Regularisierungsverfahren

Katayoun Chaman Ara

Seminar: Nichtlineare Optimierung

Übersicht

1. Einleitung
2. Moreau-Yosida-Regularisierung
3. Proximal-Punkt-Verfahren
4. Tikhonov-Regularisierung
5. Programmieraufgabe

Einleitung

Beispielproblem

Gegeben sei die nach unten halbstetige konvexe Funktion:

$$f(x) = \begin{cases} x & \text{, falls } x \geq 0 \\ \infty & \text{, falls } x < 0 \end{cases}$$

Beispielproblem

Gegeben sei die nach unten halbstetige konvexe Funktion:

$$f(x) = \begin{cases} x & \text{, falls } x \geq 0 \\ \infty & \text{, falls } x < 0 \end{cases}$$

$\Rightarrow$ Regularisierungsverfahren,
um die Kondition des Optimierungsproblems zu verbessern

Voraussetzung an Optimierungsfunktion

$\min\ f(x)$, für $x \in \mathbb{R}^n$

$f : \mathbb{R}^n \longrightarrow \mathbb{R} \bigcup \{+\infty\}$

- f ist konvex

Voraussetzung an Optimierungsfunktion

$\min f(x)$, für $x \in \mathbb{R}^n$

$f : \mathbb{R}^n \longrightarrow \mathbb{R} \bigcup \{+\infty\}$

- f ist konvex
- f ist durchgehend echt

Voraussetzung an Optimierungsfunktion

min $f(x)$, für $x \in \mathbb{R}^n$

$f : \mathbb{R}^n \longrightarrow \mathbb{R} \bigcup \{+\infty\}$

- f ist konvex
- f ist durchgehend echt
- f ist nach unten halbstetig

Voraussetzung an Optimierungsfunktion

$\min f(x)$, für $x \in \mathbb{R}^n$

$f : \mathbb{R}^n \longrightarrow \mathbb{R} \bigcup \{+\infty\}$

- f ist konvex
- f ist durchgehend echt
- f ist nach unten halbstetig

$\Rightarrow$ Ziel: Stetig, differenzierbares Optimierungsproblem

Moreau-Yosida-Regularisierung

Definition

$$f_M(x) := \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\gamma} \|y - x\|^2 \right\}$$

mit einer gegebenen Konstante $\gamma > 0$.

Definition

$$f_M(x) := \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\gamma} \|y - x\|^2 \right\}$$

mit einer gegebenen Konstante $\gamma > 0$.

$$g(x, y) := f(y) + \frac{1}{2\gamma} \|y - x\|^2$$

bezeichne die Moreau-Yosida-Funktion.

Beispiel

Zurück zum Anfangsbeispiel

$$g(x,y) = \begin{cases} y + \frac{1}{2\gamma}(y-x)^2 & \text{,falls } y \geq 0 \\ \infty & \text{, falls } y < 0 \end{cases}$$

Beispiel

Zurück zum Anfangsbeispiel

$$g(x,y) = \begin{cases} y + \frac{1}{2\gamma}(y-x)^2 & ,\text{falls } y \geq 0 \\ \infty & ,\text{ falls } y < 0 \end{cases}$$

Für festes x nimmt g Minimum im Punkt

$$p(x) = \begin{cases} 0 & ,\text{ falls } x < \gamma \\ x - \gamma & ,\text{ falls } x \geq \gamma \end{cases}$$

an.

Beispiel

Zurück zum Anfangsbeispiel

$$g(x,y) = \begin{cases} y + \frac{1}{2\gamma}(y-x)^2 & \text{,falls } y \geq 0 \\ \infty & \text{, falls } y < 0 \end{cases}$$

Für festes x nimmt g Minimum im Punkt

$$p(x) = \begin{cases} 0 & \text{, falls } x < \gamma \\ x - \gamma & \text{, falls } x \geq \gamma \end{cases}$$

an. Damit ist:

$$f_M(x) = p(x) + \frac{1}{2\gamma}(p(x) - x)^2 = \begin{cases} \frac{1}{2\gamma}x^2 & \text{, falls } x < y \\ x - \frac{\gamma}{2} & \text{, falls } x \geq \gamma \end{cases}$$

