Numerische Methoden der globalen Optimierung

Mitschrift*zur Vorlesung von Dr. Ivo Nowak

Sommersemester 2002, HU Berlin

*Gefundene Fehler bitte an vigerske@mathematik.hu-berlin.de

Inhaltsverzeichnis

1 Einführung

Definition 1.1. *Globales Optimierungsproblem*

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & h_E(x) = 0 \\ & & h_I(x) \leq 0 \qquad E \cup I = \{1, \ldots, m\} \\ & & x \in [\underline{x}, \bar{x}] \qquad \underline{x}, \bar{x} \in \bar{\mathbb{R}}^n \\ & & x_k \in \{0,1\}, k \in K \,(\text{binär}) \quad K \subseteq \{1, \ldots, n\} \end{array}$$

$\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$; $x_i \in \{a, b\} \Leftrightarrow (x_i - a)(x_i - b) = 0$

Schwierigkeiten:

- Anzahl der lokalen Minimalstellen
- Lösungen liegen eng beinander
- Finden von zulässigen Punkten

Definition 1.2. $\Omega := \{x \in [\underline{x}, \bar{x}] \,|\, x_k \in \{0,1\}, k \in K,\ h_E(x) = 0,\ h_I(x) \leq 0\}$ heißt *zulässiger Bereich* von (P).

Satz 1.3 (Weierstraß). *Falls f stetig, Ω kompakt und $\Omega \neq \emptyset$, dann existiert eine Lösung von (P).*

1.1 Anwendungen

- kombinatorische Probleme (z.B. kürzester Weg)
- Designprobleme (Min. von Kosten, Zeit, Energie, ...)
- Entscheidungsprobleme
- Netzwerkprobleme
- Optimale Kontrolle (Differentialgleichungen)

1.2 Geschichte der globalen Optimierung

1.2.1 Lokale Optimierung

~ 0: Start: Griechen

18.Jh.: Lagrange: Theorie für restingierte Optimierungsprobleme (Lagrange-Funktion)

1940: Lineare Optimierung (Dantzig / Kantorovisch)

1956: Nichtlineare Optimierung (Karush, Kuhn, Tucker)

1.2.2 Globale Optimierung

1964: Tuy's cutting plane Methode

1969: Branch+Bound (Falk+Soland)

1972: Bender's Decomposition (Geoffrian)

1985: Intervall-Arithmetik (Ratschek)

1.3 Konvexe Mengen und konvexe Funktionen

Definition 1.4. Die Summe $\lambda_1 x_1 + \ldots + \lambda_m x_m$ mit $x_i \in \mathbb{R}^n$, $\sum_{i=1}^m \lambda_i = 1$, $\lambda_i \geq 0$, heißt *Konvexkombination* und eine Menge C aus $\mathbb{R}^n$ ist *konvex*, wenn $\forall x_1, x_2 \in C, \lambda \in [0,1]$: $\lambda x_1 + (1-\lambda) x_2 \in C$.

Satz 1.5. *Falls C konvex ist, dann ist auch $\alpha \cdot C = \{\alpha \cdot x | x \in C\}$ konvex.*

Falls C_1 und C_2 konvex sind, dann ist auch $C_1 + C_2 = \{x_1 + x_2 | x_1 \in C_1, x_2 \in C_2\}$ konvex.

Definition 1.6. $H = \left\{x \in \mathbb{R}^n \,|\, c^T x = b\right\}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}$ heißt *Hyperebene.*

$H_+ = \left\{x \in \mathbb{R}^n \,|\, c^T x \geq b\right\}$ heißt *abgeschlossener Halbraum.*

Konvexe Hülle von S ist $\operatorname{conv}(S) = \bigcap \{C | \, C \text{ konvex und } S \subset C\}$

x heißt *Extrempunkt* einer konvexen Menge, falls es nicht $x_1, x_2 \in C$, $x_1 \neq x_2$, und $\lambda \in (0,1)$ gibt, so daß $x = \lambda x_1 + (1-\lambda) x_2$.

Satz 1.7.
1. *Eine nichtleere abgeschlossene konvexe Menge ist ein Durchschnitt von abgeschlossenen Halbräumen. (dual)*
2. *(Krein-Milman) Eine beschränkte, abgeschlossene Menge ist die konvexe Hülle seiner Extrempunkte. (primal)*

Definition 1.8. Der Schnitt endlich vieler Halbräume heißt *Polyeder*, $P = \{x \in \mathbb{R}^n \,|\, Ax \leq b\}$. Ein beschränktes Polyeder heißt *Polytop.*

Definition 1.9. C heißt *Kegel*, wenn $\lambda x \in C$ für alle $x \in C$, $\lambda \geq 0$ gilt.

Definition 1.10. Eine Funktion $f : C \to \mathbb{R}$ über einer konvexen Menge C heißt *konvex*, falls $f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$ für $x_1, x_2 \in C$, $\lambda \in [0,1]$. Falls die Ungleichung strikt gilt, so heißt f *streng konvex.*

Satz 1.11. *Falls $f \in C^2(C, \mathbb{R})$, C konvex und offen, dann gilt f konvex $\Leftrightarrow \nabla^2 f(x) \succeq 0$ (pos. semidefinit) $\forall x \in C$ (bzw. $\lambda_{\min}\left(\nabla^2 f(x)\right) \geq 0 \; \forall x \in C$)*

Definition 1.12. f sei konvex. Dann heißt $\partial f(x) = \left\{v \in \mathbb{R}^n \,|\, f(y) \geq f(x) + (y-x)^T v \, \forall y\right\}$ *Subdifferential* von f in x.

2 Problemklassen

"Black-box" Probleme: keine Information über Funktion, nur Funktionswerte, Unstetigkeitsstellen sind möglich, keine Gradientenauswertung

Probleme mit bekannten Funktionen: Gradientenauswertung möglich

MINLP (Mixed Integer Nonlinear Program):

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & h_E(x) = 0 \\ & & h_I(x) \leq 0 \qquad E \cup I = \{1, \ldots, m\} \\ & & x \in [\underline{x}, \bar{x}] \qquad \underline{x}, \bar{x} \in \mathbb{R}^n \\ & & x_k \in \{0,1\}, k \in K \,(\text{binär}) \quad K \subseteq \{1, \ldots, n\} \end{array}$$

Unterklassen von MINLP:

NLP (Nonlinear program): $K = \emptyset$, $f, h \in C^2$

PP (polynomial program): $K = \emptyset$, f, h sind Polynome

QQP (quadratically constrained quadratic program): $K = \emptyset$, f, h sind quadratisch

QP (quadratic program): $K = \emptyset$, f quadr. (konvex), h linear

LP (linear program): $K = \emptyset$, f, h linear

IP (integer program): $K = \{1, \ldots, n\}$, f, h linear

MILP (mixed integer linear program): $K \neq \emptyset$, f, h linear

MIQQP (mixed integer quadratically constrained quadratic program): $K \neq \emptyset$, f, h quadratisch.

3 Konvexifizierung von nichtkonvexen Optimierungsprobleme-men

3.1 Konvexifizierung der Zielfunktion

Definition. $f : S \to \mathbb{R}$ sei eine unterhalbstetige Funktion und $S \neq \emptyset$ konvex. Die *konvexe Hüllfunktion* (convex envelope) von f über S ist eine Funktion F, definiert durch

1. F ist konvex auf S.
2. $F(x) \leq f(x)\ \forall x \in S$.
3. Falls $h \leq f$ auf S und h konvex $\Rightarrow h \leq F$ auf S.

