Vorlesung
Einführung in die Mathematische Optimierung

(Wintersemester 2013/14)

Kapitel 7: Polynomiale Algorithmen für Konvexe Optimierungsprobleme

Volker Kaibel

Otto-von-Guericke Universität Magdeburg

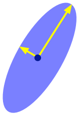
(Version vom 13. Januar 2014)

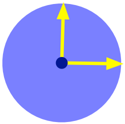
Gliederung

Wiederholung: Ellipsoide

Definition

Ist $Q \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit, und ist $z \in \mathbb{R}^n$, so heißt $\mathrm{Ell}(z, Q) := \{x \in \mathbb{R}^n : (x - z)^{\mathsf{T}} Q^{-1}(x - z) \leq 1\}$ ein **Ellipsoid** (mit Zentrum z). Für $\varrho \in \mathbb{R}$, $\varrho > 0$, heißt $\mathrm{B}(z, \varrho) := \mathrm{Ell}(z, \varrho^2 \mathbb{I}_n)$ der **Ball** um z mit Radius ϱ.





Affine Transformationen von Einheitsbällen

Ist C eine Matrix, deren Spalten bis auf Skalierung mit den Quadratwurzeln der jeweiligen Eigenwerte eine aus Eigenvektoren von Q bestehende Orthonormalbasis von $\mathbb{R}^n$ bilden, so ist

$$\mathrm{Ell}(z, Q) \;=\; C \cdot \mathrm{B}(\mathbb{O}_n, 1) + z \; .$$

Das Löwner-John Ellipsoid eines konvexen Körpers

Definition 7.1

Ein **konvexer Körper** in $\mathbb{R}^n$ ist eine kompakte konvexe Teilmenge von $\mathbb{R}^n$, die einen Ball (mit Radius $\varrho > 0$) enthält.

Satz 7.2

für jeden konvexen Körper $K \subseteq \mathbb{R}^n$ gibt es ein eindeutig bestimmtes Ellipsoid (das **Löwner-John-Ellipsoid** $\mathrm{Loew}(K) \subseteq \mathbb{R}^n$ *von K) minimalen Volumens unter allen Ellipsoiden, die K enthalten. Es gilt:*

$$z + \tfrac{1}{n}(\mathrm{Loew}(K) - z) \;\subseteq\; K \;\subseteq\; \mathrm{Loew}(K)\,,$$

wobei z der Mittelpunkt von $\mathrm{Loew}(K)$ sei.

Beweis z.B. in: Danzer, Grünbaum, Klee, *Helly's theorem and relatives*, in: V. Klee, *Convexity*. AMS 1963, S. 101–180.

Löwner-John Ellipsoide abgeschnittener Ellipsoide

Satz 7.3

Seien $Q \in \mathbb{R}^{n \times n}$ *symmetrisch und positiv definit,* $z \in \mathbb{R}^n$ *und* $a \in \mathbb{R}^n \setminus \mathbb{O}_n\}$*. Dann ist*

$$\mathrm{Loew}\left(\mathrm{Ell}(z, Q) \cap \mathrm{H}^{\leq}(a, \langle a, z\rangle)\right) \;=\; \mathrm{Ell}(z', Q')$$

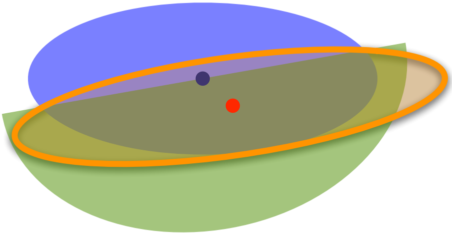
mit

$$z' \;=\; z - \tfrac{1}{n+1} b \qquad \text{und} \quad Q' \;=\; \tfrac{n^2}{n^2-1}\left(Q - \tfrac{2}{n+1} b b^{\mathsf{T}}\right) ,$$

wobei $b = \frac{1}{\sqrt{a^{\mathsf{T}} Q a}} Q a$ *ist.*

Beweis: Abschnitt nach Theorem 3.1.9 in [Grötschel, Lovasz, Schrijver]

Volumenreduktion



Satz 7.4

Sind $Q \in \mathbb{R}^{n \times n}$ *symmetrisch und positiv definit,* $z \in \mathbb{R}^n$ *und* $a \in \mathbb{R}^n \setminus \{\mathbb{O}_n\}$*, so ist*

$$\frac{\mathrm{Vol}\left(\mathrm{Loew}\left(\mathrm{Ell}(z, Q) \cap \mathrm{H}^{\leq}(a, \langle a, z \rangle)\right)\right)}{\mathrm{Vol}\left(\mathrm{Ell}(z, Q)\right)} < \frac{1}{\mathrm{e}^{1/(2n)}} < 1.$$

Beweis: Lemma 3.1.34 in [Grötschel, Lovasz, Schrijver]

Die “ideale” Ellipsoid-Methode

- Seien $K \subseteq \mathbb{R}^n$ ein konvexer Körper und $z^{(0)} \in \mathbb{R}^n$, $R > 0$ mit $K \subseteq E^{(0)} := \mathrm{B}(z, R)$ sowie $\varepsilon > 0$.
- Definiere rekursiv eine (endliche) Folge $E^{(1)}, \dots, E^{(k)} \subseteq \mathbb{R}^n$ von Ellipsoiden mit Zentren $z^{(1)}, \dots, z^{(k)}$ wie folgt:
  1. Falls $z^{(i)} \in K$ oder $\mathrm{vol}\, E^{(i)} < \varepsilon$: Setze $k := i$ (Stop)
  2. Sonst: Wähle ein beliebiges $a \in \mathbb{R}^n$ mit $K \subseteq \mathrm{H}^{\leq}(a, \langle a, z^{(i)} \rangle)$ (existiert nach Satz 2.15) und definiere

$$E^{(i+1)} \ := \ \mathrm{Loew}\left(E^{(i)} \cap \mathrm{H}^{\leq}(a, \langle a, z^{(i)} \rangle)\right) \ .$$