Schritte zum Konvergenznachweis

- Für $f = f_1 + f_2$ gilt: $x^* \in \mathbb{R}^n$ Lösung
  $\Rightarrow \nabla f_2(x^*)^T(x - x^*) + f_1(x) - f_1(x^*) \geq 0$

Schritte zum Konvergenznachweis

- Für $f = f_1 + f_2$ gilt: $x^* \in \mathbb{R}^n$ Lösung
  $\Rightarrow \nabla f_2(x^*)^T(x - x^*) + f_1(x) - f_1(x^*) \geq 0$

- $p(x)$ eindeutige Lösung
  $\Rightarrow \|p(x) - p(y)\| \leq \|x - y\|$, für alle $x, y \in \mathbb{R}^n$

Schritte zum Konvergenznachweis

- Für $f = f_1 + f_2$ gilt: $x^* \in \mathbb{R}^n$ Lösung
  $\Rightarrow \nabla f_2(x^*)^T(x - x^*) + f_1(x) - f_1(x^*) \geq 0$

- $p(x)$ eindeutige Lösung
  $\Rightarrow \|p(x) - p(y)\| \leq \|x - y\|$, für alle $x, y \in \mathbb{R}^n$

- f_M ist stetig, differenzierbar mit
  $\nabla f_M(x) = \frac{1}{\gamma}(x - p(x))$, für alle $x \in \mathbb{R}^n$

Schritte zum Konvergenznachweis

- Für $f = f_1 + f_2$ gilt: $x^* \in \mathbb{R}^n$ Lösung
  $\Rightarrow \nabla f_2(x^*)^T(x - x^*) + f_1(x) - f_1(x^*) \geq 0$
- $p(x)$ eindeutige Lösung
  $\Rightarrow \|p(x) - p(y)\| \leq \|x - y\|$, für alle $x, y \in \mathbb{R}^n$
- f_M ist stetig, differenzierbar mit
  $\nabla f_M(x) = \frac{1}{\gamma}(x - p(x))$, für alle $x \in \mathbb{R}^n$

Dann gelten:

f_M konvex und $f(x^*) = f_M(x)$ in jedem Minimum von f (bzw. f_M)

Beobachtungen

- Zur Auswertung eines Punkts x von f_M muss $p(x)$ berechnet werden

Beobachtungen

- Zur Auswertung eines Punkts x von f_M muss $p(x)$ berechnet werden
- Kann zu nicht differenzierbarem Optimierungsproblem führen

Beobachtungen

- Zur Auswertung eines Punkts x von f_M muss $p(x)$ berechnet werden
- Kann zu nicht differenzierbarem Optimierungsproblem führen

$\Rightarrow$ Verwendung bei schlecht konditionierten, differenzierbaren Problemen

Beobachtungen

- Zur Auswertung eines Punkts x von f_M muss $p(x)$ berechnet werden
- Kann zu nicht differenzierbarem Optimierungsproblem führen

$\Rightarrow$ Verwendung bei schlecht konditionierten, differenzierbaren Problemen

$\Rightarrow$ Verwendung zur Herleitung des Proximal-Punkts $p(x)$

Proximal-Punkt-Verfahren

Einführung

Algorithmus

1. Wähle $x^0 \in dom(f)$, $k = 0$
2. Ist x^k Minimum von $f \longrightarrow$ STOP
3. Wähle $\gamma_k > 0$. Bestimme x^{k+1} globales Minimum von
   $f_k(x) = f(x) + \frac{1}{2\gamma_k} \left\| x - x^k \right\|^2$
4. Setze $k + 1 \longrightarrow k$. Gehe zum Schritt 2)

Schritte zum Konvergenznachweis

- (x^k) und (γ_k) erzeugte Folgen. $s^k := \frac{x^{k-1}-x^k}{\gamma_{k-1}}$
  $\Rightarrow s^k \in \partial f(x^k)$

