Non-linear Optimization (2DME20), lecture 7

Gerhard Woeginger

Technische Universiteit Eindhoven

Fall 2015, Q1

Program for this week

- Self-concordant functions
- Generalized logarithms
- The Newton method for self-concordant functions
- The barrier method for self-concordant functions

- Second-order cone optimization
- Semi-definite programming

Recall, recall (1): the logarithmic barrier

- Indicator function $I_-(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \infty & \text{otherwise} \end{cases}$
- Approximate indicator function $\widehat{I}_-(u) = \begin{cases} -\frac{1}{t}\log(-u) & \text{if } u < 0 \\ \infty & \text{otherwise} \end{cases}$
- $\widehat{I}_-(u)$: convex; non-decreasing; differentiable; closed sublevel sets
- The problem $\min\{f_0(x) \mid f_i(x) \leq 0, i = 1\ldots, r,\ Ax = b\}$ can be approximated by $\min\{f_0(x) + \frac{1}{t}\phi(x) \mid Ax = b\}$

- Logarithmic barrier $\phi(x) := \sum_{i=1}^{r} -\log(-f_i(x))$

- Central path $x^*(t) := \arg\min\{f_0(x) + \frac{1}{t}\phi(x) \mid Ax = b\}$
- As $t \to \infty$, the central path leads to the optimum

Recall, recall (2): the barrier method

In order to approximate $p^* = \min\{f_0(x) \mid f_i(x) \leq 0, i = 1 \ldots, r,\ Ax = b\}$ within error $\varepsilon > 0$, it suffices to solve the unconstrained problem

$$y^*(t) \;=\; \arg\min\{f_0(Wy+v) + \frac{1}{t}\phi(Wy+v)\}$$

with $t = r/\varepsilon$.

The barrier method

Given a strictly feasible $y = y^{(0)}$ and $t = t^{(0)} > 0$, do

1. compute $y^*(t)$, starting the solution algorithm with y (centering)
2. put $y \leftarrow y^*(t)$ (update)
3. if $r/t < \varepsilon$ then quit; else put $t \leftarrow \mu t$ and repeat (increase)

The centering step uses the Newton method

Recall, recall (3): the barrier method

Example: Linear programming

$\min\{c^T x \mid a_i^T x \leq b_i,\ i = 1, \dots, r\}$

The logarithmic barrier for this problem is

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

Then

$$tf_0(x) + \phi(x) \;=\; tc^T x + \sum_{i} -\log(b_i - a_i^T x)$$

Self-concordant functions (1)

Definition

A convex function $f : \mathbb{R} \to \mathbb{R}$ is self-concordant if

$$|f'''(x)| \;\leq\; 2f''(x)^{3/2}$$

for all $x \in \mathrm{Dom}(f)$.

Note: This self-concordance condition is equivalent to (and motivated by)

$$\left| \frac{d}{dt} \left(f''(t)^{-1/2} \right) \right| \;\leq\; 1$$

Example

- $f(x) = -\log x$
- $f(x) = x \log x - \log x$

Self-concordant functions (2)

Self-concordance is affine-invariant:

Lemma

If f is self-concordant, then so is $g(y) := f(ay + b)$ for fixed $a, b \in \mathbb{R}$.

Definition

A convex function $f : \mathbb{R}^n \to \mathbb{R}$ is self-concordant
if $g(t) := f(x + tv)$ is self-concordant for all $x, v \in \mathbb{R}^n$.

Self-concordant functions (3)

Lemma

- *If f is self-concordant and $a \geq 1$, then af is self-concordant.*
- *If f_1, f_2 are self-concordant, then $f_1 + f_2$ is self-concordant.*
- *If f is self-concordant, then $f(Ax + b)$ is self-concordant.*

Example

- $f(x) := \sum_i -\log(b_i - a_i^T x)$ is self-concordant
- $f(X) := -\log\det X$ is self-concordant on S^n_{++}
- $f(x) := -\log(x^T P x + q^T x + r)$ is self-concordant if $P \preceq 0$

Generalized logarithms (1)

Let $K \subseteq \mathbb{R}^q$ be a proper cone.