Satz 3.1. *Es sei f^* der Optimalwert und X^* die Lösungsmenge von $\min_{x \in S} f(x)$, wobei S konvex und kompakt ist. Die konvexe Hüllenfunktion von f über S sei F. Dann gilt*

1. $f^* = \min_{x \in S} F(x)$
2. $X^* \subseteq \operatorname{argmin}_{x \in S} f(x) = \{x \in S |\, F(x) = f^*\}$

Beweis. 1. Annahme: $\exists y^* \in S$: $F(y^*) < f(x^*) = f^*$.
Definiere $H(x) = \max\{f^*, F(x)\} \Rightarrow H$ ist konvex.
Es gilt $F \leq H \leq f$ auf S und $H(y^*) > F(y^*)$. Widerspruch zu F größte konvexe Funktion mit $F \leq f$. □

$F(x) = L(x; \mu) \leq f(x)$, $\mu \in \mathcal{M}$ =zulässiger Bereich des dualen Problems

3.2 Konvexifizierung der zulässigen Menge

$$\min_{x \in S} f(x) \quad \Leftrightarrow \quad \min_{f(x) \leq t, x \in S} t \quad \Leftrightarrow \quad \min_{x \in \tilde{S}} a^T x$$

Satz 3.2. $\min_{x \in \tilde{S}} a^T x = \min_{x \in \operatorname{conv}(\tilde{S})} a^T x$

Beweis. Es sei $x^* \in \operatorname{argmin}_{x \in \tilde{S}} a^T x \Rightarrow a^T x \geq a^T x^*\ \forall x \in \tilde{S} \Rightarrow a^T (x - x^*) \geq 0$, $x \in \tilde{S}$ □

Schnittebenenverfahren, cutting planes

Definition. Der *Anziehungsbereich* $D(x^*) \subseteq \mathbb{R}^n$ einer lokalen Minimalstelle x^* ist die Menge

$$D(x^*) = \{x \in \mathbb{R}^n \,|\, x^* = \mathrm{LOCMIN}(x)\}$$

und LOCMIN(x) ist ein lokales Suchverfahren mit Startwert x.

4 Heuristische Methoden zur Lösung von globalen Optimierunsproblemen

Übersicht: Pinter

1. Globale Erweiterungen von lokalen Suchverfahren:

   (a) Generierung von Punkten im Raum (Gitterpunkte, gleichverteilte Zufallspunkte)

   (b) lokale Suche über alle Startpunkte

   Probleme: zu wenig Startpunkte, Cluster

2. Genetische (Evolutions)algorithmen

   Vorbild: Natur, biologische Evolution wird simuliert

   Ausgang: Population (Menge von Punkten)

   1. Schritt: Selektion (Auslese) über Auswertung der Zielfunktion

   2. Schritt: Mutation (cross-over)

   Ausgabe: Population

3. Simulated Annealing (simuliertes Abkühlen)

   Basiert auf dem physikalischen Prinzip des Abkühlens von Kristall-Strukturen, die plötzlich einen stabilen Zustand annehmen.

   Abhängig von einem (Temperatur)Parameter T wird die Bewegungsfreiheit der Suche immer weiter eingeschränkt.

4. Tabu search

   Verbieten von Suchrichtungen in Richtung bereits besuchter Paare (zumindest für einige Iterationsschritte), zwischendurch können auch schlechtere Punkte zugelassen werden. Ziel ist es, bisher nicht besuchte Gebiete zu finden, Vorgehensweise ist deterministisch (Glover)

5. Approximate convex global underestimation

   Konvexe Minoranten aufbauend auf Samplepunkten werden gebildet. Die Minimalstellen dieser Funktion liefern Schätzwerte für bessere Minimalstellen.

6. Homotopie (homotopy, continuation, pathfollowing, deformation)

   Stetige Deformation der originalen Funktion in eine einfachere Funktion, die weniger lokale Minimalstellen besitzt. Ausgehend von Minimalstellen der einfachen Funktion wird approximativ ein Pfad verfolgt. Wird oft kombiniert mit simulated annealing.

7. Partition + search (Bayesian search)

   Charakteristische Daten werden mit Hilfe eines statistischen Modells der Funktionen adaptiv geschätzt. Das Modell wird durch sukzessive Zerlegung verfeinert. (Mockus)

8. Local escape heuristics

   Mit Hilfe von Hilfsproblemen werden lokale Minimalstellen verbessert

   Heuristiken können nicht garantieren, daß globale Minmalstellen gefunden werden. Manchmal läßt sich beweisen, daß globale Minimalstellen mit Wahrscheinlichkeit 1 gefunden werden, wenn die Anzahl der Samplepunkte gegen unendlich geht.

Meta-Heuristiken: Genetische Algorithmen, Simulated Annealing, Tabu Search

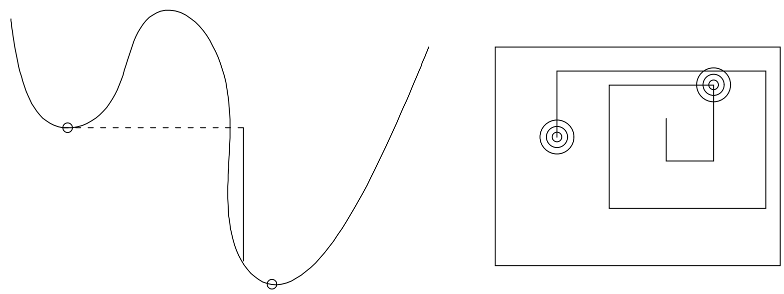

Abbildung 4.1: Local escape heuristics: tunneling und filled function

4.1 Exakte (deterministische) Methoden

Hier kann die Konvergenz gegen das globale Minimum gezeigt werden.

1. Complete enumeration

   Vollständiges Abzählen aller möglichen Lösungen.

   Eignet sich nur für kombinatorische (konkave) Probleme.

2. Vollständige Homotopie-Methoden

   Verfolgung aller Pfade eines parametrischen Optimierungsproblems

3. Successive (outer) approximation (relaxation)

   Das ursprüngliche Problem wird durch eine Folge von relaxierten Problemen gelöst, Verfeinerung der Relaxierungen durch cutting planes oder allgemein durch zulässige Restriktionen

4. Branch and Bound

   - der zulässige Bereich wird sukzessive verfeinert (branching)
   - für jeden Bereich wird die untere Schranke bestimmt (bounding)
   - untere Schranken liefern zulässige Punkte
   - Konvergenz, falls die Differenz zwischen größter unterer und kleinster oberer Schranke gegen Null geht

5. Branch and Cut Verfahren

   Kombination aus 3 und 4 (CPLEX)

5 Branch+Bound Verfahren

Wir betrachten Probleme der Form

$$\min_{x\in\Omega} f(x)$$

5.1 Basis-Operationen

5.1.1 Partitionsmengen

$S_1,\ldots,S_\nu \subseteq \mathbb{R}^n$ seien Teilmengen mit $\bigcup_{i=1}^r S_i \supseteq \Omega$, wobei $\mathrm{int}S_i \cap \mathrm{int}S_j = \emptyset$, $i \neq j$.

S_i heißt Partitionsmenge und $P = \{S_1,\ldots,S_\nu\}$ heißt Partition von Ω. Oft ist S_i eine n-Box oder n-Simplex (oder allgemeines Polyeder).

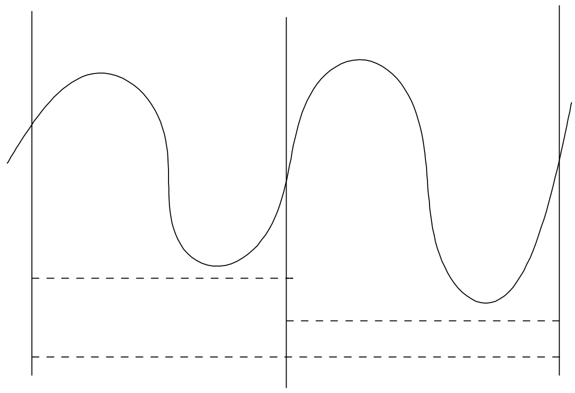

Abbildung 5.1: Branch & Bound

5.1.2 Zerlegungsmethoden

Eine Zerlegungsmethode ist eine Vorschrift, die ausgehend von einer Partition P eine neue Partition P' generiert.

