12. Interior-point methods

- inequality constrained minimization
- logarithmic barrier function and central path
- barrier method
- feasibility and phase I methods
- complexity analysis via self-concordance
- generalized inequalities
Inequality constrained minimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

(1)

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

\[
\tilde{x} \in \text{dom} \ f_0, \quad f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \quad A\tilde{x} = b
\]

hence, strong duality holds and dual optimum is attained
Examples

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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \quad Fx \leq g \\
& \quad Ax = b
\end{align*}
\]

with \( \text{dom } f_0 = \mathbb{R}_{++}^n \)

- differentiability may require reformulating the problem, e.g., piecewise-linear minimization or \( \ell_\infty \)-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:

\[
\begin{align*}
\text{minimize} & \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

where \( I_-(u) = 0 \) if \( u \leq 0 \) and \( I_-(u) = \infty \) otherwise (\( I_- \) is indicator function of \( \mathbb{R}_- \))

Approximation via logarithmic barrier

\[
\begin{align*}
\text{minimize} & \quad f_0(x) - \left(\frac{1}{t}\right) \sum_{i=1}^{m} \log(-f_i(x)) \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

- 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

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

- 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

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

(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

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

hyperplane $c^T x = c^T x^*(t)$ is tangent to level curve of $\phi$ 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 \( \lambda^*_i(t) = 1/(-tf_i(x^*(t))) \) and \( \nu^*(t) = w/t \)

• this confirms the intuitive idea that \( f_0(x^*(t)) \to p^* \) if \( t \to \infty \):

\[
p^* \geq g(\lambda^*(t), \nu^*(t)) \\
= L(x^*(t), \lambda^*(t), \nu^*(t)) \\
= f_0(x^*(t)) - m/t
\]
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: \( \lambda \geq 0 \)

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 \( \lambda_i f_i(x) = 0 \)
Force field interpretation

Centering problem (for problem with no equality constraints)

\[
\text{minimize} \quad t f_0(x) - \sum_{i=1}^{m} \log(-f_i(x))
\]

Force field interpretation

- \( t f_0(x) \) is potential of force field \( F_0(x) = -t \nabla f_0(x) \)
- \(- \log(-f_i(x))\) is potential of force field \( F_i(x) = (1/f_i(x)) \nabla f_i(x) \)
- the forces balance at \( x^*(t) \):

\[
F_0(x^*(t)) + \sum_{i=1}^{m} F_i(x^*(t)) = 0
\]
Example

minimize $c^T x$
subject to $a_i^T x \leq b_i, \quad i = 1, \ldots, m$

- objective force field is constant: $F_0(x) = -tc$
- constraint force field decays as inverse distance to constraint hyperplane:

$$F_i(x) = \frac{-a_i}{b_i - a_i^T x}, \quad \|F_i(x)\|_2 = \frac{1}{d(x, \mathcal{H}_i)}$$

where $d(x, \mathcal{H}_i)$ is distance of $x$ to hyperplane $\mathcal{H}_i = \{x \mid a_i^T x = b_i\}$

$t = 1$

$t = 3$
Barrier method

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

repeat
  1. centering step: compute \( x^*(t) \) by minimizing \( t f_0 + \phi \), 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 \( f_0(x) - p^* \leq m/t < \epsilon \)

• centering usually done using Newton’s method, starting at current \( x \)

• choice of \( \mu \) involves a trade-off: large \( \mu \) means fewer outer iterations, more inner (Newton) iterations; typical values: \( \mu = 10–20 \)

• several heuristics for choice of \( t^{(0)} \)
Convergence analysis

Number of outer (centering) iterations: exactly

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

plus the initial centering step (to compute \(x^*(t^{(0)})\))

Centering problem

\[
\text{minimize } tf_0(x) + \phi(x)
\]

see convergence analysis of Newton’s method

- \(tf_0 + \phi\) must have closed sublevel sets for \(t \geq t^{(0)}\)
- classical analysis requires strong convexity, Lipschitz condition
- analysis via self-concordance requires self-concordance of \(tf_0 + \phi\)
Examples

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

- starts with \(x\) on central path \((t^{(0)} = 1, \text{ duality gap } 100)\)
- 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)

minimize $\log(\sum_{k=1}^{5} \exp(a_{0k}^T x + b_{0k}))$

subject to $\log(\sum_{k=1}^{5} \exp(a_{ik}^T x + b_{ik})) \leq 0$, $i = 1, \ldots, m$
Family of standard LPs \((A \in \mathbb{R}^{m \times 2m})\)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b, \quad x \geq 0
\end{align*}
\]

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

![Graph showing the number of Newton iterations](graph.png)