Definition

A function $\psi : \mathbb{R}^q \to \mathbb{R}$ is a generalized logarithm for the cone K, if

- ψ is concave and twice continuously differentiable
- ψ has closed sublevel sets
- $\mathrm{Dom}(\psi) = \mathrm{int}\, K$, and $\nabla^2\psi(y) \prec 0$ for $y \in \mathrm{int}\, K$
- There exists a constant $\theta > 0$, so that for all $y \succ_K 0$ and all $s > 0$ $\psi(sy) = \psi(y) + \theta \log s$

The constant θ is called the degree of ψ.

Example

The function $\psi : x \mapsto \sum_{i=1}^{r} \log(x_i)$ is a generalized logarithm of degree r for the non-negative orthant cone $\mathbb{R}^r_+$.

Generalized logarithms (2)

Example

For the semi-definite cone S_+^q we have the generalized logarithm

$$\psi(X) := \log\det X$$

- The degree of ψ is q, since for all $s > 0$

$$\log\det(sX) \;=\; \log\det X + q\log s$$

- (Boyd & Vandenberghe, pp 641–642)
  The gradient of ψ at $X \in S_{++}^n$ is

$$\nabla\psi(X) = X^{-1}$$

- Moreover, $-\psi$ is self-concordant.

Generalized logarithms (3)

Lorentz cone $L^{q+1} = \{(x,t) \in \mathbb{R}^{q+1} \mid \|x\| \le t,\ x \in \mathbb{R}^q, t \in \mathbb{R}\}$

Example

For the second-order cone L^{q+1} we have the generalized logarithm

$$\psi(x,t) = \log(t^2 - \|x\|^2)$$

- The degree of ψ is 2, since for all $s > 0$

$$\psi(sx, st) \ = \ \log(s^2(t^2 - \|x\|^2)) \ = \ \psi(x,t) + 2\log s$$

- The gradient of ψ at (x,t) is

$$\nabla_{x,t}\psi(x,t) = \frac{2}{t^2 - \|x\|^2}(-x_1, \ldots, -x_n, t)$$

- Moreover, $-\psi$ is self-concordant.

The Newton method on self-concordant functions (1)

Unconstrained scenario:
Given a convex, self-concordant $f : \mathbb{R}^n \to \mathbb{R}$,
we want to determine the value $p^* = \min_x f(x)$

Definition

The Newton decrement at x is $\lambda(x) := (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$

Lemma (Boyd & Vandenberghe, pp 501–502)

If f is self-concordant and $\lambda(x) < 0.68$, then $p^* \geq f(x) - \lambda(x)^2$.

The Newton method on self-concordant functions (2)

The Newton Algorithm

Given: a precision bound $\varepsilon > 0$, and a starting point $x^{(0)}$

Initialize $k = 0$

Repeat

1. Determine $\Delta x^{(k)} = -(\nabla^2 f(x^{(k)}))^{-1} \nabla f(x^{(k)})$
2. Determine $t^{(k)}$ by backtracking line search
3. Put $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$ and $k \leftarrow k + 1$.

until $\lambda(x)^2 \leq \varepsilon$

So the Newton algorithm terminates with an ε-approximation of p^*.

The Newton method on self-concordant functions (3)

Theorem (Boyd & Vandenberghe, pp 503–505)

There exist real numbers η, γ with $0 < \eta \leq 1/4$ and $0 < \gamma$,
so that the Newton search direction Δx,
the value t found by backtracking line search,
and the new point $x^+ = x + t\Delta x$
satisfy the following statements:

- If $\lambda(x) > \eta$, then $f(x^+) < f(x) - \gamma$
- If $\lambda(x) < \eta$, then $2\lambda(x^+) \leq (2\lambda(x))^2$

