12. Interior-point methods

- inequality constrained minimization
- logarithmic barrier function and central path
- barrier method
- feasibility and phase I methods

- generalized inequalities

Inequality constrained minimization

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

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

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

(Slater's condition)

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 $\mathbf{dom}\, f_0 = \mathbf{R}^n_{++}$

min_x   max_{i=1…m} { x' a_i + b_i }

min_{x,t}   t
s.t.   x' a_i + b_i <= t,   i=1…m

- 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

min fo(x)
s.t. f_i(x) <= 0 , i=1…m
Ax=b

reformulation of (1) via indicator function:

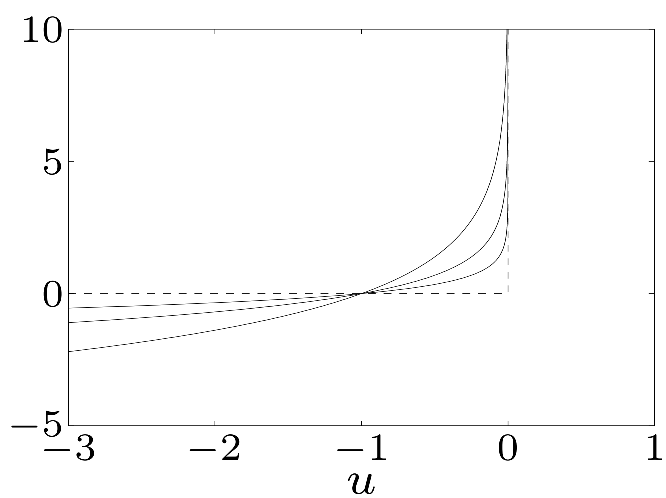
$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

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

approximation via logarithmic barrier

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

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



logarithmic barrier function

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

(Slater's condition)

min fo(x) + 1/t ϕ(x)
s.t. Ax=b

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

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

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

(Useful for KKT analysis and Newton's method)

Central path

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

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

min fo(x) + 1/t ϕ(x)
s.t. Ax=b

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

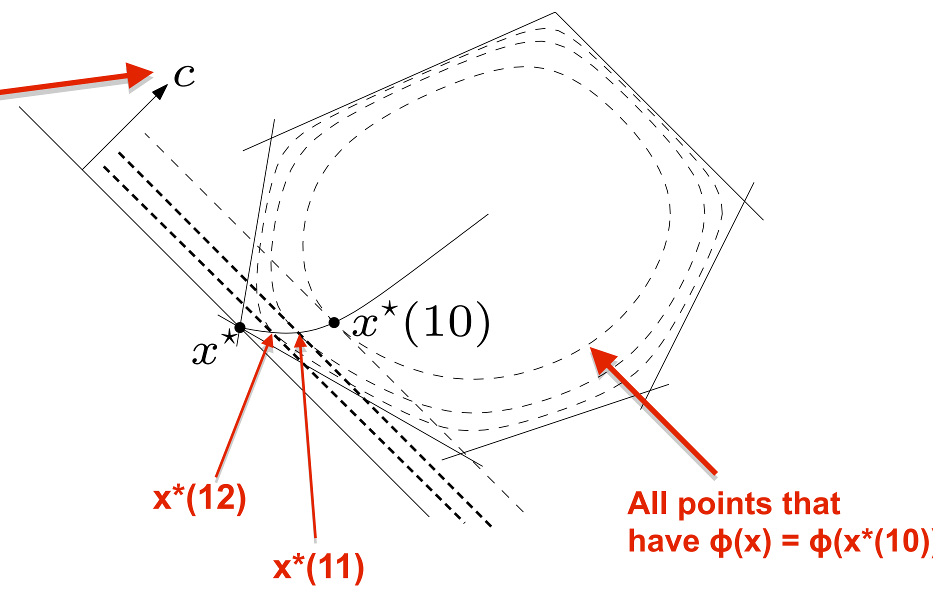
- central path is $\{x^\star(t) \mid t > 0\}$

Also, as t increases, we obtain x*(t) approaches the optimal of the original problem

example: central path for an LP

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

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



Dual points on central path

min t fo(x) + ϕ(x)
s.t. Ax-b=0

L(x,w) = t fo(x) + ϕ(x) + w'(Ax-b)