- Wegen Satz 7.4 ist $\mathrm{vol}\, E^{(i)} \leq \mathrm{e}^{-\frac{i}{2n}} \mathrm{vol}\, E^{(0)} \ \leq \ \mathrm{e}^{-\frac{i}{2n}} (2R)^n$
- Also ist die Folge in der Tat endlich mit

$$k \ \leq \ 2 \ln(2R)\, n^2 + (2 \ln \tfrac{1}{\varepsilon}) n + 1 \ .$$

Binärsuche

- Ein **Zulässigkeitsproblem** ist die Aufgabe, zu entscheiden, ob eine gegebene Menge $X \in \mathbb{R}^n$ leer ist, und falls $X \neq \varnothing$ ein beliebiges $x \in X$ zu bestimmen.
- Sind $X \subseteq \mathbb{R}^n$, $f : X \to \mathbb{R}$ und $L, U \in \mathbb{R}$ mit

$$L \;\leq\; \min\{f(x) \,:\, x \in X\} \;\leq\; U$$

  gegeben, so kann man mit Hilfe einer Folge von Zulässigkeitsproblemen das Optimierungsproblem approximativ lösen.
- Solange $U - L$ noch nicht klein genug ist:

$$\{x \in \mathbb{R}^n \,:\, x \in X, f(x) \leq \tfrac{L+U}{2}\} \begin{cases} = \varnothing & : L \leftarrow (L+U)/2 \\ \neq \varnothing & : U \leftarrow (L+U)/2 \end{cases}$$

- "Ideale" Ellipsoid-Methode: Approximative Lösung konvexer Zulässigkeitsprobleme

Algorithmische Repräsentation konvexer Körper

Definition 7.5

Ein **(starkes) Separationsorakel** für eine konvexe Menge $X \subseteq \mathbb{R}^n$ ist ein Algorithmus, der für jedes $y \in \mathbb{Q}^n$ entscheidet, ob $y \in X$ ist, und falls $y \notin X$ ein $a \in \mathbb{Q}^n$ und $\beta \in \mathbb{Q}$ bestimmt mit:

$$X \subseteq \mathrm{H}^{\leq}(a, \beta) \quad \text{aber} \quad \langle a, y \rangle > \beta$$

- $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ Polyeder mit explizit gegebener Matrix $A \in \mathbb{Q}^{m \times n}$ und Vektor $b \in \mathbb{Q}^m$: Implementation eines Separationsorakels mittels Ausrechnen von Ay und Vergleich mit b.

Algorithmische Repräsentation konvexer Körper

Definition 7.6

Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ und $\varepsilon > 0$ definiere

$$X^{[\varepsilon]} := \{y \in \mathbb{R}^n : \|y - x\| \leq \varepsilon \text{ für ein } x \in X\}$$

und

$$X^{[-\varepsilon]} := \{x \in X : x + \mathrm{B}(\mathbb{O}_n, \varepsilon) \subseteq X\} .$$

Definition 7.7

Ein **schwaches Separationsorakel** für eine konvexe Menge $X \subseteq \mathbb{R}^n$ ist ein Algorithmus, der für $y \in \mathbb{Q}^n$ und $\varepsilon \in \mathbb{Q}$ ($\varepsilon > 0$) feststellt, dass $y \in X^{[\varepsilon]}$ ist oder ein $a \in \mathbb{Q}^n$ mit $\|a\|_\infty = 1$ bestimmt, für das $X^{[-\varepsilon]} \subseteq \mathrm{H}^{\leq}(a, \langle a, y \rangle + \varepsilon)$ ist.

Das schwache Optimierungsproblem

Definition 7.8

Das **schwache (konvexe) Optimierungproblem** ist die Aufgabe, für einen konvexen Körper $K \subseteq \mathbb{R}^n$, $c \in \mathbb{Q}^n$ und $\varepsilon \in \mathbb{Q}$ $(\varepsilon > 0)$ einen Punkt $x^\star \in \mathbb{Q}^n$ zu bestimmen mit $x^\star \in K^{[\varepsilon]}$ und

$$\langle c, x\rangle \;\leq\; \langle c, x^\star\rangle + \varepsilon \quad \text{für alle } x \in K^{[-\varepsilon]} \;,$$

oder $K^{[-\varepsilon]} = \varnothing$ festzustellen.

- $x^\star$ ist also fast in K und maximiert $\langle c, x\rangle$ fast über $K^{[-\varepsilon]}$.
- Minimieren von $\langle c, x\rangle$ via Maximieren von $\langle -c, x\rangle$
- Konvexe Zielfunktion f:
  $\tilde{K} := \{(x, \zeta) \in \mathbb{R}^{n+1} \,:\, x \in K, f(x) \leq \zeta\}$ ist konvex mit

$$\min\{f(x) \,:\, x \in K\} \;=\; \min\{\zeta \,:\, (x, \zeta) \in \tilde{K}\} \;.$$

Kodierungslängen

Definition 7.9

Wir definieren folgende **Kodierungslängen** $\langle \cdot \rangle$ für rationale Zahlen, Vektoren und Matrizen:

- Für $n \in \mathbb{Z}$: $\langle n \rangle := 1 + \lceil \log_2(|n| + 1) \rceil$
- Für $r \in \mathbb{Q}$: $\langle r \rangle := \langle p \rangle + \langle q \rangle$, wobei $r = \frac{p}{q}$ mit $p, q \in \mathbb{Z}$ teilerfremd, $q > 0$
- Für $a \in \mathbb{Q}^n$: $\langle a \rangle := \sum_{i=1}^{n} \langle a_i \rangle$
- Für $A = (a_{ij}) \in \mathbb{Q}^{m \times n}$: $\langle A \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} \langle a_{ij} \rangle$

Komplexitätsresultate für konvexe Minimierung

Satz 7.10

Es gibt einen Algorithmus und eine Polynomfunktion p*, der das schwache Optimierungsproblem für durch schwache Separationsorakel gegebene konvexe Körper mit höchstens* $p(\langle c\rangle + \langle \varepsilon\rangle + \langle R\rangle)$ *Orakelaufrufen und Rechenschritten löst. Dabei ist* $K \subseteq \mathbb{R}^n$ *und* $R \in \mathbb{Q}$ *eine positive Zahl mit* $K \subset \mathrm{B}(\mathbb{O}_n, R)$*, die dem Algorithmus übergeben werden muss;* $c \in \mathbb{Q}^n$ *und* $\varepsilon > 0$ *sind wie in Def. 7.8.*

