Regularisierung

Alexander Steinicke
Montanuniversität Leoben

Vorlesung Numerische Methoden 1, 18.01.2024

Gewisse Beobachtungen, Probleme anschaulich in der Realität gestellt, führen (mittels mathematischer Modellierung) zu Operationen zwischen Funktionenräumen

Gewisse Beobachtungen, Probleme anschaulich in der Realität gestellt, führen (mittels mathematischer Modellierung) zu Operationen zwischen Funktionenräumen

$$\mathcal{A}\colon X \to Y, f \mapsto \mathcal{A}f = m,$$

Gewisse Beobachtungen, Probleme anschaulich in der Realität gestellt, führen (mittels mathematischer Modellierung) zu Operationen zwischen Funktionenräumen

$$\mathcal{A} \colon X \to Y, f \mapsto \mathcal{A}f = m,$$

dabei: $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ sind meistens normierte (siehe 2. VL) Vektorräume (klick) deren Elemente Funktionen sind (siehe nächste Folie)

Gewisse Beobachtungen, Probleme anschaulich in der Realität gestellt, führen (mittels mathematischer Modellierung) zu Operationen zwischen Funktionenräumen

$$\mathcal{A}\colon X \to Y, f \mapsto \mathcal{A}f = m,$$

dabei: $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ sind meistens normierte (siehe 2. VL) Vektorräume (klick) deren Elemente Funktionen sind (siehe nächste Folie)
$\mathcal{A}$...Operator, in vielen Fällen linear
m...Messung, Lösung des direkten Problems
f...Ursache von m, Lösung des inversen Problems

Die Räume $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ werden fast immer als Banachräume angenommen. Ein Banachraum $(E, \|\cdot\|_E)$ ist ein Vektorraum E mit Norm $\|\cdot\|_E$, in dem jede Cauchy-Folge [klick] auch gegen einen Grenzwert in E konvergiert (Vollständigkeit).

Die Räume $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ werden fast immer als Banachräume angenommen. Ein Banachraum $(E, \|\cdot\|_E)$ ist ein Vektorraum E mit Norm $\|\cdot\|_E$, in dem jede Cauchy-Folge [klick] auch gegen einen Grenzwert in E konvergiert (Vollständigkeit).

Beispiele: $(\mathbb{R}^n, \|\cdot\|_p), 1 < p \leq \infty$, $\Big(C([0, T]), h \mapsto \sup_{t\in[0,T]} |h|\Big)$, u.v.m.
$C(U)$ = stetige Funktionen auf $U \subseteq \mathbb{R}^n$.

Die Räume $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ werden fast immer als Banachräume angenommen. Ein Banachraum $(E, \|\cdot\|_E)$ ist ein Vektorraum E mit Norm $\|\cdot\|_E$, in dem jede Cauchy-Folge [klick] auch gegen einen Grenzwert in E konvergiert (Vollständigkeit).

Beispiele: $(\mathbb{R}^n, \|\cdot\|_p), 1 < p \leq \infty$, $\left(C([0,T]), h \mapsto \sup_{t\in[0,T]} |h|\right)$, u.v.m.
$C(U)$ = stetige Funktionen auf $U \subseteq \mathbb{R}^n$.

Nicht-Beispiele: $(\mathbb{Q}, |\cdot|)$, $\left(C([0,T]), h \mapsto \int_0^T |h(s)| ds\right)$
Bei beiden Beispielen ist die Vollständigkeit verletzt.

Ein Hilbertraum H ist ein spezieller Banachraum, in dem es zusätzlich noch ein Skalarprodukt (klick)

$$\langle \cdot, \cdot \rangle_H \colon H \times H \to \mathbb{R} \text{ oder } \mathbb{C}, (f, g) \mapsto \langle f, g \rangle_H$$

gibt, sodass $\|f\|_H^2 = \langle f, f \rangle_H$ gilt.
Diese Funktionenräume haben besonders schöne Eigenschaften.

Beispiele für Hilberträume: $(\mathbb{R}^n, (x, y) \mapsto \sum_{k=1}^n x_k y_k)$,

Beispiele für Hilberträume: $(\mathbb{R}^n, (x,y) \mapsto \sum_{k=1}^n x_k y_k)$,

$L^2(\mathbb{R}^d, \mathbb{C}) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty \right\}$, Beispiele für Hilberträume: $(\mathbb{R}^n, (x, y) \mapsto \sum_{k=1}^n x_k y_k)$,

$L^2(\mathbb{R}^d, \mathbb{C}) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty \right\}$,
mit $\langle f, g \rangle_L^2 := \int_{\mathbb{R}^d} f(x_1, \ldots, x_d) \overline{g(x_1, \ldots, x_d)} dx_1 \ldots dx_d$ ergibt sich ein Skalarprodukt, das $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ zu einem Hilbertraum macht.