Eine Folge von Partitionsmengen heißt *isoton*, falls $S_{i+1} \subset S_i$.

Eine isotone Folge von Partitionsmengen heißt *ausschöpfend* (exhaustive), falls S_i zu einem Punkt schrumpft, d.h.

$$\bigcap_{i=1}^{\infty} S_i = \{x\}.$$

Eine Zerlegungsmethode heißt *ausschöpfend*, falls jede isotone Teilfolge die von der Zerlegungsmethode generiert wird, ausschöpfend ist.

Beispiel für eine ausschöpfende Zerlegungsmethode ist die Bisektion entlang der längsten Kante eine n-Box oder n-Simplex.

Falls S_i ausschöpfend ist, dann gilt $\operatorname{diam}(S_i) \to 0$.

5.1.3 Untere Schranken

$$f^*(S) = \begin{cases} \min_{x \in \Omega \cap S} f(x) & \text{falls } \Omega \cap S \neq \emptyset \\ \infty & \text{sonst} \end{cases}$$

$\left(= \min_{x\in\Omega} f(x) + i_S(x),\ i_S(x) = \begin{cases} 0, & x \in S \\ \infty, & \text{sonst} \end{cases} \right)$ heißt *Optimalwert* über einer Partitionsmenge.

Eine *untere Schranke* von $f^*(S)$ wird mit $\beta(S)$ bezeichnet.

Es sei S_i eine ausschöpfende Teilfolge mit $\bigcap_{i=1}^{\infty} S_i = \{\hat{x}\}$.

$\beta(S)$ heißt *dicht* (tight), falls

$$\lim_{i\to\infty} \beta(S_i) = \begin{cases} f(\hat{x}), & \hat{x} \in \Omega, \\ \infty & \text{sonst.} \end{cases}$$

$\beta(S)$ heißt *lokal exakt*, falls ein N existiert mit $\beta(S_i) = f(\hat{x})\ \forall i \geq N$ und $\hat{x}$ eine lokale Minimalstelle ist mit $\hat{x} \in S_i$.

Untere Schranken für diskrete Optimierungsprobleme sind normalerweise exakt.

5.1.4 Obere Schranken

$\alpha(S)$ heißt *obere Schranke*, falls $\alpha(S) \geq f^*(S)$ ist.

Falls $x(S) \in \mathbb{R}^n$ eine Schätzung für eine globale Minimalstelle in S ist, dann

$$\alpha(S) = \begin{cases} f(x(S)), & x(S) \in \Omega, \\ \infty & \text{sonst} \end{cases}$$

Die Heuristik $x(S)$ heißt *lokal exakt*, falls $x(S) = \hat{x}$ für alle S mit $\mathrm{diam}(S)$ genügend klein, $\hat{x} \in S$ und $\hat{x}$ ist eine lokale Minimalstelle.

Dies gilt u.a. wenn $x(S)$ durch ein lokales Suchverfahren bestimmt wird.

5.1.5 Selektion der aktiven Partitionsmenge

Normalerweise: $\hat{S} = \operatorname{argmin}\{\beta(S) | S \in P\}$

5.1.6 Schnitte (Cuts)

Schnitte können optional zur Beschleunigung des Verfahrens verwendet werden.

1. Feasibility cuts (zulässige Schnitte) sind Restriktionen, die ein Stück des zulässigen Bereiches abschneiden.
2. Optimality cuts (Optimalitätsschnitte) schneiden gegebene lokale Minimalstellen ab.
3. Globale Optimalitätskriterien: geben Auskunft, ob eine lokale Minimalstelle auch globale Minmalstelle ist

5.2 Branch and Bound Algorithmus

1. (Initialisierung)
   Bestimme $S_1 \supset \Omega$ und $\beta(S_1)$, setze $P = \{S_1\}$ und $\alpha = \infty$
2. (Hauptschleife)
   Wähle $\hat{S} \in P$ und setze $P = P \setminus \left\{\hat{S}\right\}$
3. Bestimme $x\left(\hat{S}\right)$ und setze $\alpha = \min\left\{\alpha, \alpha\left(\hat{S}\right)\right\}$ (optional: füge Schnitte hinzu)
4. Zerlege $\hat{S}$ in $S_1, \ldots, S_r$ (branching)
5. Bestimme $\beta(S_i), i = 1, \ldots, r$ (bounding)
6. Setze $P = P \cup \{S_1, \ldots, S_r\}$
7. Setze $P = P \setminus \{S \in P | \beta(S) \geq \alpha\}$
8. Setze $\beta = \min\{\beta(S) | S \in P\}$
9. Falls $\alpha - \beta > 0$, gehe zu 2.

Satz 5.1. *Falls die Zerlegungsmethode ausschöpfend (exhaustive) ist, die unteren Schranken dicht (tight, consistent) und die Selektion die Schranken verbessert (boundimproving), dann konvergiert β gegen den Optimalwert des Problems.*

Beweis. Es sei $\{S_i\}$ eine infinite isotone Folge mit $\beta(S_i) = \beta$. Da die Zerlegungsmethode ausschöpfend ist, folgt $\bigcap S_i = \{\hat{x}\}$.

1. $\hat{x} \in \Omega$ (zulässig) $\Rightarrow$ $\lim \beta(S_i) = f(\hat{x})$ (wegen dichten Schranken)
   Es gilt $\beta = \beta(S_i) \leq f^* \leq f(\hat{x})$. $\Rightarrow$ Behauptung.
2. $\hat{x} \notin \Omega$ $\Rightarrow$ $\lim \beta(S_i) \to \infty$.
   Da $\beta(S_i) \leq f^*$ $\Rightarrow$ $f^* = \infty$, Problem ist nicht zulässig. □

Falls die Heuristik $x(S)$ lokal exakt ist, oder $x(S)$ zulässig ist, falls $S \cap \Omega \neq \emptyset$ und $\mathrm{diam}(S)$ genügend klein, dann konvergiert $\alpha - \beta$ gegen Null. In dem Fall findet der Branch and Bound Algorithmus in endlich vielen Schritten eine ε-genaue Lösung.

Falls die Schranken und die Heuristik lokal exakt sind, konvergiert der Algorithmus in endlich vielen Schritten gegen den Optimalwert des Problems.