- Corollary 4.2.7 in [Grötschel, Lovasz, Schrijver]
- "Vernünftig gestellte" konvexe Minimierungsprobleme (konvexe Zielfunktion über konvexer Menge) können also in polynomialer Zeit approximativ gelöst werden.
- Insbesondere: Semidefinite Optimierungsprobleme (Trennhyperebene für nicht positiv-semidefinite symmetrische Matrizen aus Eigenvektor zu negativem Eigenwert)

Komplexitätsresultate für lineare Optimierung

Satz 7.11

Es gibt einen Algorithmus und eine Polynomfunktion q, der lineare Optimierungsprobleme

$$\max\{\langle c, x\rangle \,:\, Ax \leq b\} \qquad (LP)$$

für gegebene $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$ und $c \in \mathbb{Q}^n$ in durch $q(\langle A\rangle + \langle b\rangle + \langle c\rangle)$ beschränkter Laufzeit **exakt** *löst, d.h., eine Optimallösung $x^\star \in \mathbb{Q}^n$ von (LP) bestimmt oder feststellt, dass (und ob) (LP) unzulässig oder unbeschränkt ist.*

- Theorem 6.4.9 in [Grötschel, Lovasz, Schrijver]
- Der Beweis (Exakte Lösung!) braucht zusätzlich Strukturresultate über Polyeder (später).
- Insbesondere: Lineare Optimierungsprobleme haben rationale Optimallösungen (wenn sie nicht unbeschränkt oder unzulässig sind).

Bemerkungen zur Ellipsoid-Methode

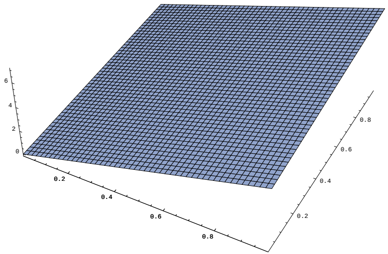
Bemerkung 7.12

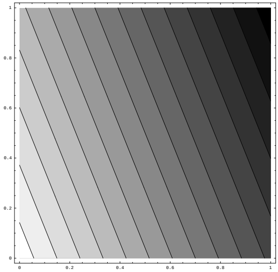
Sind die Ungleichungssysteme $Ax \leq b$ nicht explizit gegeben, sondern hat man für eine Klasse von linearen Optimierungsproblemen nur (starke) Separationsorakel für die Polyeder der zulässigen Lösungen, so kann man die LPs trotzdem in polynomialer Laufzeit lösen, sofern die Separationsorakel polynomiale Laufzeit haben (bezüglich der Kodierungslängen der relevanten Problemdaten).

- Die Ellipsoid-Methode liefert für die komplexitätstheoretische Klassifikation von Optimierungsproblemen eminent wichtige Ergebnisse...
- ...jedoch keine praktisch brauchbaren Algorithmen für die lineare oder konvexe Optimierung
- Technische Bedingung an das Orakel für Satz 7.10: Die Kodierungslänge der Ausgabe muss polynomial beschränkt sein in der Kodierungslänge der Eingabe.

Das Setup

- $\min\{\langle c, x\rangle \;:\; x \in X\}$
- $X \in \mathbb{R}^n$: konvexe (beschränkte) Menge, $c \in \mathbb{R}^n$
- (Konvexe Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$: $\min\{t \;:\; f(x) \leq t, x \in X\}$)



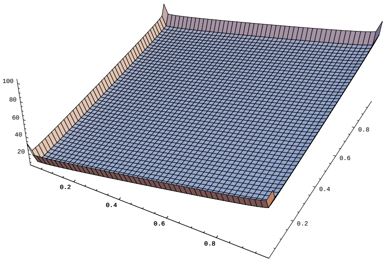


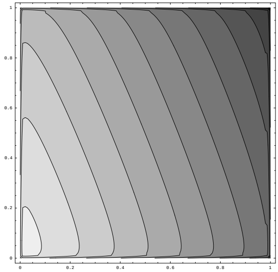
Beispiel: $\min\{\langle c, x\rangle \;:\; x \in X\}$, $X = [0,1]^2$, $c = (5,2)$

Beispiel

- Für $\eta > 0$ definiere $f_\eta : \mathrm{int}(X) \to \mathbb{R}$ mittels:

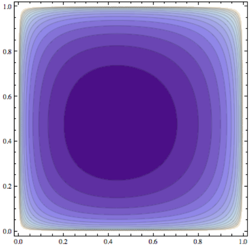
$$f_\eta(x) \;=\; \eta \cdot \langle c, x\rangle + \big(-\ln x_1 - \ln x_2 - \ln(1-x_1) - \ln(1-x_2)\big)$$



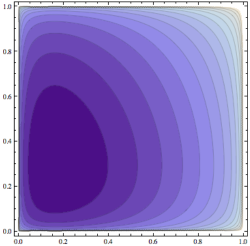


- f_η streng konvex, Minimum eindeutig in $z(\eta) \in \mathrm{int}(X)$
- Für großes η: $z(\eta)$ in der Nähe einer Optimallösung von $min\{\langle c, x\rangle \,:\, x \in X\}$.
- Newton-Verfahren für f_η zur Approximation von $z(\eta)$

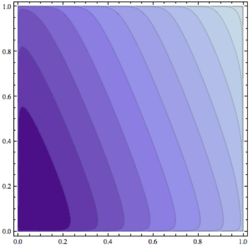
Verschiedene η



$\eta = 0.1$



$\eta = 1$



$\eta = 10$

- Wichtig für Newton-Verfahren: Start nah bei $z(\eta)$
- Ansatz: Folge $\eta_1 < \eta_2 < \dots$
  - Bestimme η_0 und Punkt $x^{(0)} \in \mathrm{int}(X)$ in der Nähe von $z(\eta_0)$.
  - Für $k > 0$: Bestimme $\eta_k > \eta_{k-1}$ und mache einen Newton-Schritt für f_{η_k} von $x^{(k-1)}$ zu $x^{(k)} \in \mathrm{int}(X)$.