Beispiele für Hilberträume: $(\mathbb{R}^n, (x,y) \mapsto \sum_{k=1}^n x_k y_k)$,

$L^2(\mathbb{R}^d, \mathbb{C}) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty \right\}$,
mit $\langle f, g \rangle_L^2 := \int_{\mathbb{R}^d} f(x_1, \ldots, x_d)\overline{g(x_1, \ldots, x_d)} dx_1 \ldots dx_d$ ergibt sich ein Skalarprodukt, das $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ zu einem Hilbertraum macht.

Die Norm auf $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ ist dann

$$\|f\|_{L^2} = \sqrt{\langle f, f \rangle} = \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d,$$

also genau, was in der Definition des Raumes steckt.

Beispiele für Hilberträume: $(\mathbb{R}^n, (x,y) \mapsto \sum_{k=1}^n x_k y_k)$,

$L^2(\mathbb{R}^d, \mathbb{C}) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty \right\}$,
mit $\langle f, g \rangle_L^2 := \int_{\mathbb{R}^d} f(x_1, \ldots, x_d) \overline{g(x_1, \ldots, x_d)} dx_1 \ldots dx_d$ ergibt sich ein Skalarprodukt, das $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ zu einem Hilbertraum macht.

Die Norm auf $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ ist dann

$$\|f\|_{L^2} = \sqrt{\langle f, f \rangle} = \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d,$$

also genau, was in der Definition des Raumes steckt.

Strenggenommen sollte man noch nicht integrierbare Funktionen ausschließen, sowie nichtnegative Funktionen, die die Norm 0 ergäben 'vernachlässigen'. Ist für unsere praktische Verwendung später aber nicht so wichtig.

Beispiele für Hilberträume: $(\mathbb{R}^n, (x,y) \mapsto \sum_{k=1}^n x_k y_k)$,

$L^2(\mathbb{R}^d, \mathbb{C}) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty \right\}$,
mit $\langle f, g \rangle_L^2 := \int_{\mathbb{R}^d} f(x_1, \ldots, x_d)\overline{g(x_1, \ldots, x_d)} dx_1 \ldots dx_d$ ergibt sich ein Skalarprodukt, das $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ zu einem Hilbertraum macht.

Die Norm auf $(L^2(\mathbb{R}^d, \mathbb{C}, \langle \cdot, \cdot_{L^2} \rangle)$ ist dann

$$\|f\|_{L^2} = \sqrt{\langle f, f \rangle} = \int_{\mathbb{R}^d} |f(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d,$$

also genau, was in der Definition des Raumes steckt.

Strenggenommen sollte man noch nicht integrierbare Funktionen ausschließen, sowie nichtnegative Funktionen, die die Norm 0 ergäben 'vernachlässigen'. Ist für unsere praktische Verwendung später aber nicht so wichtig.

Nicht-Beispiele: $(\mathbb{R}^2, \|\cdot\|_p), p \neq 2$, $\left(C([0, T]), h \mapsto \sup_{t \in [0,T]} |h| \right)$ sind Banachräume, aber keine Hilberträume

Die praktische Implementierung von Funktionenräumen geschieht meist durch hochdimensionale Vektoren, $(N \gg 3)$, die Funktionen approximieren sollten.

Die praktische Implementierung von Funktionenräumen geschieht meist durch hochdimensionale Vektoren, ($N \gg 3$), die Funktionen approximieren sollten.

Direkt als Vektor von Funktionswerten:
$E \to \mathbb{R}^N, \quad f \to (f_1, \dots, f_N)$

Die praktische Implementierung von Funktionenräumen geschieht meist durch hochdimensionale Vektoren, $(N \gg 3)$, die Funktionen approximieren sollten.