6 Relaxierungen (untere und obere Schranken)

6.1 LP-Schranken für MILP

MILP:

$$\begin{array}{ll} \min & a_0^T x \\ \text{s.t.} & l_i \leq a_i^T x \leq u_i \quad i = 1, \ldots, m \\ & x_i \in \mathbb{Z} \quad i \in I \subseteq \{1, \ldots, n\} \end{array}$$

LP-Relaxierung entsteht durch Weglassen der Ganzzahligkeitsbedingungen

LP:

$$\begin{array}{ll} \min & a_0^T x \\ \text{s.t.} & l_i \leq a_i^T x \leq u_i \quad i = 1, \ldots, m \end{array}$$

- Branch über ganzzahlige Variablen
- endliche Konvergenz von B+B-Algorithmus

6.2 LP-Schranken für QQP

QQP:

$$\begin{array}{ll} \min & q_0(x) \\ \text{s.t.} & q_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

$$q_i(x) = \underbrace{x^T A_i x}_{\sum_{i,j} a_{i,j} x_i x_j} + 2b_i^T x + c_i = \underbrace{\langle X, A_i \rangle}_{\sum_{i,j} a_{i,j} X_{i,j}} + 2b_i^T x + c_i$$

mit $\langle A, B \rangle := \sum_{i,j} a_{i,j} b_{i,j}$, $X_{i,j} := x_i x_j$.
LP:

$$\begin{array}{ll} \min & \langle A_0, X \rangle + 2b_0^T x + c_0 \\ \text{s.t.} & \langle A_i, X \rangle + 2b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & (\underline{X} \leq X \leq \bar{X}) \end{array}$$

6.3 SDP-Schranken für QQP

QQP:

$$\begin{array}{ll} \min & \langle A_0, X\rangle + 2b_0^T x + c_0 \\ \text{s.t.} & \langle A_i, X\rangle + 2b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & X = x \cdot x^T \end{array}$$

$y^T X y = y^T x \cdot x^T y = \left(x^T y\right)^2 \geq 0 \Rightarrow X \succeq 0$

SDP (semidef. Opt.Problem):

$$\begin{array}{ll} \min & \langle A_0, X\rangle + 2b_0^T x + c_0 \\ \text{s.t.} & \langle A_i, X\rangle + 2b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

6.4 Separable Relaxierungen von faktorisierbaren Problemen

Definition. *Faktorisierbare Funktionen* sind rekursive Zusammensetzungen von Summen und Produkten von univariaten Funktionen.

Beispiel.

$$f(x,y) = \sqrt{\underbrace{\underbrace{\exp\underbrace{\left(\underbrace{xy}_{x_1} + \underbrace{z \cdot \underbrace{\ln(w)}_{x_2}}_{x_3}\right)}_{x_4}}_{x_5} \cdot \underbrace{z^3}_{x_6}}_{x_7}}$$

$$\begin{aligned} f(x) &= \sqrt{x_7} \\ x_7 &= x_5 \cdot x_6 \\ x_6 &= z^3 \\ x_5 &= \exp(x_4) \\ x_4 &= x_1 + x_3 \\ x_3 &= z \cdot x_2 \\ x_2 &= \ln(w) \\ x_1 &= x \cdot y \end{aligned}$$

Definition. $f(x)$ heißt *separabel*, falls

$$f(x) = \sum_{i=1}^{n} f_i(x_i)$$

Ein Optimierungsproblem heißt *separabel*, falls alle Funktionen separabel sind.

6.4.1 Bilineare Terme

1. $xy = \frac{1}{4}\left\{\left(\underbrace{x+y}_{z_1}\right)^2 - \left(\underbrace{x-y}_{z_2}\right)^2\right\} = \frac{1}{4}\left(z_1^2 + z_2^2\right)$

2. $xy \geq \max\left\{\bar{x}y + \bar{y}x - \bar{x}\bar{y}, \underline{x}y + \underline{y}x - \underline{xy}\right\},\ x \in [\underline{x}, \bar{x}]$
   $xy \leq \min\left\{\bar{x}y + \underline{y}x - \bar{x}\underline{y}, \underline{x}y + \bar{y}x - \underline{x}\bar{y}\right\},\ y \in [\underline{y}, \bar{y}]$

6.4.2 Relaxierungen von univariaten Funktionen

gegeben: $f(x)$ auf $x = [\underline{x}, \bar{x}] \subset \mathbb{R}$

gesucht: $f_r(x)$ mit $f_r(x) \leq f(x)$ auf X und f_r "einfach"

1. f_r ist stückweise linear.
   Falls alle f_i's in (P) so ersetzt werden, dann ist (P) äquivalent zu einem MILP.

2. f_r ist konvex und aus C^2

$$\begin{aligned}
f_r(x) &= f(x) + \alpha \cdot \overbrace{\underbrace{(x-\underline{x})(x-\bar{x})}_{\leq 0\ \forall x \in X}}^{q(x)} \\
f_r''(x) &= f''(x) + \alpha \cdot 2 \\
\rho &:= \min_{x \in X} f''(x) \\
\Rightarrow f_r''(x) &\geq \rho + 2\alpha \\
&\geq 0 \text{ für } \alpha = \max\left(0, -\frac{1}{2}\rho\right)
\end{aligned}$$

D.h. $f_r(x) = f(x) + \max\left(0, -\frac{1}{2}\rho\right) q(x)$ ist eine konvexe untere Schätzfunktion von f auf X. Ersetzt man in (P) alle f_i's durch $f_{i,r}$, dann ist das Problem konvex.

6.5 Intervall-Methoden

Intervall: $X = [\underline{x}, \bar{x}] \subset \mathbb{R}^n$

Arithmetik:

$$\begin{aligned}
[a,b] + [c,d] &= [a+c, b+d] \\
[a,b] - [c,d] &= [a-d, b-c] \\
[a,b] \cdot [c,d] &= [\min\{ac, ad, bc, bd\}, \max\{ac, ad, bc, bd\}] \\
[a,b] / [c,d] &= [a,b] \cdot \left[\frac{1}{d}, \frac{1}{c}\right], \text{ falls } 0 < c \text{ oder } d < 0
\end{aligned}$$

Inklusionsfunktionen: $F(X)$ so, daß $f(x) \in F(X)\ \forall x \in X$ und $F(X)$ ist ein Intervall und $\mathrm{diam}(F(X)) \to 0$, wenn $\mathrm{diam}(X) \to 0$.

F und G Inklusionsfunktionen zu f und g, dann sei $(F \circ G)(X) = F(G(X))$

Anwendung zur Bestimmung von $\rho = \min_{x \in X} f''(x)$:

$$\left.\begin{array}{r} [\underline{y}, \bar{y}] = Y = F''(X) \\ \underline{\rho} = \underline{y} \end{array}\right\} \Rightarrow \underline{\rho} \leq \rho$$

6.6 Primale Konvexifizierung

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & g(x) \leq 0 \end{array}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$; $h(x) = 0 \Leftrightarrow h(x) \leq 0,\ -h(x) \leq 0$

Falls $f \in C^2$ und $\nabla^2 f(x) \succeq 0$ für $x \in X$, dann ist f auf X konvex.

$$\Leftrightarrow \lambda_1\left(\nabla^2 f(x)\right) \geq 0\,\forall x \in X$$

Konvexe untere Schätzfunktionen auf X:

$$\begin{aligned} f_r(x) &= f(x) + \alpha \cdot q(x) \\ \text{mit } q(x) &= \sum_{i=1}^{n} (x_i - \underline{x}_i)(x_i + \bar{x}_i) \leq 0\,\forall x \in X \\ \rho &= \min_{x \in X} \lambda_1\left(\nabla^2 f(x)\right) \\ \Rightarrow \nabla^2 f_r(x) &= \nabla^2 f(x) + \alpha \cdot \nabla^2 q(x) \text{ und } \nabla^2 q(x) = 2 \cdot I \\ \lambda_1\left(\nabla^2 f_r(x)\right) &= \lambda_1\left(\nabla^2 f(x) + 2\alpha I\right) \\ &= \lambda_1\left(\nabla^2 f(x)\right) + 2\alpha \\ &\geq 0 \text{ für } \alpha = \max\left\{0, -\frac{1}{2}\rho\right\} \end{aligned}$$

Satz 6.1 (Floudas et. al.). *Es sei $A = \left[\underline{A}, \bar{A}\right]$ mit $a_{i,j} = \left[\underline{a}_{i,j}, \bar{a}_{i,j}\right]$ die Intervall-Hessematrix von f auf X, d.h. $\partial^2_{i,j} f(x) \in a_{i,j}$ $\forall x \in X$ und*

$$\alpha_i := \max\left\{0, -\frac{1}{2}\left(\underline{a}_{i,i} - \sum_{j \neq i} |a|_{i,j} \frac{d_j}{d_i}\right)\right\},$$

wobei $|a|_{i,j} = \max\left\{\left|\underline{a}_{i,j}\right|, \left|\bar{a}_{i,j}\right|\right\}$. Dann gilt

$$H + 2\mathrm{Diag}(\alpha) \succeq 0\ \forall H \in A.$$

A heißt *diagonaldominant*, falls $a_{ii} > \sum_{i \neq j} |a_{i,j}|$, $i = 1, \ldots, n$. Falls A diagonaldominant ist, dann ist A positiv definit.

