

Inverse Probleme

Inhaltsverzeichnis

1 Einführung	2
2 Lineare Integralgleichungen und kompakte Operatoren	6
2.1 Eigenschaften der Operatoren	7
2.2 Singulärwertzerlegung kompakter Operatoren	8
2.3 Verallgemeinerte Inverse	11
3 Allgemeine lineare Regularisierung mit SVD	12
3.1 Konvergenz von $R_\alpha f$	14
3.2 Praktische Realisierung von linearen Regularisierungsverfahren	17
4 Stochastische Modellierung	19
4.1 Beispiele für Datenterme	21
4.2 Beispiele für Regularisierungsterme	23
5 Variationsmethoden zur Regularisierung	27
5.1 Existenz von Minima	27
5.2 Eindeutigkeit von Minimierern	31
5.3 Stabilität von Minimierern	31
5.4 Ableitung von Funktionalen und Operatoren in Banachräumen	33
5.5 Konvexe Analysis	34
5.6 Anwendung auf Variationsmethoden zur Regularisierung	37
5.7 Fehlerabschätzungen	39
6 Iterative Regularisierung basierend auf Variationsmethoden	45
6.1 Bregman-Iteration	45
6.2 Konstruktion weiterer Iterationsverfahren	51

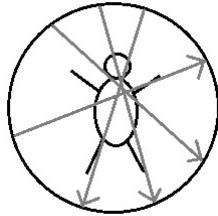
Inverse Probleme

1 Einführung

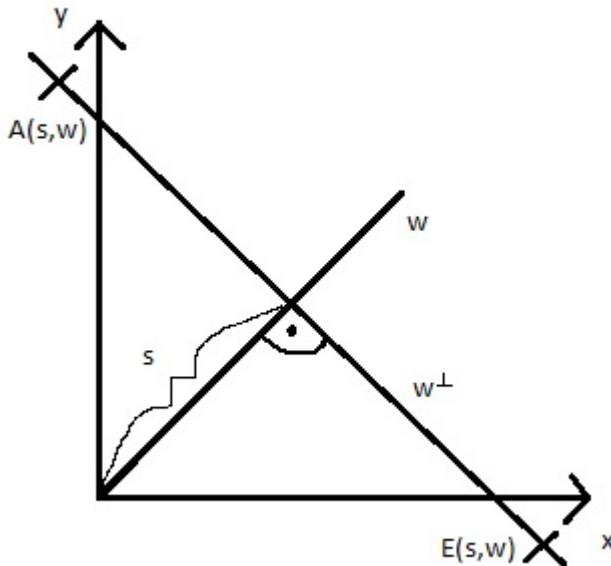
1.1 Metadefinition inverses Problem: Die Suche nach der Ursache für eine beobachtbare oder gewünschte Wirkung über ein mathematisches Modell. Das Modell beschreibt, wie dann aus der Ursache die Wirkung entsteht. In den Ingenieurwissenschaften spricht man auch vom *Inverse Engineering* oder *Inverse Modelling*.

Die Unterscheidung in beobachtbare und gewünschte Wirkungen hat mathematische Konsequenzen: Liegt ein beobachteter Effekt vor, so sucht man eine eindeutige Lösung des Problems. Handelt es sich um einen gewünschten Effekt, so ist Eindeutigkeit weniger wichtig, denn dann kann das angestrebte Ziel auf verschiedene Weisen erreicht werden, sodass man zusätzliche Freiheit beim Lösen des Problems erhält, z.B. in Form zusätzlicher Anforderungen an die Lösung.

1.2 Erstes Beispiel (Computertomographie): Bei der Computertomographie werden Röntgenstrahlen durch verschiedenen Richtungen geschickt:



Es wird dann die Intensität (Energie) eines Strahls am Ende der Strecke gemessen, wobei die Intensität I_0 am Anfang der Strecke gegeben ist. Zur mathematischen Modellierung betrachte:



Dabei ist w der normierte Orthogonalvektor auf den Röntgenstrahl ($|w| = 1$). Jeder Strahl wird dann beschrieben durch $w \in S^1$ und $s \in \mathbb{R}$. Bezeichnet $t \in \mathbb{R}$ die lokale Koordinate am Strahl, so können die Punkte entlang des Röntgenstrahls beschrieben werden durch

$$(x(t), y(t))^T = sw + tw^\perp, \quad t \in \mathbb{R},$$

wobei letztendlich nur die Punkte für $t \in [A(s, w), E(s, w)]$ relevant sind.

Die Intensität des Strahls bei t sei mit $I(t; s, w)$ bezeichnet. Setze für die Emission bei A :

$$I_0 := I(A(s, w); s, w).$$

Die Änderung von I in Abhängigkeit von der Dichte $\rho(x, y)$ im Körper kann für ein kleines Intervall Δt approximiert werden durch:

$$I(sw + (t + \Delta t)w^\perp) - I(sw + tw^\perp) \approx -\rho(sw + tw^\perp)\Delta t I(sw + tw^\perp).$$

Damit ergibt sich:

$$\lim_{\Delta t \rightarrow 0} \frac{I(sw + (t + \Delta t)w^\perp) - I(sw + tw^\perp)}{\Delta t} = \frac{d}{dt} I(sw + tw^\perp) = -\rho(sw + tw^\perp)I(sw + tw^\perp).$$

Seien nun s, w fix. Setze dann:

$$i(t) = I(sw + tw^\perp), \quad r(t) = \rho(sw + tw^\perp).$$

Damit ergibt sich folgende gewöhnliche Differentialgleichung:

$$i'(t) = -r(t)i(t), \quad i(A) = I_0.$$

Die Lösung ist bekanntermaßen gegeben durch

$$i(t) = I_0 e^{-\int_A^t r(\tau) d\tau},$$

womit wir für allgemeine s, w die Gleichung

$$I(sw + tw^\perp) = I_0 e^{-\int_{A(s, w)}^t \rho(sw + \tau w^\perp) d\tau}$$

erhalten. Insbesondere gilt dann:

$$I(sw + E(s, w)w^\perp) = I_0 e^{-\int_{A(s, w)}^{E(s, w)} \rho(sw + \tau w^\perp) dt}.$$

Für die gemessenen Daten y erhalten wir:

$$y(s, w) = -\ln \left(\frac{I(sw + E(s, w)w^\perp)}{I_0} \right) = \int_{A(s, w)}^{E(s, w)} \rho(sw + \tau w^\perp) dt.$$

Damit besteht das inverse Problem aus der Rekonstruktion einer Dichte $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$ aus den Daten $y : \mathbb{R} \times S^1 \rightarrow \mathbb{R}$.

Das mathematische Problem hinter der Computertomographie ist die Inversion der Radon-Transformation

$$\mathcal{R}f(s, w) = \int_{\mathbb{R}} f(sw + tw^\perp) dt, \quad f : \mathbb{R}^2 \rightarrow \mathbb{R}, s \in \mathbb{R}^+, w \in S^1.$$

Zusätzliche praktische Schwierigkeiten sind:

- nur endlich viele Richtungen (Abtastung/Sampling).
- Messfehler \Rightarrow Ist die Rekonstruktion robust?

1.3 Zweites Beispiel (Differentiation): Sei $u : I \rightarrow \mathbb{R}$ mit $I = [a, b] \subseteq \mathbb{R}$. Sei weiter

$$y(t) := (Ku)(t) := \int_a^t u(s) \, ds.$$

Dann ist K ein linearer Operator zwischen unendlich-dimensionalen Funktionenräumen, z.B. $K : L^\infty(I) \rightarrow L^\infty(I)$ oder $K : \mathcal{C}(I) \rightarrow \mathcal{C}(I)$ mit $\|\cdot\|_\infty$. Die Umkehrung von K entspricht der ersten Ableitung:

$$u(t) = \frac{dy}{dt}(t).$$

Wir wollen nun den Effekt des Messfehlers untersuchen. Dazu modellieren wir für $\delta \ll 1$ und unbekanntes $n \in \mathbb{N}$:

$$y^\delta(t) := y(t) + \delta \sin(n(t - a)).$$

Dann gilt $\|y - y^\delta\|_\infty = \delta$. Für fehlerhafte Daten erhalten wir:

$$u^\delta(t) = \frac{dy^\delta}{dt}(t) = \frac{dy}{dt}(t) + \delta \cos(n(t - a))n = u(t) + n\delta \cos(n(t - a)).$$

Damit gilt $\|u - u^\delta\|_\infty = n\delta$ für beliebig groSSes $n \in \mathbb{N}$. Der Rekonstruktionsfehler kann also selbst bei beliebig kleinem Messfehler δ beliebig groSS werden!

Als mathematische Alternative könnten wir $K : \mathcal{C}(I) \rightarrow \mathcal{C}^1(I)$ mit $\|y\|_{\mathcal{C}^1} = \max\{\|y\|_\infty, \|y'\|_\infty\}$ betrachten, allerdings sind in der Praxis die Funktionenräume nicht frei wählbar.

Als Botschaft dieses Beispiels erhalten wir, dass die Art der Messfehler (Schranke) die Norm/den Raum für die Daten bestimmt.

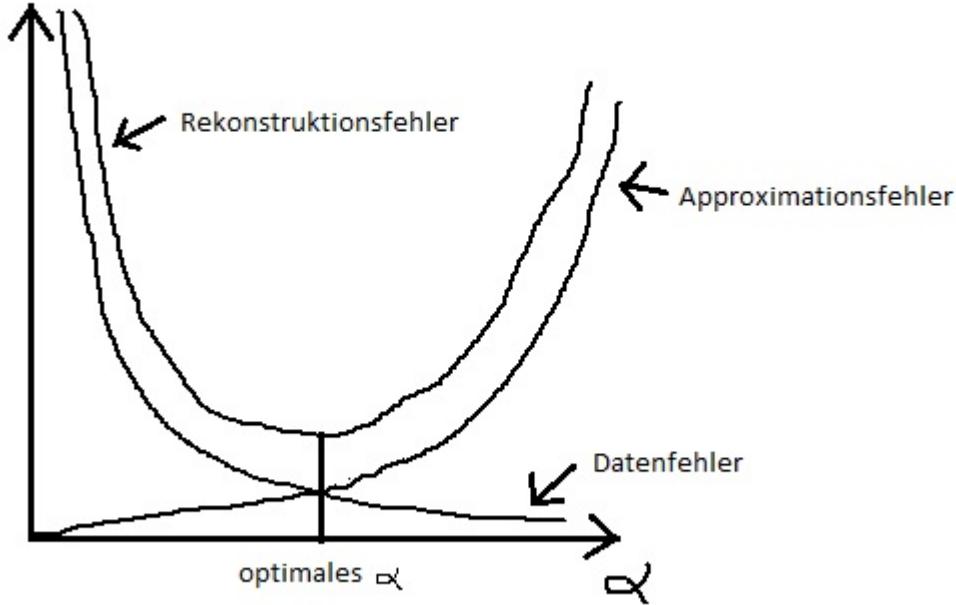
1.4 Definition: Ein Problem heiSSt schlecht gestellt, wenn beliebig kleine Messfehler zu beliebig groSSen Rekonstruktionsfehlern führen können.