Consequently, the Newton algorithm terminates after at most

$$\frac{1}{\gamma}\left(f(x^{(0)}) - p^*\right) + \log_2 \log_2(1/\epsilon) \quad \text{iterations}$$

Barrier method for self-concordant functions (1)

Minimization scenario under constraints:

$\min\{f_0(x) \mid f_i(x) \le 0, i = 1\ldots, r,\ Ax = b\}$

Our next goal is to analyze the performance of the Newton method in the centering step of the barrier method

Assumptions

1. the function $tf_0 + \phi$ is self-concordant for every $t \ge t^{(0)}$
2. the sublevel sets $\{x \in \mathbb{R}^n \mid f_0(x) \le u,\ f_i(x) \le 0,\ Ax = b\}$ are bounded for any u

Barrier method for self-concordant functions (2)

In the centering step,
we move from point x (for value t) to point x^+ (for value μt),
and apply Newton to solve $x^+ = \arg\min\{\mu t f(x) + \phi(x) \mid Ax = b\}$

The number of iterations in the centering step is bounded by

$$\frac{1}{\gamma}\left(tf_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)\right) + \log_2\log_2(1/\epsilon)$$

Lemma (Boyd & Vandenberghe, pp 590–592)

$$tf_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+) \;\le\; r(\mu - 1 - \log(\mu))$$

By choosing $\mu = 1 + 1/\sqrt{r}$,
the number of Newton iterations can be bounded by $O(\sqrt{r})$

Barrier method for self-concordant functions (3)

The barrier method also applies to problems

$$\min\{f_0(x) \mid f_i(x) \preceq_{K_i} 0,\ Ax = b\}$$

with generalized inequalities, by using

$$\phi(x) := -\sum \psi_i(-f_i(x))$$

as barrier, where ψ_i is a generalized logarithm for K_i of degree θ_i.

- As before, we put $x^*(t) = \arg\min\{tf_0(x) + \phi(x) \mid Ax = b\}$
- For $\nu^*(t)$ with $\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\nu^*(t) = 0$ and for $\lambda_i^*(t) := \frac{1}{t}\nabla\psi_i(-f_i(x^*(t)))$ we then get

$$g(\lambda^*(t), \nu^*(t)) = f_0(x^*(t)) - \frac{1}{t}\sum_i \theta_i$$

- As before, the central path leads to the optimum

Homework 8

Read the following sections in Boyd & Vandenberghe
9.1, 9.2, 9.5, 9.6; 11.1–11.6

Recommended exercises (Boyd & Vandenberghe):
4.40, 4.43; 9.2, 9.13, 9.15; 11.3, 11.4, 11.5. 11.6, 11.15, 11.16

Attention!
Weeks 6-7 (Oct 6; Oct 9; Oct 13; Oct 16):

- all lectures and instructions in flux 1.06

- tuesday: 1+2 instructions; 3+4 lecture
- friday: 5+6 lecture

Second-order cone optimization (SOCO)

- Lorentz cone $L^{n+1} = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \le t,\ x \in \mathbb{R}^n, t \in \mathbb{R}\}$

Second-order cone optimization

For matrices $A_1, \ldots, A_r$, vectors $b_1, \ldots, b_r, c_1, \ldots, c_r$, reals $d_1, \ldots, d_r$, for matrix F, and vectors f, g solve

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \le c_i^T x + d_i \qquad i = 1, \ldots, r \\ & Fx = g \\ & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

SOCO (1): Rotated second-order cone

Definition

The rotated second-order cone in $\mathbb{R}^{n+2}$ is the set

$$L_r^{n+2} := \{(x, y, z) \mid x \in \mathbb{R}^n,\ y, z \in \mathbb{R},\ x^T x \le 2yz,\ y, z \ge 0\}$$

- Note that $\|x\|^2 \le 2yz$ and $y, z \ge 0$ is equivalent to

$$\left\| \begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y - z) \end{bmatrix} \right\| \ \le\ \frac{1}{\sqrt{2}}(y + z)$$

