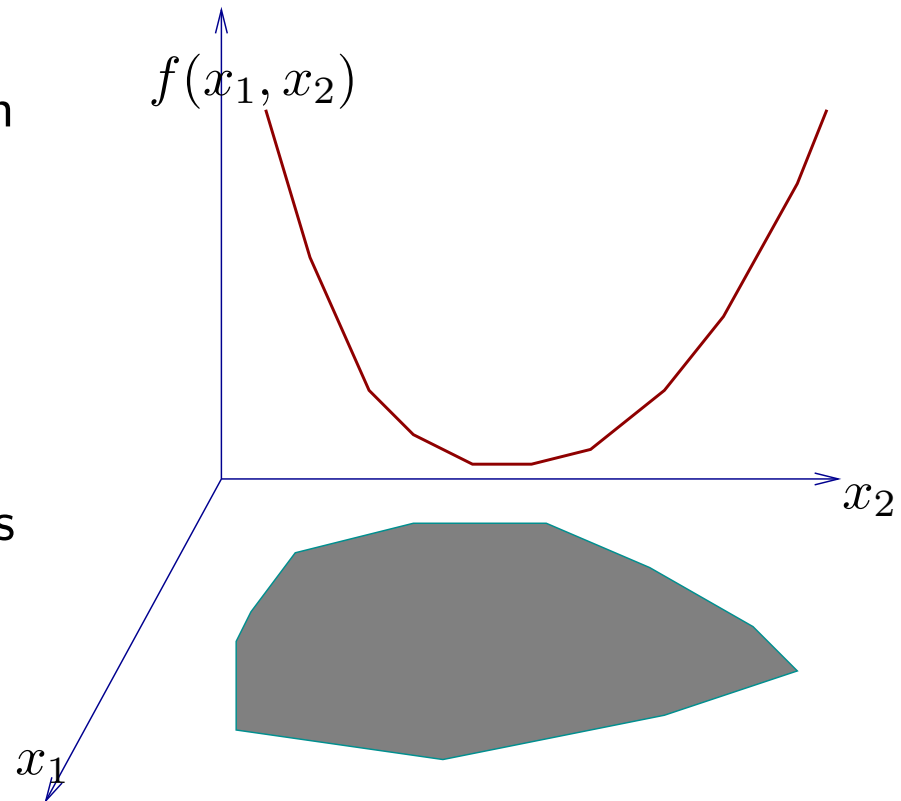
. . . the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.

— *R. Rockafellar, SIAM Review 1993*

convex problems are **fundamentally tractable**

Advantages of convex optimization

- local optimum \Rightarrow *global* optimum
- convergence rate guarantees
- very efficient numerical methods
- exact stopping criterion
- easy initialization & unambiguous feasibility detection
- handle *large number of variables*



50's examples: Linear Programming, Quadratic Programming

90's examples: Geometric Programming, Semidefinite Programming, etc

Convex optimization problem

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

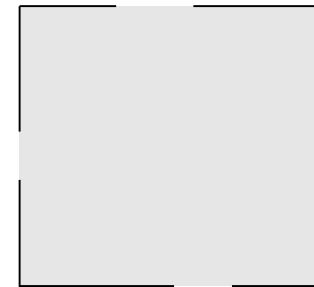
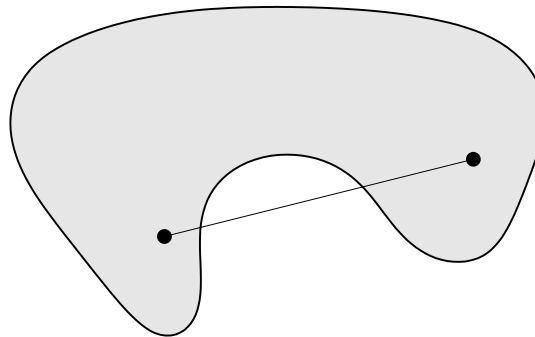
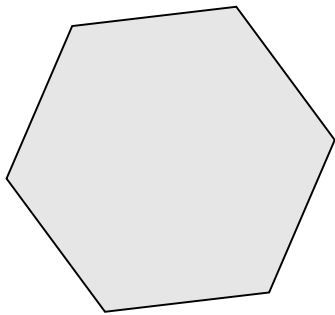
where $A \in \mathbf{R}^{p \times n}$

- f_0, f_1, \dots, f_m convex
- equality constraints $Ax = b$ are linear (affine)
- feasible set is convex
(intersection of 0-sublevel sets of f_i with hyperplane $\{x \mid Ax = b\}$)