Im zweiten Beispiel ist K^{-1} nur auf $\mathcal{C}^1(I) \subseteq \mathcal{C}(I)$, also einem nicht abgeschlossenen Teilraum definiert. Zudem ist K^{-1} unstetig.

Diese Problematik führt zur Regularisierung. Grundsätzlich wollen wir K^{-1} durch eine stetige Approximationen R_α ersetzen, sodass in einem gewissen Sinne $R_\alpha \rightarrow K^{-1}$ für $\alpha \rightarrow 0$ gilt. Der Rekonstruktionsfehler setzt sich dann zusammen aus dem (fortgepflanzten) Datenfehler und dem Approximationsfehler:

$$R_\alpha y^\delta - K^{-1}y = R_\alpha(y^\delta - y) + R_\alpha y - K^{-1}y.$$

Folgende Grafik veranschaulicht die Aufteilung in die beiden Fehlerarten:



Ein einfaches Beispiel für eine Regularisierung des Differentiationsoperators ist

$$(R_\alpha y)(t) = \frac{y(t + \alpha) - y(t)}{\alpha}, \quad \alpha > 0.$$

1.5 Charakteristika inverser Probleme:

- Ein typisches inverses Problem ist eine Operatorgleichung $K(u) = f$ mit einem linearen oder nicht-linearen Operator K zwischen Funktionenräumen. Meistens stecken Integriertionen in K , z.B. in Form eines Faltungsintegrals.
- Die Art des Informationsverlustes ist unterschiedlich:
 - teilweise zu wenige Messungen;
 - teilweise zu viel Rauschen auf den Daten;
 - unterschiedliche Glättung von u durch den Vorförtsoperator.
- Wir brauchen immer a-priori Informationen, um sinnvolle Lösungen zu erhalten. Meist sind diese nicht mathematisch formuliert, z.B. in Form von:
 - Vorwissen: keine grossen Löcher, keine grossen lokalen Änderungen;
 - Anforderungen: keine Texturen, Objekte kleiner als 3mm sind uninteressant, grosse Übergänge sind interessant.

1.6 Definition (nach Hadamard): Ein Problem heiSt korrekt gestellt, falls:

- (1) Für alle (zulässigen) Daten existiert eine Lösung.
- (2) Für alle (zulässigen) Daten ist die Lösung eindeutig.
- (3) Die Lösung hängt stabil bzw. stetig von den gegebenen Daten ab.

Ein Problem heiSt schlecht gestellt, wenn eine der drei Bedingungen verletzt ist. Im Allgemeinen wird die dritte Bedingung Probleme bereiten.

2 Lineare Integralgleichungen und kompakte Operatoren

Wir wollen zunächst Integraloperatoren mit einem Integralkern $k : \Sigma \times \Omega \rightarrow \mathbb{R}$ betrachten:

$$(Ku)(x) := \int_{\Omega} k(x, y) u(y) dy.$$

Dabei ist $u : \Omega \rightarrow \mathbb{R}$ eine Funktion aus einem näher zu spezifizierenden Funktionenraum auf Ω . Ku ist dann eine reelle Funktion auf Σ . Wir wollen zwei Beispiele betrachten:

- (1) Stammfunktion: Sei $\Omega = \Sigma = [0, 1]$. Wir definieren einen Integralkern durch:

$$k(x, y) = \begin{cases} 1, & x \geq y, \\ 0, & x < y. \end{cases}$$

Damit erhalten wir den Operator

$$(Ku)(x) = \int_0^x u(y) dy = \int_0^1 k(x, y) u(y) dy,$$

der einer Funktion u eine Stammfunktion zuordnet.

- (2) Entzerrung/Entfaltung: Sei $\Omega = \Sigma = \mathbb{R}^2$ und $k(x, y) = G(x - y)$. Dabei kann G z.B. die Gauss-Kurve sein.

2.1 Definition: Sei Ω kompakt und $\mathcal{C}(\Omega) := \{u : \Omega \rightarrow \mathbb{R} \mid u \text{ stetig}\}$ mit $\|u\|_{\mathcal{C}(\Omega)} = \max_{x \in \Omega} |u(x)|$. Dann können wir Operatoren $K : \mathcal{C}(\Omega) \rightarrow \mathcal{C}(\Sigma)$ betrachten.

Allerdings stellt der Funktionenraum $\mathcal{C}(\Omega)$ eine groSSe Einschränkung dar, denn oft haben wir es mit unstetigen Funktionen zu tun, beispielsweise in der Bildrekonstruktion, wo Kanten Unstetigkeitsstellen entsprechen. Deshalb:

2.2 Definition: Für $p \in [1, \infty)$ definieren wir die Räume

$$L^p(\Omega) := \left\{ u : \Omega \rightarrow \mathbb{R} \mid u \text{ messbar}, \|u\|_{L^p} := \left(\int_{\Omega} |u(x)|^p dx \right)^{\frac{1}{p}} < \infty \right\}.$$

Dann können wir Operatoren $K : L^p(\Omega) \rightarrow L^q(\Sigma)$ für $p, q \in [0, \infty)$ betrachten. Oft haben wir den Fall $p = q = 2$, denn $L^2(\Omega)$ ist ein Hilbertraum.

2.3 Link zu Gleichungen im \mathbb{R}^n : Seien $\{\varphi_i : \Sigma \rightarrow \mathbb{R} \mid i = 1, \dots, n\}$, $\{\psi_i : \Omega \rightarrow \mathbb{R} \mid i = 1, \dots, n\}$ jeweils linear unabhängig. Damit erhalten wir einen degenerierten Kern:

$$k(x, y) = \sum_{i=1}^n \varphi_i(x) \psi_i(y)$$

und den zugehörigen Integraloperator:

$$(Ku)(x) = \int_{\Omega} \left(\sum_{i=1}^n \varphi_i(x) \psi_i(y) \right) u(y) dy = \sum_{i=1}^n \varphi_i(x) \underbrace{\int_{\Omega} \psi_i(y) u(y) dy}_{=: c_i}.$$

Es gilt dann $\dim(\mathcal{R}(K)) = n$. Zur Lösung der Gleichung $Ku = f$ verwende die Darstellung $f = \sum_{i=1}^n c_i \varphi_i$. Mit dem Galerkin-Verfahren (Multiplikation mit φ_j und Integration über Σ auf beiden Seiten) erhalten wir dann:

$$\int_{\Sigma} f(x) \varphi_j(x) dx = \int_{\Sigma} \left(\sum_{i=1}^n c_i \varphi_i(x) \right) \varphi_j(x) dx = \sum_{i=1}^n c_i \int_{\Sigma} \varphi_i(x) \varphi_j(x) dx, \quad j = 1, \dots, n.$$

also ein $n \times n$ lineares Gleichungssystem für die Koeffizienten $c_i, i = 1, \dots, n$. Damit erhalten wir:

$$\int_{\Omega} \psi_i(y) u(y) dy = c_i, \quad i = 1, \dots, n \quad (\text{N.B.}),$$

also ein unterbestimmtes Gleichungssystem für u , denn u liegt i.A. nicht in einem endlich-dimensionalen Funktionenraum. Wir suchen nun die Lösung minimaler Norm in $L^2(\Omega)$, d.h. wir lösen:

$$\frac{1}{2} \int_{\Omega} u(y)^2 dy \longrightarrow \min_{u \in L^2(\Omega)} \quad \text{unter (N.B.)}$$

Betrachte dazu für $\lambda \in \mathbb{R}^n$ das Lagrange-Funktional:

$$\mathcal{L}(u, \lambda) := \frac{1}{2} \int_{\Omega} u(x)^2 dx + \sum_{i=1}^n \lambda_i \left(\int_{\Omega} u(y) \psi_i(y) dy - c_i \right).$$

u minimiert \mathcal{L} dann für passendes λ . Betrachte weiter:

$$\begin{aligned} \mathcal{L}(u, \tilde{\lambda}) &= \int_{\Omega} \frac{1}{2} u(x)^2 + \left(\sum_{i=1}^n \tilde{\lambda}_i \psi_i(x) \right) u(x) + \frac{1}{2} \left(\sum_{i=1}^n \tilde{\lambda}_i \psi_i(x) \right)^2 dx - \frac{1}{2} \int_{\Omega} \left(\sum_{i=1}^n \tilde{\lambda}_i \psi_i(x) \right)^2 dx - \sum_{i=1}^n \tilde{\lambda}_i c_i \\ &= \frac{1}{2} \int_{\Omega} \left(u(x) + \sum_{i=1}^n \tilde{\lambda}_i \psi_i(x) \right)^2 dx + \text{const.} \end{aligned}$$

Somit ist das Minimum \tilde{u} gegeben als Linearkombination der $\psi_i, i = 1, \dots, n$:

$$\tilde{u}(x) = - \sum_{i=1}^n \tilde{\lambda}_i \psi_i(x).$$

Aus der Nebenbedingung (N.B.) erhalten wir somit ein $n \times n$ Gleichungssystem für die $\tilde{\lambda}_i$.

Wir wollen die Idee für die numerische Lösung noch einmal zusammenfassen: Approximiere einen Kern k durch degenerierte Kerne k_n (d.h. $k_n \rightarrow k$ für $n \rightarrow \infty$), und löse dann die Gleichung für k_n durch zwei $n \times n$ Gleichungssysteme wie oben. Im Optimalfall wähle $\{\varphi_i\}, \{\psi_i\}$ als Orthonormalbasen, denn dann ist $(\int \varphi_i \varphi_j)_{i,j}$ die Einheitsmatrix, sodass die $n \times n$ -Systeme trivial sind.

2.1 Eigenschaften der Operatoren

2.4 Definition: Sei $K : X \rightarrow Y$ ein linearer Operator zwischen Banachräumen X und Y . Dann heiSt K stetig bzw. beschränkt, wenn eine Konstante $c > 0$ existiert mit $\|Ku\|_Y \leq c\|u\|_X$ für alle $u \in X$. Die kleinste Konstante, die diese Eigenschaft erfüllt, heiSt Norm des Operators. Es gilt dann:

$$\|K\| = \sup_{u \in X \setminus \{0\}} \frac{\|Ku\|_Y}{\|u\|_X} = \sup_{\|u\|_X=1} \|Ku\|_Y.$$

2.5 Lemma: Seien Ω, Σ offen und beschränkt. Sei weiter $k \in L^2(\Sigma \times \Omega)$ und

$$K : L^2(\Omega) \rightarrow L^2(\Sigma); \quad (Ku)(x) = \int_{\Omega} k(x, y) u(y) dy.$$

Dann ist K wohldefiniert und stetig.