Direkt als Vektor von Funktionswerten:
$E \to \mathbb{R}^N, \quad f \to (f_1, \dots, f_N)$

per Basisdarstellung mit speziellen Basisfunktionen $\phi_1, \phi_2, \dots$ (z.B. sin und cos-Schwingungen zu steigenden Frequenzen wie bei den Fourierreihen, oder orthogonale Polynome wie etwa bei der numerischen Integration):

$$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$$

und $E \to \mathbb{R}^N, f \to (c_1, \dots, c_N)$

Nun können wir auf solchen Räumen Beispiele für Operatoren betrachten:

- Fourier-Transformation: $\mathcal{F}\colon L^2(\mathbb{R}^d, \mathbb{C}) \to L^2(\mathbb{R}^d, \mathbb{C})$

$$\underbrace{(x \mapsto f(x))}_{f} \mapsto \underbrace{\left(y \mapsto \int_{\mathbb{R}^d} f(x) e^{i\langle x,y\rangle}\, dx\right)}_{\mathcal{F}f}$$

Nun können wir auf solchen Räumen Beispiele für Operatoren betrachten:

- Fourier-Transformation: $\mathcal{F}\colon L^2(\mathbb{R}^d, \mathbb{C}) \to L^2(\mathbb{R}^d, \mathbb{C})$

$$\underbrace{(x \mapsto f(x))}_{f} \mapsto \underbrace{\left(y \mapsto \int_{\mathbb{R}^d} f(x) e^{i\langle x,y \rangle} dx\right)}_{\mathcal{F}f}$$

- Radon-Transformation: $\mathcal{R}\colon C(D) \to C([0,\pi) \times (-1,1))$
  (D hier: Unit disk)

$$f \mapsto \underbrace{\left((\varphi, s) \mapsto \int_{-\infty}^{\infty} f(s\theta + \tau\theta^{\perp}) d\tau\right)}_{\mathcal{R}f}$$

(θ ist hier eine von φ abhängige Richtung)

- Gravimetrie: $T: C([0,a]) \to C([0,a])$

$$u \mapsto w = Tu = \left(s \mapsto \int_0^a g \frac{hu(t)}{((s-t)^2 + h^2)^{3/2}} dt \right)$$

w...auf der Erde gemessene Vertikalkomponente der Schwerkraft entlang einer Strecke $[0,a]$
u...Masse entlang der Strecke $[0,a]$ in der Tiefe h

- Gravimetrie: $T: C([0,a]) \to C([0,a])$

$$u \mapsto w = Tu = \left(s \mapsto \int_0^a g \frac{hu(t)}{((s-t)^2 + h^2)^{3/2}} dt \right)$$

  w...auf der Erde gemessene Vertikalkomponente der Schwerkraft entlang einer Strecke $[0,a]$
  u...Masse entlang der Strecke $[0,a]$ in der Tiefe h

- Linear Seismic Tomography:

$$f \mapsto V_d = Tf = \left(t \mapsto -\frac{1}{\sigma_0} \int_o^t g(t-2s) f'(s) ds \right)$$

- Gravimetrie: $T: C([0,a]) \to C([0,a])$

$$u \mapsto w = Tu = \left(s \mapsto \int_0^a g \frac{hu(t)}{((s-t)^2 + h^2)^{3/2}} dt \right)$$

  w...auf der Erde gemessene Vertikalkomponente der Schwerkraft entlang einer Strecke $[0,a]$
  u...Masse entlang der Strecke $[0,a]$ in der Tiefe h
- Linear Seismic Tomography:

$$f \mapsto V_d = Tf = \left(t \mapsto -\frac{1}{\sigma_0} \int_o^t g(t-2s) f'(s) ds \right)$$

- Kanalschätzung, Signal-Entzerrung:

$$u \mapsto w = Tu = \left(s \mapsto \int_0^\ell g(t) u(s-t) dt = \int_s^{s-\ell} g(s-t) u(t) dt \right)$$

Allgemeine Form: f gegeben auf D

$$m(x) = (\mathcal{A}f)(x) = \int_D k(x,y)f(y)dy, \quad x \in D$$

(oder anderer Integrationsbereich)