- In other words,
  $(x, y, z) \in L_r^{n+2}$ if and only if $(x, (y - z)/\sqrt{2}, (y + z)/\sqrt{2}) \in L^{n+2}$

- These two sets are related by a rotation matrix R:

$$\begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y - z) \\ \frac{1}{\sqrt{2}}(y + z) \end{bmatrix} = R \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{with} \quad R = \begin{bmatrix} I_n & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

SOCO (2): Linear and convex quadratic constraints

For $s \in \mathbb{R}^n$ and $t \in \mathbb{R}$,
consider the linear constraint $s^T x \leq t$

For $A = 0$ and $b = 0$, this constraint is equivalent to
$\|Ax + b\| \leq t - s^T x$

For $Q \in S^n_+$, $c \in \mathbb{R}^n$ and $t \in \mathbb{R}$,
consider the convex quadratic constraint $x^T Qx + c^T x \leq t$

For $Q = P^T P$, this constraint is equivalent to
$w = Px, \;\; y = t - c^T x, \;\; z = 1/2, \;$ and $\; w^T w \leq 2yz$

Observation

Linear programs & convex quadratic programs are special cases of SOCO.

SOCO (3): Quadratic-over-linear

Example: Quadratic-over-linear problem

For matrices $A_1, \ldots, A_r$, for vectors $b_1, \ldots, b_r, c_1, \ldots, c_r$ and for numbers $d_1, \ldots, d_r \in \mathbb{R}$, consider the problem

$$\min\{ \sum_{i=1}^{r} \frac{\|A_i x - b_i\|^2}{c_i^T x + d_i} \mid c_i^T x + d_i > 0 \text{ for } i = 1, \ldots, r \}$$

$$\begin{array}{lll} \text{minimize} & \sum_{i=1}^{r} t_i & \\ \text{subject to} & \|A_i x - b_i\|^2 \leq (c_i^T x + d_i) t_i & i = 1, \ldots, r \\ & c_i^T x + d_i > 0 & i = 1, \ldots, r \end{array}$$

$$\begin{array}{lll} \text{minimize} & \sum_{i=1}^{r} t_i & \\ \text{subject to} & \left\| \begin{bmatrix} 2(A_i x - b_i) \\ c_i^T x + d_i - t_i \end{bmatrix} \right\| \leq c_i^T x + d_i + t_i & i = 1, \ldots, r \\ & c_i^T x + d_i > 0, \quad t_i \geq 0 & i = 1, \ldots, r \end{array}$$

SOCO (4)

Example

For $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$,
determine $\min\{a^T x + bt \mid \|x\| \le t,\ x \in \mathbb{R}^n,\ t \in \mathbb{R}\}$.

- If $\|a\| \le b$, then $a^T x + bt \ \ge\ a^T x + \|a\|\|x\| \ \ge\ 0$
  Hence opt $= 0$, by choosing $x = 0$ and $t = 0$

- If $\|a\| > b$, then choose $x = -\gamma a$ and $t = \|x\|$ for $\gamma \in \mathbb{R}_+$.
  Then $a^T x + bt = -\gamma a^T a + b\gamma\|a\| \ = \ \gamma\|a\|(-\|a\| + b)$.
  As $\gamma \to \infty$, the objective value goes to $-\infty$.
  Hence opt $= -\infty$.

Exercise

For $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, determine $\min\{a^T x + b\|x\| \mid x \in \mathbb{R}^n\}$.

Facility location

Example

A group of n remote communities wants to build a central warehouse. The location of the ith community is given by coordinates (a_i, b_i). Goods will be delivered by plane from the warehouse to the communities, and the ith community needs d_i deliveries per month.

The goal is to locate the warehouse so that the total travel distance of all deliveries is minimized.

