7c. Nichtlineare Optimierung
Algorithmen

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Penalty-Verfahren
- Augmented-Lagrangian-Verfahren
- Barriere-Verfahren

Quadratische Straffunktion

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $h_i(x) = 0, \quad i = 1, \dots, m$

Strafparameter > 0

Minimiere $Q(x;\beta) = f(x) + \frac{\beta}{2} \sum_{i=1}^{m} h_i^2(x)$ über $x \in \mathbb{R}^n$

Quadratische Straffunktion

- Bedingungsverletzungen werden bestraft
- Geht $\beta \to \infty$, so nähert sich das unrestringierte Problem an das ursprüngliche restringierte Problem an

Quadratische Straffunktion

Minimiere $Q(x;\beta) = f(x) + \frac{\beta}{2}\sum_{i=1}^{m} h_i^2(x)$ über $x \in \mathbb{R}^n$

Sei $\beta^1, \beta^2, \ldots$ eine Folge mit $\beta^k \to \infty$ und $x^k = \operatorname{argmin}_x Q(x;\beta^k)$

- Liegt β^k nah an β^{k-1}, so kann man x^k schnell ausgehend von x^{k-1} finden
- Ist die Berechnung von $\operatorname{argmin}_x Q(x;\beta^k)$ aufwendig, so wählt man β^k ein bisschen größer als β^{k-1}, z.B. $\beta^k = 1.5\beta^{k-1}$
- Ansonsten kann man größere Schritte wählen, z.B. $\beta^k = 10\beta^{k-1}$

Satz 7.13. Konvergenz

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $h_i(x) = 0, \ \ i = 1, \dots, m$

$f, h_i \in C(\mathbb{R}^n)$

$$Q(x;\beta) = f(x) + \frac{\beta}{2}\sum_{i=1}^{m} h_i^2(x)$$

- Angenommen, das restringierte Problem besitzt eine optimale Lösung x^*
- Sei $\beta^1, \beta^2, \dots$ eine Folge mit $\beta^k \to \infty$ und sei $x^k = \operatorname{argmin}_x Q(x;\beta^k)$
- Gelte $x^k \to \bar{x}$, so ist $\bar{x}$ eine optimale Lösung

Beweis

$$x^k = \mathrm{argmin}_x Q(x, \beta^k)$$

$$Q(x^k, \beta^k) \leq Q(x^*, \beta^k)$$

$$f(x^k) + \frac{\beta^k}{2} \sum_{i=1}^m h_i^2(x^k) \leq f(x^*) + \frac{\beta^k}{2} \underbrace{\sum_{i=1}^m h_i^2(x^*)}_{} = f(x^*)$$

$= 0$

$k \to \infty$

$$f(\bar{x}) \leq f(x^*)$$

$$\sum_{i=1}^m h_i^2(x^k) \leq \frac{2}{\beta^k}\left(f(x^*) - f(x^k)\right)$$

$k \to \infty$

$$\sum_{i=1}^m h_i^2(\bar{x}) = 0$$

$\square$

Bemerkung

- Man kann die Konvergenz gegen eine optimale Lösung nur dann garantieren, wenn $x^k = \operatorname{argmin}_x Q(x, \beta^k)$ für alle k
- Es ist einfacher, $\left\|\nabla_x Q\left(x^k, \beta^k\right)\right\|_2 \leq \tau^k$ mit $\tau^k \to 0$ zu garantieren
- Wir werden die Bedingungen herleiten, die die Konvergenz von x^k (und gewissen Lagrange-Multiplikatoren) gegen einen KKT-Punkt garantieren