Allgemeine Form: f gegeben auf D

$$m(x) = (\mathcal{A}f)(x) = \int_D k(x,y)f(y)dy, \quad x \in D$$

(oder anderer Integrationsbereich)

oft: Faltungsoperator (spezielle Form des Kerns k)

$$m(x) = (\mathcal{A}f)(x) = \int_D k(x-y)f(y)dy, \quad x \in D$$

Allgemeine Form: f gegeben auf D

$$m(x) = (\mathcal{A}f)(x) = \int_D k(x,y) f(y) dy, \quad x \in D$$

(oder anderer Integrationsbereich)

oft: Faltungsoperator (spezielle Form des Kerns k)

$$m(x) = (\mathcal{A}f)(x) = \int_D k(x-y) f(y) dy, \quad x \in D$$

Falls $\mathcal{A}\colon L^2(D) \to L^2(D)$ und $k \in L^2(D \times D)$
$\Rightarrow$ $\mathcal{A}$ so ist $\mathcal{A}$ ein sogenannter kompakter Operator

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*
- inverses Problem $\mathcal{A}f = m$ *ill-posed*,

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*
- inverses Problem $\mathcal{A}f = m$ *ill-posed*,
  d.h. inverser Operator $\mathcal{A}^{-1}$ falls existent linear aber unstetig ($\Leftrightarrow$ unbeschränkt) oder existiert nicht eindeutig

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*
- inverses Problem $\mathcal{A}f = m$ *ill-posed*,
  d.h. inverser Operator $\mathcal{A}^{-1}$ falls existent linear aber unstetig ($\Leftrightarrow$ unbeschränkt) oder existiert nicht eindeutig
- f anfällig für Störungen: Bei leicht gestörter Messung m^δ mit $\|m - m^\delta\|_Y < \delta$ ist für zugehörige Lösungen $\mathcal{A}f = m, \mathcal{A}f^\delta = m^\delta$ möglicherweise $\|f - f^\delta\|_X$ groß

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*
- inverses Problem $\mathcal{A}f = m$ *ill-posed*,
  d.h. inverser Operator $\mathcal{A}^{-1}$ falls existent linear aber unstetig
  ($\Leftrightarrow$ unbeschränkt) oder existiert nicht eindeutig
- f anfällig für Störungen: Bei leicht gestörter Messung m^δ mit
  $\|m - m^\delta\|_Y < \delta$ ist für zugehörige Lösungen
  $\mathcal{A}f = m, \mathcal{A}f^\delta = m^\delta$ möglicherweise $\|f - f^\delta\|_X$ groß

$\mathcal{A}\colon U \subseteq X \to Y$ kompakt hat zur Folge, dass:

- direktes Problem $f \mapsto m = \mathcal{A}f$ stetig, *well-posed*
- inverses Problem $\mathcal{A}f = m$ *ill-posed*,
  d.h. inverser Operator $\mathcal{A}^{-1}$ falls existent linear aber unstetig ($\Leftrightarrow$ unbeschränkt) oder existiert nicht eindeutig
- f anfällig für Störungen: Bei leicht gestörter Messung m^δ mit $\|m - m^\delta\|_Y < \delta$ ist für zugehörige Lösungen $\mathcal{A}f = m, \mathcal{A}f^\delta = m^\delta$ möglicherweise $\|f - f^\delta\|_X$ groß

Fazit: Je *regulärer* der Operator $\mathcal{A}$ ist,
je besser die Integrierbarkeitseigenschaften von $(x, y) \mapsto k(x, y)$,
bei Faltungsoperatoren: je mehr Glattheit k bei 0 besitzt
umso anfälliger ist das inverse Problem

Daher: Regularisierung

Gegeben: gestörte Messungen $\mathcal{A}f + e^{\delta} = m^{\delta}$

e^{δ}...Störung, $\|e^{\delta}\|_Y \leq \delta$

Finde f oder eine gute Näherung!

Daher: Regularisierung
Gegeben: gestörte Messungen $\mathcal{A}f + e^\delta = m^\delta$
e^δ...Störung, $\|e^\delta\|_Y \leq \delta$
Finde f oder eine gute Näherung!

Definition

Seien X, Y Hilberträume und $\mathcal{A}\colon X \to Y$ ein Operator. Die Familie von stetigen Operatoren

$$\{\mathcal{R}_\alpha\colon Y \to X \mid \alpha > 0\}$$

heißt Regularisierung zu $\mathcal{A}$, falls gilt

$$\lim_{\alpha \to 0} \mathcal{R}_\alpha \mathcal{A} f = f.$$

Daher: Regularisierung

Gegeben: gestörte Messungen $\mathcal{A}f + e^\delta = m^\delta$

e^δ...Störung, $\|e^\delta\|_Y \leq \delta$

Finde f oder eine gute Näherung!