Es sei $H + 2\mathrm{Diag}(\alpha) \succeq 0$ und $\bar{\alpha} = \max_i \alpha_i$, dann ist $H + 2\bar{\alpha} I \succeq 0$.

Optimale Konvexifizierung

bei Mengen: konvexe Hülle, $\mathrm{conv}(S)$

bei Funktionen: konvexe Einhüllende, $\mathrm{conv}_X(f)$, größte konvexe Funktion, die kleiner-gleich einer gegebenen Funktion über X ist, wobei X konvex ist.

Gegeben

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & g(x) \leq 0 \\ & & x \in X \end{array}$$

mit $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$ konvex.

Optimale primale Konvexifizierung von (P)

$$\begin{array}{lll} (P_c) & \min & \mathrm{conv}_X f(x) \\ & \text{s.t.} & \mathrm{conv}_X g(x) \leq 0 \\ & & x \in X \end{array}$$

6.7 Duale Konvexifizierung

Lagrangefunktion

$$L(x,\mu) = f(x) + \langle \mu, g(x) \rangle = f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$

Dual zulässiger Bereich:

$$\mathcal{M} = \mathbb{R}_+^m$$

Lagrange Relaxation

$$\min_{x \in X} L(x,\mu) \tag{R_μ}$$

Duale Funktion

$$D(\mu) = \mathrm{val}(R_\mu) \quad (\text{Optimalwert von } (R_\mu))$$

Duales Problem zu (P)

$$\max_{\mu \in \mathcal{M}} D(\mu) \tag{D}$$

Satz 6.2. *Für alle primal zulässigen Punkte x und dual zulässigen Punkte μ gilt $L(x;\mu) \le f(x)$.*

Beweis. Falls x primal zulässig ist, gilt $g(x) \le 0$ und daher $\langle \mu, g(x) \rangle \le 0$ für alle $\mu \in \mathcal{M}$. □

Falls f und g stetig sind, dann gilt

$$D(\mu) \le \mathrm{val}(P) = f(x^*) \text{ mit } g(x^*) \le 0$$

Dualitätslücke: $\mathrm{val}(P) - \mathrm{val}(D) \ge 0$ (schwache Dualität)

starke Dualität: $\mathrm{val}(P) - \mathrm{val}(D) = 0$

Satz 6.3. *Falls (P) konvex ist (f, g, X konvex), und eine Slater-Bedingung gilt ($\exists x$: $g(x) > 0$ und $x \in \mathrm{int}X$), dann ist*

$$\mathrm{val}(P) = \mathrm{val}(D).$$

Beispiel (starke Dualität bei dem trust-region problem).

$$\begin{array}{ll} (P) \quad \min & x^T A x + b^T x \\ \|x\|_2 & \le \rho \end{array}$$

Es gilt $\mathrm{val}(P) = \mathrm{val}(D)$, auch wenn A nicht positiv semidefinit ist.

Satz 6.4. *Der Definitionsbereich* $\mathrm{dom}D = \{\mu \in \mathcal{M} \mid D(\mu) > -\infty\}$ *ist konvex und D ist über* $\mathrm{dom}D$ *konkav. (D.h. (D) ist konvex.)*

Beweis. Es sei $\mu(t) = t\mu + (1-t)\bar{\mu}$ und $\mu, \bar{\mu} \in \mathcal{M}$ und $t \in [0,1]$.

$$\begin{aligned} D(\mu(t)) &= \inf_{x \in X} L(x, \mu(t)) = \inf_{x \in X} t \cdot L(x;\mu) + (1-t) \cdot L(x, \bar{\mu}) \\ &\le t \inf_{x \in X} L(x,\mu) + (1-t) \inf_{x \in X} L(x, \bar{\mu}) \\ &= tD(\mu) + (1-t) D(\bar{\mu}) \end{aligned}$$

Falls $\mu, \bar{\mu} \in \mathrm{dom}D$, d.h. $D(\mu) > -\infty, D(\bar{\mu}) > -\infty$, dann gilt $D(\mu(t)) > -\infty$ □

Definition. Ein *Supergradient* $g \in \mathbb{R}^m$ einer Funktion $D : \mathbb{R}^m \to \mathbb{R}$ an der Stelle $\bar{\mu}$ erfüllt

$$D(\mu) \le D(\bar{\mu}) + \langle g, \mu - \bar{\mu} \rangle \quad \forall \mu \in \mathbb{R}^m.$$

(Falls g Supergradient von D ist, dann ist $-g$ Subgradient von $-D$)

Satz 6.5. *Für alle* $x_\mu \in \operatorname{argmin}_{x\in X} L(x,\mu)$ *ist der Vektor* $g(\mu) = g(x_\mu)$ *ein Supergradient von* D *an der Stelle* μ.

Beweis.

$$D(\pi) = L(x_\pi, \pi) \leq \underbrace{L(x_\mu, \pi)}_{=f(x)+\langle \pi, g(x_\mu)\rangle} = \underbrace{L(x_\mu, \mu)}_{D(\mu)} + \langle g(x_\mu), \pi - \mu \rangle \qquad \square$$

Beispiel.

1. MILP: f, g sind linear, $X = [\underline{x}, \bar{x}] \times \{0,1\}^{n_2}$, $n = n_1 + n_2$

$$f(x) = b_f^T \cdot x + c_f, \qquad g(x) = B_g \cdot x + c_g$$

$$L(x;\mu) = b(\mu)^T x + c(\mu) := b_f^T \cdot x + c_f + \mu^T (B_g \cdot x + c_G)$$

$$\begin{array}{lll} (R_\mu) & \min & b(\mu)^T x + c(\mu) \\ & \text{s.t.} & x^1 \in [\underline{x}, \bar{x}], \quad x^2 \in \{0,1\}^{n_2} \end{array}$$

$$(R_\mu) = \sum_{i=1}^{n} \min_{x_i \in [\underline{x}_i, \bar{x}_i]} b_i(\mu) \cdot x_i + c_i(\mu)$$

2. Dekomposition: $x_1 \in X_1$, $x_2 \in X_2$,

$$\begin{array}{rcl} f(x) & = & f_1(x) + f_2(x), \\ g(x) & = & g_1(x) + g_2(x). \end{array}$$

$$\begin{array}{rcl} L_i(x;\mu) & := & f_i(x_i) + \mu^T g_i(x_i), \quad i = 1,2 \\ L(x:\mu) & := & L_1(x_1,\mu) + L_2(x_2,\mu) \end{array}$$

$$(R_\mu) \min_{x_1 \in X_1, x_2 \in X_2} L(x,\mu) = \min_{x_1\in X_1} L_1(x_1;\mu) + \min_{x_2 \in X_2} L_2(x_2;\mu)$$

3. Allgemeine quadratische Probleme

$$\begin{array}{ll} \min & x^T A_0 x + c_0 \\ \text{s.t.} & x^T A_i x + c_i \leq 0 \\ & x \in [-e, e] \Leftrightarrow x_i \in [-1,1] \Leftrightarrow x_i^2 - 1 \leq 0 \Leftrightarrow x^T E_i x - 1 \leq 0 \end{array}$$

   Es gilt $\|x\|^2 \leq n$.