Schritte zum Konvergenznachweis

- (x^k) und (γ_k) erzeugte Folgen. $s^k := \frac{x^{k-1}-x^k}{\gamma_{k-1}}$
  $\Rightarrow s^k \in \partial f(x^k)$
- $\Rightarrow \left(\left\|s^k\right\|\right)$ ist monoton fallend

Schritte zum Konvergenznachweis

- (x^k) und (γ_k) erzeugte Folgen. $s^k := \frac{x^{k-1}-x^k}{\gamma_{k-1}}$
  $\Rightarrow s^k \in \partial f(x^k)$
- $\Rightarrow \left(\left\|s^k\right\|\right)$ ist monoton fallend
- $(\sigma_k) : \sigma_k := \sum_{j=0}^{k} \gamma_j$
  $\Rightarrow f(x^k) - f(x) \leq \frac{\left\|x-x^0\right\|^2}{2\sigma_{k-1}} - \frac{\left\|x-x^k\right\|^2}{2\sigma_{k-1}} - \frac{\sigma_{k-1}}{2}\left\|s^k\right\|^2$

Schritte zum Konvergenznachweis

- (x^k) und (γ_k) erzeugte Folgen. $s^k := \frac{x^{k-1} - x^k}{\gamma_{k-1}}$
  $\Rightarrow s^k \in \partial f(x^k)$
- $\Rightarrow \left(\left\|s^k\right\|\right)$ ist monoton fallend
- $(\sigma_k) : \sigma_k := \sum_{j=0}^{k} \gamma_j$
  $\Rightarrow f(x^k) - f(x) \leq \frac{\left\|x - x^0\right\|^2}{2\sigma_{k-1}} - \frac{\left\|x - x^k\right\|^2}{2\sigma_{k-1}} - \frac{\sigma_{k-1}}{2} \left\|s^k\right\|^2$

Satz

Die Lösungmenge $S := \{x^* \in \mathbb{R}^n | f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ sei nichtleer und $\sigma_k \longrightarrow \infty$ für $k \longrightarrow \infty$.

Dann konvergiert (x^k) gegen ein Element aus S.

Beobachtungen

- $\gamma = \gamma_k$ konstant erfüllt $\sigma_k \longrightarrow \infty$

Beobachtungen

- $\gamma = \gamma_k$ konstant erfüllt $\sigma_k \longrightarrow \infty$
- (x^k) konvergiert unter Voraussetzungen gegen ein Minimum von f (nicht nur: jeder Häufungspunkt von (x_k) ist Minimum von f).

Tikhonov-Regularisierung

Einführung

Algorithmus

1. Wähle $x^0 \in dom(f)$, $k = 0$
2. Ist x^k Minimum von $f \longrightarrow$ STOP
3. Wähle $\epsilon_k > 0$. Bestimme x^{k+1} globales Minimum von $f_k(x) = f(x) + \frac{\epsilon_k}{2} \|x\|^2$
4. Setze $k + 1 \longrightarrow k$. Gehe zum Schritt 2)

Schritte zum Konvergenznachweis

- (x^k) und (ϵ_k) erzeugte Folgen. Gilt $\epsilon_k \downarrow 0$
  $\Rightarrow$ Jeder Häufungspunkt von (x^k) ist Lösung des Optimierungsproblems

Schritte zum Konvergenznachweis

- (x^k) und (ϵ_k) erzeugte Folgen. Gilt $\epsilon_k \downarrow 0$
  $\Rightarrow$ Jeder Häufungspunkt von (x^k) ist Lösung des Optimierungsproblems

- Sei $s^k := -\epsilon_{k-1} x^k$
  $\Rightarrow s^k \in \partial f(x^k)$

Schritte zum Konvergenznachweis

- (x^k) und (ϵ_k) erzeugte Folgen. Gilt $\epsilon_k \downarrow 0$
  $\Rightarrow$ Jeder Häufungspunkt von (x^k) ist Lösung des Optimierungsproblems
- Sei $s^k := -\epsilon_{k-1} x^k$
  $\Rightarrow s^k \in \partial f(x^k)$

Satz

Die Lösungmenge $S := \{x^* \in \mathbb{R}^n | f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ sei nichtleer und $\epsilon_k \downarrow 0$.