Definition

Seien X, Y Hilberträume und $\mathcal{A}\colon X \to Y$ ein Operator. Die Familie von stetigen Operatoren

$$\{\mathcal{R}_\alpha\colon Y \to X \mid \alpha > 0\}$$

heißt Regularisierung zu $\mathcal{A}$, falls gilt

$$\lim_{\alpha \to 0} \mathcal{R}_\alpha \mathcal{A} f = f.$$

Man beachte: $\mathcal{R}_\alpha$ sind stetig, (nicht unstetig wie z.B. $\mathcal{A}^{-1}$)!

Interessant: Wahl von α in Abhängigkeit von f.

Definition

Seien X, Y Hilberträume und $\mathcal{A}\colon X \to Y$ ein Operator. Die Familie von Operatoren

$$\{\mathcal{R}_\alpha\colon Y \to X \mid \alpha > 0\}$$

heißt Regularisierung zu $\mathcal{A}$, falls gilt

$$\lim_{\alpha \to 0} \mathcal{R}_\alpha \mathcal{A} f = f.$$

$\|\mathcal{A}f - m\|_Y \leq \delta$ gegeben. Finde $\alpha(\delta)$ so, dass

- $\alpha(\delta) \to 0$ falls $\delta \to 0$
- $\sup\{\|\mathcal{R}_{\alpha(\delta)}\mathcal{A}f - f\|_X : \|\mathcal{A}f - m\|_Y < \delta\} \to 0$ falls $\delta \to 0$, für alle $f \in X$.

Wie findet man Regularisierungen?

Wie findet man Regularisierungen?

$\rightarrow$ Durch Ausnützen von Zusatzinformation über die gesuchte Lösung f.

Wie findet man Regularisierungen?

$\rightarrow$ Durch Ausnützen von Zusatzinformation über die gesuchte Lösung f.

Bsp:

- f hat gewisse Nullstellen
- f erfüllt gewisse Relationen (Gleichungen z.B. $G(f) = 0, L \cdot f = 0$)
- f besitzt eine gewisse Glattheit (f', f'',... existiert)
- f hat keine hohen (tiefen) Frequenzen
- f ist nur an gewissen Stellen $\neq 0$ (sparsity)
- $\vdots$

Ausnützen der Zusatzinformation durch penalty function P_α:

Ausnützen der Zusatzinformation durch penalty function P_α:

Anstatt $\|\mathcal{A}f - m^\delta\|_Y$ für f zu minimieren, minimiere

$$\|\mathcal{A}f - m^\delta\|_Y + P_\alpha(f)$$

Ausnützen der Zusatzinformation durch penalty function P_α:

Anstatt $\|\mathcal{A}f - m^\delta\|_Y$ für f zu minimieren, minimiere

$$\|\mathcal{A}f - m^\delta\|_Y + P_\alpha(f)$$

P_α *bestraft* Entfernung von der Zusatzinformation, $P_\alpha(f) = 0$, wenn f diese bereits erfüllt.

Ausnützen der Zusatzinformation durch penalty function P_α:

Anstatt $\|\mathcal{A}f - m^\delta\|_Y$ für f zu minimieren, minimiere

$$\|\mathcal{A}f - m^\delta\|_Y + P_\alpha(f)$$

P_α *bestraft* Entfernung von der Zusatzinformation, $P_\alpha(f) = 0$, wenn f diese bereits erfüllt.

$\mathcal{R}_\alpha\colon m \mapsto \mathcal{R}_\alpha(m) = \arg\min_f \left(\|\mathcal{A}f - m\|_Y + P_\alpha(f)\right)$ ist eine Regularisierung (für geeignetes P_α).

Ausnützen der Zusatzinformation durch penalty function P_α:

Anstatt $\|\mathcal{A}f - m^\delta\|_Y$ für f zu minimieren, minimiere

$$\|\mathcal{A}f - m^\delta\|_Y + P_\alpha(f)$$

P_α *bestraft* Entfernung von der Zusatzinformation, $P_\alpha(f) = 0$, wenn f diese bereits erfüllt.

$\mathcal{R}_\alpha \colon m \mapsto \mathcal{R}_\alpha(m) = \arg\min_f \left(\|\mathcal{A}f - m\|_Y + P_\alpha(f)\right)$ ist eine Regularisierung (für geeignetes P_α).