   Dualisiere bzgl. $X = \left\{ x | \, \|x\|^2 \leq n \right\}$

$$L(x;\mu) := x^T A(\mu) x + c(\mu)$$

$$(R_\mu) \min_{\|x\|^2 \leq n} x^T A(\mu) x + c(\mu) = n \cdot \min\{\lambda_{\min}(A), 0\} + c(\mu),$$

   $\lambda_{\min}(A) = \min_{\|x\|^2=1} x^T A x$

Satz 6.6. *Gegeben seien*

$$\begin{array}{lll} (P) & \min & f(x) \\ & s.t. & g(x) \leq 0 \\ & & x \in X \end{array}$$

$$
\begin{array}{lll}
(P') & \min & f(x) \\
 & s.t. & g_{1:m-1}(x) \leq 0 \\
 & & x \in X' := \{x \in X \,|\, g_m(x) \leq 0\}
\end{array}
$$

und (D) und (D') seien die dazugehörigen dualen Probleme. Dann gilt:

$$
\operatorname{val}(D) \leq \operatorname{val}(D') \leq \operatorname{val}(P) = \operatorname{val}(P')
$$

Beweis.

$$
D'(\mu') = \min_{x \in X'} L'(x;\mu') = \min_{x \in X, g_m(x) \leq 0} L'(x;\mu') \geq \min_{x \in X} L'(x;\mu') + \mu_m g_m(x)
$$

für $\mu_m \geq 0$. $\Rightarrow \max_{\mu' \in M'} D'(\mu') \geq \max_{\mu \in M} D(\mu)$. □

6.8 Verfahren für nichtdifferenzierbare Optimierungsprobleme

$D(\cdot)$ konkav und stetig:

$$
\max_{\mu \in \mathbb{R}^m_+} D(\mu) \tag{D}
$$

1. Subgradientenverfahren

   $$
   \mu^{k+1} = \Pi_{\mathbb{R}^m_+}\left(\mu^k + t_k \cdot g^k\right),
   $$

   wobei g^k ein Supergradient von D in μ^k ist.
   Falls $t_k = \frac{\lambda_k}{\|g^k\|}$, $\lambda_k \to 0$, $\sum \lambda_k = \infty$, z.B. $\lambda_k = \frac{1}{k}$, dann konvergiert das Subgradientenverfahren linear (Poljak).

2. Schnittebenenverfahren (cutting plane)
   Es sei $\bar{D}_k(\mu) = D\left(\mu^k\right) + \left(g^k\right)^T\left(\mu - \mu^k\right)$.

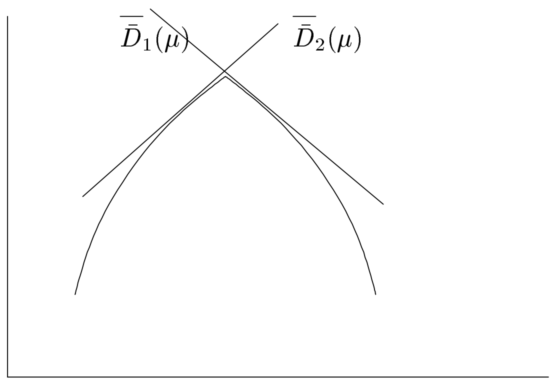

Abbildung 6.1: Schnittebenenverfahren

$$
\mu^{k+1} = \operatorname{argmax}_{\mu \in \mathbb{R}^m_+} \min_{i=1,\ldots,k} \bar{D}_i(\mu) = \operatorname{argmax}_{\mu \in \mathbb{R}^m_+} \left\{ t \,|\, t \leq \bar{D}_i(\mu),\, i = 1,\ldots,k \right\}
$$

3. Bündelverfahren (NOA von Kiwiel): für $I_k \subseteq \{1,\ldots,k\}$

$$
\begin{aligned}
\mu^{k+1} &= \operatorname{argmax}_{\mu \in \mathbb{R}^m_+} \min_{i \in I_k} \bar{D}_i(\mu) - \frac{u^k}{2}\left\|\mu - \mu^k\right\|^2 \qquad \left\|\mu - \mu^k\right\|^2 \leq \rho^k \\
&= \operatorname{argmax}_{\mu \in \mathbb{R}^m_+} \left\{ \min_{i \in I_k} \bar{D}_i(\mu) \,\middle|\, \left\|\mu - \mu^k\right\| \leq \rho^k \right\}
\end{aligned}
$$

Alle Verfahren sind linear konvergent.

6.9 Konvexifizierungsmethoden / Abschätzung des Relaxationsfehlers

Relaxationsfehler: $\mathrm{val}\,(P) - \mathrm{val}\,(R)$ (=Dualitätslücke, falls (R) das duale Problem ist)

- $\mathrm{conv}\,(S)$ =konvexe Hülle von S
- $\mathrm{conv}_X\,(f)$ =größte konvexe Funktion über X, die kleiner gleich f ist.
- Epigraph von f über X ist $\mathrm{epi}_X f = \{\,(x,t)|\, x \in X, t \geq f\,(x)\}$

$$\begin{array}{lll} (G) & \min & x_1 \\ & \text{s.t.} & g\,(x) \leq 0 \\ & & x \in X \end{array}$$

zulässige Menge von (G): $\Omega = \{\,x \in X|\, g\,(x) \leq 0\}$, $g : \mathbb{R}^n \to \mathbb{R}^m$

exakte konvexe Relaxierung von (G), exakte Konvexifizierung

$$\begin{array}{lll} (R_e) & \min & x_1 \\ & \text{s.t.} & x \in \mathrm{conv}\,(\Omega) \end{array}$$

duale Konvexifizierung: $D\,(\mu) = \min_{x \in X} L\,(x,\mu) = \min_{x \in X} (x_1 + \langle \mu, g\,(x)\rangle)$

$$\begin{array}{lll} (R_d) & \min & L\,(x;\mu^*) \\ & \text{s.t.} & x \in X \\ & & \mu^* \in \underset{\mu \in \mathbb{R}^m_+}{\mathrm{argmax}}\, D\,(\mu) \end{array}$$

primale Konvexifizierung:

$$\begin{array}{lll} (R_p) & \min & x_1 \\ & \text{s.t.} & \mathrm{conv}_X\, g\,(x) \leq 0 \\ & & x \in X \end{array}$$

Epigraph von g_i über X in $\mathbb{R}^n \times \mathbb{R}^m$ eingebettet: $G_i = \{\,(x, t \cdot e_i)|\, x \in X,\, t \geq g_i\,(x)\}$

Es sei $H = \mathbb{R}^n \times \{0 \in R^m\}$.

$$S = \bigcap_{i=1}^m G_i \cap H = \{\,(x,0)|\, x \in X, 0 \geq g_i\,(x)\,, i = 1, \ldots, m\} = \Omega \times \{0\}$$

$$\begin{aligned} S_e &= \mathrm{conv}\,(S) = \mathrm{conv}\,(\Omega) \times \{0\} \\ S_d &= \mathrm{conv}\left(\bigcap_{i=1}^m G_i\right) \cap H \\ S_p &= \left(\bigcap_{i=1}^m \mathrm{conv}\, G_i\right) \cap H \end{aligned}$$

Satz 6.7. $S_e \subseteq S_d \subseteq S_p$.