Dann konvergiert (x^k) gegen das eindeutig bestimmte kleinste Element in S.

Beobachtungen

- Folge (x^k) kann nicht beschränkt sein, wenn $S = \emptyset$

Beobachtungen

- Folge (x^k) kann nicht beschränkt sein, wenn $S = \emptyset$
- Wissen a priori gegen welches Element (x^k) aus S konvergiert

Beobachtungen

- Folge (x^k) kann nicht beschränkt sein, wenn $S = \emptyset$
- Wissen a priori gegen welches Element (x^k) aus S konvergiert
- Kondition der Teilprobleme unter Umständen beliebig schlecht, da $\epsilon_k \downarrow 0$

Beobachtungen

- Folge (x^k) kann nicht beschränkt sein, wenn $S = \emptyset$
- Wissen a priori gegen welches Element (x^k) aus S konvergiert
- Kondition der Teilprobleme unter Umständen beliebig schlecht, da $\epsilon_k \downarrow 0$

$\Rightarrow$ Theorie von Tikhonov-Verfahren geht für nichtglatte Probleme durch

Beobachtungen

- Folge (x^k) kann nicht beschränkt sein, wenn $S = \emptyset$
- Wissen a priori gegen welches Element (x^k) aus S konvergiert
- Kondition der Teilprobleme unter Umständen beliebig schlecht, da $\epsilon_k \downarrow 0$

$\Rightarrow$ Theorie von Tikhonov-Verfahren geht für nichtglatte Probleme durch

$\Rightarrow$ Praktische Bedeutung liegt in glatten, schlecht konditionierten Problemen

Programmieraufgabe

Aufgabe

- Implementiere das Gradientenverfahren und die Proximal-Punkt- bzw. Tikhonov-Regularisierung zur Lösung der Aufgabe zur optimalen Aufheizung

Aufgabe

- Implementiere das Gradientenverfahren und die Proximal-Punkt- bzw. Tikhonov-Regularisierung zur Lösung der Aufgabe zur optimalen Aufheizung
- Als Schrittweitenstrategie soll die exakte Schrittweite verwendet werden

Aufgabe

- Implementiere das Gradientenverfahren und die Proximal-Punkt- bzw. Tikhonov-Regularisierung zur Lösung der Aufgabe zur optimalen Aufheizung
- Als Schrittweitenstrategie soll die exakte Schrittweite verwendet werden
- Als Temperaturprofil soll

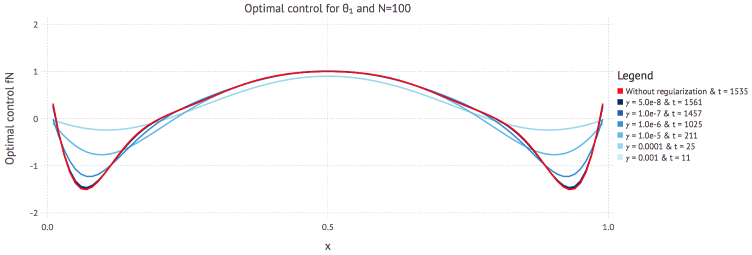
  1. $\theta(x) := x^2(1-x)^2$ und 2. $\theta(x) := \begin{cases} 1 & , 0.25 \leq x \leq 0.75 \\ 0 & , \text{sonst} \end{cases}$