Fragen

1. Welche Funktionen können i.A. die Rolle von
$$-\ln x_1 - \ln x_2 - \ln(1 - x_1) - \ln(1 - x_2)$$
spielen? ($\rightsquigarrow$ *(selbstkonkordante) Barrier-Funktionen*)
2. Konvergiert $z(\eta)$ für $\eta \to \infty$ gegen eine Optimallösung?
3. Wie findet man η_0 und $x^{(0)}$?
4. Wie groß darf $\eta_k - \eta_{k-1}$ sein, damit $x^{(k)}$ in $\operatorname{int}(X)$ und ebenfalls nah an $z(\eta_k)$ ist (vorausgesetzt, dass $x^{(k-1)}$ nah an $z(\eta_{k-1})$ ist)?
5. Was heißt "nah"?
6. Wie groß muss k sein, damit (für vorgegebenes $\varepsilon > 0$)
$$\langle c, x^{(k)} \rangle - \min\{\langle c, x \rangle \, : \, x \in X\} < \varepsilon$$
ist?
7. Welche arithmetischen Operationen müssen durchgeführt werden?

Intrinsische Skalarprodukte

Generalvoraussetzungen von jetzt an:

- $D \subseteq \mathbb{R}^n$ offen und konvex
- $f : D \to \mathbb{R}$ zweimal stetig differenzierbar
- $\operatorname{hess}_x f$ positiv definit für alle $x \in D$ (f streng konvex)

Definition 7.13

Für jedes $x \in D$ sei $\langle \cdot, \cdot \rangle_x : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ das durch

$$\langle v, w \rangle_x \;=\; \langle v, (\operatorname{hess}_x f) w \rangle = \langle (\operatorname{hess}_x f) v, w \rangle = v^{\mathsf{T}} (\operatorname{hess}_x f) w$$

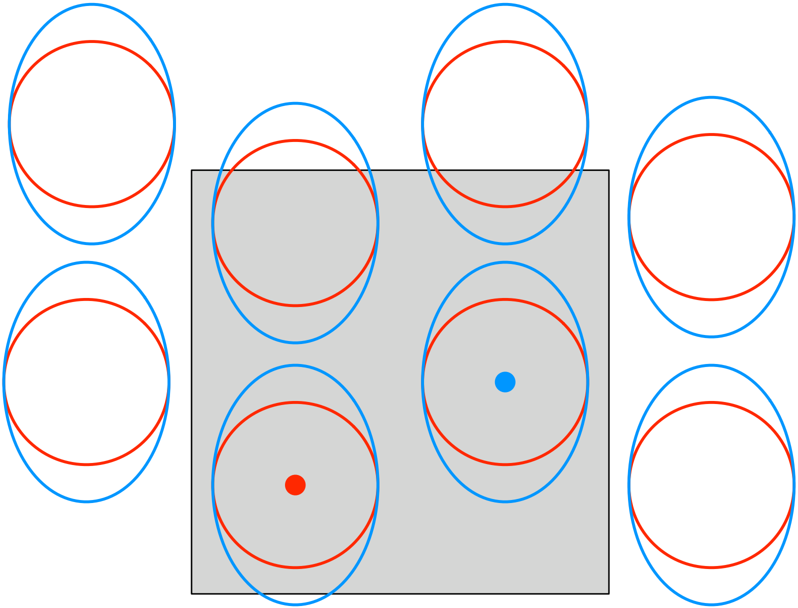
definierte Skalarprodukt auf $\mathbb{R}^n$. Die von diesem Skalarprodukt induzierte Norm bezeichnen wir mit $\|\cdot\|_x : \mathbb{R}^n \to \mathbb{R}$:

$$\|v\|_x \;=\; \sqrt{\langle v, v \rangle_x}$$

Der *offene* Ball vom Radius $\varrho > 0$ bzgl. $\|\cdot\|_x$ um v ist

$$\mathrm{B}_x(v, \varrho) \;=\; \{ w \in \mathbb{R}^n \,:\, \|w - v\|_x < \varrho \} \,.$$

Beispiele



Selbstkonkordante Funktionen

- $D \to \mathbb{S}^n$, $x \mapsto \operatorname{hess}_x f$ ist stetig
- Für alle $v \in \mathbb{R}^n \setminus \{\mathbb{O}_n\}$ und für jedes $x \in D$ und $\varepsilon > 0$ gibt es eine Umgebung $U \subseteq D$ von x mit:

$$1 - \varepsilon \;\leq\; \frac{\|v\|_x}{\|v\|_y} \;\leq\; 1 + \varepsilon \quad \text{für alle } y \in U$$

Definition 7.14

Die Funktion f heißt **selbstkonkordant**, wenn für alle $x \in D$ gilt:

- $\mathrm{B}_x(x, 1) \subseteq D$ und
- Für alle $y \in \mathrm{B}_x(x, 1)$ ist

$$1 - \|y - x\|_x \;\leq\; \frac{\|v\|_x}{\|v\|_y} \;\leq\; \frac{1}{1 - \|y - x\|_x} \text{ für alle } v \in \mathbb{R}^n \setminus \{\mathbb{O}_n\}$$

(Literatur: *strongly non-degenerate* self-concordant function)

In der Nähe des Randes

Satz 7.15

Sind f *selbstkonkordant und* $(x^{(k)})_{k\in\mathbb{N}}$ *eine Folge in* D *mit*

$$\lim_{k\to\infty} x^{(k)} \in \mathrm{bd}(D) \ ,$$

(mit $\mathrm{bd}(D) = \mathrm{cl}(D) \setminus D$*) so gelten*

$$\lim_{k\to\infty} f(x^{(k)}) = \infty$$

und

$$\lim_{k\to\infty} \|\mathrm{grad}_{x^{(k)}} f\| = \infty \ .$$

Beweis: [Renegar, Thm. 2.2.9]

James Renegar, A Mathemtical View of Interior-Point Methods in Convex Optimization. MPS-SIAM Series on Optimization , SIAM 2001.

Newton-Schritte

Definition 7.16

Wir bezeichnen mit

$$\mathrm{new}_x(f) \;:=\; -(\mathrm{hess}_x\, f)^{-1}\,\mathrm{grad}_x\, f \;\in\; \mathbb{R}^n$$

den **Newton-Schritt** für f in $x \in D$.