In other words: find a point $x = (a_x, b_x)$ that
minimizes the objective value $\sum_i d_i \|(a_x, b_x) - (a_i, b_i)\|$

With real variables a_x, b_x and z_i for $i = 1, \ldots, n$, this becomes

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} d_i z_i \\ \text{subject to} & \|(a_x, b_x) - (a_i, b_i)\| \leq z_i \quad i = 1, \ldots, n \end{array}$$

Semi-definite programming (SDP)

Semi-definite programming

For matrices $A_0, \ldots, A_n$ and vector c, solve

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0 \\ & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

Recall, recall, recall (1)

Recall that $A \in S^n$ is positive semi-definite
if and only if each eigenvalue of A is ≥ 0
if and only if there is some real matrix Z such that $A = Z^T Z$
if and only if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$

Recall that $A \in S^n$ is positive semi-definite,
if and only if $U^T A U$ with non-singular U is positive semi-definite

Recall that $A, B \in S^n$ are both positive semi-definite,
if and only if $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is positive semi-definite

Recall, recall, recall (2)

For a matrix $A \in S^n$ and an integer k with $1 \leq k \leq n$,
a principal minor of order k is the determinant of a submatrix of A obtained by considering a k-element subset $J \subseteq \{1, \ldots, n\}$ of its rows and columns.

Lemma

$A \succeq 0$ if and only if for every k,
the sum $S_k(A)$ of the principal minors of order k is non-negative.

Example

$$\begin{bmatrix} 1+x & x-y & x \\ x-y & 1-y & 0 \\ x & 0 & 1+y \end{bmatrix}$$ is positive semi-definite, if and only if

$$\begin{aligned} S_1(A) &= \operatorname{tr}(A) = x+3 \geq 0 \\ S_2(A) &= 3+2x+2xy-2x^2-2y^2 \geq 0 \\ S_3(A) &= \det(A) = 1+x-2x^2-2y^2+2xy+xy^2-y^3 \geq 0 \end{aligned}$$

SDP (1a): Linear programs

For matrix P and vectors q, r
consider the linear program $\min\{q^T x \mid Px \le r\}$

The ordinary affine inequalities $p_i^T x \le r_i$ for $i = 1, \ldots, m$
can be cast as a single semi-definite constraint $A(x) \succeq 0$

$$
\begin{aligned}
A(x) &= \operatorname{diag}(r_1 - p_1^T x,\ \ldots,\ r_m - p_m^T x) \\
&= \begin{bmatrix} r_1 - p_1^T x & & \\ & \ddots & \\ & & r_m - p_m^T x \end{bmatrix} \succeq 0
\end{aligned}
$$

SDP (1b): Second-order cone constraints

Lemma

For a real number t, the second-order cone constraint $\|x\| \le t$ is equivalent to

$$\begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0$$

$$t \cdot \begin{bmatrix} y \\ s \end{bmatrix}^T \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \begin{bmatrix} y \\ s \end{bmatrix} = \|sx + ty\|^2 + s^2(t^2 - x^T x)$$

- If $\|x\| \le t$, then right hand side ≥ 0
- If $\|x\| > t$, then pick $y = x$ and $s = -t$

SDP (2): Eigenvalues

Example

For $A \in S^n$, show that $\min\{t \mid tI_n - A \succeq 0\} = \lambda_{\max}(A)$.

$$
\begin{aligned}
L(t, Z) &= t - \mathrm{tr}(Z(tI_n - A)) = t(1 - \mathrm{tr}Z) + \mathrm{tr}(ZA) \\
g(Z) &= \begin{cases} \mathrm{tr}(ZA) & \text{if } \mathrm{tr}Z = 1 \\ -\infty & \text{otherwise} \end{cases}
\end{aligned}
$$

Lagrange dual problem: $\max\{\mathrm{tr}(ZA) \mid \mathrm{tr}Z = 1,\ Z \succeq 0\}$

Eigenvalue shift rule

For $A \in S^n$ and $t \in \mathbb{R}$, $\lambda_i(A + tI_n) = \lambda_i(A) + t$ for $i = 1, \ldots, n$

- $\lambda_{\max}(A) = \min\{t : \ tI_n \succeq A\}$
- $\lambda_{\min}(A) = \max\{t : \ A \succeq tI_n\}$