2.6 Definition: Ein linearer Operator $K : X \rightarrow Y$ heiSt kompakt, wenn eine der folgenden Bedingungen erfüllt ist:

- (1) Für jede beschränkte Menge $B \subseteq X$ ist das Bild $K(B)$ präkompakt in Y ist, d.h. $\overline{K(B)}$ ist kompakt.
- (2) Für jede beschränkte Folge $(x_n)_{n \in \mathbb{N}}$ in X besitzt die Folge $(Kx_n)_{n \in \mathbb{N}}$ eine konvergente Teilfolge.
- (3) Falls X, Y reflexiv: Aus $x_n \rightarrow x$ folgt $Kx_n \rightarrow Kx$. Erinnerung: Es gilt $x_n \rightarrow x$, wenn $\langle \varphi, x_n \rangle \rightarrow \langle \varphi, x \rangle$ für alle $\varphi \in X'$ gilt.

Aus der Definition kompakter Operatoren folgt direkt, dass jeder beschränkte lineare Operator, der in einen endlich-dimensionalen Raum abbildet, kompakt ist. Damit sind insbesondere Integraloperatoren mit degeneriertem Kern wie in 2.3 kompakt.

2.7 Lemma: Seien X, Y Banachräume. Dann gelten:

- (1) Sind $K_1, K_2 : X \rightarrow Y$ kompakt, so ist auch $K_1 + K_2$ kompakt.
- (2) Ist $K : X \rightarrow Y$ kompakt und $\lambda \in \mathbb{R}$, so ist auch λK kompakt.
- (3) Seien $K_1 : X \rightarrow Y$, $K_2 : Y \rightarrow Z$ beschränkt und linear und K_1 oder K_2 kompakt. Dann ist $K_2 \circ K_1$ ebenfalls kompakt.

2.8 Satz: Sei $K_n : X \rightarrow Y$ eine Folge kompakter Operatoren und $K : X \rightarrow Y$ mit $\|K - K_n\| \rightarrow 0$ für $n \rightarrow \infty$. Dann ist auch K kompakt.

Lemma 2.7 und Satz 2.8 liefern insbesondere, dass die kompakten Operatoren ein abgeschlossenes Ideal in den beschränkten linearen Operatoren sind.

2.9 Satz: Sei $K : L^2(\Omega) \rightarrow L^2(\Sigma)$ wie in Lemma 2.5. Dann ist K kompakt.

2.10 Lemma von Riesz: Sei X ein Banachraum und $U \subsetneq X$ ein abgeschlossener Teilraum. Dann gibt es für jedes $\varepsilon > 0$ ein $x \in X$ mit $\|x\|_X = 1$, sodass $\text{dist}(x, U) = \inf_{y \in U} \|x - y\|_X \geq 1 - \varepsilon$ gilt.

2.11 Satz: Seien X, Y unendlich-dimensionale Banachräume und $K : X \rightarrow Y$ kompakt, so dass $\mathcal{R}(K)$ unendlich-dimensional ist. Dann besitzt K keine stetige Inverse. Insbesondere ist $Ku = f$ mit K kompakt immer ein schlecht gestelltes Problem.

Beweis: Wähle $y_1 \in Y$ mit $\|y_1\|_Y = 1$. Nach Lemma 2.10 gibt es ein $y_2 \in Y$ mit $\|y_2\|_Y = 1$ und $\text{dist}(y_2, LH\{y_1\}) \geq \frac{1}{2}$. Weiter gibt es ein $y_3 \in Y$ mit $\|y_3\|_Y = 1$ und $\text{dist}(y_3, LH\{y_1, y_2\}) \geq \frac{1}{2}$. Insgesamt erhalten wir eine Folge $(y_k)_{k \in \mathbb{N}}$ mit $\|y_k\|_Y = 1$ für alle $k \in \mathbb{N}$ und $\|y_k - y_l\|_Y \geq \frac{1}{2}$ für $k \neq l$. Insbesondere besitzt $(y_k)_{k \in \mathbb{N}}$ keine konvergente Teilfolge. Angenommen, $K^{-1} : Y \rightarrow X$ existiert und ist stetig. Setze $x_k := K^{-1}y_k$. Es gilt $\|x_k\|_X \leq \|K^{-1}\| \|y_k\| = \|K^{-1}\| \leq c$. Damit ist $(x_k)_{k \in \mathbb{N}}$ beschränkt. Laut Annahme besitzt $(Kx_k)_{k \in \mathbb{N}}$ eine konvergente Teilfolge, was im Widerspruch zu Obigem steht. \square

2.2 Singulärwertzerlegung kompakter Operatoren

2.12 Definition: Sei X ein normierter Raum und $K \in \mathcal{L}(X)$. Dann ist das Spektrum von K definiert als

$$\sigma(K) = \{\lambda \in \mathbb{C} \mid (\lambda I - K) : X \rightarrow X \text{ besitzt keine stetige Inverse}\}.$$

Ein Eigenwert von K ist eine komplexe Zahl $\lambda \in \mathbb{C}$ mit $\mathcal{N}(\lambda I - K) \neq \{0\}$, d.h. $\lambda x - Kx = f$.

2.13 Satz: Sei X ein normierter Raum und $K : X \rightarrow X$ kompakt. Dann gelten:

- (1) Ist $\dim(X) = \infty$, so folgt $0 \in \sigma(K)$.
- (2) Ist $\lambda \in \sigma(K) \setminus \{0\}$, so ist λ ein Eigenwert von K mit endlicher geometrischer Vielfachheit, d.h. $\dim(\mathcal{N}(\lambda I - K)) < \infty$.
- (3) Das Spektrum $\sigma(K)$ ist höchstens abzählbar und 0 ist der einzige mögliche Häufungspunkt.

Für den Beweis benötigt man den ersten Rieszschen Satz:

2.14 Satz: Sei $L = I - K$ für $K : X \rightarrow X$ kompakt. Dann ist $\mathcal{N}(L)$ endlich-dimensional.

Im folgenden wollen wir X, Y als Hilberträume annehmen. Ist dann $K : X \rightarrow Y$ kompakt, so sind auch $K^*K : X \rightarrow X$ und $KK^* : Y \rightarrow Y$ kompakt. Wir betrachten die Eigenwertprobleme $K^*Ku = \lambda u$ und $KK^*v = \lambda v$. Wegen $(K^*K)^* = K^*K$ und $(KK^*)^* = KK^*$ sind K^*K und KK^* symmetrisch, womit $\lambda \in \mathbb{R}$ folgt. Außerdem sind beide positiv semidefinit, denn es gilt $\langle K^*Ku, u \rangle = \|Ku\|_Y^2 \geq 0$. Damit folgt weiter $\lambda \geq 0$.

2.15 Definition: Sei $K \in \mathcal{L}(X, Y)$ kompakt. Dann heiSt $(\sigma_n, u_n, v_n)_{n \in \mathbb{N}}$ ein singuläres System, wenn für alle $n \in \mathbb{N}$ gilt: $\sigma_n > 0$, $(\sigma_n^2, u_n)_{n \in \mathbb{N}}$ ist ein Eigensystem zu K^*K und $v_n = \frac{Ku_n}{\|Ku_n\|}$.

2.16 Satz: Sei $(\sigma_n, u_n, v_n)_{n \in \mathbb{N}}$ ein singuläres System für einen kompakten Operator $K \in \mathcal{L}(X, Y)$. Dann gelten:

- (1) $\sigma(K^*K) \setminus \{0\} = \{\sigma_n^2 \mid n \in \mathbb{N}\}$ und die Eigenräume zu σ_n^2 für K^*K und KK^* haben dieselbe Dimension.
- (2) Für alle $n \in \mathbb{N}$ gilt $Ku_n = \sigma_n v_n$ und $K^*v_n = \sigma_n u_n$.
- (3) $\{u_n\}$ bzw. $\{v_n\}$ sind vollständige ONS in

$$\overline{\mathcal{R}(K^*K)} = \overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp \text{ bzw. } \overline{\mathcal{R}(KK^*)} = \overline{\mathcal{R}(K)}$$

und $(\sigma_n^2, v_n)_{n \in \mathbb{N}}$ ist ein Eigensystem zu KK^* .

Aus diesem Satz ergeben sich die Darstellungen

$$K^*Kx = \sum_{n=1}^{\infty} \sigma_n^2 \langle x, u_n \rangle u_n, \quad KK^*y = \sum_{n=1}^{\infty} \sigma_n^2 \langle y, v_n \rangle v_n.$$

Außerdem erhalten wir die Singulärwertzerlegung von K :

$$Kx = K \left(\sum_{n=1}^{\infty} \langle x, u_n \rangle u_n \right) = \sum_{n=1}^{\infty} \langle x, u_n \rangle Ku_n = \sum_{n=1}^{\infty} \sigma_n \langle x, u_n \rangle v_n.$$

Analog erhalten wir für den adjungierten Operator K^* :

$$K^*y = \sum_{n=1}^{\infty} \sigma_n \langle y, v_n \rangle u_n.$$

Mithilfe der Singulärwertentwicklung wollen wir nun untersuchen, ob eine Lösung von $Ku = f$ existiert. Zunächst gilt für alle $n \in \mathbb{N}$:

$$f = Ku \Leftrightarrow \langle f, v_n \rangle = \langle Ku, v_n \rangle = \langle u, K^*v_n \rangle = \sigma_n \langle u, u_n \rangle.$$

Damit ergibt sich:

2.17 Picard-Kriterium: Wir suchen eine Lösung $u \in X \Leftrightarrow \|u\|_X < \infty$ von $Ku = f$. Es gilt dann:

$$\|u\|_X^2 = \sum_{n=1}^{\infty} \langle u, u_n \rangle^2 = \sum_{\sigma_n \neq 0} \frac{\langle f, v_n \rangle^2}{\sigma_n^2} < \infty.$$

Beachte jedoch: Es gilt

$$f \in Y \Leftrightarrow \|f\|_Y^2 < \infty \Leftrightarrow \sum_{n=1}^{\infty} \langle f, v_n \rangle^2 < \infty,$$

aber das Picard-Kriterium

$$\sum_{\sigma_n \neq 0} \frac{\langle f, v_n \rangle^2}{\sigma_n^2} < \infty$$

ist nicht automatisch erfüllt für $f \in Y$. Die Existenz einer Lösung $u \in X$ hängt also davon ab, wie schnell die Singulärwerte σ_n gegen 0 konvergieren. Dies führt zu folgenden Begriffen:

2.18 Definition: Wir nennen das inverse Problem $Ku = f$ mildly ill-posed, wenn $\sigma_n = \mathcal{O}(n^{-\alpha})$ für ein $\alpha > 0$ gilt (d.h. die Singulärwerte konvergieren höchstens so schnell wie ein Polynom gegen 0: Es existieren $C, \alpha > 0$ mit $\sigma_n \geq Cn^{-\alpha}$) und severely ill-posed, wenn $\sigma_n = \mathcal{O}(e^{-n})$ gilt (d.h. die Singulärwerte konvergieren schneller als jedes Polynom gegen 0: $\sigma_n \leq Cn^{-\alpha}$ für alle $C, \alpha > 0$).