Aufgabe

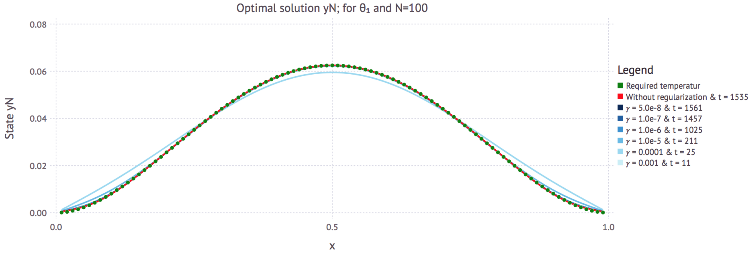
- Implementiere das Gradientenverfahren und die Proximal-Punkt- bzw. Tikhonov-Regularisierung zur Lösung der Aufgabe zur optimalen Aufheizung
- Als Schrittweitenstrategie soll die exakte Schrittweite verwendet werden
- Als Temperaturprofil soll

  1. $\theta(x) := x^2(1-x)^2$ und 2. $\theta(x) := \begin{cases} 1 & , 0.25 \leq x \leq 0.75 \\ 0 & , \text{sonst} \end{cases}$
- Teste für $n = 10, 100, 1000$

Plots

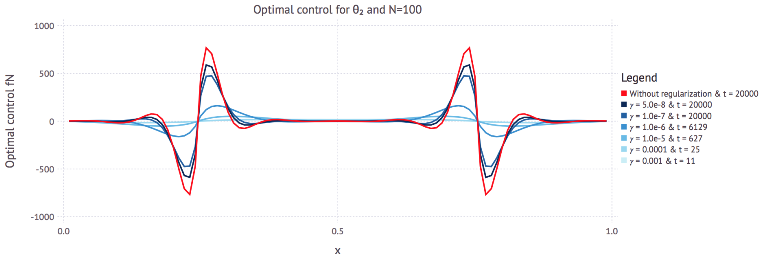


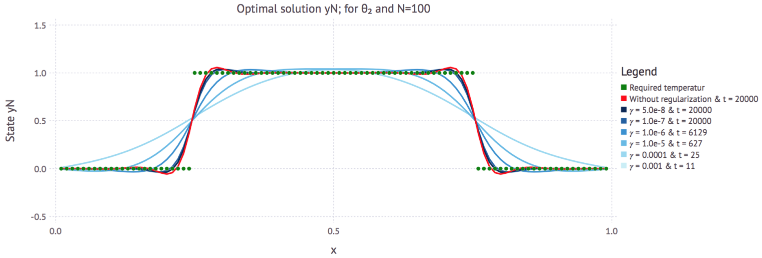
Optimal control for θ1 and N=100
Optimal control fN
2
1
0
-1
-2
0.0
0.5
1.0
x
Legend
Without regularization & t = 1535
γ = 5.0e-8 & t = 1561
γ = 1.0e-7 & t = 1457
γ = 1.0e-6 & t = 1025
γ = 1.0e-5 & t = 211
γ = 0.0001 & t = 25
γ = 0.001 & t = 11



Optimal solution yN; for θ1 and N=100
State yN
0.08
0.06
0.04
0.02
0.00
0.0
0.5
1.0
x
Legend
Required temperatur
Without regularization & t = 1535
γ = 5.0e-8 & t = 1561
γ = 1.0e-7 & t = 1457
γ = 1.0e-6 & t = 1025
γ = 1.0e-5 & t = 211
γ = 0.0001 & t = 25
γ = 0.001 & t = 11

Plots


Optimal control for θ_2 and N=100
Optimal control fN
1000
500
0
-500
-1000
0.0
0.5
1.0
x
Legend
Without regularization & t = 20000
γ = 5.0e-8 & t = 20000
γ = 1.0e-7 & t = 20000
γ = 1.0e-6 & t = 6129
γ = 1.0e-5 & t = 627
γ = 0.0001 & t = 25
γ = 0.001 & t = 11


Optimal solution yN; for θ_2 and N=100
State yN
1.5
1.0
0.5
0.0
-0.5
0.0
0.5
1.0
x
Legend
Required temperatur
Without regularization & t = 20000
γ = 5.0e-8 & t = 20000
γ = 1.0e-7 & t = 20000
γ = 1.0e-6 & t = 6129
γ = 1.0e-5 & t = 627
γ = 0.0001 & t = 25
γ = 0.001 & t = 11

Ende

Danke für die Aufmerksamkeit