The 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

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

Phase I: computes strictly feasible starting point for barrier method

Basic phase I method

minimize (over $x$, $s$) $\quad s$
subject to $\quad f_i(x) \leq s, \quad i = 1, \ldots, m$
$Ax = b$  \hfill (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
Sum of infeasibilities phase I method

\[
\begin{align*}
\text{minimize} & \quad 1^T s \\
\text{subject to} & \quad s \geq 0, \quad f_i(x) \leq s_i, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

for infeasible problems, produces a solution that satisfies many more inequalities than basic phase I method

Example (infeasible set of 100 linear inequalities in 50 variables)

left: basic phase I solution; satisfies 39 inequalities  
right: sum of infeasibilities phase I solution; satisfies 79 inequalities
Example: family of linear inequalities $Ax \leq b + \gamma \Delta b$

- data chosen to be strictly feasible for $\gamma > 0$, infeasible for $\gamma \leq 0$
- use basic phase I, terminate when $s < 0$ or dual objective is positive

Number of iterations roughly proportional to $\log(1/|\gamma|)$
Complexity analysis via self-concordance

same assumptions as on page 12.2, plus:

• sublevel sets (of $f_0$, on the feasible set) are bounded
• $t f_0 + \phi$ is self-concordant with closed sublevel sets

second condition

• holds for LP, QP, QCQP
• may require reformulating the problem, e.g.,

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \quad Fx \leq g
\end{align*}
\]

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & \quad Fx \leq g, \quad x \geq 0
\end{align*}
\]

• needed for complexity analysis; barrier method works even when self-concordance assumption does not apply
Newton iterations per centering step: from self-concordance theory

\[
\text{#Newton iterations} \leq \frac{\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)}{\gamma} + c
\]

- bound on effort of computing \( x^+ = x^*(\mu t) \) starting at \( x = x^*(t) \)
- \( \gamma, c \) are constants (depend only on Newton algorithm parameters)
- from duality (with \( \lambda = \lambda^*(t), \nu = \nu^*(t) \)):

\[
\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)
= \mu t f_0(x) - \mu t f_0(x^+) + \sum_{i=1}^{m} \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu
\]

\[
\leq \mu t f_0(x) - \mu t f_0(x^+) - \mu t \sum_{i=1}^{m} \lambda_i f_i(x^+) - m - m \log \mu
\]

\[
\leq \mu t f_0(x) - \mu t g(\lambda, \nu) - m - m \log \mu
\]

\[
= m(\mu - 1 - \log \mu)
\]
Total number of Newton iterations (excluding first centering step)

\[ \#\text{Newton iterations} \leq N = \left\lfloor \frac{\log(m/(t^{(0)}\varepsilon))}{\log \mu} \right\rfloor \left( \frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right) \]

figure shows \( N \) for typical values of \( \gamma, c \),

\[ m = 100, \quad \frac{m}{t^{(0)}\varepsilon} = 10^5 \]

- confirms trade-off in choice of \( \mu \)
- in practice, \#iterations is in the tens; not very sensitive for \( \mu \geq 10 \)
Polynomial-time complexity of barrier method

- for $\mu = 1 + 1/\sqrt{m}$:

$$N = O \left( \sqrt{m} \log \left( \frac{m/t^{(0)}}{\epsilon} \right) \right)$$

- number of Newton iterations for fixed gap reduction is $O(\sqrt{m})$

- multiply with cost of one Newton iteration (a polynomial function of problem dimensions), to get bound on number of flops

this choice of $\mu$ optimizes worst-case complexity; in practice we choose $\mu$ fixed ($\mu = 10, \ldots, 20$)
Generalized inequalities

minimize $f_0(x)$
subject to $f_i(x) \leq_{K_i} 0$, $i = 1, \ldots, m$
$Ax = b$

- $f_0$ convex, $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$, $i = 1, \ldots, m$, convex with respect to proper cones $K_i \in \mathbb{R}^{k_i}$
- $f_i$ twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with rank $A = p$
- we assume $p^*$ 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
Generalized logarithm for proper cone

\( \psi : \mathbb{R}^q \to \mathbb{R} \) is generalized logarithm for proper cone \( K \subseteq \mathbb{R}^q \) if:

- \( \text{dom} \psi = \text{int} K \) and \( \nabla^2 \psi(y) < 0 \) for \( y >_K 0 \)
- \( \psi(sy) = \psi(y) + \theta \log s \) for \( y >_K 0, \ s > 0 \) (\( \theta \) is the degree of \( \psi \))