Möglichkeit für $P_\alpha(f)$: $\alpha\|f\|$, wobei $\|\cdot\|$ eine 'stärkere' Norm als $\|\cdot\|_X$ ist (Zusatzinformationen führen zu stärkeren Bedingungen als $\|\cdot\|_X$ liefert).

Klassische Tikhonov-Regularisierung (Andrey Nikolayevich Tikhonov):

Definition

Die Tikhonov-Regularisierung $T_\alpha(m)$, ist gegeben durch

$$T_\alpha(m) := \operatorname{argmin}_f \|\mathcal{A}f - m\|_Y^2 + \alpha\|f\|_X^2.$$

Klassische Tikhonov-Regularisierung (Andrey Nikolayevich Tikhonov):

Definition

Die Tikhonov-Regularisierung $T_\alpha(m)$, ist gegeben durch

$$T_\alpha(m) := \operatorname{argmin}_f \|\mathcal{A}f - m\|_Y^2 + \alpha\|f\|_X^2.$$

$\|\mathcal{A}T_\alpha(m) - m\|_Y^2 \ldots T_\alpha(m)$ nähert die Messdaten an

$\alpha\|T_\alpha(m)\|_X^2$...hält die Norm der Lösung $T_\alpha(m)$ klein

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In diesem Fall: Operatoren $\mathcal{A}$ realisierbar als $M \times N$-Matrix A.

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In diesem Fall: Operatoren $\mathcal{A}$ realisierbar als $M \times N$-Matrix A.

Direktes Problem:

$$f \mapsto \mathcal{A}f = m \leftrightarrow \hat{f} \mapsto \hat{m} \Leftrightarrow c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \mapsto A \cdot c = b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}$$

Direktes Problem:

$$f \mapsto \mathcal{A}f = m \leftrightarrow \hat{f} \mapsto \hat{m} \Leftrightarrow c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \mapsto A \cdot c = b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}$$

Bsp: $f \mapsto f'$ per Standardbasisvektoren und Matrix

$$\frac{1}{\Delta t} \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 \dots & 0 \\ \vdots & \dots & \ddots & & \end{pmatrix}$$

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In diesem Fall: Operatoren $\mathcal{A}$ realisierbar als $M \times N$-Matrix A.

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In diesem Fall: Operatoren $\mathcal{A}$ realisierbar als $M \times N$-Matrix A.

Indirektes Problem: Minimiere $\|Ac - b\|_2$

In der Praxis: Funktionen f, m: lange Vektoren

z.B: per Basisdarstellung
$f = \sum_{n=1}^{\infty} c_n \phi_n \to \sum_{n=1}^{N} c_n \phi_n = \hat{f}$,
$m = \sum_{n=1}^{\infty} b_n \psi_n \to \sum_{n=1}^{M} b_n \psi_n = \hat{m}$

In diesem Fall: Operatoren $\mathcal{A}$ realisierbar als $M \times N$-Matrix A.

Indirektes Problem: Minimiere $\|Ac - b\|_2$

Regularisierung: Minimiere $\|Ac - b\|_2 + P_\alpha(c)$

Tikhonov Regularisierung für (diskretisierten) Operator A und $\alpha > 0$:

Tikhonov Regularisierung für (diskretisierten) Operator A und $\alpha > 0$:
Um die Minimierung von $\|Ac - b\|_2^2 + \alpha\|c\|_2^2$ zu lösen, finde Singulärwertzerlegung von $A = USV^T$,
$S = diag(\sigma_1, \ldots, \sigma_{\mathrm{rank}(A)}, 0, \ldots, 0)$

($\Rightarrow \hat{c} = V \cdot diag(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_{\mathrm{rank}(A)}}) \cdot U^T \cdot b$ löst Minimierung von $\|Ac - b\|_2^2$ (Regression; z.B. in MatLab mit `A\b`))

Tikhonov Regularisierung für (diskretisierten) Operator A und $\alpha > 0$:
Um die Minimierung von $\|Ac - b\|_2^2 + \alpha\|c\|_2^2$ zu lösen, finde Singulärwertzerlegung von $A = USV^T$,
$S = diag(\sigma_1, \ldots, \sigma_{\mathrm{rank}(A)}, 0, \ldots, 0)$

($\Rightarrow \hat{c} = V \cdot diag(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_{\mathrm{rank}(A)}}) \cdot U^T \cdot b$ löst Minimierung von $\|Ac - b\|_2^2$ (Regression; z.B. in MatLab mit `A\b`))