Stationarity:
dL/dx = t dfo(x) + dϕ(x) + A'w = 0

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

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

(Primal feasibility)

min fo(x)
s.t. f_i(x) <= 0, i=1…m
Ax-b=0

L(x,λ,v) = fo(x) + Σ_i λ_i f_i(x) + v'(Ax-b)

- therefore, $x^\star(t)$ minimizes the Lagrangian

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

Make dL/dx=0 and get same as above

where we define $\lambda_i^\star(t) = 1/(-t f_i(x^\star(t)))$ and $\nu^\star(t) = w/t$

> 0 since t>0 and f_i(x*(t)) < 0

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

$$\begin{aligned} p^\star &\geq g(\lambda^\star(t), \nu^\star(t)) \\ &= L(x^\star(t), \lambda^\star(t), \nu^\star(t)) \end{aligned}$$

… for any (λ,v) so we can plug (λ*(t),v*(t))

= fo(x*(t)) + Σ_i λ*_i(t) f_i(x*(t)) + v*(t)(A x*(t)-b)
= fo(x*(t)) + Σ_i f_i(x*(t)) / (-t f_i(x*(t))) … since A x*(t)=b
= fo(x*(t)) - m/t … m terms

As t -> ∞, m/t -> 0 and then p* = fo(x*(t))

Interpretation via KKT conditions

$x = x^\star(t)$, $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$ 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:

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

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

We said before:
λ_i(t) = 1/(-t f_i(x))

Recall original problem:
min fo(x)
s.t. f_i(x) <= 0, i=1…m
Ax=b

Barrier method

---

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

repeat

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

- terminates with $f_0(x) - p^\star \le \epsilon$ (stopping criterion follows from $f_0(x^\star(t)) - p^\star \le m/t$)
- centering usually done using Newton's method, starting at current x

The gradient at the current x is d = t dfo(x) + dϕ(x)
The Hessian at the current x is H = t d^2fo(x) + d^2ϕ(x)
[H A'] [Δx] = [-d]
[A 0] [v]   [0]

Feasibility and phase I methods

feasibility problem: find x such that

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

phase I: computes strictly feasible starting point for barrier method

basic phase I method

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

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

Generalized inequalities

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

fo : R^n -> R

- f_0 convex, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}$, $i = 1, \ldots, m$, convex with respect to proper cones $K_i \in \mathbf{R}^{k_i}$
- f_i twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible; hence strong duality holds and dual optimum is attained

examples of greatest interest: SOCP, SDP

Similar to log z
which is undefined
for z=0

Generalized logarithm for proper cone

$\psi : \mathbf{R}^q \to \mathbf{R}$ is generalized logarithm for proper cone $K \subseteq \mathbf{R}^q$ if:

- $\mathbf{dom}\,\psi = \mathbf{int}\,K$ and $\nabla^2\psi(y) \prec 0$ for $y \succ_K 0$
- $\psi(sy) = \psi(y) + \theta \log s$ for $y \succ_K 0$, $s > 0$

Example: for positive semidefinite cone:
behaves like strictly concave when matrix
is positive definite

(θ>0 is the degree of ψ)

Take K = {z in R | z >= 0}: ψ(z) = log z
For y > 0, s > 0: ψ(s y) = ψ(y) + θ log s, where θ=1

examples

- nonnegative orthant $K = \mathbf{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$, with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}^n_+$:

ψ(s y) = Σ_{i=1…n} log(s y_i)
= Σ_{i=1…n} { log y_i + log s }
= Σ_{i=1…n} log y_i + n log s
= ψ(y) + n log s

$$\psi(Y) = \log\det Y \qquad (\theta = n)$$

ψ(s Y) = log det (s Y)
= log(s^n det Y)
= log det Y + n log s
= ψ(Y) + n log s

- second-order cone $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \le y_{n+1}\}$:

$$\psi(y) = \log(y_{n+1}^2 - y_1^2 - \cdots - y_n^2) \qquad (\theta = 2)$$

ψ(s y) = log(s^2 (y_{n+1}^2 - y_1^2 … - y_n^2))
= log(y_{n+1}^2 - y_1^2 … - y_n^2) + 2 log s
= ψ(y) + 2 log s