2.19 Beispiel: Wir betrachten

$$K : L^2([0, 1]) \rightarrow L^2([0, 1]); \quad (Ku)(x) = \int_0^x u(y) dy.$$

Es gilt dann für alle $u, v \in L^2([0, 1])$:

$$\langle Ku, v \rangle_{L^2} = \int_0^1 v(x)(Ku)(x) dx = \int_0^1 \int_0^x v(x)u(y) dy dx = \int_0^1 \left(\int_y^1 v(x) dx \right) u(y) dy = \langle u, K^*v \rangle_{L^2},$$

also $(K^*v)(y) = \int_y^1 v(x) dx$. Damit gilt dann weiter:

$$(K^*Ku)(x) = \int_x^1 (Ku)(y) dy = \int_x^1 \int_0^y u(z) dz dy.$$

Wir suchen nun $u \in L^2([0, 1])$ und $\sigma \geq 0$ mit $K^*Ku = \sigma u$. Mit $w(x) := (K^*Ku)(x)$ gilt dann:

$$w = \sigma u, \quad w(1) = 0, \quad w'(x) = - \int_0^x u(z) dz, \quad w'(0) = 0, \quad w''(x) = -u(x), \quad -w''(x) = \frac{1}{\sigma^2} w.$$

Daraus folgt:

$$w(x) = \alpha \sin\left(\frac{x}{\sigma}\right) + \beta \cos\left(\frac{x}{\sigma}\right), \quad w'(x) = \frac{\alpha}{\sigma} \cos\left(\frac{x}{\sigma}\right) - \frac{\beta}{\sigma} \sin\left(\frac{x}{\sigma}\right).$$

Wegen $0 = w'(0) = \frac{\alpha}{\sigma} \cos(0)$ folgt $\alpha = 0$. Es gilt weiter $0 = w(1) = \beta \cos\left(\frac{1}{\sigma}\right)$, woraus $\frac{1}{\sigma} = \frac{(2n-1)\pi}{2}$ folgt. Damit gilt $\sigma_n = \frac{2}{(2n-1)\pi} = \mathcal{O}(\frac{1}{n})$, womit die Differentiation mildly ill-posed ist.

2.20 Fehlerverstärkung: Sei $u^* \in X$ die echte Lösung von $Ku = f$ und seien $Ku^* \in Y$ die echten/fehlerfreien Daten. Gegeben seien weiter die gemessenen Daten $f \neq Ku^*$. Wie gross

ist der Fehler $\|u - u^*\|_X$, wenn wir $Ku = f$ lösen? Betrachte dazu:

$$\begin{aligned}\|u - u^*\|_X^2 &= \sum_{n=1}^{\infty} \langle u - u^*, u_n \rangle^2 \\ &= \sum_{n=1}^{\infty} \left(\frac{1}{\sigma_n} \langle f, v_n \rangle - \langle u^*, u_n \rangle \right)^2 \\ &\stackrel{u_n = \frac{1}{\sigma_n} K^* v_n}{=} \sum_{n=1}^{\infty} \frac{1}{\sigma_n^2} \langle f - Ku^*, v_n \rangle^2.\end{aligned}$$

Damit werden Störungen in Koeffizienten von v_n mit kleinem σ_n am meisten verstärkt. Die Sprechweise ist hier analog zu Beispiel 2.19: $\frac{1}{\sigma_n}$ heiSt die Frequenz und entsprechende Störungen heiSSten hochfrequent.

2.3 Verallgemeinerte Inverse

Es können im Allgemeinen zwei Probleme hinsichtlich der Invertierbarkeit von $K : X \rightarrow Y$ auftauchen:

- (1) K ist nicht surjektiv, d.h. $\mathcal{R}(K) \subsetneq Y$ bzw. $\overline{\mathcal{R}(K)} \subsetneq Y$.
- (2) K ist nicht injektiv, d.h. $\mathcal{N}(K) \neq \{0\}$.

Im ersten Fall scheint es sinnvoll ein $u \in X$ zu suchen, sodass Ku minimalen Abstand zu $f \in Y$ hat (denn eventuell gilt $f \in Y \setminus \mathcal{R}(K)$). Im zweiten Fall können mehrere Lösungen existieren. Dann ist es sinnvoll diejenige mit der kleinsten Norm auszuwählen. Diese Überlegungen motivieren folgende Definition:

2.21 Definition: Ein $u \in X$ heiSt bestapproximierende Lösung (least-squares solution), falls $\|Ku - f\|_Y \leq \|Kx - f\|_Y$ für alle $x \in X$ gilt. Dabei ist die Existenz einer bestapproximierenden Lösung unklar. Weiter heiSt eine bestapproximierende Lösung $u \in X$ Lösung minimaler Norm (minimal norm solution), wenn $\|u\|_X \leq \|\tilde{u}\|_X$ für alle bestapproximierenden Lösungen \tilde{u} gilt.

Wir wollen nun die verallgemeinerte Inverse K^\dagger konstruieren. Setzen wir den Definitionsbereich von K^\dagger als

$$\mathcal{D}(K^\dagger) := \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp,$$

so besitzt $f \in \mathcal{D}(K^\dagger)$ eine eindeutige Zerlegung als $f = Ku + z$ für ein $u \in X$ und $z \in \mathcal{R}(K)^\perp$. Eine bestapproximierende Lösung ist dann ein $u \in X$ mit $Ku = P_{\mathcal{R}(K)}(f)$, denn $f = Ku + z$ mit $z \in \mathcal{R}(K)^\perp$ impliziert, dass u eine bestapproximierende Lösung ist. Insbesondere gelten:

- (1) Für $f = Ku$, d.h. $f \in \mathcal{R}(K)$, ist u eine bestapproximierende Lösung.
- (2) Ist $f \in \mathcal{R}(K)^\perp$, so folgt $P_{\mathcal{R}(K)}(f) = 0$, d.h. jedes $u \in \mathcal{N}(K)$ ist bestapproximierende Lösung. Insbesondere ist $u = 0 \in \mathcal{N}(K)$ die Lösung minimaler Norm.

Im Fall (1) gilt $u = u_0 + w$ mit $u_0 \in \mathcal{N}(K)$, $w \in \mathcal{N}(K)^\perp$ (denn $\mathcal{N}(K)$ ist abgeschlossen, also $\mathcal{N}(K) \oplus \mathcal{N}(K)^\perp = X$), $Ku = K(u_0 + w) = Ku_0 + Kw = Kw$ und $\|u\|^2 = \|u_0\|_X^2 + \|w\|_X^2 \geq \|w\|_X^2$ mit Gleichheit genau dann, wenn $u_0 = 0$. Betrachte nun

$$\tilde{K} : \mathcal{N}(K)^\perp \rightarrow \mathcal{R}(K); \quad \tilde{K}u := Ku.$$

Dann ist \tilde{K} bijektiv, d.h.

$$\tilde{K}^{-1} : \mathcal{R}(K) \rightarrow \mathcal{N}(K)^\perp \subseteq X$$

existiert. Setze dann

$$K^\dagger : \mathcal{D}(K^\dagger) \rightarrow X; \quad K^\dagger f := \begin{cases} \tilde{K}^{-1}f, & \text{falls } f \in \mathcal{R}(K), \\ 0, & \text{falls } f \in \mathcal{R}(K)^\perp. \end{cases}$$

Es lässt sich zeigen: Für jedes $y \in \mathcal{D}(K^\dagger)$ besitzt die Gleichung $Kx = y$ eine eindeutige Lösung minimaler Norm, gegeben durch $x^\dagger = K^\dagger y$. Die Menge aller bestapproximierenden Lösungen ist dann gegeben durch $\{x^\dagger\} + \mathcal{N}(K)$.

Wir wollen noch eine Reihendarstellung von K^\dagger herleiten. Betrachte dazu:

$$\sigma_n = 0 \Rightarrow Ku_n = 0, \quad K^*v_n = 0 \Rightarrow u_n \in \mathcal{N}(K), \quad v_n \in \mathcal{N}(K^*);$$

$$Kv_n = 0 \Leftrightarrow \langle K^*v_n, u \rangle = 0 \quad \forall u \in X \Leftrightarrow \langle v_n, Ku \rangle = 0 \quad \forall u \in X \Leftrightarrow v_n \in \mathcal{R}(K)^\perp.$$

Weiter gilt für $x^\dagger = K^\dagger f$:

$$\sum_{n=1}^{\infty} \sigma_n^2 \langle x^\dagger, u_n \rangle u_n = K^*Kx^\dagger = K^*f = \sum_{n=1}^{\infty} \sigma_n \langle f, v_n \rangle u_n.$$

Durch Koeffizientenvergleich ergibt sich $\langle x^\dagger, u_n \rangle u_n = \frac{1}{\sigma_n} \langle f, v_n \rangle u_n$, also insgesamt:

$$x^\dagger = K^\dagger f = \sum_{\sigma_n \neq 0} \frac{\langle f, v_n \rangle}{\sigma_n} u_n.$$

Im folgenden wollen wir $\sigma_n \neq 0$ für alle $n \in \mathbb{N}$ annehmen.