Examples

- nonnegative orthant \( K = \mathbb{R}^n_+ \): \( \psi(y) = \sum_{i=1}^n \log y_i \), with degree \( \theta = n \)
- positive semidefinite cone \( K = S^n_+ \):
  \[
  \psi(Y) = \log \det Y \quad (\theta = n)
  \]
- second-order cone \( K = \{ y \in \mathbb{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1} \} \):
  \[
  \psi(y) = \log(y_{n+1}^2 - y_1^2 - \cdots - y_n^2) \quad (\theta = 2)
  \]
Properties (without proof): for $y >_K 0$,

\[ \nabla \psi(y) \preceq_K 0, \quad y^T \nabla \psi(y) = \theta \]

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

\[ \nabla \psi(y) = (1/y_1, \ldots, 1/y_n), \quad y^T \nabla \psi(y) = n \]

- positive semidefinite cone $\mathbb{S}_+^n$: $\psi(Y) = \log \det Y$

\[ \nabla \psi(Y) = Y^{-1}, \quad \text{tr}(Y \nabla \psi(Y)) = n \]

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

\[ \nabla \psi(y) = \frac{2}{y_{n+1}^2 - y_1^2 - \cdots - y_n^2} \begin{bmatrix} -y_1 \\ \vdots \\ -y_n \\ -y_{n+1} \end{bmatrix}, \quad y^T \nabla \psi(y) = 2 \]
Logarithmic barrier and central path

Logarithmic barrier for \( f_1(x) \leq K_1 0, \ldots, f_m(x) \leq K_m 0 \):

\[
\phi(x) = -\sum_{i=1}^{m} \psi_i(-f_i(x)), \quad \text{dom } \phi = \{x \mid f_i(x) < K_i 0, \ i = 1, \ldots, m\}
\]

- \( \psi_i \) is generalized logarithm for \( K_i \), with degree \( \theta_i \)
- \( \phi \) is convex, twice continuously differentiable

Central path: \( \{x^*(t) \mid t > 0\} \) where \( x^*(t) \) solves

\[
\begin{align*}
\text{minimize} & \quad tf_0(x) + \phi(x) \\
\text{subject to} & \quad Ax = b
\end{align*}
\]
Dual points on central path

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

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

($D f_i(x) \in \mathbb{R}^{k_i \times n}$ is derivative matrix of $f_i$)

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

$$
\lambda^*_i(t) = \frac{1}{t} \nabla \psi_i(-f_i(x^*(t))), \quad \nu^*(t) = \frac{w}{t}
$$

- from properties of $\psi_i$: $\lambda^*_i(t) > K_i^* 0$, with duality gap

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

minimize $c^T x$
subject to $F(x) = \sum_{i=1}^{n} x_i F_i + G \leq 0$

with $F_i \in S^p$

- logarithmic barrier: $\phi(x) = \log \det(-F(x)^{-1})$
- central path: $x^*(t)$ minimizes $tc^T x - \log \det(-F(x))$; hence

$$tc_i - \text{tr}(F_i F(x^*(t))^{-1}) = 0, \quad i = 1, \ldots, n$$

- dual point on central path: $Z^*(t) = -(1/t)F(x^*(t))^{-1}$ is feasible for

$$\text{maximize} \quad \text{tr}(GZ)$$
$$\text{subject to} \quad \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n$$
$$Z \succeq 0$$

- duality gap on central path: $c^T x^*(t) - \text{tr}(GZ^*(t)) = p/t$

Interior-point methods
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 \( (\sum_i \theta_i)/t < \epsilon \)
4. increase \( t \): \( t := \mu t \)

- only difference is duality gap \( m/t \) on central path is replaced by \( \sum_i \theta_i/t \)
- number of outer iterations:
  \[
  \left\lceil \frac{\log((\sum_i \theta_i)/(\epsilon t^{(0)}))}{\log \mu} \right\rceil
  \]
- complexity analysis via self-concordance applies to SDP, SOCP
Examples

**Second-order cone program** (50 variables, 50 SOC constraints in $\mathbb{R}^6$

- \( \mu = 50 \) \( \mu = 200 \) \( \mu = 2 \)

**Semidefinite program** (100 variables, LMI constraint in $S^{100}$)

- \( \mu = 150 \) \( \mu = 50 \) \( \mu = 2 \)
Family of SDPs \((A \in S^n, x \in \mathbb{R}^n)\)

\[
\begin{align*}
\text{minimize} & \quad 1^T x \\
\text{subject to} & \quad A + \text{diag}(x) \succeq 0
\end{align*}
\]

\(n = 10, \ldots, 1000\), for each \(n\) solve 100 randomly generated instances

![Graph showing Newton iterations vs. \(n\)]
Primal-dual interior-point methods

more efficient than barrier method when high accuracy is needed

- update primal and dual variables at each iteration; no distinction between inner and outer iterations
- often exhibit superlinear asymptotic convergence
- search directions can be interpreted as Newton directions for modified KKT conditions
- can start at infeasible points
- cost per iteration same as barrier method