Convex set

$C \subseteq \mathbf{R}^n$ is convex if

$$x, y \in C, \theta \in [0, 1] \implies \theta x + (1 - \theta)y \in C$$



Geometrically, convex \implies no dents & always bulging outward

Properties of convex sets

- Convexity is preserved under *intersection*
if $\{S_\alpha \mid \alpha \in \mathcal{A}\}$ is collection of sets, then:

$$S_\alpha \text{ is } \begin{pmatrix} \text{subspace} \\ \text{affine} \\ \text{convex} \\ \text{convex cone} \end{pmatrix} \implies \bigcap_{\alpha \in \mathcal{A}} S_\alpha \text{ is } \begin{pmatrix} \text{subspace} \\ \text{affine} \\ \text{convex} \\ \text{convex cone} \end{pmatrix}$$

Every closed convex set S is (usually infinite) intersection of halfspaces

- Convexity is preserved under *linear transformation*

$$S \text{ convex} \implies f^{-1}(S) = \{x \mid Ax + b \in S\} \text{ convex}$$

$$S \text{ convex} \implies f(S) = \{Ax + b \mid x \in S\} \text{ convex}$$

eg: projections, slices, . . .

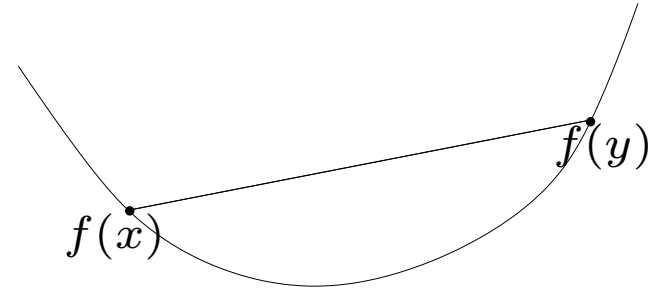
Convex function

$f : \mathbf{R}^n \rightarrow \mathbf{R}$ is convex if

$$x, y \in \mathbf{R}^n, \theta \in [0, 1]$$

↓

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$



examples

- $f(x) = x^2$
- $f(x) = -\log(x), \quad x > 0$
- norms, positive quadratics, etc . . .

Properties of convex functions

- f convex \Rightarrow α -**sublevel sets** $\{x \mid f(x) \leq \alpha\}$ convex
 - f convex $\Rightarrow f(Ax + b)$ convex
convexity preserved under linear transformation
 - f_1, f_2 convex $\Rightarrow f_1 + f_2$ convex
 - f_1, f_2 convex $\Rightarrow \max f_1, f_2$ convex
 - composition rules: if h cvx & increasing, f cvx, then $g(x) = h(f(x))$
cvx
 - minimization: if $f(x, y)$ cvx, then $g(x) = \inf_y f(x, y)$ cvx
 - many others
- ... and many, many others

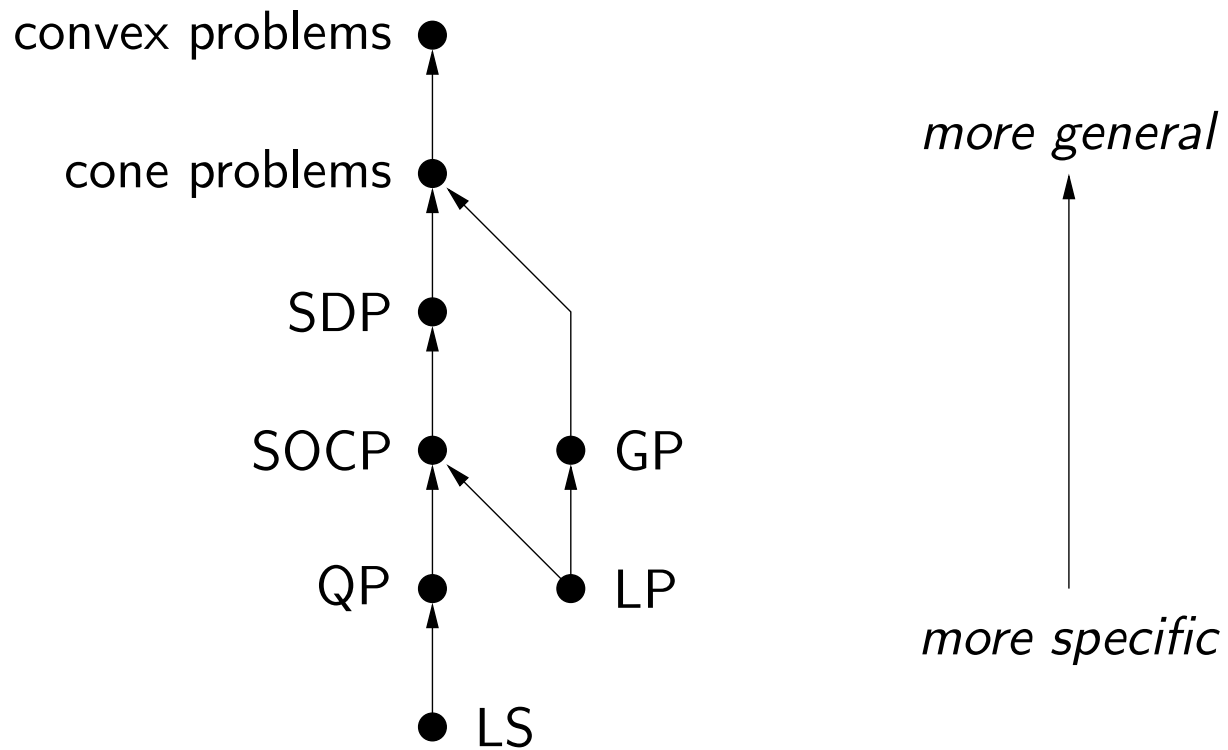
Some new standard convex problem classes