- Wir haben: $\mathrm{new}_x(f) = \mathbb{O}_n \;\Leftrightarrow\; \mathrm{grad}_x\, f = \mathbb{O}_n$
- Also wegen der Konvexität von f:

Beobachtung 7.17

Ein Punkt $x \in D$ ist genau dann der Minimierer von f (über D), wenn $\mathrm{new}_x(f) = \mathbb{O}_n$ gilt.

Abschätzungen für selbstkonkordante Funktionen

Satz 7.18

Sind f selbstkonkordant und $x \in D$ mit $\gamma := \|\mathrm{new}_x(f)\|_x < 1$, so ist $x' = x + \mathrm{new}_x(f) \in D$ und es gilt:

$$\|\mathrm{new}_{x'}(f)\|_{x'} \ \leq \ \left(\tfrac{\gamma}{1-\gamma}\right)^2 .$$

Beweis: [Renegar, Thm. 2.2.4]

Satz 7.19

Sind f selbstkonkordant und $x \in D$ mit $\gamma := \|\mathrm{new}_x(f)\|_x \leq \frac{1}{4}$, so hat f einen eindeutigen Minimierer $z \in D$, und es gilt

$$\|x - z\|_x \ \leq \ \gamma + \tfrac{3\gamma^2}{(1-\gamma)^3} .$$

Beweis: [Renegar, Thm. 2.2.5]

Barrier-Funktionen

Definition 7.20

Die Funktion $f : D \to \mathbb{R}$ ist eine **Barrier-Funktion**, wenn sie selbstkonkordant ist und ihr **Komplexitätswert**

$$\vartheta_f \;:=\; \sup\{\|\mathrm{new}_x(f)\|_x \;:\; x \in D\} \;<\; \infty$$

endlich ist.

Satz 7.21

Ist f eine Barrier-Funktion mit Komplexitätsparameter ϑ_f, so gilt

$$\langle \mathrm{grad}_x\, f, y - x\rangle < \vartheta_f$$

für alle $x, y \in D$.

Beweis: [Renegar, Thm. 2.2.3]

Addition von Barrier-Funktionen

Satz 7.22

Sind $f_1 : D_{f_1} \to \mathbb{R}$ *und* $f_2 : D_{f_2} \to \mathbb{R}$ *Barrier-Funktionen mit* $D := D_{f_1} \cap D_{f_2} \neq \varnothing$ *und Komplexitätsparametern* ϑ_1 *bzw.* ϑ_2*, so ist* $f_1 + f_2 : D \to \mathbb{R}$ *eine Barrier-Funktion mit* $\vartheta_f \leq \vartheta_1 + \vartheta_2$*.*

Beweis: [Renegar, Thm. 2.2.6 und 2.3.1]

Bemerkung 7.23

Durch $f(x) := -\ln x$ *wird eine Barrierfunktion auf* $\mathbb{R}_{++} := \{x \in \mathbb{R} \,:\, x > 0\}$ *definiert mit* $\vartheta_f = 1$*.*

Korollar 7.24

Durch $f(x) := -\sum_{i=1}^{n} \ln x_i$ *ist eine Barrier-Funktion auf* $\mathbb{R}^n_{++}$ *mit* $\vartheta_f \leq n$ *definiert.*

Barrier-Funktionen für Polyeder

Satz 7.25

Ist $f : \mathbb{R}_{++} \subseteq \mathbb{R}$ *eine Barrierfunktion und sind* $a \in \mathbb{R}^n \setminus \{\mathbb{O}_n\}$, $\beta \in \mathbb{R}$, *so ist auch*

$$f_{a,\beta} \ : \ \{x \in \mathbb{R}^n \ : \ \langle a, x \rangle < \beta\} \to \mathbb{R} \quad \textit{mit } f_{a,\beta}(x) = f(\beta - \langle a, x \rangle)$$

eine Barrier-Funktion mit $\vartheta_{f_{a,\beta}} \leq \vartheta_f$.

Beweis: [Renegar, Theoreme 2.2.7 und 2.3.2]

Korollar 7.26

Sind $A \in \mathbb{R}^{m \times n}$ *und* $b \in \mathbb{R}^m$ *so, dass*

$$\{x \in \mathbb{R}^n \ : \ Ax < b\} = \mathrm{int}(\mathrm{P}^{\leq}(A, b)) \ \neq \ \varnothing$$

ist, so ist $f : \mathrm{int}(\mathrm{P}^{\leq}(A, b)) \to \mathbb{R}$ *mit*

$$f(x) \ = \ -\sum_{i=1}^{m} \ln(b_i - A_{i,\star}x)$$

eine Barrier-Funktion mit $\vartheta_f \leq m$.

Affine Unterräume

- Sei $\mathcal{A} \subseteq \mathbb{R}^n$ affiner Unterraum.
- Affiner Isomorphismus $\mathcal{A} \leftrightarrow \mathbb{R}^p$ liefert Differenzierbarkeitsbegriff für Funktionen auf $\mathcal{A}$
- $\rightsquigarrow$ Begriff der *Selbstkonkordanz* von auf (relativ offenen) Teilmengen von $\mathcal{A}$ definierten Funktionen
- Unabhängig von Wahl des Ismorphismus $\mathcal{A} \leftrightarrow \mathbb{R}^p$

Bemerkung 7.27

Sind f eine Barrier-Funktion und $\mathcal{A} \subseteq \mathbb{R}^n$ ein affiner Unterraum mit $D \cap \mathcal{A} \neq \varnothing$, so ist die Einschränkung $f|_{\mathcal{A}}$ von f auf $\mathcal{A}$ eine Barrier-Funktion auf $D \cap \mathcal{A}$ mit $\vartheta_{f|_{\mathcal{A}}} \leq \vartheta_f$.

(Siehe [Renegar, Abschn. 2.2.1 und 2.3.1])

Korollar 7.28

Sind $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$ so, dass es ein $\tilde{x} \in \mathbb{R}^n_{++}$ gibt mit $A\tilde{x} = b$, so ist durch $f(x) := -\sum_{i=1}^n \ln x_i$ eine Barrier-Funktion auf $\operatorname{relint}\{x \in \mathbb{R}^n_+ \ : \ Ax = b\}$ mit $\vartheta_f \leq n$ definiert

Eine Barrier-Funktion für $\mathbb{S}_+^k$

Satz 7.29

Die Funktion $f : \operatorname{int}(\mathbb{S}_+^k) \to \mathbb{R}$ *mit* $f(X) = -\ln(\det X)$ *ist eine Barrier-Funktion mit* $\vartheta_f = k$.