3 Allgemeine lineare Regularisierung mit SVD

Wir haben oben gesehen, dass die ersten beiden Bedingungen aus Definition 1.6 größtenteils erfüllt werden können, indem man verallgemeinerte Inverse betrachtet. Um der dritten Bedingung gerecht zu werden, verwenden wir Regularisierungen. Diese modifizieren im Wesentlichen die kleinsten Singulärwerte. Für $f \in \mathcal{D}(K^\dagger)$ wollen wir also

$$K^\dagger f = \sum_{n=1}^{\infty} \frac{1}{\sigma_n} \langle f, v_n \rangle u_n$$

durch

$$R_\alpha f := \sum_{n=1}^{\infty} g_\alpha(\sigma_n) \langle f, v_n \rangle u_n$$

approximieren, wobei $g_\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ eine für fixes α beschränkte Funktion mit $g_\alpha(t) \xrightarrow{\alpha \rightarrow 0} \frac{1}{t}$ (punktweise für $t > 0$) ist. Setze $C_\alpha := \sup_{t>0} g_\alpha(t) < \infty$. Es gilt dann $C_\alpha \approx \frac{1}{\alpha}$ und damit $\lim_{\alpha \rightarrow 0} C_\alpha = \infty$. Beachte die Abschätzung:

$$\|R_\alpha f\|^2 = \sum_{n=1}^{\infty} (g_\alpha(\sigma_n))^2 |\langle f, v_n \rangle|^2 \leq C_\alpha^2 \sum_{n=1}^{\infty} |\langle f, v_n \rangle|^2 \leq C_\alpha^2 \|f\|^2.$$

3.1 Beispiele: Im folgenden werden wir näher untersuchen:

- (1) Die abgeschnittene Singulärwertzerlegung (Truncated Singular Value Decomposition, TS-VD):

$$g_\alpha(t) = \begin{cases} 0, & t \leq \alpha, \\ \frac{1}{t}, & t > \alpha. \end{cases}$$

Es gilt offensichtlich $C_\alpha = \frac{1}{\alpha}$.

(2) Die Lavrentiev-Regularisierung:

$$g_\alpha(t) = \frac{1}{t + \alpha},$$

d.h. alle Singulärwerte werden um α verschoben. Wegen $\frac{1}{\sigma+\alpha} \leq \frac{1}{\alpha}$ erhalten wir $C_\alpha = \frac{1}{\alpha}$.

(3) Die Tikhonov-Regularisierung:

$$g_\alpha(t) = \frac{t}{t^2 + \alpha}.$$

Es gilt $C_\alpha = \frac{1}{2\sqrt{\alpha}}$.

Frage: Ist R_α eine vernünftige Regularisierung? Was ist eigentlich eine Regularisierung bzw. ein Regularisierungsoperator?

Sei u^* die exakte Lösung, d.h. es gilt $Ku^* = f$. Weiter seien gestörte Daten f^δ gegeben. Dann gilt $f^\delta = f + (f^\delta - f)$, wobei $f^\delta - f$ die Störung der exakten Daten ist mit $\|f^\delta - f\|_\infty \leq \delta$ ist. Wir berechnen $R_\alpha f^\delta = u_\alpha^\delta$. Im Optimalfall soll dann $u_\alpha^\delta \approx u^*$ gelten, d.h. $u_\alpha^\delta \rightarrow u^*$ für $\alpha \rightarrow 0, \delta \rightarrow 0$. Beachte allerdings, dass α nicht beliebig gegen 0 gehen darf.

Für den Fehler gilt nun:

$$u_\alpha^\delta - u^* = R_\alpha f^\delta - u^* = R_\alpha(f^\delta - f) + (R_\alpha f - u^*).$$

Dabei bezeichnet man $R_\alpha(f^\delta - f)$ als den fortgepflanzten Datenfehler und $R_\alpha f - u^*$ als den Approximations- bzw. systematischen Fehler. Damit gilt dann auch:

$$\|u_\alpha^\delta - u^*\| \leq \|R_\alpha(f^\delta - f)\| + \|R_\alpha f - u^*\|.$$

Für den Datenfehler haben wir die Abschätzung

$$\|R_\alpha(f^\delta - f)\| \leq \|R_\alpha\| \|f^\delta - f\|_\infty \leq C_\alpha \delta,$$

sodass für als zusätzliche Bedingung für Konzergenz

$$C_\alpha \delta \rightarrow 0 \text{ für } \alpha \rightarrow 0, \delta \rightarrow 0$$

erhalten. Wir wählen α somit in Abhängigkeit von δ . Diese Überlegungen führen uns zu folgenden Definitionen:

3.2 Definition: Eine Familie von beschränkten linearen Operatoren $(R_\alpha)_{\alpha>0}$, $R_\alpha : Y \rightarrow X$, zusammen mit einer Parameterwahl

$$A : \mathbb{R}^+ \times Y \rightarrow \mathbb{R}^+; (\delta, f^\delta) \mapsto \alpha,$$

heiSt eine Regularisierung von K^\dagger , wenn

$$R_{\alpha(\delta, f^\delta)} f^\delta \xrightarrow{\delta \rightarrow 0} K^\dagger f$$

für alle f^δ mit $\|f^\delta - f\| \leq \delta$ gilt.

3.3 Definition: Eine Parameterwahl heiSt a-priori, wenn $\alpha = \alpha(\delta)$ unabhängig von f^δ ist und a-posteriori, wenn $\alpha = \alpha(\delta, f^\delta)$ echt von den Daten abhängt.

Eine Parameterwahl $\alpha = \alpha(f^\delta)$ kann nur für korrekt gestellte Probleme funktionieren.

3.1 Konvergenz von $R_\alpha f$

Für den fortgepflanzten Datenfehler kennen wir bereits die Abschätzung $\|R_\alpha(f^\delta - f)\| \leq C_\alpha \delta$ ohne weitere Annahmen. Wir wollen nun den Approximationsfehler genauer untersuchen: gilt $\|R_\alpha f - K^\dagger f\| \leq \varphi(\alpha)$ für eine stetige Funktion $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ mit $\varphi(0) = 0$?

Zunächst kennen wir die Darstellung

$$R_\alpha f - K^\dagger f = \sum_{n=1}^{\infty} (g_\alpha(\sigma_n) - \frac{1}{\sigma_n}) \langle f, v_n \rangle u_n.$$

Betrachten wir den Spezialfall $u^* = u_m$, so gilt $f = Ku_m = \sigma_m v_m$ und damit:

$$R_\alpha f - K^\dagger f = (g_\alpha(\sigma_m) - \frac{1}{\sigma_m}) \sigma_m u_m \text{ bzw. } \|R_\alpha f - K^\dagger f\| = |\sigma_m g_\alpha(\sigma_m) - 1|.$$

Für m groß genug bzw. σ_m klein genug gilt $\sigma_m g_\alpha(\sigma) < \frac{1}{2}$ (da $\sigma g_\alpha(\sigma) \rightarrow 0$ für $\sigma \rightarrow 0$), also $\|R_\alpha f - K^\dagger f\| > \frac{1}{2}$. Damit ist die Konvergenz für $\alpha \rightarrow 0$ beliebig langsam. Damit wird der systematische Fehler nur beliebig langsam kleiner.

3.4 Satz: Sei $g_\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ stückweise stetig, beschränkt für fixes $\alpha > 0$ und $g_\alpha(\sigma) \xrightarrow{\alpha \rightarrow 0} \frac{1}{\sigma}$ für σ fix. Weiter existiere eine Konstante γ mit

$$\sup_{\alpha > 0, \sigma > 0} \sigma g_\alpha(\sigma) \leq \gamma.$$

Dann gilt für R_α wie oben und $f \in \mathcal{D}(K^\dagger)$:

$$R_\alpha f \rightarrow K^\dagger f \text{ für } \alpha \rightarrow 0.$$

3.5 Satz: Sei g_α wie oben und $u_\alpha = R_\alpha f, u_\alpha^\delta = R_\alpha f^\delta$. Dann gelten:

- (1) $u_\alpha \rightarrow K^\dagger f$ für $\alpha \rightarrow 0$ (Satz 3.4).
- (2) $\|u_\alpha - u_\alpha^\delta\| \leq C_\alpha \delta$, $\|Ku_\alpha - Ku_\alpha^\delta\| \leq \gamma \delta$.
- (3) Wenn $C_\alpha \delta \rightarrow 0$, so folgt $u_\alpha^\delta \rightarrow K^\dagger f$.

Wir wollen nun eine weitere Sichtweise auf den systematischen Fehler gewinnen, nämlich eine Art erweiterte Picard-Bedingung. Betrachte dazu für $\vartheta > 0$:

$$\begin{aligned} \|K^\dagger f - R_\alpha f\|^2 &= \sum_{n=1}^{\infty} (\frac{1}{\sigma_n} - g_\alpha(\sigma_n))^2 \langle f, v_n \rangle^2 \\ &= \sum_{n=1}^{\infty} (1 - \sigma_n g_\alpha(\sigma_n))^2 \frac{\langle f, v_n \rangle^2}{\sigma_n^2} \\ &= \sum_{n=1}^{\infty} (1 - \sigma_n g_\alpha(\sigma_n))^2 \sigma_n^{2\vartheta} \frac{\langle f, v_n \rangle^2}{\sigma_n^{2+2\vartheta}}. \end{aligned}$$

Als zusätzliche Bedingung fordern wir, dass $\frac{\langle f, v_n \rangle}{\sigma_n^{1+\vartheta}}$ quadratsummierbar ist und setzen dann

$$z = \sum_{n=1}^{\infty} \frac{\langle f, v_n \rangle}{\sigma_n^{1+\vartheta}} u_n \in X.$$

Damit gilt:

$$\|K^\dagger f - R_\alpha f\|^2 \leq \max_{\sigma} (\sigma^\vartheta - \sigma^{1+\vartheta} g_\alpha(\sigma))^2 \sum_{n=1}^{\infty} \frac{\langle f, v_n \rangle^2}{\sigma_n^{2+2\vartheta}}.$$

Definiere dann eine Funktion in Abhängigkeit von α :

$$\varphi_\vartheta(\alpha) := \max_{\sigma} |\sigma^\vartheta - \sigma^{1+\vartheta} g_\alpha(\sigma)|.$$

Damit erhalten wir schlieSSlich:

$$\|K^\dagger f - R_\alpha f\| \leq \varphi_\vartheta(\alpha) \|z\|.$$

Wir wollen nun φ_ϑ für Standardbeispiele von Regularisierungen berechnen.