- second-order cone program (SOCP)
- geometric program (GP) (and entropy problems)
- semidefinite program (SDP)
- sum of squares program (SOS)

for these new problem classes we have

- complete duality theory, similar to LP
- good algorithms, and robust, reliable software
- wide variety of new applications

Convex optimization hierarchy



2. Duality

Key practical uses of duality are

- Lower bounding hard (nonconvex) problems
- Stopping criterion for numerical optimization
- Decomposition of large problems
- Sensitivity analysis

Lagrangian

standard form problem (not necessarily convex)

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

variable $x \in \mathbf{R}^n$, domain \mathcal{D} , optimal value p^*

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$, with $\text{dom } L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right) \end{aligned}$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if \tilde{x} is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible \tilde{x} gives $p^* \geq g(\lambda, \nu)$

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \dots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) &= \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^* \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^* \geq n\lambda_{\min}(W)$

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on p^* , obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted d^*
- λ, ν are dual feasible if $\lambda \succeq 0, (\lambda, \nu) \in \mathbf{dom} g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom} g$ explicit

example: standard form LP and its dual

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

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Weak and strong duality

weak duality: $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems
for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

gives a lower bound for the two-way partitioning problem on page 2–4

strong duality: $d^* = p^*$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

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

if it is strictly feasible, *i.e.*,

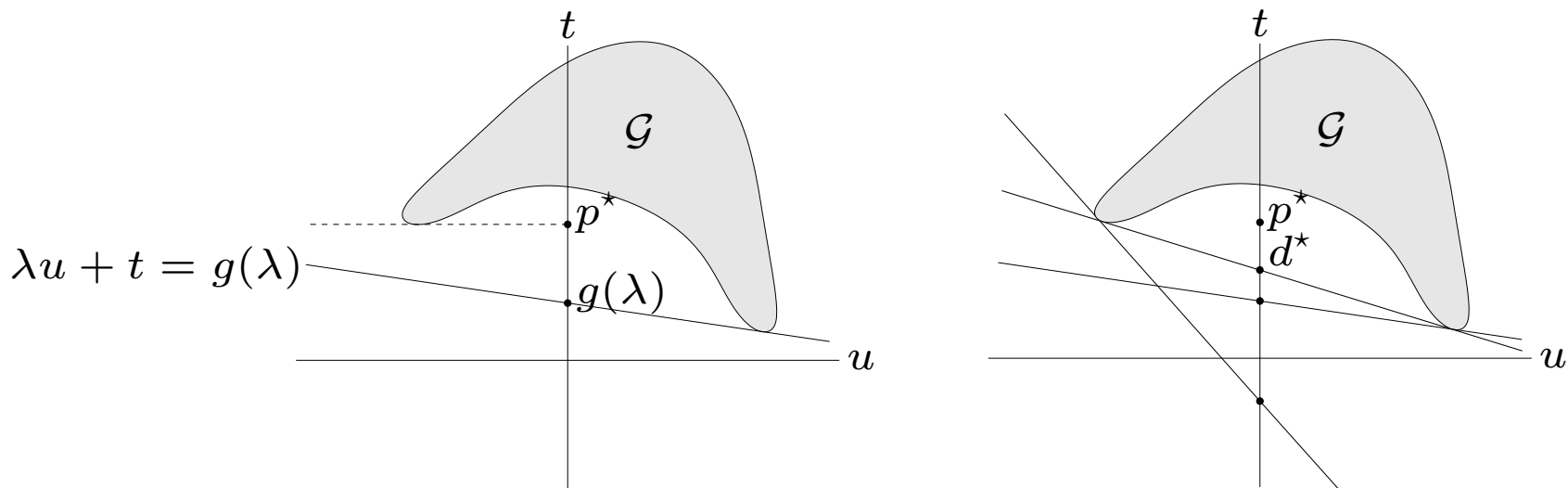
$$\exists x \in \mathbf{int} \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^* > -\infty$)
- can be sharpened: *e.g.*, can replace $\mathbf{int} \mathcal{D}$ with $\mathbf{relint} \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$
interpretation of dual function:

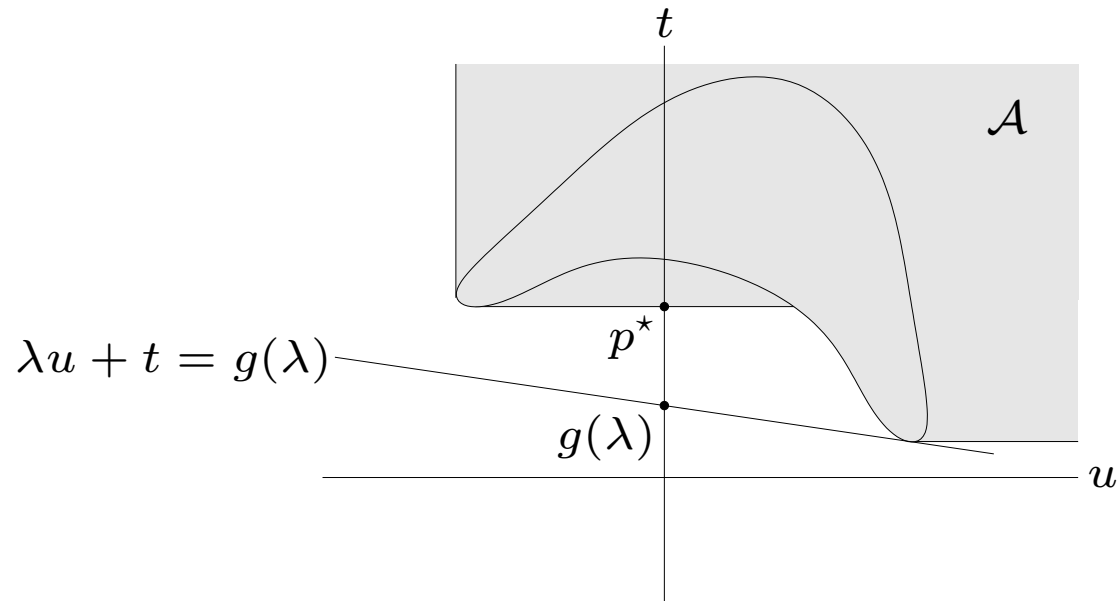
$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \quad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$



- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to \mathcal{G}
- hyperplane intersects t -axis at $t = g(\lambda)$

epigraph variation: same interpretation if \mathcal{G} is replaced with

$$\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$



strong duality

- holds if there is a non-vertical supporting hyperplane to \mathcal{A} at $(0, p^*)$
- for convex problem, \mathcal{A} is convex, hence has supp. hyperplane at $(0, p^*)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^*)$ must be non-vertical

Complementary slackness

assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \dots, m$ (known as complementary slackness):

$$\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$$

from page 2–10: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if \tilde{x} , $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ , ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

example: water-filling (assume $\alpha_i > 0$)

$$\begin{aligned} & \text{minimize} && -\sum_{i=1}^n \log(x_i + \alpha_i) \\ & \text{subject to} && x \succeq 0, \quad \mathbf{1}^T x = 1 \end{aligned}$$

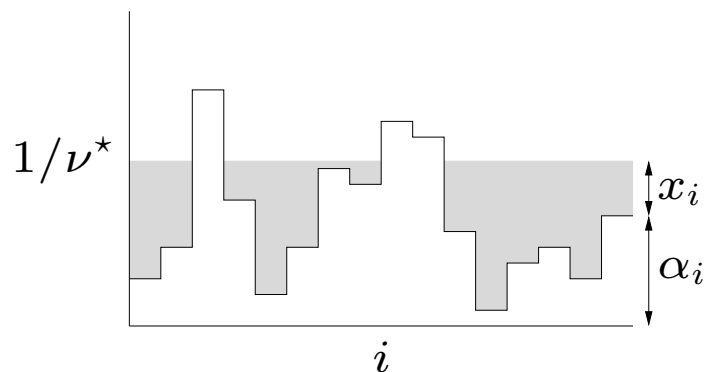
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

$$\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^*$



Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

perturbed problem and its dual

$$\begin{array}{ll} \text{min.} & f_0(x) \\ \text{s.t.} & f_i(x) \leq u_i, \quad i = 1, \dots, m \\ & h_i(x) = v_i, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \text{max.} & g(\lambda, \nu) - u^T \lambda - v^T \nu \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- x is primal variable; u, v are parameters
- $p^*(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^*(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual

global sensitivity result

assume strong duality holds for unperturbed problem, and that λ^* , ν^* are dual optimal for unperturbed problem

apply weak duality to perturbed problem:

$$\begin{aligned} p^*(u, v) &\geq g(\lambda^*, \nu^*) - u^T \lambda^* - v^T \nu^* \\ &= p^*(0, 0) - u^T \lambda^* - v^T \nu^* \end{aligned}$$

sensitivity interpretation

- if λ_i^* large: p^* increases greatly if we tighten constraint i ($u_i < 0$)
- if λ_i^* small: p^* does not decrease much if we loosen constraint i ($u_i > 0$)
- if ν_i^* large and positive: p^* increases greatly if we take $v_i < 0$;
if ν_i^* large and negative: p^* increases greatly if we take $v_i > 0$
- if ν_i^* small and positive: p^* does not decrease much if we take $v_i > 0$;
if ν_i^* small and negative: p^* does not decrease much if we take $v_i < 0$

local sensitivity: if (in addition) $p^*(u, v)$ is differentiable at $(0, 0)$, then

$$\lambda_i^* = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu_i^* = -\frac{\partial p^*(0, 0)}{\partial v_i}$$