Beweis: [Renegar, Abschn. 2.2.1]

Korollar 7.30

Sind $A^{(1)}, \ldots, A^{(m)} \in \mathbb{S}^k$, $b \in \mathbb{R}^m$ *so, dass es eine positiv definite Matrix* $\tilde{X} \in \mathbb{S}^k$ *gibt mit* $\langle A^{(i)}, \tilde{X} \rangle = b_i$ *für alle* $i \in [m]$ *, so definiert* $f(X) = -\ln(\det(X))$ *eine Barrier-Funktion auf*

$$\operatorname{relint}\{X \in \mathbb{S}_+^k \;:\; \langle A^{(i)}, X \rangle = b_i \text{ für alle } i \in [m]\}$$

mit $\vartheta_f \le k$.

Der zentrale Pfad

Generalvoraussetzungen von jetzt an:

- $D \subseteq \mathbb{R}^n$ offen und konvex
- $f : D \to \mathbb{R}$ Barrier-Funktion
- $\mathcal{A} \subseteq \mathbb{R}^n$ affiner Unterraum ($\mathcal{A} = \mathbb{R}^n$ möglich)
- $D' = D \cap \mathcal{A} \neq \varnothing$ beschränkt, $X = \mathrm{cl}(D')$
- $c \in \mathbb{R}^n$, $\omega := \min\{\langle c, x\rangle \; : \; x \in X\}$

Definition 7.31

Für $\eta \in \mathbb{R}_+$ definiere streng konvexe Funktionen $f_\eta : D' \to \mathbb{R}$ durch

$$f_\eta(x) \; := \; \eta\langle c, x\rangle + f(x)$$

für alle $x \in D'$. Der Minimierer von f_η sei $z(\eta) \in D'$. Der **zentrale Pfad** von $(f_\eta)_{\eta\in\mathbb{R}_+}$ ist $\{z(\eta) \; : \; \eta \in \mathbb{R}_+\}$.

Beobachtung 7.32

Die intrinsischen Skaparprodukte bzgl. der Funktionen f_η sind die intrinsischen Skalarprodukte bzgl. f; insbesondere sind alle f_η selbstkonkordant.

Auf dem zentralen Pfad

- Es gilt: $\operatorname{grad}_{z(\eta)} f_\eta = \mathbb{O}_n$ für alle $\eta \in \mathbb{R}_+$
- Andererseits: $\operatorname{grad}_{z(\eta)} f_\eta = \eta c + \operatorname{grad}_{z(\eta)} f$
- Also: $\operatorname{grad}_{z(\eta)} f = -\eta c$
- Wegen Satz 7.21 für alle $y \in D'$: $\langle \operatorname{grad}_{z(\eta)} f, y - z(\eta)\rangle < \vartheta_f$
- Also für alle $y \in D'$: $\langle c, z(\eta)\rangle - \langle c, y\rangle < \frac{\vartheta_f}{\eta}$

Bemerkung 7.33

Für alle $\eta, \delta \in \mathbb{R}_+$ *gilt* $\langle c, z(\eta)\rangle \;\leq\; \omega + \delta + \frac{1}{\eta}\vartheta_f$. *Insbesondere (Frage 2):*

$$\lim_{\eta\to\infty} \langle c, z(\eta)\rangle = \omega$$

Außerdem gilt für alle $y \in \mathbb{R}^n$*:*

$$\langle c, y\rangle \leq \omega + \frac{1}{\eta}\vartheta_f(1 + \|y - z(\eta)\|_{z(\eta)})$$

(Beweis: [Renegar, Abschnitt 2.4.1])

Der Short-Step Barrier Algorithmus

1. Bestimme eine Iterationszahl K.
2. Bestimme $\eta_0 > 0$ und $x^{(0)} \in D'$ mit $\|\mathrm{new}_{x^{(0)}}(f_{\eta_0})\|_{x^{(0)}} \leq \frac{1}{9}$.
3. Für $k = 1, \ldots, K$:
   3.1 Setze $\eta_k := (1 + \frac{1}{8\sqrt{\vartheta_f}})\eta_{k-1}$.
   3.2 Berechne $\mathrm{new}_{x^{(k-1)}}(f_{\eta_k})$
   3.3 Setze $x^{(k)} := x^{(k-1)} + \mathrm{new}_{x^{(k-1)}}(f_{\eta_k})$

Lemma 1

Für alle k gilt:

1. $\|\mathrm{new}_{x^{(k-1)}}(f_{\eta_k})\|_{x^{(k-1)}} < 1$ *(also $x^{(k)} \in D'$ für alle k)*
2. $\|\mathrm{new}_{x^{(k)}}(f_{\eta_k})\|_{x^{(k)}} \leq \frac{1}{9}$

Beweis: [Renegar, Abschnitt 2.4.2] (Fragen 4,5)

Analyse

- Nach Satz 7.19 also:

$$\|x^{(k)} - z(\eta_k)\|_{x^{(k)}} \;\leq\; \frac{\frac{1}{9}(\frac{8}{9})^3 + 3(\frac{1}{9})^2}{(\frac{8}{9})^3} \;\leq\; \frac{1}{6}\;.$$

- Definition der Selbstkonkordanz (mit $x = x^{(k)}$, $y = z(\eta)$, $v = x - y$):

$$\frac{\|x^{(k)} - z(\eta_k)\|_{x^{(k)}}}{\|x^{(k)} - z(\eta_k)\|_{z(\eta_k)}} \;\geq\; \frac{5}{6}$$

- Also für alle k: $\|x^{(k)} - z(\eta_k)\|_{z(\eta_k)} \leq \frac{1}{5}$
- Nach Bem 7.33 also: $\langle c, x^{(k)}\rangle \leq \omega + \frac{1}{\eta_k}\frac{6\vartheta_f}{5}$

Lemma 2

Für $\varepsilon > 0$ *und* $K \geq 10\sqrt{\vartheta_f}\ln\frac{6\vartheta_f}{5\eta_0\varepsilon}$ *ist* $x^{(K)} \in D'$ *mit*

$$\langle c, x^{(K)}\rangle \leq \omega + \varepsilon.$$

(Frage 6)