3.6 Beispiel (TSVD): Wir betrachten

$$g_\alpha(\sigma) = \begin{cases} \frac{1}{\sigma}, & \sigma > \alpha, \\ 0, & \sigma \leq \alpha. \end{cases}$$

Dann gilt:

$$\sigma^\vartheta - \sigma^{1+\vartheta} g_\alpha(\sigma) = \begin{cases} 0, & \sigma > \alpha, \\ \sigma^\vartheta, & \sigma \leq \alpha, \end{cases}$$

also $\varphi_\vartheta(\alpha) = \max_{\sigma} |\sigma^\vartheta - \sigma^{1+\vartheta} g_\alpha(\sigma)| = \alpha^\vartheta$. Für den Gesamtfehler erhalten wir (mit $C_\alpha = \frac{1}{\alpha}$):

$$\|K^\dagger f - R_\alpha f^\delta\| \leq \frac{\delta}{\alpha} + \|z\| \alpha^\vartheta.$$

Wir wollen nun den optimalen Parameter α (bei δ, ϑ, z gegeben) bestimmen, d.h. wir betrachten:

$$f(\alpha) = \frac{\delta}{\alpha} + \|z\| \alpha^\vartheta \rightarrow \min_{\alpha > 0}.$$

Es gilt $f'(\alpha) = -\frac{\delta}{\alpha^2} + \|z\| \vartheta \alpha^{\vartheta-1}$ und damit:

$$f'(\tilde{\alpha}) = 0 \Leftrightarrow \tilde{\alpha}^{1+\vartheta} = \frac{\delta}{\vartheta \|z\|} \Rightarrow \tilde{\alpha} = \frac{\delta^{\frac{1}{1+\vartheta}}}{(\vartheta \|z\|)^{\frac{1}{1+\vartheta}}}.$$

Damit ergibt sich als optimale Fehlerschranke:

$$\|K^\dagger f - R_\alpha f^\delta\| \leq \delta^{\frac{\vartheta}{1+\vartheta}} (\vartheta \|z\|)^{\frac{1}{1+\vartheta}} + \|z\| \frac{\delta^{\frac{\vartheta}{1+\vartheta}}}{(\vartheta \|z\|)^{\frac{\vartheta}{1+\vartheta}}} = \text{const}(\vartheta) \delta^{\frac{\vartheta}{1+\vartheta}} \|z\|^{\frac{1}{1+\vartheta}}.$$

Zur Interpretation:

- (1) Der Fehler wird kleiner, wenn ϑ grōSSer wird.
- (2) Der Fehler hat immer eine kleinere Ordnung als δ , was eine Konsequenz aus der Schlechtgestelltheit ist.
- (3) Der Fehler wächst mit z , wir können ein f , das die Bedingung nicht erfüllt, nicht durch gute f_n approximieren.

3.7 Beispiel (Lavrentiev): Wir betrachten $g_\alpha(\sigma) = \frac{1}{\sigma+\alpha}$. Es gilt wieder $C_\alpha = \frac{1}{\alpha}$. Dann gilt:

$$\varphi_\vartheta(\alpha) = \max_{\sigma} \left| \sigma^\vartheta - \frac{\sigma^{1+\vartheta}}{\sigma + \alpha} \right| = \max_{\sigma} \frac{\alpha \sigma^\vartheta}{\sigma + \alpha} = \alpha \max_{\sigma} \frac{\sigma^\vartheta}{\sigma + \alpha}.$$

Für $\vartheta > 1$ existiert das Maximum nicht ($\sup = \infty$), also ist nur $\vartheta \in [0, 1]$ interessant. Sei nun $f(\sigma) = \frac{\sigma^\vartheta}{\sigma + \alpha}$. Dann gilt:

$$f'(\sigma) = \frac{\vartheta \sigma^{\vartheta-1}}{\sigma + \alpha} - \frac{\sigma^\vartheta}{(\sigma + \alpha)^2} = \frac{\sigma^{\vartheta-1}(\vartheta(\sigma + \alpha) - \sigma)}{(\sigma + \alpha)^2}.$$

Es gilt dann $f'(\tilde{\sigma}) = 0 \Leftrightarrow \tilde{\sigma} = \frac{\vartheta\alpha}{1-\vartheta}$. Damit folgt dann:

$$\varphi_\vartheta(\alpha) = \alpha f(\tilde{\sigma}) = \alpha \left(\frac{\vartheta\alpha}{1-\vartheta} \right)^\vartheta \frac{1}{\frac{\vartheta\alpha}{1-\vartheta} + \alpha} = const(\vartheta)\alpha^\vartheta.$$

Eine analoge Rechnung zu oben liefert:

$$\|K^\dagger f - R_\alpha f^\delta\| \leq const(\vartheta)\delta^{\frac{\vartheta}{1+\vartheta}} \|z\|^{\frac{1}{1+\vartheta}},$$

allerdings nur für $\vartheta \leq 1$. Wir erhalten die bestmögliche Schranke $\delta^{\frac{1}{2}}$ für $\vartheta = 1$ und sagen dann, dass das Verfahren die Qualifikation 1 hat.

3.8 Definition: Die Qualifikation eines Regularisierungsverfahrens ist der größtmögliche Wert ϑ , der zu einer endlichen Fehlerabschätzung führt. (Tikhonov hat Qualifikation 0.)

Wir wollen noch eine Interpretation der Bedingung $z = \sum_{n=1}^{\infty} \frac{\langle f, v_n \rangle}{\sigma_n^{1+\vartheta}} u_n \in X$ für verschiedene ϑ angeben.

Im Fall $\vartheta = 0$ erhalten wir die Picard-Bedingung.

Im Fall $\vartheta = 1$ gilt:

$$z = \sum_{n=1}^{\infty} \frac{\langle f, v_n \rangle}{\sigma_n^2} u_n \stackrel{u^* = K^\dagger f}{=} \sum_{n=1}^{\infty} \frac{\langle u^*, u_n \rangle}{\sigma_n} u_n.$$

Sei nun $w = \sum_{n=1}^{\infty} \frac{\langle u^*, u_n \rangle}{\sigma_n} v_n \in Y$. Wegen $K^* v_n = \sigma_n u_n$ gilt dann:

$$K^* w = \sum_{n=1}^{\infty} \frac{\langle u^*, u_n \rangle}{\sigma_n} K^* v_n = \sum_{n=1}^{\infty} \langle u^*, u_n \rangle u_n = u^*.$$

Damit erhalten wir:

$$z \in X \Leftrightarrow w \in Y \Leftrightarrow u^* = K^* w.$$

Dies ist eine abstrakte Glattheitsbedingung an die Minimum-Norm-Lösung u^* ($u^* \in \mathcal{R}(K^*)$). Diese Bedingung heiSt Quellbedingung. Es gilt dann $\|w\| = \|z\|$. Das Element w (bzw. z) heiSt Quellelement.

Eine weitere Interpretation dieses Falls sieht folgendermaSSen aus: u^* ist die Lösung minimaler Norm, d.h. u^* löst für $f \in \mathcal{R}(f)$:

$$\begin{aligned} \frac{1}{2} \|u\|^2 \rightarrow \min_{u \in X, Ku=f} &\Leftrightarrow \min_{u \in X} \underbrace{\sup_{w \in Y} \left\{ \frac{1}{2} \|u\|^2 + \langle w, f - Ku \rangle \right\}}_{\begin{cases} \frac{1}{2} \|u\|^2, & Ku = f, \\ +\infty, & Ku \neq f. \end{cases}} \\ &= \begin{cases} \frac{1}{2} \|u\|^2, & Ku = f, \\ +\infty, & Ku \neq f. \end{cases} \end{aligned}$$

Wähle dann $w = t(f - Ku)$ für $t \rightarrow \infty$. Es gilt weiter:

$$\text{Ein Sattelpunkt } (u^*, w^*) \text{ existiert } \Leftrightarrow \min_{u \in X} \sup_{w \in Y} = \min_{u \in X} \max_{w \in Y} = \max_{w \in Y} \min_{u \in X}.$$

Berechne nun:

$$\begin{aligned} \min_{u \in X} \frac{1}{2} \|u\|^2 + \langle w, f - Ku \rangle &= \min_{u \in X} \frac{1}{2} \|u\|^2 - \langle u, K^* w \rangle + \frac{1}{2} \|K^* w\|^2 - \frac{1}{2} \|K^* w\|^2 + \langle w, f \rangle \\ &= \min_{u \in X} \frac{1}{2} \|u - K^* w\|^2 - \frac{1}{2} \|K^* w\|^2 + \langle w, f \rangle. \end{aligned}$$

Damit erhalten wir ein Minimum bei $u = K^*w$, d.h. ein Sattelpunkt (u^*, w^*) existiert nur, wenn $u^* = K^*w^*$.

Im Fall $\vartheta = 2$ gilt:

$$z = \sum_{n=1}^{\infty} \frac{\langle f, v_n \rangle}{\sigma_n^3} u_n = \sum_{n=1}^{\infty} \frac{\langle u^*, u_n \rangle}{\sigma_n^2} u_n$$

und weiter:

$$K^*Kz = \sum_{n=1}^{\infty} \frac{\langle u^*, u_n \rangle}{\sigma_n^2} K^*Ku_n = \sum_{n=1}^{\infty} \langle u^*, u_n \rangle u_n = u^*,$$

denn $K^*Ku_n = \sigma_n^2 u_n$. Damit erhalten wir erneut eine Quellbedingung:

$$u^* \in \mathcal{R}(K^*K) \Leftrightarrow \text{Es existiert ein Sattelpunkt f\"ur } \min_{u \in X} \sup_z \left\{ \frac{1}{2} \|u\|^2 - \langle z, K^*(Ku - f) \rangle \right\}.$$

F\"ur allgemeines ϑ ergibt sich die Quellbedingung $u^* \in \mathcal{R}((K^*K)^{\frac{\vartheta}{2}})$ mit

$$(K^*K)^{\frac{\vartheta}{2}} x := \sum_{n=1}^{\infty} \langle x, u_n \rangle \sigma_n^{2\vartheta} u_n.$$

Quellbedingungen sind zusätzliche Glattheitsbedingungen an die L\"ösung minimaler Norm (und damit an die exakten Daten f).

3.2 Praktische Realisierung von linearen Regularisierungsverfahren

3.9 Beispiel (TSVD): Wir betrachten:

$$g_{\alpha}(\sigma) = \begin{cases} \frac{1}{\sigma}, & \sigma \geq \alpha, \\ 0, & \sigma < \alpha. \end{cases}$$

Dies ist nur mit der Singulärwertzerlegung berechenbar:

$$R_{\alpha}y = \sum_{n=1}^{N(\alpha)} \frac{1}{\sigma_n} \langle y, v_n \rangle u_n,$$

wobei $N(\alpha) = \max\{n \mid \sigma_n > \alpha\}$. Damit werden nur endlich viele Singulärvektoren ben\"otigt. Ist δ nicht sehr klein, so ist α nicht sehr klein und $N(\alpha)$ eventuell sehr klein. Es werden eventuell also nur sehr wenige Singulärvektoren ben\"otigt.