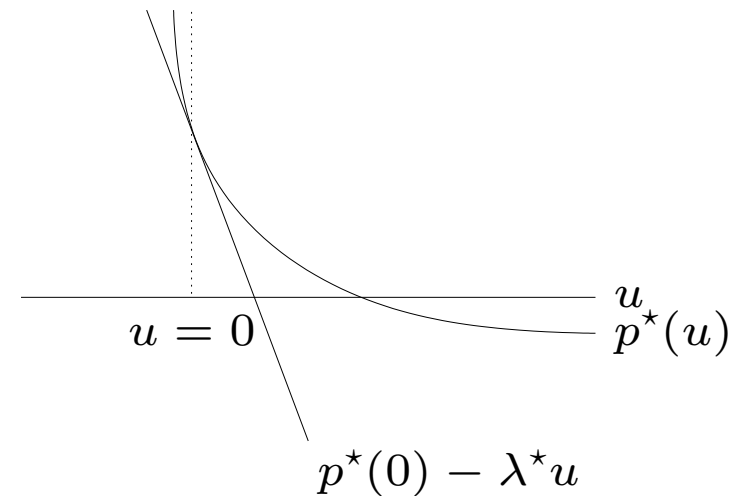
proof (for λ_i^*): from global sensitivity result,

$$\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \geq -\lambda_i^*$$

$$\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \leq -\lambda_i^*$$

hence, equality

$p^*(u)$ for a problem with one (inequality) constraint:



Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions

e.g., replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

3. Interior Point Methods

- backtracking linesearch for convex functions
- Newton's method for problems with equality constraints
- Barrier method for problems with equality and inequality constraints

Descent methods

- produce sequence of points $x^{(k)}$, with $f(x^{(k)}) \rightarrow p^*$

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- other notations: $x^+ = x + t\Delta x$, $x := x + t\Delta x$
- Δx is the *step*, or *search direction*; t is the *step size*, or *step length*
- from convexity, $f(x^+) < f(x)$ implies $\nabla f(x)^T \Delta x < 0$
(*i.e.*, Δx is a *descent direction*)

General descent method.

given a starting point $x \in \text{dom } f$.

repeat

1. Determine a descent direction Δx .
2. *Line search.* Choose a step size $t > 0$.
3. *Update.* $x := x + t\Delta x$.

until stopping criterion is satisfied.

Line search types

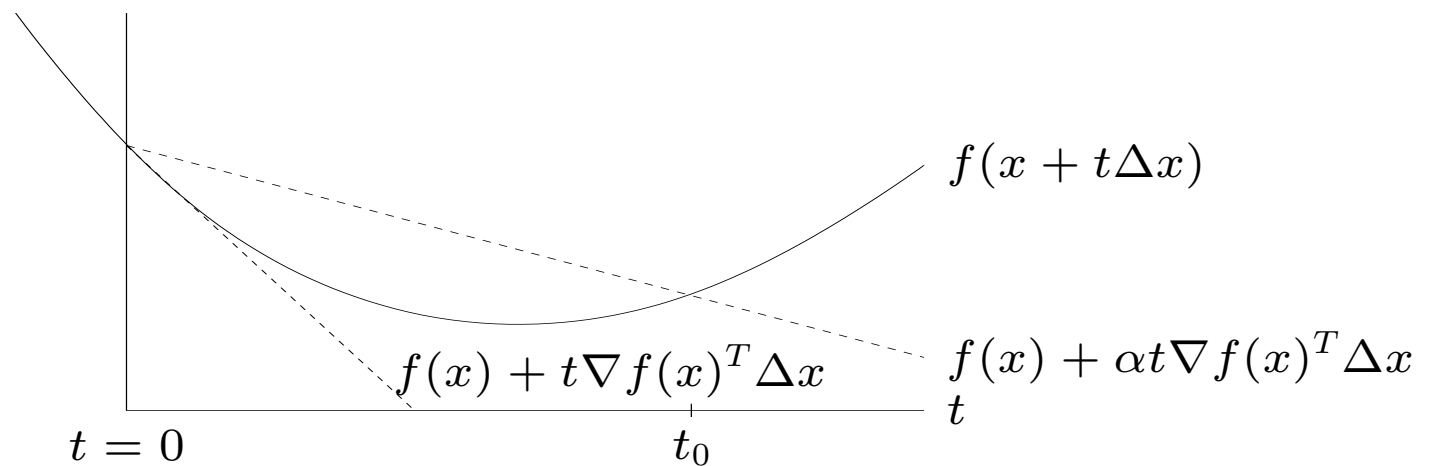
exact line search: $t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$

backtracking line search (with parameters $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$)