Die Bestimmung von η_0 und $x^{(0)}$

- (Frage 3)
- $\mathrm{new}_x(f_\eta) = -\eta(\mathrm{hess}_x\, f)^{-1}c + \mathrm{new}_x(f)$
- Ziel: Finde ein $\tilde{x} \in D'$ mit $\|\mathrm{new}_{\tilde{x}}(f)\|_{\tilde{x}} \leq \frac{1}{6}$ (Z)
- ($\mathrm{new}_{\tilde{x}}(f) = \mathbb{O}_n \Leftrightarrow \tilde{x}$ minimiert f über D')
- Dann setze:

$$\eta_0 \;:=\; \frac{1}{12\|(\mathrm{hess}_{\tilde{x}}\, f)^{-1}c\|_{\tilde{x}}}$$

- Dann ist $\|\mathrm{new}_{\tilde{x}}(f_{\eta_0})\|_{\tilde{x}} \leq \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$
- Für

$$x^{(0)} \;:=\; \tilde{x} + \mathrm{new}_{\tilde{x}}(f_\eta)$$

gilt dann wegen Satz 7.18 (mit $\gamma = \frac{1}{4}$):

$$\|\mathrm{new}_{x^{(0)}}(f_{\eta_0})\|_{x^{(0)}} \;\leq\; \left(\frac{1/4}{3/4}\right)^2 \;=\; \frac{1}{9}$$

Erreichen des Ziels (Z)

- Gegeben: Ein beliebiger Punkt $\tilde{x}^{(0)} \in D'$
- Sei $g := -\operatorname{grad}_{\tilde{x}^{(0)}} f$
- Für $0 < \nu \leq 1$:
  - Definiere $\tilde{f}_\nu : D' \to \mathbb{R}$ via $\tilde{f}_\nu(x) := \nu\langle g, x\rangle + f(x)$
  - Der Minimierer von $\tilde{f}_\nu$ sei $\tilde{z}(\nu)$
- $\operatorname{grad}_{\tilde{x}^{(0)}} \tilde{f}_1 = g + \operatorname{grad}_{\tilde{x}^{(0)}} f = \mathbb{O}_n$
- Also: $\tilde{z}(1) = \tilde{x}^{(0)}$ und $\operatorname{new}_{\tilde{x}^{(0)}}(\tilde{f}_1) = \mathbb{O}_n$
- Algorithmus:
  1. $\ell := 0$, $\nu_0 := 1$
  2. Solange $\|\operatorname{new}_{\tilde{x}^{(\ell)}}(f)\|_{\tilde{x}^{(\ell)}} > \frac{1}{6}$ ist:
     - 2.1 Inkrementiere ℓ
     - 2.2 Setze $\nu_\ell := (1 - \frac{1}{8\sqrt{\vartheta_f}})\nu_{\ell-1}$
     - 2.3 Berechne $\operatorname{new}_{\tilde{x}^{(\ell-1)}}(f_{\nu_\ell})$
     - 2.4 Setze $\tilde{x}^{(\ell)} := \tilde{x}^{(\ell-1)} + \operatorname{new}_{\tilde{x}^{(\ell-1)}}(f_{\nu_\ell})$
- Analog zu Lemma 1 zeigt man für alle ℓ:
  - $\|\operatorname{new}_{\tilde{x}^{(\ell-1)}}(\tilde{f}_{\nu_\ell})\|_{\tilde{x}^{(\ell-1)}} < 1$ (also $\tilde{x}^{(\ell)} \in D'$ für alle ℓ)
  - $\|\operatorname{new}_{\tilde{x}^{(\ell)}}(\tilde{f}_{\nu_\ell})\|_{\tilde{x}^{(\ell)}} \leq \frac{1}{9}$

Wie klein muss ν_ℓ werden?

- Gilt: $\mathrm{new}_{\tilde{x}(\ell)}(\tilde{f}_{\nu_\ell}) = -\nu_\ell(\mathrm{hess}_{\tilde{x}(\ell)}\, f)^{-1} g + \mathrm{new}_{\tilde{x}(\ell)}(f)$
- Also (wegen $\|\mathrm{new}_{\tilde{x}(\ell)}(\tilde{f}_{\nu_\ell})\|_{\tilde{x}(\ell)} \leq \frac{1}{9}$):

$$\begin{aligned}\|\mathrm{new}_{\tilde{x}(\ell)}(f)\|_{\tilde{x}(\ell)} &\leq \frac{1}{9} + \nu_\ell \|(\mathrm{hess}_{\tilde{x}(\ell)}\, f)^{-1}\, \mathrm{grad}_{\tilde{x}(0)}\, f\|_{\tilde{x}(\ell)} \\ &= \frac{1}{9} + \nu_\ell \sqrt{\mathrm{grad}_{\tilde{x}(0)}\, f^{\mathsf{T}} (\mathrm{hess}_{\tilde{x}(\ell)}\, f)^{-1}\, \mathrm{grad}_{\tilde{x}(0)}\, f}\end{aligned}$$

Satz 7.34

Ist $f : D' \to \mathbb{R}$ eine Barrier-Funktion, so gilt für alle $x, y \in D'$:

$$\sqrt{\mathrm{grad}_x\, f^{\mathsf{T}} (\mathrm{hess}_y\, f)^{-1}\, \mathrm{grad}_x\, f} \leq \left(1 + \frac{1}{\mathrm{sym}(x, D)}\right)\vartheta_f$$

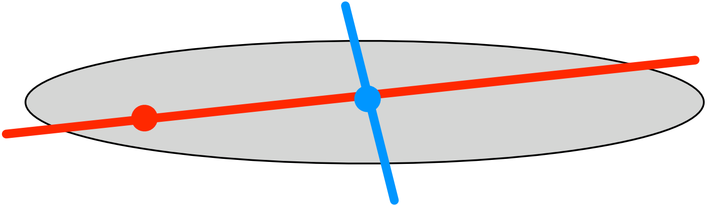
Beweis: [Renegar, Proposition 2.3.7]