3.10 Beispiel (Lavrentiev): Wir haben $g_{\alpha}(\sigma) = \frac{1}{\sigma + \alpha}$ und damit:

$$R_{\alpha}y = \sum_{n=1}^{\infty} \frac{\langle y, v_n \rangle}{\sigma_n + \alpha} u_n = x_{\alpha}.$$

Damit gilt dann:

$$y = \sum_{n=1}^{\infty} (\sigma_n + \alpha) \langle x_{\alpha}, u_n \rangle v_n \stackrel{\sigma_n v_n = Ku_n}{=} \sum_{n=1}^{\infty} \langle y, v_n \rangle v_n,$$

und weiter:

$$\begin{aligned} y &= \sum_{n=1}^{\infty} \langle x_{\alpha}, u_n \rangle (Ku_n + \alpha v_n) \\ &= K \left(\sum_{n=1}^{\infty} \langle x_{\alpha}, u_n \rangle u_n \right) + \alpha \left(\sum_{n=1}^{\infty} \langle x_{\alpha}, u_n \rangle v_n \right) \\ &= Kx_{\alpha} + \alpha \left(\sum_{n=1}^{\infty} \langle x_{\alpha}, u_n \rangle v_n \right). \end{aligned}$$

Im Spezialfall $u_n = v_n$ ($X = Y$) gilt dann:

$$y = Kx_\alpha + \alpha x_\alpha = (K + \alpha I)x_\alpha.$$

Dieser Spezialfall tritt genau dann ein, wenn $K : X \rightarrow X$ symmetrisch ($K^* = K$) und positiv semidefinit ist ($\langle Ku, u \rangle \geq 0 \forall u \in X$), denn $Ku_n = \sigma_n u_n$ und $K^*u_n = \sigma_n u_n$, also haben K und K^* die gleiche SVD $\Leftrightarrow K = K^*$ und $\sigma \geq 0 \Leftrightarrow K$ positiv semidefinit.

3.11 Beispiel (Tikhonov): Wir haben $g_\alpha(\sigma) = \frac{\sigma}{\sigma^2 + \alpha}$ und damit:

$$R_\alpha y = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha} \langle y, v_n \rangle u_n = x_\alpha.$$

Umgekehrt gilt:

$$Kx_\alpha = \sum_{n=1}^{\infty} \frac{\sigma_n \langle y, v_n \rangle}{\sigma_n^2 + \alpha} Ku_n = \sum_{n=1}^{\infty} \frac{\sigma_n^2 \langle y, v_n \rangle}{\sigma_n^2 + \alpha} v_n,$$

und weiter:

$$K^*Kx_\alpha = \sum_{n=1}^{\infty} \frac{\sigma_n^2 \langle y, v_n \rangle}{\sigma_n^2 + \alpha} K^*v_n = \sum_{n=1}^{\infty} \frac{\sigma_n^3 \langle y, v_n \rangle}{\sigma_n^2 + \alpha} u_n.$$

Schließlich gilt:

$$\begin{aligned} K^*Kx_\alpha + \alpha x_\alpha &= \sum_{n=1}^{\infty} \frac{\sigma_n^3 + \alpha \sigma_n}{\sigma_n^2 + \alpha} \langle y, v_n \rangle u_n \\ &= \sum_{n=1}^{\infty} \langle y, v_n \rangle \sigma_n u_n \\ &\stackrel{\sigma_n u_n = K^*v_n}{=} K^* \left(\sum_{n=1}^{\infty} \langle y, v_n \rangle v_n \right) = K^*y. \end{aligned}$$

Damit ist die Tikhonov-Regularisierung realisierbar durch:

$$(K^*K + \alpha I)x_\alpha^\delta = K^*y^\delta.$$

Eine weitere Formulierung ist über ein Minimierungsproblem möglich:

$$K^*Kx = K^*y \Leftrightarrow \min_{x \in X} \frac{1}{2} \|Kx - y\|^2.$$

Setzen wir

$$J_\alpha(u) := \frac{1}{2} \|Ku - y\|^2 + \frac{\alpha}{2} \|u\|^2,$$

so gilt:

3.12 Satz: Ein $x_\alpha \in X$ minimiert J_α genau dann, wenn $K^*Kx_\alpha + \alpha x_\alpha = K^*y$. Dies ist ein Kompromiss zwischen der bestapproximierenden Lösung und der Lösung minimaler Norm.

3.13 Beispiel (Fixpunktiteration): Wir haben:

$$Ku = y \leftrightarrow K^*Ku = K^*y.$$

In Fixpunktform erhalten wir $u = u - \tau(K^*Ku - K^*y)$, also

$$u^{k+1} = u^k - \tau(K^*Ku^k - K^*y).$$

3.14 Beispiel (Asymptotische Regularisierung): Wir interpretieren den Parameter τ in Beispiel 3.13 als Zeitschritt und $t = k\tau$ und $u^k \approx u(t)$. Dann erhalten wir durch Umformen:

$$\frac{u(t + \tau) - u(t)}{\tau} = -(K^* K u(t) - K^* y).$$

Für $\tau \rightarrow 0$ ergibt sich:

$$\frac{du}{dt} = -(K^*(Ku - y)).$$

Sei weiter $u(0) = 0, t \in [0, T]$. Dann ist der Regularisierungsparameter quasi T , d.h. $\alpha = \frac{1}{T}$. Durch Lösen dieser gewöhnlichen Differentialgleichung ergibt sich:

$$g_\alpha(\sigma) = \frac{1}{\sigma}(1 - e^{-\frac{\sigma^2}{\alpha}}) \xrightarrow{\alpha \rightarrow 0} \frac{1}{\sigma}.$$

3.15 Praktische Wahl des Regularisierungsparameters: Theoretisch gilt $\alpha \sim \delta^{\frac{1}{1+\vartheta}}$. Dies ist jedoch unpraktikabel, da ϑ nicht bekannt ist. Für optimales α zu gegebenem ϑ haben wir:

$$\|Kx_\alpha^\delta - y^\delta\| \leq (1 + \gamma)\delta,$$

wobei $\gamma = \max(\sigma g_\alpha(\sigma))$. Wir wissen zum Einen, dass der Output-Fehler die Ordnung δ hat, aber zum Anderen auch, dass es nicht notwendig ist, eine regularisierte Lösung zu suchen, sodass $\|Kx_\alpha^\delta - y^\delta\| < \delta$ gilt. Dies führt uns zum

Diskrepanzprinzip: Wähle $\alpha = \alpha(\delta, y^\delta)$ so, dass für $x_\alpha^\delta = R_\alpha y^\delta$ $\|Kx_\alpha^\delta - y^\delta\| \approx \delta$ gilt, d.h.

$$\alpha(\delta, y^\delta) = \max\{\alpha > 0 \mid \|Kx_\alpha^\delta - y^\delta\| \leq \tau\delta\}$$

für ein $\tau \geq 1$. Im Fall $\tau = 1$ gilt dann für die exakte Lösung x^* :

$$\|y - y^\delta\| = \|Kx^* - y^\delta\| \leq \delta.$$

Das α aus dem Diskrepanzprinzip liefert eine optimale Rate, wenn $x^* = (K^* K)^\vartheta w$.

4 Stochastische Modellierung

Wir betrachten das Problem $Ku = f$ mit $K : \mathbb{R}^N \rightarrow \mathbb{R}^M$. Dabei sind die auftretenden Objekte allesamt unbekannt bzw. nur ungenau als Zufallsvariable bekannt. Im Einzelnen bedeutet dies:

- (1) Rauschen: Wir haben eine Zufallsvariable F mit $\mathbb{E}[F] = f = Ku^*$.
- (2) Parameter in K (oder K selbst): Beispielsweise kann K als Integral mit Faltungskern modelliert werden. Bei ungenauem Faltungskern entstehen Fehler in der Modellierung.
- (3) Variable u : Die exakte Lösung u^* ist die Realisierung einer Zufallsvariable U . Die Wahrscheinlichkeitsverteilung für U modelliert das Vorwissen (A-priori-Verteilung, A-priori-Wahrscheinlichkeit).

Wir betrachten nun die Likelihood, das heißt die Wahrscheinlichkeit f zu messen, wenn u^* eine exakte Lösung ist. Beides sind Realisierungen von Zufallsvariablen F, U . Die Likelihood l ist definiert als bedingte Wahrscheinlichkeit $l = P(F|U)$.

4.1 Beispiel: Wir betrachten additives, unabhängig normalverteiltes Rauschen. Sei dazu F eine Zufallsvariable in \mathbb{R}^M . Dann haben wir für $1 \leq i \leq M$:

$$F_i = (KU)_i + n_i, \quad n_i \sim \mathcal{N}(0, \sigma^2) \quad (\sigma \sim \delta).$$

Dann gilt:

$$l = P(F|U) = P(n = F - KU) \stackrel{\text{unabh.}}{=} \prod_{i=1}^M P(n_i = F_i - (KU)_i) \stackrel{\text{z.B.}}{=} \prod_{i=1}^M \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(F_i - (KU)_i)^2}{2\sigma^2}\right).$$

Die Schätzung der Maximum Likelihood ergibt:

$$\hat{u} = \arg \max_U P(F|U) = \arg \max_U \log P(F|U) = \arg \min_U -\log(P(F|U)).$$

Für obiges Beispiel erhalten wir:

$$-\log P(F|U) = -M \underbrace{\log \frac{1}{\sqrt{2\pi}\sigma}}_{\text{unabh. von } u} + \underbrace{\sum_{i=1}^M \frac{((KU)_i - F_i)^2}{2\sigma^2}}_{\frac{1}{2\sigma^2} \|KU - F\|_2^2 + \text{const.}},$$

also

$$\hat{u} = \arg \max_U P(F|U) = \arg \min_U \|KU - F\|_2^2,$$

d.h. die Maximum Likelihood wird bei der bestapproximierenden Lösung angenommen. Wir brauchen damit zusätzlich eine Regularisierung!

4.2 Beispiel: Wir betrachten die Penalized Maximum Likelihood:

$$\hat{u} = \arg \min -\log P(F|u) + \alpha \|u\|^2.$$

Die entsprechende Likelihood ist dann:

$$\tilde{l} = cP(F|U)e^{-\alpha\|u\|^2}, \quad c := \frac{1}{\int_{\mathbb{R}^M} P(F|U)e^{-\alpha\|u\|^2} du}.$$

4.3 Satz von Bayes: Wir nehmen an, dass wir $P(F|U)$ und a-priori $P(U)$ kennen. Dann liefert der Satz von Bayes die a-posteriori Wahrscheinlichkeit

$$P(U|F) = \frac{P(F|U)P(U)}{P(F)}.$$

$P(U|F)$ bezeichnet also die Wahrscheinlichkeit, dass U die zugrunde liegende Lösung ist, wenn ich die Daten F messe.