- starting at $t = 1$, repeat $t := \beta t$ until

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

- graphical interpretation: backtrack until $t \leq t_0$



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 $\text{rank } A = p$
- we assume p^* is finite and attained

optimality conditions: x^* is optimal iff there exists a ν^* such that

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

equality constrained quadratic minimization (with $P \in \mathbf{S}_+^n$)

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

optimality condition:

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

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

$$Ax = 0, \quad x \neq 0 \quad \implies \quad x^T P x > 0$$

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

Newton step

Newton step of f at feasible x is given by (1st block) of solution of

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

interpretations

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

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

- equations follow from linearizing optimality conditions

$$\nabla f(x + \Delta x_{\text{nt}}) + A^T w = 0, \quad A(x + \Delta x_{\text{nt}}) = b$$

Newton decrement

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

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:

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

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

repeat

1. Compute the Newton step and decrement $\Delta x_{\text{nt}}, \lambda(x)$.
 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

Inequality constrained minimization

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

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

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

hence, strong duality holds and dual optimum is attained

Examples

- LP, QP, QCQP, GP
- entropy maximization with linear inequality constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g \\ & Ax = b \end{array}$$

with $\text{dom } f_0 = \mathbf{R}_{++}^n$

- differentiability may require reformulating the problem, *e.g.*, piecewise-linear minimization or ℓ_∞ -norm approximation via LP
- SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

Logarithmic barrier

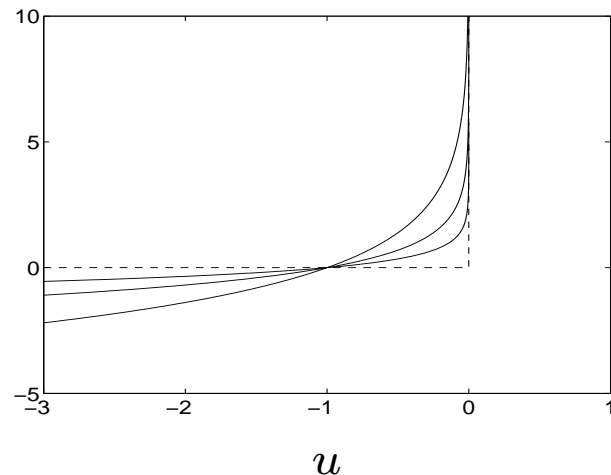
reformulation of (1) via indicator function:

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

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

$$\begin{aligned} & \text{minimize} && f_0(x) - (1/t) \sum_{i=1}^m \log(-f_i(x)) \\ & \text{subject to} && Ax = b \end{aligned}$$

- an equality constrained problem
- for $t > 0$, $-(1/t) \log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \rightarrow \infty$



logarithmic barrier function

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

- convex (follows from composition rules)
- twice continuously differentiable, with derivatives

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

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

Central path

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

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

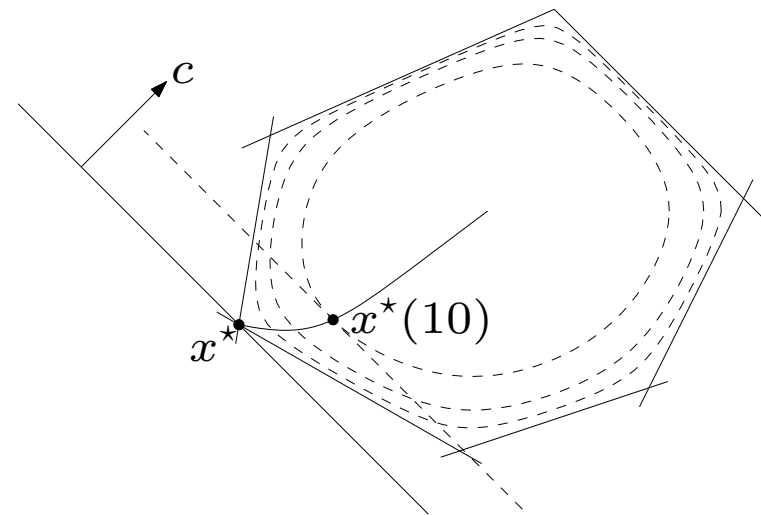
(for now, assume $x^*(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^*(t) \mid t > 0\}$

example: central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \dots, 6 \end{array}$$

hyperplane $c^T x = c^T x^*(t)$ is tangent to level curve of ϕ through $x^*(t)$



Dual points on central path

$x = x^*(t)$ if there exists a w such that

$$t \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, \quad Ax = b$$

- therefore, $x^*(t)$ 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 we define $\lambda_i^*(t) = 1/(-t f_i(x^*(t)))$ and $\nu^*(t) = w/t$

- this confirms the intuitive idea that $f_0(x^*(t)) \rightarrow p^*$ if $t \rightarrow \infty$:

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

Interpretation via KKT conditions

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

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \dots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

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

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

Barrier method

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

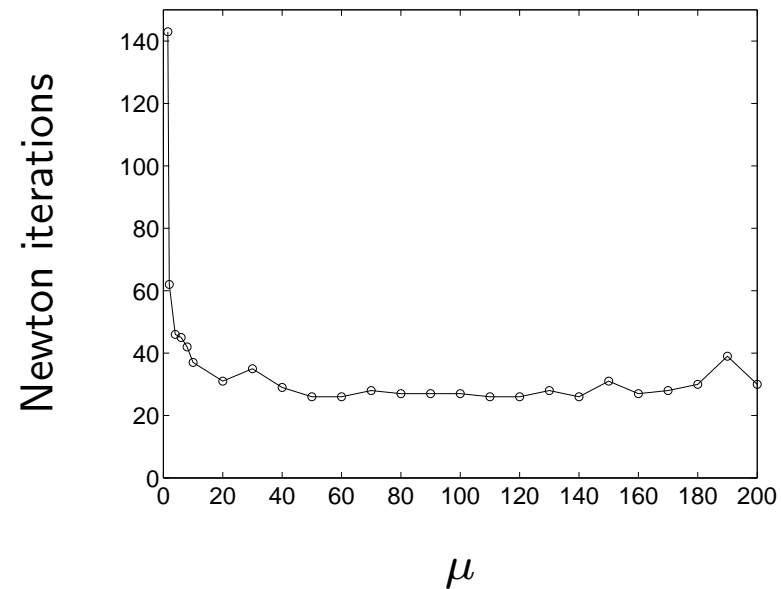
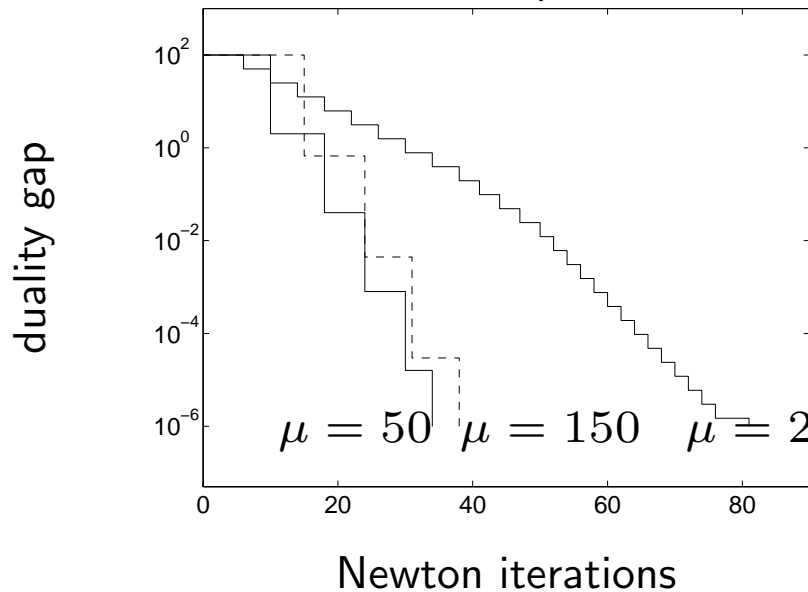
repeat

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
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10\text{--}20$
- several heuristics for choice of $t^{(0)}$

Examples

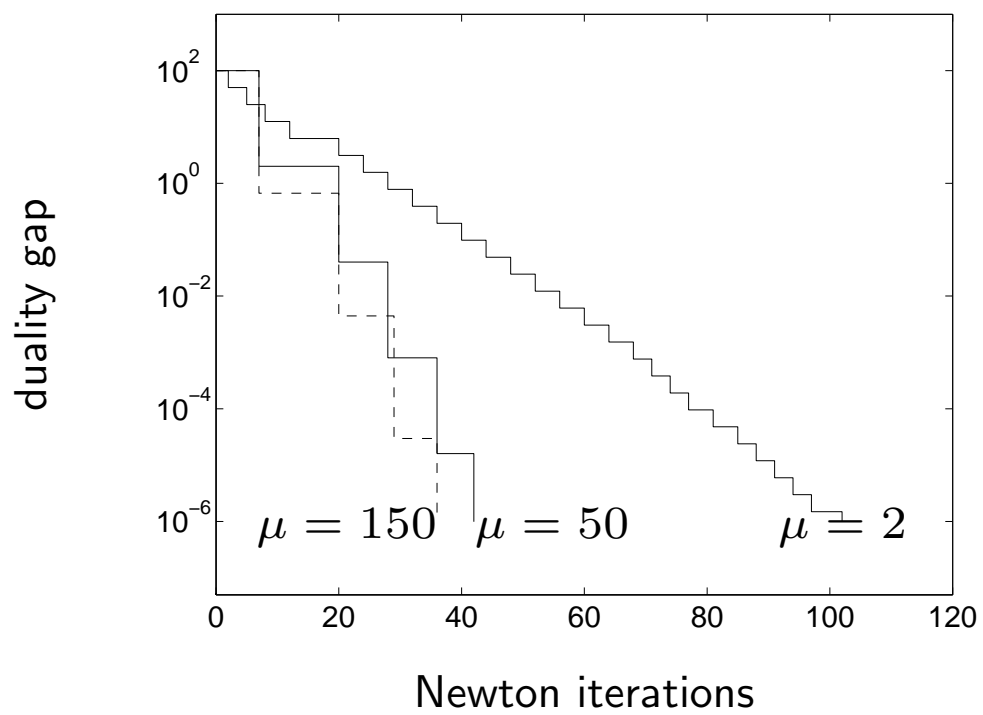
inequality form LP ($m = 100$ inequalities, $n = 50$ variables)