Beweis. $\mathrm{conv}\,(A \cap B) \subseteq \mathrm{conv}\,(A) \cap \mathrm{conv}\,(B)$ □

Satz 6.8 (Lemárechal - Renauld 2001). *1.* $\mathrm{val}\,(G) = \mathrm{val}\,(R_e) = \min\{x_1 | \, x \in S_e\}$

2. $\mathrm{val}\,(R_d) = \min\{x_1 | \, x \in S_d\}$

3. $\mathrm{val}\,(R_p) = \min\{x_1 | \, x \in S_p\}$ $(\mathrm{conv}\, G_i = \{(x, t \cdot e) | \, x \in X, \, \mathrm{conv}_X\, g_i(x) \le t\})$

$$
\begin{array}{lll}
(R_c) & \min & \sum_{k=1}^{n+1} \tau_k \cdot x_1^k \\
 & \text{s.t.} & \sum_{k=1}^{n+1} \tau_k \cdot g\left(x^k\right) \le 0 \\
 & & x^k \in X, \quad k = 1, \ldots, n+1 \\
 & & \tau \in \Delta_{n+1} = \left\{ \tau \in \mathbb{R}^{n+1} \,\middle|\, \tau \ge 0, \sum_{i=1}^{n+1} \tau_i = 1 \right\}
\end{array}
$$

Satz 6.9 (Kiwiel - Feltenmark). *(R_c) ist konvex und* $\mathrm{val}\,(R_d) = \mathrm{val}\,(R_c)$.

Beweis. L_c sei die Lagrangefunktion zu (R_c)

$$
\begin{aligned}
D_c(\mu) &= \min_{x^k \in X, \tau \in \Delta_{n+1}} L_c\left(\underbrace{x}_{(x^1, \ldots, x^{n+1})}, \tau; \mu \right) \\
&= \min_{\tau \in \Delta_{n+1}} \sum_{k=1}^{n+1} \cdot \underbrace{\min_{x^k \in X} L\left(x^k; \mu\right)}_{= D(\mu)} \\
&= D(\mu) \cdot \underbrace{\min_{\tau \in \Delta_{n+1}} \sum_{k=1}^{n+1} \tau_k}_{=1} \\
&= D(\mu)
\end{aligned}
$$

□

7 Schnitte und globale Optimalitätskriterien

7.1 Zulässige Schnitte

1. (Sahinidis et al)

$$
\begin{aligned}
\underline{f} = \min_{x \in X = [\underline{x}, \bar{x}]} L(x; \mu) &\le L(x; \mu) = f(x) + \mu^T g(x) \\
&\le f(\hat{x}) + \mu^T g(x) \text{ für zulässige } x \text{ mit } f(x) \le f(\hat{x}) =: f \\
&\le \bar{f} + \mu_i g_i(x) \\
\Leftrightarrow g_i(x) &\ge -\frac{1}{\mu_i}\left(\bar{f} - \underline{f}\right).
\end{aligned}
$$

2. $L(x; \mu) \le f(x) \le \bar{f}$, $f(x) \le \bar{f}$
3. $-g_i(x) \cdot g_j(x) \le 0$ (interessant für SDP)
4. MILP-Schnitte (Nemhauser, Wolsey)

7.2 Box Reduktion

Beispiel (Neumeier).

$$\begin{array}{ll} \min & f(x) = -x_1 - 2x_2 \\ \text{s.t.} & h(x) = (x_1-1)^2 + (x_2-1)^2 = 1 \\ & x_1, x_2 \in [-1,1] \end{array}$$

$f(x) \leq f(\hat{x}) = -2 \Rightarrow x_1 \in [0,1]$, $x_2 \in [0.5,1]$; aus $h(x) = 0 \Rightarrow x_1 \in [0, 0.14]$ ...

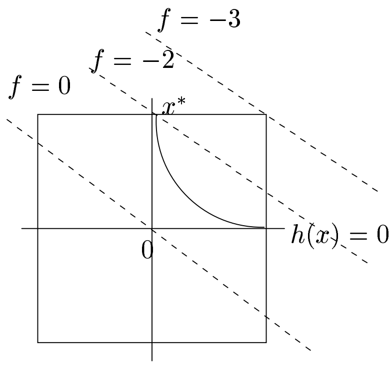

Abbildung 7.1: $\min_{x_i \in [-1,1], (x_1-1)^2+(x_2-1)^2=1} -x_1 - 2x_2$

Allgemein:

$$(B) \quad \tilde{x} := \begin{array}{ll} \min & x_i \\ \text{s.t.} & g(x) \leq 0 \\ & x \in [\underline{x}, \bar{x}] \end{array}$$

Falls $\tilde{x}_i > \bar{x}_i$, kann $\bar{x}_i$ durch $\tilde{x}_i$ ersetzt werden.

Spezialfall: $g(x) = a^T x + b$

Ansatz:

$$a^T x + b = 0 \Leftrightarrow x_i = -\frac{1}{a_i}\left(\sum_{j \neq i} a_j x_j + b\right)$$

(B) mit Gleichungsrestriktionen

$$\Leftrightarrow \begin{array}{ll} \min & -\frac{1}{a_i}\left(\sum_{i \neq j} a_j x_j + b\right) \\ \text{s.t.} & x_j \in \left[\underline{x_j}, \bar{x}_j\right], j \neq i \end{array}$$

$$\Leftrightarrow \sum_{j \neq i} \begin{array}{ll} \min & -\frac{a_j}{a_i} x_j - \frac{b}{a_i} \\ \text{s.t.} & x_j \in \left[\underline{x_j}, \bar{x}_j\right], j \neq i \end{array}$$

Constraint programming: erzeugen von neuen Restriktionen + Box-Reduzierung

Der Prozess heißt constraint-propagation.

Eine Box heißt konsistent bzgl. einer constraint-propagation, wenn sie ein Fixpunkt der Iteration ist.

7.3 Globale Optimalitätskriterien

1. starke Dualität (Dualitätslücke = 0)
2. Monotonie: Falls $\underbrace{f'(x) \geq 0 \, \forall x \in [\underline{x}, \overline{x}]}_{\min_{x\in[\underline{x},\bar{x}]} f'(x)\geq 0} \Rightarrow \underline{x}$ ist optimal für $\min \{ f(x) | \, x \in [\underline{x}, \bar{x}]\}$
3. Konvexität: Falls $\nabla^2 f(x) \succeq 0 \, \forall x \in [\underline{x}, \bar{x}]$, dann $\text{locmin}_{x\in[\underline{x},\bar{x}]} f(x) = \text{globmin}_{x\in[\underline{x},\bar{x}]} f(x)$
   z.B. mit α-Methoden von Floudas: $\nabla^2 f(x) + 2\alpha I \succeq 0$, $x \in [\underline{x}, \bar{x}]$
4. Prüfe Eindeutigkeit der Gleichung $\nabla L(x; \mu) = 0$ für $x \in [\underline{x}, \bar{x}]$, $\mu \in \mathcal{M}$
   $\Leftrightarrow \nabla^2 L(x; \mu)$ regulär, $x \in [\underline{x}, \bar{x}]$, $\mu \in \mathcal{M}$

7.4 Optimalitätsschnitte

Falls $\nabla^2 L(x; \mu) \succeq 0$, $x \in X$ und $\nabla L(\hat{x}; \mu) = 0$, $\hat{x}, \mu$ zulässig $\Rightarrow L(\hat{x}, \mu) = f(\hat{x})$.
$\hat{x} = \text{argmin}_{x\in[\underline{x},\bar{x}]} L(x; \mu) \Rightarrow \hat{x}$ ist globale Minimalstelle auf $[\underline{x}, \bar{x}]$.

8 Relaxierungsbasierte Heuristiken

8.1 Lagrange Heuristiken

1. Löse ein duales Problem $(D) \rightarrow$ duale Punkte, μ, und primale Punkte, x_μ (Lösung der Lagrange-Probleme)
2. Generiere Lösungen des primalen Problems (P) durch obige Lösungen.