Das Penalty-Verfahren

1. *Initialisierung:* Startwert $x^0 \in \mathbb{R}^n$, *Anfangsstrafparameter* $\beta^1 > 0$
   Toleranzwerte (τ^k) mit $\tau^k \to 0$
2. **for** $k = 1,2,3,\ldots$ **do**
3.   **if** *Abbruchskriterium ist erfüllt* **then** STOP
4.   bestimme x^k, sodass $\left\|\nabla_x Q\left(x^k, \beta^k\right)\right\|_2 \leq \tau^k$, ausgehend von x^{k-1}
5.   bestimme einen Strafparameter $\beta^{k+1} \geq \beta^k$
6. **end for**

Satz 7.14. Bestimmung eines KKT-Punktes

Seien $f, h_i \in C^1(\mathbb{R}^n)$. Angenommen:

- Seien $(\tau^k), (\beta^k), (x^k)$ die Folgen mit $\tau^k \to 0, \beta^k \to \infty$ so, dass:
$$\left\| \nabla_x Q\left(x^k; \beta^k\right) \right\|_2 \leq \tau^k$$
- $x^k \to \bar{x}$, wobei $\bar{x}$ ist so, dass $\{\nabla h_i(\bar{x})\}_{i=1,\dots,m}$ linear unabhängig sind

Dann $\lambda^k = \beta^k h\left(x^k\right) \to \bar{\lambda}$, sodass $(\bar{x}, \bar{\lambda})$ ein KKT-Punkt ist

Beweis: $\bar{x}$ ist zulässig

$$\nabla_x Q(x^k, \beta^k) = \nabla f(x^k) + \beta^k \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k)$$

$$\left\| \nabla f(x^k) + \beta^k \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k) \right\|_2 \leq \tau^k$$

$$\left\| \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k) \right\|_2 \leq \frac{1}{\beta^k} \left(\tau^k + \left\| \nabla f(x^k) \right\|_2 \right)$$

$k \to \infty$

$$\sum_{i=1}^m h_i(\bar{x}) \nabla h_i(\bar{x}) = 0$$

$\{\nabla h_i(\bar{x})\}_i$ linear unabhängig

$$h_i(\bar{x}) = 0, \; i = 1, \dots, m$$

$\Rightarrow$ $\bar{x}$ ist zulässig

Beweis: (x^k, λ^k) strebt gegen einen KKT-Punkt $(\bar{x}, \bar{\lambda})$

$$\nabla_x Q(x^k, \beta^k) = \nabla f(x^k) + \beta^k \sum_{i=1}^n h_i(x^k)\, \nabla h_i(x^k)$$

$$A(x)^T := [\nabla h_1(x), \cdots, \nabla h_m(x)]$$

$$\lambda^k = \beta^k h(x^k)$$

$$\nabla_x Q(x^k, \beta^k) = \nabla f(x^k) + A(x^k)^T \lambda^k$$

$$\operatorname{rang}(A(\bar{x})^T) = m$$

$$\lambda^k = \left(A(x^k)A(x^k)^T\right)^{-1}\left(\nabla Q_x(x^k, \beta^k) - \nabla f(x^k)\right) \text{ für großes } k$$

$$\lambda^k \to \bar{\lambda} = -(A(\bar{x})A(\bar{x})^T)^{-1}\nabla f(\bar{x})$$

$$\nabla f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla h_i(\bar{x}) = 0$$

$\square$

Allgemeine Nebenbedingungen

$$\begin{aligned} \text{Minimiere} \;\; & f(x) \;\; \text{über } x \in \mathbb{R}^n \\ \text{u.d.N.} \;\; & h_i(x) = 0, \;\; i = 1, \dots, m_G \\ & g_i(x) \leq 0, \;\; i = 1, \dots, m_U \end{aligned}$$

$$\text{Minimiere} \;\; Q(x;\beta) = f(x) + \frac{\beta}{2} \sum_{i=1}^{m_G} h_i^2(x) + \frac{\beta}{2} \sum_{i=1}^{m_U} ([g_i(x)]^-)^2$$

$$[y]^- = \max\{-y, 0\}$$

- Die Straffunktion kann weniger glatt als die Zielfunktion und die Beschränkungsfunktionen sein

$$\max\{-x, 0\} \notin C^1(\mathbb{R})$$

Plan

- Penalty-Verfahren
- Augmented-Lagrangian-Verfahren
- Barriere-Verfahren

Penalty-Verfahren und systematischer Fehler

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $h_i(x) = 0, \quad i = 1, \dots, m$