Die Symmetrie von D

Definition 7.35

Für $x \in D'$ sei $\mathcal{L}(x, D')$ die Menge aller affinen Geraden $L \subseteq \mathcal{A}$ mit $x \in L$. Für $L \in \mathcal{L}(x, D')$ ist $L \cap D'$ eine (offene) Strecke, die von x in zwei Teile der Längen $\lambda_1, \lambda_2 > 0$ geteilt wird; sei $r(L) := \min\{\frac{\lambda_1}{\lambda_2}, \frac{\lambda_2}{\lambda_1}\}$. Die **Symmetrie** von D bzgl. x ist

$$\operatorname{sym}(x, D) \; := \; \inf\{r(L) \, : \, L \in \mathcal{L}(x, D)\} \, .$$



Zurück zu ν_ℓ

- Also mit Satz 7.34:

$$\|\mathrm{new}_{\tilde{x}^{(\ell)}}(f)\|_{\tilde{x}^{(\ell)}} \;\leq\; \tfrac{1}{9} + \nu_\ell\Big(1 + \frac{1}{\mathrm{sym}(\tilde{x}^{(0)}, D)}\Big)\vartheta_f$$

- Daher gilt $\|\mathrm{new}_{\tilde{x}^{(\ell)}}(f)\|_{\tilde{x}^{(\ell)}} \leq \frac{1}{6}$, sobald

$$\nu_\ell \geq \Big(18\vartheta_f(1 + \frac{1}{\mathrm{sym}(\tilde{x}^{(0)}, D)})\Big)^{-1}$$

  ist.

- Mit $\Omega := \sup\{\langle c, x\rangle \;:\; x \in D'\}$ kann man zeigen:

$$\|(\mathrm{hess}_x f)^{-1} c\|_x \leq \Omega - \omega \text{ für alle } x \in D'$$

  [Renegar, Abschnitt 2.4.1]

- Also

$$\eta_0 = \frac{1}{12\|(\mathrm{hess}_{\tilde{x}} f)^{-1} c\|_{\tilde{x}}} \geq \frac{1}{12(\Omega - \omega)}$$

Das Ergebnis

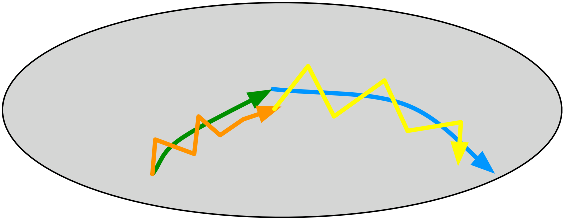
Satz 7.36

Für einen gegebenen Startpunkt $\tilde{x}^{(0)} \in D'$ *und* $0 < \varepsilon < 1$ *berechnet der Short-Step Barrier Algorithmus in*

$$O\left(\sqrt{\vartheta_f} \ln \frac{\vartheta_f}{\varepsilon \operatorname{sym}(\tilde{x}^{(0)}, D')}\right)$$

Newton-Iterationen einen Punkt $x^{(K)} \in D'$ *mit*

$$\frac{\langle c, x^{(K)} \rangle - \omega}{\Omega - \omega} \leq \varepsilon .$$



Arithmetischer Aufwand

- (Frage 7)
- Ein Newton-Schritt erfordert dabei die Berechnung von $\mathrm{new}_x(f_\eta)$ bzw. $\mathrm{new}_x(\tilde{f}_\nu)$.
- Da $\mathrm{hess}_x f_\eta = \mathrm{hess}_x \tilde{f}_\nu = \mathrm{hess}_x f$ ist, muss in jedem Schritt für ein $g \in \mathbb{R}^n$ die Lösung $v \in \mathbb{R}^n$ eines linearen Gleichungssystems $(\mathrm{hess}_x f)v = g$ gefunden werden.
- Beispiel $f(x) = -\sum_{i=1}^m \ln(b_i - \langle A_{i,\star}, x\rangle)$, $D = \mathrm{int}(\mathrm{P}^{\leq}(A, b))$:
  - Sei $\Delta(x) \in \mathbb{R}^{m\times m}$ die Diagonalmatrix mit $\Delta_{i,i} = (b_i - \langle A_{i,\star}, x\rangle)^{-1}$
  - Dann ist $\mathrm{hess}_x f = \Delta(x)A^\mathsf{T} A\Delta(x)$.
  - $A^\mathsf{T} A$ positiv semidefinit, also gibt es Cholesky-Zerlegung $A^\mathsf{T} A = CC^\mathsf{T}$ mit oberer Dreiecksmatrix C.
  - Hat man die Cholesky-Zerlegung einmal berechnet, kann man danach in jeder Iteration das Gleichungssystem $(\Delta(x)C)(\Delta(x)C)^\mathsf{T} = g$ in $\mathrm{O}(n^2)$ arithmetischen Operationen lösen.

Primal-duale Methoden für konische Probleme

- Für $f : D \to \mathbb{R}$ definiert $f^\star(u) = -\inf\{\langle u, x\rangle + f(x) : x \in D\}$ die zu f *konjugierte* Funktion auf $\mathbb{R}^n$.
- Sind $K \subseteq \mathbb{R}^n$ ein konvexer Kegel, $D = \operatorname{int}(K)$ und f eine Barrier-Funktion, so ist $f^\star$ eine Barrier Funktion für $\operatorname{int}(K^\circ)$.
- Primal-duale Methoden verfolgen die zentralen Pfade für primales und duales Problem simultan.
- Geschickte Schrittwahl (nicht Newton) durch Ausnutzen der dualen Informationen (z.B. *Nesterov-Todd Richtungen*)
- Sehr erfolgreich in der Praxis.
- Insbesondere: Stets aktuelle Gütegarantie über dual zulässige Lösung.

Komplexitätsresultate

- Lineare Optimierungsprobleme können mit Barrier-Methoden in polynomialer Zeit exakt gelöst werden.
- Für $\min\{\langle c, x\rangle : Ax \leq b\}$ hängt die Laufzeit polynomial von $\langle A\rangle + \langle b\rangle + \langle c\rangle$ ab.
- Schwierigkeiten:
  - Wie findet man überhaupt einen zulässigen Punkt?
  - Wie findet man eine exakte Lösung?
- Semidefinite Optimierungsprobleme (unter Regularitätsannahmen) können in polynomialer Zeit (approximativ) gelöst werden.
- Geeignete Varianten von Innere-Punkte Methoden sind (im Gegensatz zur Ellipsoid-Methode) praktisch effizient.