Viele Anwendungen bei MILPs unter Verwendung von Dekomposition. Funktioniert gut, wenn die Dualitätslücken, $\text{val}(P) - \text{val}(D)$ klein sind.

8.2 "Smoothing" Heuristiken

1. Löse ein einfaches (glattes) Problem (H_0).
2. Löse (H_t) für $t \in \{t_1, \ldots, t_{\max}\}$ wobei $\text{val}(H_{t_{\max}}) \approx \text{val}(P)$ ist.

8.3 Multipath

Kombination aus Lagrange Heuristik und Smoothing Heuristik.

$$H(x; \mu; t) := (1 - t) L(x; \mu) + tP(x; t)$$

$$\min_{x\in[\underline{x},\bar{x}]} P(x; t) \xrightarrow{t\rightarrow 1} \text{val}(P)$$

Parametrisches Optimierungsproblem:

$$\begin{array}{lll} (H_{\mu,t}) & \min & H(x; \mu; t) \\ & \text{s.t.} & x \in [\underline{x}, \bar{x}] \end{array}$$

Es gilt $\text{val}(H_{\mu,t}) \xrightarrow{t\rightarrow 1} \text{val}(P)$

Algorithmus 8.1 (Multipath). 1. Bestimme $S = \{(\mu, x_\mu)\}$ durch Lösen von (D), wobei (D) ein einfaches duales Problem ist.

2. Vergrößere S durch Hinzunahme von neuen Punkten.

3. Löse (H_{t_1}) mit Startwerten aus S.

4. Lösche schlechte Kandidaten aus S.

5. Verfolge einen Pfad zu $(H_{\mu,t})$ von t_k bis t_{k+1} mit Startwerten aus S

6. (branching) Vergrößere S

7. Lösche schlechte Kandidaten

8. $k := k + 1$, gehe zu (5), falls Abbruchkriterium nicht erfüllt ist.

9. (lokale Minimierung) Für alle $x \in S$:

   (a) runde binäre Variablen

   (b) löse (P) lokal mit gerundetem x und Festhaltung der binären Variablen.

A MaxCut

Complexity (introduction)

Problem:

$$\begin{aligned} \text{LP} : z^* := \quad \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

size$(LP) = n + m + r$, r =#bits that represent $A, b, c = \log_2 a_j$, A a $n \times m$-Matrix, $x \in \mathbb{R}, \mathbb{Q}$, "$x \in \mathbb{Z}$"

$\mathcal{P}$: $P \in \mathcal{P}$ can be solved in polynomial time / with polynomial space in the size of P: size$(P)^\gamma$ ($\exists$ polytime Alg.)

$\mathcal{NP}$: Yes/No-Questions. $\mathcal{P} \subseteq \mathcal{NP}$, Quesition $\rightarrow$ Yes with (polynomiell) "certificate" or No

Beispiel.

1. Is $z^* \leq r$?

2. $G = (N, E)$ N: nodes, vertices, E: edges, arcs; Is there an Hamiltoian circuit?

Algorithms

- Exact deterministic: sure, that you get z^*, $O(e^n)$, useful for $\mathcal{NP}$ instances
- Approximation: $rz^a \leq z^* \leq z^a$ sure of quality of z^a with approximation factor r; usually: $r = \frac{1}{2}$
- Heuristics: not sure

Graph $G=(N,E)$, $S\subseteq N$, $\delta : P(N) \to E$ incidence function

$$\delta(S)=\{(i,j)|\, i\in S, j\notin S\}$$

A_G incidency matrix: $a_e=1$, if $e\in E$

W_G weight matrix: w_e cost of edge e

MaxCut:

$$\max_{S\subseteq N}\sum_{e\in\delta(S)} a_e$$

bipartite: $\exists S:\ \delta(S)=E$

$x\in\{-1,1\}^n$, $x_i^S=\begin{cases} 1 & i\in S\\ -1 & i\notin S\end{cases}$

$$\mathrm{cut}(S)=\frac{1}{2}\sum_{x_ix_j<0} a_{i,j}$$

$$x^TAx=\sum_{i,j}a_{i,j}x_ix_j=\sum_{x_ix_j>0}a_{i,j}-\sum_{x_ix_j<0}a_{i,j}=\sum_{i,j}a_{i,j}-2\sum_{x_ix_j<0}a_{i,j}=\underbrace{e^TAe}_{2|E|}-4\mathrm{cut}(S)$$

$$\begin{aligned}\mathrm{cut}(S) &= \frac{2|E|-x^TAx}{4}\\ \Rightarrow \max\mathrm{cut}_{S\subseteq N}(G) &= \frac{1}{4}\left(2|E|-\min_{x\in\{-1,1\}^n}x^TAx\right)\end{aligned}$$

$$\begin{array}{lll} \mathrm{Max-Cut}: & \min & x^TAx \qquad \to \text{ not convex, because } a_{i,i}=0\forall i\\ & s.t. & Xx=e,\quad X=\mathrm{diag}(x) \qquad \to \text{ non conex}\\ & & \text{or } x_i^2=1\forall i \Leftrightarrow x_i\in\{-1,1\}\end{array}$$

first relaxation: $Xx\le e$

SDP:

$$\begin{array}{ll}\min & A_0\bullet Z\\ s.t. & A_i\bullet Z=b_i\\ & Z\succeq 0\end{array}$$

MC-SDP: $E_{i,j}=e_ie_j^T$, $Z=xx^T$, $A\cdot Z=\sum_{i,j}a_{i,j}z_{i,j}=\sum_{i,j}a_{i,j}x_ix_j=x^TAx$

$$\begin{array}{ll}\min & A\bullet Z\\ s.t. & E_{i,i}\bullet Z=1\,\forall i\\ & \mathrm{rank}(Z)=1\\ & Z\succeq 0\end{array}$$

relaxation: removing rank$(Z)=1$: MC-SDP-R:

$$\begin{array}{ll}\min & A\cdot Z\\ s.t. & E_{i,i}\cdot Z=1\,\forall i\\ & Z\succeq 0\end{array}$$

Z^* optimal SDP-R, $Z^* = V^T V$ Choleski-factorization

$$u \sim U\left[B\left(0,1\right)\right], \hat{x} = \text{sign}\left(V^T u\right)$$

$$0.878 \cdot E\left[\text{cut}\left(\hat{x}\right)\right] \le \text{cut}\left(x^*\right), \text{cut}\left(x^*\right) \le \text{cut}\left(\hat{x}\right), x^* \text{ optimal cut}$$

back to MaxCut:

$$\begin{array}{rcl}
 & \min & x^T A x \\
 & s.t. & Xx = e\ (\mu) \\
L\left(x;\mu\right) & = & x^T A x + \mu\left(Xx - e\right) \\
 & = & x^T\left(A+M\right)x - e^T\mu \quad M = \text{Diag}\left(\mu\right) \\
D\left(\mu\right) & = & \inf_{x\in\mathbb{R}} x^T\left(A+M\right)x - e^T\mu \quad A+M \not\succeq 0 \to \exists v : v^T\left(A+M\right)v < 0 \\
 & = & \begin{cases} -\infty & \text{if } A+M \not\succeq 0 \\ -e^T\mu & \end{cases} \\
 & \max_{\mu\in\mathbb{R}^n} & D\left(\mu\right) \\
\to & \max & -e^T\mu \\
 & s.t. & A+M \succeq 0
\end{array}$$

D-SDP:

$$\begin{array}{rcl}
 & \min & \sum_i \mu_i \\
 & s.t. & \underbrace{\sum E_{i,i}\mu_i + A}_{M+A} \succeq 0 \\
\mu^* e & = & A \cdot Z^*
\end{array}$$