- Das Penalty-Verfahren erzeugt eine Folge (x^k, λ^k) die gegen einen KKT-Punkt $(\bar{x}, \bar{\lambda})$ konvergiert

$$\lambda^k = \beta^k h(x^k), \quad \beta^k \to \infty$$

$$h(x^k) = \lambda^k / \beta^k$$ *systematischer Fehler*

- x^k wird im Allgemeinen nur für $\beta^k \to \infty$ zulässig sein

Penalty-Verfahren und systematischer Fehler

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $h_i(x) = 0, \;\; i = 1, \ldots, m$

- Sei x^* eine optimale Lösung, die die LICQ erfüllt
- Sei λ^* der zugehörige Vektor der Lagrange-Multiplikatoren, sodass (x^*, λ^*) ein KKT-Punkt ist (Sätze 7.5, 7.7)

Augmented-Lagrange-Funktion

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $h_i(x) = 0, \quad i = 1, \ldots, m$

Strafparameter

Minimiere $\mathcal{L}_A(x, \lambda^k; \beta^k) = f(x) + \sum_{i=1}^m \lambda_i^k h_i(x) + \frac{\beta^k}{2} \sum_{i=1}^m h_i^2(x)$ über $x \in \mathbb{R}^n$

aktuelle Approximation an λ^*

Augmented-Lagrange-Funktion

Augmented-Lagrange-Funktion

Minimiere $\mathcal{L}_A(x, \lambda^k; \beta^k) = f(x) + \sum_{i=1}^m \lambda_i^k h_i(x) + \frac{\beta^k}{2} \sum_{i=1}^m h_i^2(x)$ über $x \in \mathbb{R}^n$

Sei x^{k+1} eine optimale Lösung

$$\nabla \mathcal{L}_A(x^{k+1}, \lambda^k; \beta^k) = \nabla f(x^{k+1}) + \sum_{i=1}^m \left(\underbrace{\lambda_i^k + \beta^k h_i(x^{k+1})}\right) \nabla h_i(x^{k+1}) = 0$$

- Der KKT-Punkt (x^*, λ^*) erfüllt:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0$$

$$\lambda_i^{k+1} := \lambda_i^k + \beta^k h_i(x^{k+1})$$

Neue Approximation

$$\lambda_i^{k+1} = \lambda_i^k + \beta^k h_i(x^{k+1})$$

$$h_i\left(x^{k+1}\right) = (\lambda_i^{k+1} - \lambda_i^k)/\beta_k$$

- Ist $\lambda_i^k \to \lambda_i^*$, so $h_i\left(x^k\right) = o(1/\beta_k)$

  schneller als für das Penalty-Verfahren

- Unter gewissen allgemeinen Voraussetzungen $\exists \bar{\beta} > 0$ so, dass x^* ein striktes lokales Minimum von $\mathcal{L}_A(x, \lambda^*; \beta)$ ist $\forall \beta \geq \bar{\beta}$ [NW, Thm. 17.5]

  $\beta^k \to \infty$ *ist nun nicht erforderlich*

  λ^* *ist in der Realität nicht bekannt*

Konvergenz [NW, Thm. 17.6]

- Unter gewissen allgemeinen Voraussetzungen und für alle (x^0, λ^0) hinreichend nah an (x^*, λ^*) gibt es $M > 0$ so, dass:

$$\left\|\lambda^{k+1} - \lambda^*\right\|_2 \leq M\left\|\lambda^k - \lambda^*\right\|_2 / \beta_k \qquad k = 1,2, \ldots$$

$$\left\|x^k - x^*\right\|_2 \leq M\left\|\lambda^k - \lambda^*\right\|_2 / \beta_k \qquad k = 1,2, \ldots$$