Setze dann

$$T_\alpha(b) = V \cdot diag\left(\frac{\sigma_1}{\sigma_1^2 + \alpha}, \ldots, \frac{\sigma_{\mathrm{rank}(A)}}{\sigma_{\mathrm{rank}(A)}^2 + \alpha}\right) \cdot U^T \cdot b$$

Einige weitere Möglichkeiten:

- $\|Ac - b\|_2^2 + \alpha \|Lc\|_2^2$
  Zusatzinformation: c sollte orthogonal auf die Zeilen von L sein, $\langle L_{i-}^T, c \rangle = 0$ (Lineare Relationen von c). Bsp: Filtern hoher Frequenzen.

  Äquivalent: Finde $\operatorname{argmin}_c \left\| \begin{pmatrix} A \\ \sqrt{\alpha} L \end{pmatrix} c - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$

Einige weitere Möglichkeiten:

- $\|Ac - b\|_2^2 + \alpha \|Lc\|_2^2$
  Zusatzinformation: c sollte orthogonal auf die Zeilen von L sein, $\langle L_{i-}^T, c \rangle = 0$ (Lineare Relationen von c). Bsp: Filtern hoher Frequenzen.
  Äquivalent: Finde $\operatorname{argmin}_c \left\| \begin{pmatrix} A \\ \sqrt{\alpha} L \end{pmatrix} c - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$
- $\|Ac - b\|_2 + \alpha \|c\|_1$
  $\|\cdot\|_1$ bestraft viele kleine Einträge – bei sparser Besetzung von c

Einige weitere Möglichkeiten:

- $\|Ac - b\|_2^2 + \alpha\|Lc\|_2^2$
  Zusatzinformation: c sollte orthogonal auf die Zeilen von L sein, $\langle L_{i-}^T, c\rangle = 0$ (Lineare Relationen von c). Bsp: Filtern hoher Frequenzen.
  Äquivalent: Finde $\operatorname{argmin}_c \left\| \begin{pmatrix} A \\ \sqrt{\alpha} L \end{pmatrix} c - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$
- $\|Ac - b\|_2 + \alpha\|c\|_1$
  $\|\cdot\|_1$ bestraft viele kleine Einträge – bei sparser Besetzung von c
- $\|Ac - b\|_2 + \alpha\|c\|_{TV}$
  $\|\cdot\|_{TV}$ ist gegeben durch $\|\Delta c\|_1$, wobei $\Delta c_i = c_{i+1} - c_i$, bestraft *viele verschiedene Einträge in kleinen Bereichen*, macht die Lösung *lokal ähnlich* (sehr gut z.B. bei Bildern sichtbar)

Einige weitere Möglichkeiten:

- $\|Ac - b\|_2^2 + \alpha\|Lc\|_2^2$
  Zusatzinformation: c sollte orthogonal auf die Zeilen von L sein, $\langle L_{i-}^T, c\rangle = 0$ (Lineare Relationen von c). Bsp: Filtern hoher Frequenzen.

  Äquivalent: Finde $\operatorname{argmin}_c \left\| \begin{pmatrix} A \\ \sqrt{\alpha} L \end{pmatrix} c - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$
- $\|Ac - b\|_2 + \alpha\|c\|_1$
  $\|\cdot\|_1$ bestraft viele kleine Einträge – bei sparser Besetzung von c
- $\|Ac - b\|_2 + \alpha\|c\|_{TV}$
  $\|\cdot\|_{TV}$ ist gegeben durch $\|\Delta c\|_1$, wobei $\Delta c_i = c_{i+1} - c_i$, bestraft *viele verschiedene Einträge in kleinen Bereichen*, macht die Lösung *lokal ähnlich* (sehr gut z.B. bei Bildern sichtbar)

$\rightarrow$ Schwierigere Algorithmen nötig

Wie findet man α?

Wie findet man α?

$\rightarrow$ im Allgemeinen schwierig

Wie findet man α?

$\rightarrow$ im Allgemeinen schwierig

$$\|Ac - b\|_2^2 + \alpha\|c\|^2$$

Möglichkeit:
Für eine *typische*, bekannte, rein vorliegende Störung e, berechne $\|e\|_2^2$.
Für eine bekannte Lösung c des inversen Problems (Testdaten), berechne $\|c\|$.
Wähle α so, dass $\|e\|_2^2 \approx \alpha\|c\|$