- 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 $\mu \geq 10$

geometric program ($m = 100$ inequalities and $n = 50$ variables)

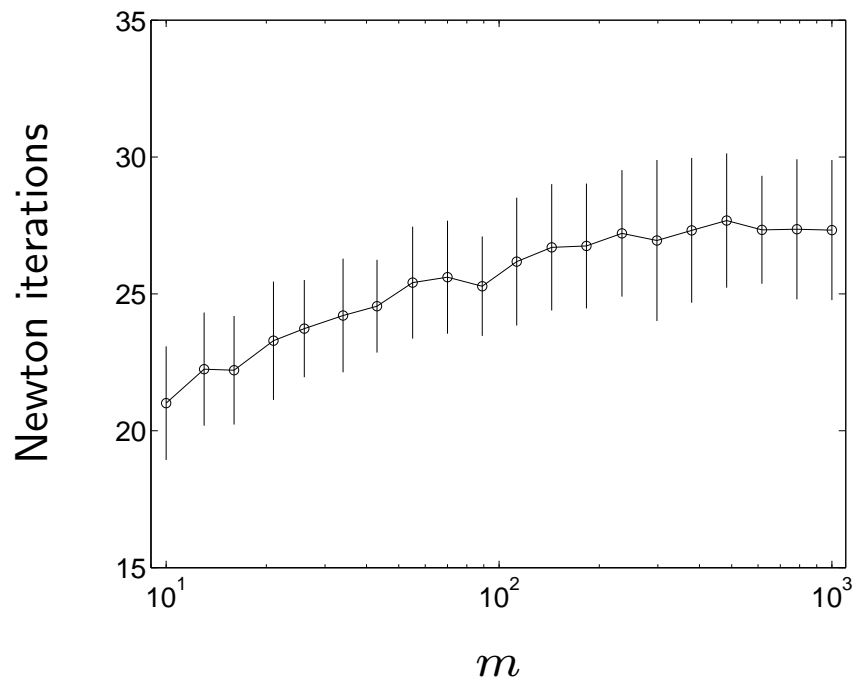
$$\begin{aligned} &\text{minimize} && \log \left(\sum_{k=1}^5 \exp(a_{0k}^T x + b_{0k}) \right) \\ &\text{subject to} && \log \left(\sum_{k=1}^5 \exp(a_{ik}^T x + b_{ik}) \right) \leq 0, \quad i = 1, \dots, m \end{aligned}$$



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

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

$m = 10, \dots, 1000$; 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 methods

feasibility problem: find x such that

$$f_i(x) \leq 0, \quad i = 1, \dots, m, \quad Ax = b \quad (2)$$

phase I: computes strictly feasible starting point for barrier method

basic phase I method

$$\begin{array}{ll} \text{minimize (over } x, s) & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \dots, m \\ & Ax = b \end{array} \quad (3)$$

- 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

Conclusion

convex optimization

- Enables solving large class of practical problems
- Duality plays critical and practical role in theory and algorithms
- Interior point methods can be used to solve convex optimization problems at cost of about 30 least-squares problems, but far more expressive
- lots of topics we didn't cover: convex relaxation, subgradient & cutting plane methods, generalized convexity, multiobjective optimization, ...

References & Software

- Convex Optimization,
Cambridge 2004, Boyd & Vandenberghe
complete text available now (and in future)
at www.stanford.edu/~boyd/cvxbook.html
- Convex Analysis and Optimization,
Athena Scientific 2003, Bertsekas, Nedic & Ozdaglar
- Lectures on Modern Convex Optimization,
SIAM 2001, Ben Tal & Nemirovsky
- YALMIP free solver
Johan Lofberg, ETHZ
- CVX free solver
Michael Grant, Stanford

Next ...

- Real-world Applications of Semidefinite Programming in Systems & Control
Prof. Mario Rotea, Purdue
- Tutorial on Duality-Based Internet Congestion Control Mechanisms
Dr. Maryam Fazel, CalTech
- YALMIP Solver for Convex & Nonconvex Optimization Problems
Dr. Johan Lofberg, ETHZ