Das Augmented-Lagrangian-Verfahren

1. *Initialisierung:* Startwerte $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$, *Strafparameter* $\beta^0 > 0$
2. **for** $k = 0,1,2,\dots$ **do**
3.   **if** *Abbruchskriterium ist erfüllt* **then** STOP
4.   bestimme $x^{k+1} = \mathrm{argmin}_x \mathcal{L}_A(x, \lambda^k; \beta^k)$ ausgehend von x^k
5. $\lambda^{k+1} := \lambda^k + \beta^k h(x^{k+1})$
6.   bestimme einen Strafparameter $\beta^{k+1} \geq \beta^k$
7. **end for**

Plan

- Penalty-Verfahren
- Augmented-Lagrangian-Verfahren
- Barriere-Verfahren

Ungleichungsrestringiertes Problem

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $g_i(x) \leq 0, \ i = 1, \dots, m$

$f, g_i \in C^2(\mathbb{R}^n)$ konvex

- Sei die Slater-Bedingung erfüllt:

  $\exists \bar{x} \in \mathbb{R}^n$ mit $g_i(\bar{x}) < 0, \mathrm{i} = 1, \dots, m$

- Sei der optimale Funktionswert endlich: $f^* > -\infty$

Logarithmische-Barriere-Funktion

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $g_i(x) \leq 0, \; i = 1, \ldots, m$

$$\phi(x) = -\sum_{i=1}^{m} \log\left(-g_i(x)\right)$$

Zentrierungsproblem

$$U = \{x : g_i(x) < 0, \mathrm{i} = 1, \ldots, m\}$$

offen, nichtleer

Minimiere $f(x) + \frac{1}{t}\phi(x)$ über $x \in U$

Anpassungsparameter > 0

Ableitungen der Barriere-Funktion

$$\phi(x) = -\sum_{i=1}^{m} \log\big(-g_i(x)\big)$$

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

Zentraler Pfad

Minimiere $f(x) + (1/t)\phi(x)$ über $x \in U$

- Angenommen, es gibt eine eindeutige Lösung $x^*(t)$ für alle $t > 0$
- $x^*(t)$ heißt zentraler Punkt und $\{x^*(t): t > 0\}$ zentraler Pfad
- Optimalitätsbedingung:

$$\nabla f\big(x^*(t)\big) + (1/t)\nabla\phi\big(x^*(t)\big) = 0$$

$$\nabla f\big(x^*(t)\big) + \sum_{i=1}^{m} \underbrace{\boxed{\frac{1}{-t g_i\big(x^*(t)\big)}}}_{=:\ \mu_i^*(t)} \nabla g_i\big(x^*(t)\big) = 0$$

Erinnerung: Dualität

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $g_i(x) \leq 0,\ i = 1, \dots, m$

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

$$G(\mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu)$$

Maximiere $G(\mu)$ über $\mu \in \mathbb{R}^m$

u.d.N. $G(\mu) > -\infty$ $\quad\mu \geq 0$

Slater-Bedingung

Satz 7.15. ε-Optimalität des zentralen Punkts

Sei $x^*(t)$ ein zentraler Punkt und setze:

$$\mu_i^*(t) = \frac{1}{-t g_i(x^*(t))}, \; i = 1, \dots m$$

Es gelten die folgenden Aussagen:

- $x^*(t), \mu^*(t)$ sind zulässige Primal- bzw. Dualpunkte und es gilt:

$$f\big(x^*(t)\big) - G\big(\mu^*(t)\big) = m/t$$

- Folglich sind $x^*(t), \mu^*(t)$ ε-optimal mit $\varepsilon = m/t$:

$$f\big(x^*(t)\big) - f^* \leq m/t$$
$$f^* - G\big(\mu^*(t)\big) \leq m/t$$

Beweis: $\mu^*(t)$ ist dual zulässig, d.h. $G(\mu^*(t)) > -\infty$

$$\nabla f(x^*(t)) + \sum_{i=1}^{m} \underbrace{\boxed{\frac{1}{-t g_i(x^*(t))}}}_{=\,\mu_i^*(t)} \nabla g_i(x^*(t)) = 0$$

f, g_i konvex

$$\nabla f(x^*(t)) + \sum_{i=1}^{m} \mu_i^*(t) \nabla g_i(x^*(t)) = 0$$

$x^*(t)$ minimiert $\mathcal{L}(x, \mu^*(t)) = f(x) + \sum_{i=1}^{m} \mu_i^*(t) g_i(x)$

$$G(\mu^*(t)) = \mathcal{L}(x^*(t), \mu^*(t)) > -\infty$$

$\mu^*(t)$ ist dual zulässig

Beweis

$$G\big(\mu^*(t)\big) = \mathcal{L}\big(x^*(t), \mu^*(t)\big)$$

$$= f(x^*(t)) + \textstyle\sum_{i=1}^{m} \mu_i^*(t) g_i(x^*(t))$$

$$\mu_i^*(t) = \frac{1}{-t g_i\big(x^*(t)\big)}$$

$$= f\big(x^*(t)\big) - \frac{m}{t}$$

gewünschte Formel

Schwache Dualität (Satz 5.13):

$$f\big(x^*(t)\big) \geq f^* \geq G(\mu^*(t))$$

$$f\big(x^*(t)\big) - f^* \leq m/t$$

$$f^* - G\big(\mu^*(t)\big) \leq m/t$$

□

Das Barriere-Verfahren

1. *Initialisierung:* Slater-Punkt $x^0 \in \mathbb{R}^n$, Parameter $t^0 > 0$, Toleranzwert ε
2. **for** $k = 0,1,2,\ldots$ **do**
3. bestimme $x^{k+1} = \operatorname{argmin}_x(t^k f(x) + \varphi(x))$ ausgehend von x^k
4. **if** $m/t^k < \varepsilon$ **then**: STOP
5. wähle $t^{k+1} > t^k$
6. **end for**

Zentrierungsschritt

Das Barriere-Verfahren

1. *Initialisierung:* Slater-Punkt $x^0 \in \mathbb{R}^n$, Parameter $t^0 > 0$, Toleranzwert ε
2. **for** $k = 0,1,2,\ldots$ **do**
3. bestimme $x^{k+1} = \operatorname{argmin}_x (t^k f(x) + \varphi(x))$ ausgehend von x^k
4. **if** $m/t^k < \varepsilon$ **then**: STOP
5. wähle $t^{k+1} > t^k$
6. **end for**

Zentrierungsschritt

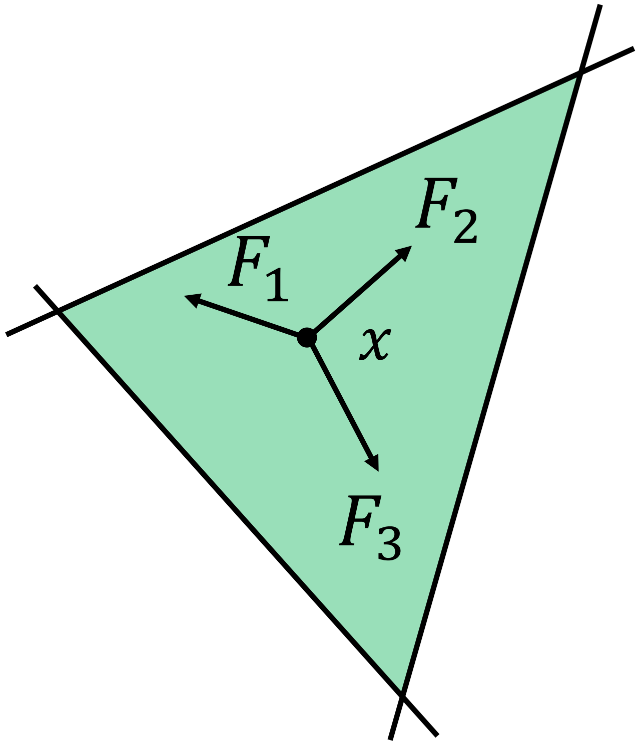
für große t^k kann schlecht konditioniert sein