SDP (3a): Schur complement

Lemma (Schur complement)

For $A \in S^n_{++}$, $X \in \mathbb{R}^{n \times k}$ and $Y \in S^k$,

$$Y - X^T A^{-1} X \succeq 0 \quad \text{if and only if} \quad \begin{bmatrix} A & X \\ X^T & Y \end{bmatrix} \succeq 0$$

(Note: left hand side quadratic in X; right hand side linear in X)

$$\begin{bmatrix} A & 0 \\ 0 & Y - X^T A^{-1} X \end{bmatrix} =$$

$$= \begin{bmatrix} I & 0 \\ -X^T A^{-1} & I \end{bmatrix} \begin{bmatrix} A & X \\ X^T & Y \end{bmatrix} \begin{bmatrix} I & -A^{-1}X \\ 0 & I \end{bmatrix}$$

- $Y - X^T A^{-1} X$ is the Schur complement of A in $\begin{bmatrix} A & X \\ X^T & Y \end{bmatrix}$

SDP (3b): Schur complement

For matrices $X, Y \in S^n$,

the constraint $X^T X \preceq Y$ is equivalent to $\begin{bmatrix} I_n & X \\ X^T & Y \end{bmatrix} \succeq 0$

For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$,

the constraint $x^T x \leq y$ is equivalent to $\begin{bmatrix} I_n & x \\ x^T & y \end{bmatrix} \succeq 0$

For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$,

the constraint $x^T x \leq y^2$ is equivalent to $\begin{bmatrix} yI_n & x \\ x^T & y \end{bmatrix} \succeq 0$

SDP (4): Spectral norm

Definition

For a matrix A, the spectral norm is $\|A\| := \sqrt{\lambda_{\max}(A^T A)}$.

Lemma

Let $A \in \mathbb{R}^{n\times m}$ be a matrix and let $t > 0$.

Then $\|A\| \le t$ if and only if $\begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0$.

$$\begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0 \quad \text{if and only if} \quad tI - A^T(\tfrac{1}{t}I)A \succeq 0$$

$$\text{if and only if} \quad t^2 I \succeq A^T A$$

SDP (5): Finsler's lemma

Finsler's lemma

For $A \in S^n$ and $B \in \mathbb{R}^{m \times n}$, the following statements are equivalent:

1. $x^T A x > 0$ for all x with $Bx = 0$ and $x \neq 0$
2. $B_\perp^T A B_\perp \succ 0$, where $B_\perp$ is a matrix of maximum rank with $BB_\perp = 0$ (that is, $B_\perp$ contains by columns a basis for the null space of B)
3. There exists $Y \in \mathbb{R}^{m \times n}$, such that $A + Y^T B + B^T Y \succ 0$

Minimizing a uni-variate polynomial (1)

We will see how to find the global minimum of a
polynomial $p \in \mathbb{R}[x]$ by means of semi-definite optimization

Example

A quadratic polynomial $p(x) = ax^2 + bx + c$
satisfies $p(x) \geq 0$ for all $x \in \mathbb{R}$ (in other words: p is PSD)
if and only if $a \geq 0$ and $b^2 - 4ac \leq 0$.

These conditions hold, if and only if $\begin{bmatrix} c & b/2 \\ b/2 & a \end{bmatrix} \succeq 0$

Similarly, one sees that $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \succeq 0$ implies that the polynomial
$p(x) = a + (b+d)x + (c+e+g)x^2 + (f+h)x^3 + ix^4$ is PSD.

Minimizing a uni-variate polynomial (2)

Theorem #1

Let $p \in \mathbb{R}[x]$. Then p is positive semi-definite, if and only if there exist two polynomials $q, r \in \mathbb{R}[x]$ with $p = q^2 + r^2$.