Recall proper cones (2-21):
z >=_K* 0 if and only if
y' z >= 0 for all y >=_K 0

Make z = dψ(y):
dψ(y) >=_K* 0 if and only if
y' dψ(y) >= 0 for all y >=_K 0

Indeed y' dψ(y) = θ > 0

properties

for $y \succ_K 0$,

$$\nabla\psi(y) \succeq_{K^*} 0, \qquad y^T\nabla\psi(y) = \theta$$

(2) ← … → (1)

Recall: ψ(s y) = ψ(y) + θ log s
from left: dψ(s y)/ds = y' dψ(s y)
from right: dψ(s y)/ds = θ/s
thus: y' dψ(s y) = θ/s

take s=1: y' dψ(y) = θ

- nonnegative orthant $\mathbf{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$

$$\nabla\psi(y) = (1/y_1, \ldots, 1/y_n), \qquad y^T\nabla\psi(y) = n$$

y' dψ(y) = 1+…+1

- positive semidefinite cone $\mathbf{S}^n_+$: $\psi(Y) = \log\det Y$

$$\nabla\psi(Y) = Y^{-1}, \qquad \mathbf{tr}(Y\nabla\psi(Y)) = n$$

tr(Y' dψ(Y)) = I

- second-order cone $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1}\}$:

$$\nabla\psi(y) = \frac{2}{y_{n+1}^2 - y_1^2 - \cdots - y_n^2}\begin{bmatrix} -y_1 \\ \vdots \\ -y_n \\ y_{n+1} \end{bmatrix}, \qquad y^T\nabla\psi(y) = 2$$

Logarithmic barrier and central path

logarithmic barrier for $f_1(x) \preceq_{K_1} 0, \ldots, f_m(x) \preceq_{K_m} 0$:

$$\phi(x) = -\sum_{i=1}^{m} \psi_i(-f_i(x)), \qquad \mathbf{dom}\,\phi = \{x \mid f_i(x) \prec_{K_i} 0,\ i = 1, \ldots, m\}$$

- ψ_i is generalized logarithm for K_i, with degree θ_i
- ϕ is convex, twice continuously differentiable

central path: $\{x^\star(t) \mid t > 0\}$ where $x^\star(t)$ solves

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

Dual points on central path

min t fo(x) + ϕ(x)
s.t. Ax-b=0

L(x,w) = t fo(x) + ϕ(x) + w'(Ax-b)

Stationarity:
dL/dx = t dfo(x) + dϕ(x) + A'w = 0

$x = x^\star(t)$ if there exists $w \in \mathbf{R}^p$,

$$t\nabla f_0(x) + \sum_{i=1}^{m} Df_i(x)^T \nabla\psi_i(-f_i(x)) + A^T w = 0$$

min fo(x)
s.t. f_i(x) <=_{K_i} 0, i=1…m
Ax-b=0

L(x,λ,v) = fo(x) + Σ_i λ_i' f_i(x) + v'(Ax-b)

$Df_i(x) \in \mathbf{R}^{k_i \times n}$ is derivative matrix of f_i : R^n -> R^{k_i}

- therefore, $x^\star(t)$ minimizes Lagrangian $L(x, \lambda^\star(t), \nu^\star(t))$, where

$$\lambda_i^\star(t) = \frac{1}{t}\nabla\psi_i(-f_i(x^\star(t))), \qquad \nu^\star(t) = \frac{w}{t}$$

Make dL/dx=0 and get same as above

- from properties of ψ_i: $\lambda_i^\star(t) \succ_{K_i^*} 0$, with duality gap

As t -> ∞, 1/t Σ_i θ_i -> 0 and then p* = fo(x*(t))

$$f_0(x^\star(t)) - g(\lambda^\star(t), \nu^\star(t)) = (1/t)\sum_{i=1}^{m}\theta_i$$

p* >= g(λ*(t),v*(t)) … for any (λ,v) so we can plug (λ*(t),v*(t))
= L(x*(t),λ*(t),v*(t))
= fo(x*(t)) + Σ_i λ*_i(t)' f_i(x*(t)) + v*(t)(A x*(t)-b)
= fo(x*(t)) - 1/t Σ_i y_i' dψ_i(y_i) … since A x*(t)=b, and letting y_i = -f_i(x*(t))
= fo(x*(t)) - 1/t Σ_i θ_i … since y_i' dψ_i(y_i) = θ_i

Barrier method

---

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

repeat

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

- only difference is duality gap m/t on central path is replaced by $\sum_i \theta_i/t$