x^{k+1} muss hinreichend nah an x^k sein

Interpretation: Kräfteausgleich

$$\text{Minimiere } f(x)$$
$$\text{u.d.N. } g_i(x) \leq 0,\ i = 1, \dots, m$$

$$\text{Minimiere } tf(x) + \underbrace{\textstyle\sum_{i=1}^{m} -\log\big(-g_i(x)\big)}_{\varphi}$$



- i-ter UNB entspricht die Kraft:

$$F_i(x) = \nabla \log\big(-g_i(x)\big) = \frac{\nabla g_i(x)}{g_i(x)}$$

- Die logarithmische Barriere-Funktion φ ist das skalare Potenzial, das die resultierende Kraft der UNB erzeugt

Interpretation: Kräfteausgleich

Minimiere $tf(x) + \sum_{i=1}^{m} -\log(-g_i(x))$

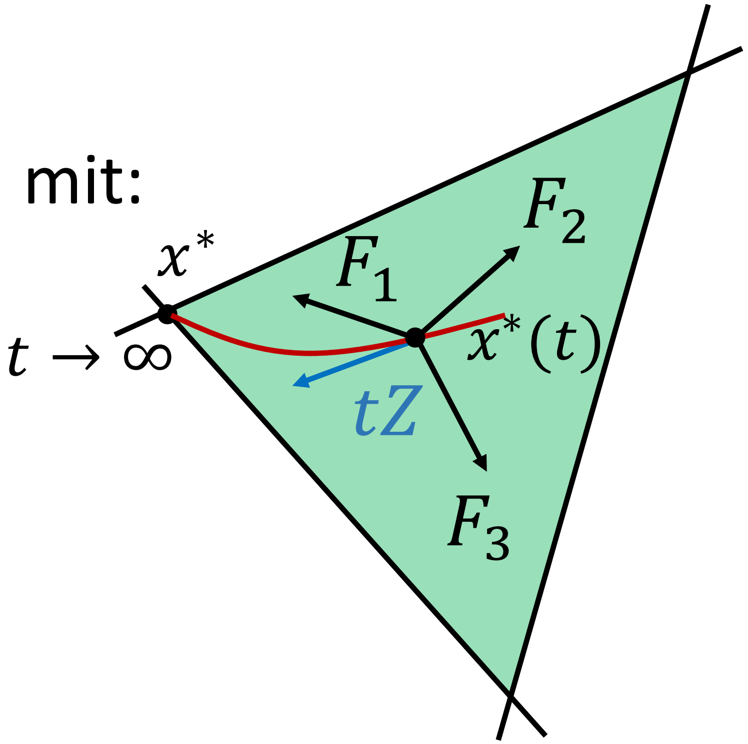
- Der Zielfunktion entspricht das Kraftfeld $tZ(x)$ mit:

$$Z(x) = -\nabla f(x)$$

- Optimalitätsbedingung:

$$\underbrace{-t\nabla f(x^*(t))}_{tZ(x^*(t))} + \underbrace{\sum_{i=1}^{m_U} \frac{\nabla g_i(x^*(t))}{g_i(x^*(t))}}_{F_i(x^*(t))} = 0$$

Die Kräfte sind im Gleichgewicht



Zusammenfassung

- Penalty-Verfahren
- Augmented-Lagrangian-Verfahren
- Barriere-Verfahren

Nächstes Video

- 8a. Diskrete Optimierung: Netzwerkflussprobleme