Theorem #2

Let $d \geq 1$ be an integer, and let $c_0, \ldots, c_{2d} \in \mathbb{R}$.
Then the polynomial $p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{2d} x^{2d}$ is the sum of squares of several polynomials, if and only if there exists $A \in S_+^{d+1}$ such that $c_k = \sum_{i+j=k+2} a_{ij}$ for $k = 0, \ldots, 2d$.

For $p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{2d} x^{2d}$, these theorems imply that

$$\begin{aligned}
&\min\{p(x) \mid x \in \mathbb{R}\} \;= \\
&= \max\{t \mid p(x) - t \ \text{ is positive semi-definite}\} \\
&= \max\{t \mid p(x) - t \ \text{ is a sum of squares of polynomials}\} \\
&= \max\{c_0 - a_{11} \mid \sum_{i+j=k+2} a_{ij} = c_k \ \text{ for all } k \text{ and } (a_{ij}) \in S_+^{d+1}\}
\end{aligned}$$

Minimizing a uni-variate polynomial (3)

Proof of Theorem #2

The polynomial $p(x) = c_0 + c_1 x + \cdots + c_{2d} x^{2d}$ is a sum of squares, if and only if $c_k = \sum_{i+j=k+2} a_{ij}$ holds for all k, with $(a_{ij}) \in S_+^{d+1}$.

- Assume that p is a sum of squares: $p(x) = \sum_{i=1}^{m} q_i(x)^2$ where $q_i(x) = z_{1,i} + z_{2,i}x + z_{3,i}x^2 + \cdots + z_{d+1,i}x^d$ for $i = 1, \ldots, m$
- Consider the $m \times (d+1)$ matrix $Z = (z_{ij})$. Then $A := Z^T Z$ is PSD with $c_k = \sum_{i+j=k+2} a_{ij}$

- Assume that there is a $(d+1) \times (d+1)$ matrix $A \succeq 0$ that satisfies $c_k = \sum_{i+j=k+2} a_{ij}$ for $k = 0, \ldots, 2d$.
- Then $A = Z^T Z$ for some matrix $Z = (z_{ij})$.
- Define $q_i(x) = z_{1,i} + z_{2,i}x + z_{3,i}x^2 + \cdots + z_{d+1,i}x^d$ for all i.
- Then $p(x) = \sum_{i=1}^{m} q_i(x)^2$

Minimizing a uni-variate polynomial (4)

Example

Consider the polynomial $p(x) = x^4 - 10x^3 + 6x^2 + 14x + 3$.

- Then $\min\{p(x) \mid x \in \mathbb{R}\}$ equals
  the maximum value of $3 - a_{11}$ subject to the constraints

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \succeq 0 \quad \text{and} \quad \begin{aligned} a_{12} + a_{21} &= 14 \\ a_{13} + a_{22} + a_{31} &= 6 \\ a_{23} + a_{32} &= -10 \\ a_{33} &= 1 \end{aligned}$$

- An optimal solution with objective value $3 - a_{11} = -634$ is given by

$$\begin{bmatrix} 637 & 7 & -14 \\ 7 & 34 & -5 \\ -14 & -5 & 1 \end{bmatrix} \succeq 0$$

Minimizing a uni-variate polynomial (5)

Attention!

This approach for minimizing uni-variate polynomials
does NOT generalize to polynomials in $k \geq 2$ variables.

Example

Consider the polynomial $p(x,y) = x^2y^2(x^2+y^2-3)+1$.

- Arithmetic-geometric mean inequality for three variables yields

$$\frac{1}{3}\left(x^2+y^2+\frac{1}{x^2y^2}\right) \;\geq\; \sqrt[3]{x^2y^2\frac{1}{x^2y^2}} \;=\; 1$$

- If $p(x,y) = q_1(x,y)^2 + q_2(x,y)^2 + \cdots + q_m(x,y)^2$,
  then every $q_i(x,y)$ must be of the form $a + bxy + cx^2y + dxy^2$.
- Then the coefficient of x^2y^2 in $p(x,y)$ must be non-negative.

EOC

Question hour / Vraagenuur:
Tuesday, October 20, 9:45, laplace-gebouw -1.19