Anstelle eines Maximum Likelihood Schätzers kann man auch den Maximum A-posteriori Probability (MAP) Schätzer betrachten:

$$\begin{aligned} \hat{u} &= \arg \max_u P(u|F) \\ &= \arg \min_u -\log(P(u|F)) \\ &\stackrel{\text{Bayes}}{=} \arg \min_u -\log P(F|U) - \log P(u) + \underbrace{\log P(F)}_{\text{unabh. von } u} \\ &= \arg \min_u -\log(P(F|U)) - \log P(u). \end{aligned}$$

Wir erhalten dann einen Datenterm $D(u, f) = -\log P(f|u)$ und einen Regularisierungsterm $R(u) = -\log P(u)$. Häufig modellieren wir die a-priori Wahrscheinlichkeitsverteilung mithilfe eines Gibbs-Priors, d.h. $P(u) = \exp(-\alpha R(u))$ für eine Konstante $\alpha > 0$ und einen (konvexen) Regularisierungsterm R , sodass sich $-\log P(u) = \alpha R(u)$ ergibt. (Manchmal arbeiten wir zusätzlich mit einer Normalisierungskonstante c , d.h. $P(u) = c \exp(-\alpha R(u))$.) Schließlich erhalten wir für den MAP-Schätzer das Problem

$$J_\alpha(u) := D(u, f) + \alpha R(u) \rightarrow \min_u .$$

4.1 Beispiele für Datenterme

4.4 Beispiel: Sei $n_i \sim \mathcal{N}(0, \sigma^2)$ GauSSsches i.i.d. Rauschen und $f_i^M = (K^M u)_i + n_i$. Wir stellen uns vor, dass K^M die Diskretisierung eines Integraloperators und f^M die Diskretisierung einer Funktion ist. Sei Beispielsweise $\Omega = [0, 1]$ und $\Omega_i^M = [\frac{i-1}{M} - \frac{i}{M}]$ für $i = 1, \dots, M$. Ist dann K ein Integraloperator nach $L^2([0, 1])$, so haben wir:

$$(K^M u)_i = \frac{1}{|\Omega_i^M|} \int_{\Omega_i^M} (Ku)(x) dx, \quad f_i^M \approx \frac{1}{|\Omega_i^M|} \int_{\Omega_i^M} f(x) dx.$$

Für den Datenterm gilt dann:

$$\begin{aligned} D(u, f) &= \frac{1}{2\sigma^2} \sum_{i=1}^M (f_i^M - (K^M u)_i)^2 \\ &= \frac{1}{2\sigma^2} \sum_{i=1}^M \frac{1}{|\Omega_i^M|^2} \left(\int_{\Omega_i^M} f(x) - (Ku)(x) dx \right)^2 \\ &\approx \frac{1}{2\sigma^2} \sum_{i=1}^M \frac{1}{|\Omega_i^M|} \int_{\Omega_i^M} (f(x) - Ku(x))^2 dx \\ &= \frac{M}{2\sigma^2} \int_0^1 (f(x) - Ku(x))^2 dx. \end{aligned}$$

Damit folgt: Die Art des Rauschens impliziert, welche Norm (oder Allgemeineres) man für die Definition der bestapproximierenden Lösung verwenden sollte.

4.5 Beispiel: Sei $f = Ku + n$ mit $n \sim \mathcal{N}(0, \sigma^2(\eta^T \eta)^{-1})$, wobei

$$\eta = \begin{pmatrix} 1 & -1 & \dots & 0 \\ \vdots & 1 & \ddots & \vdots \\ \vdots & & \ddots & -1 \\ 0 & \dots & \dots & 1 \end{pmatrix}.$$

Dann folgt:

$$P(f|u) = \text{const } e^{-\frac{(f-Ku)^T \eta^T \eta (f-Ku)}{2\sigma^2}},$$

und weiter

$$D(u, f) = \frac{1}{2\sigma^2} \|\eta^M (f^M - K^M u)\|^2.$$

Wegen $\eta^M f^M \approx f'$ gilt schlieSSlich:

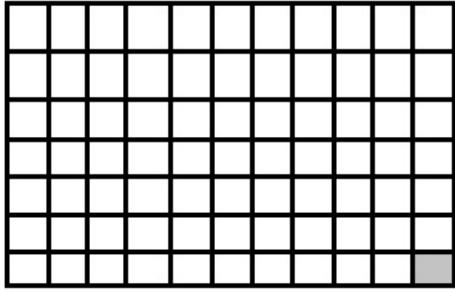
$$D(u, f) = \frac{1}{2\sigma^2} \int_0^1 (f'(x) - (Ku)'(x))^2 dx.$$

Wir halten noch einmal fest: Bei Normalverteilung bzw. GauSS haben wir:

$$D(u, f) = \frac{1}{2\sigma^2} \|Ku - f\|^2, \quad R(u) = \frac{1}{2} \|u\|^2.$$

Im folgenden Beispiel wollen wir die Poissonverteilung betrachten:

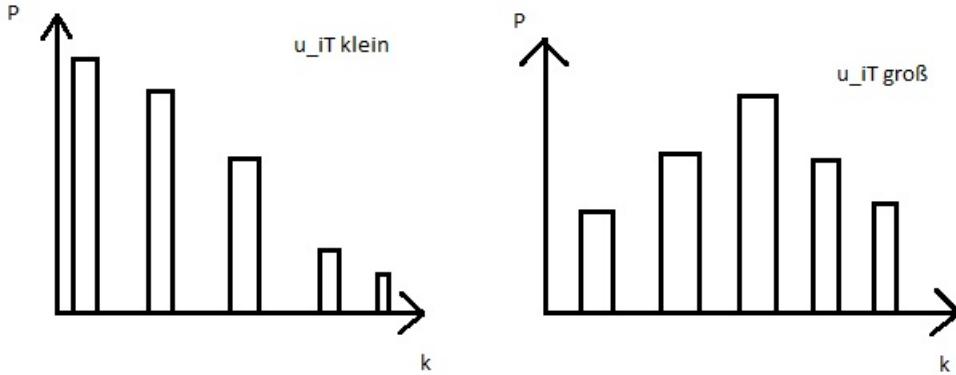
4.6 Beispiel (Photonenrauschen): Wir betrachten N Detektoren und N Unbekannte für:



Sei eine Farbstoffdichte u und ein diskretes $u \in \mathbb{R}^N$ gegeben, wobei u_i das Integral der Dichte über Pixel i ist. Die erzeugte Messung ist die Anzahl f_i von Photonen im i -ten Pixel, also $f_i \in \mathbb{N}_0$. Für ein sehr kleines Zeitintervall der Länge τ ist die Wahrscheinlichkeit ein Photon im i -ten Pixel zu messen gegeben durch $\tau u_i + \mathcal{O}(\tau^2)$. Außerdem nehmen wir an, dass ein Markov-Prozess vorliegt bzw., dass der Prozess sogar zeitunabhängig ist, d.h. die Wahrscheinlichkeit für eine Messung in einem gewissen Zeitintervall hängt nur von u_i und der Länge des Zeitintervalls ab und nicht davon, was vorher passiert ist. Ist die Messzeit T fix, so haben wir:

$$P(\{f_i = k\}|u_i) = \frac{(u_i T)^k}{k!} e^{-u_i T}, \quad k = 0, 1, 2, \dots,$$

d.h. f_i ist Poissonverteilt mit Mittelwert $u_i T$ ($f_i \sim Poi((Ku)_i T)$).



Im Folgenden benötigen wir die

Stirling-Formel: Sie erlaubt eine Approximation für $n!$, wenn n groß ist:

$$\log(n!) \approx n \log(n) - n.$$

Betrachten wir nun also den Datenterm:

$$\begin{aligned} D(u, f) &= -\log(P(f, u)) \\ &= \sum_{i=1}^N u_i T - \log\left(\frac{(u_i T)^{f_i}}{f_i!}\right) \\ &= \sum_{i=1}^N u_i T - f_i \log(u_i T) + \log(f_i!) \\ &\approx \sum_{i=1}^N u_i T + f_i \log(u_i T) + f_i \log(f_i) - f_i. \end{aligned}$$

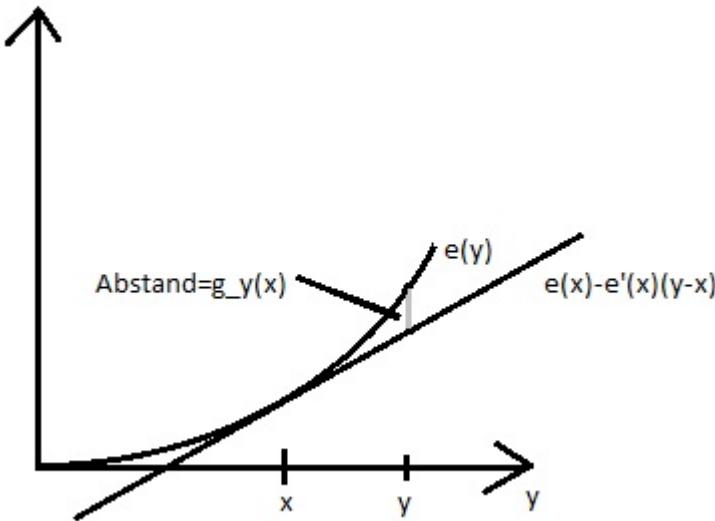
Dann folgt $D(f, f) = 0$ (wobei $u_i T = f_i$). Sei jetzt $T = 1$. Wir wollen untersuchen, ob

$$D(f, u) \geq 0 \Leftrightarrow g_y(x) = x - y \log(y) + y \log(y) - y \geq 0 \quad \forall x, y > 0$$

gilt. Die Ableitung $g'_y(x) = 1 - \frac{y}{x}$ hat nur bei $x = y$ eine Nullstelle. Weiter gilt $g''_y(x) = \frac{y}{x^2} > 0$, da $x, y > 0$. Damit ist g_y konvex und hat ein eindeutiges Minimum bei $x = y$ mit $g_y(x) = 0$. Wir nennen $g_y(x)$ dann Kullback-Leibler-Distanz. Betrachte nun die Entropie $e(x) = x \log(x) - x$. Dann ist e konvex, denn es gilt $e''(x) = \frac{1}{x}$. Für die Taylorentwickelung von e gilt:

$$\begin{aligned} e(y) - \underbrace{(e(x) + e'(x)(y-x))}_{\text{Taylorentwickelung an der Stelle } y} &= y \log(y) - y - x \log(x) + x - \log(x)(y-x) \\ &= x - y \log(x) - y + y \log(y) \\ &= g_y(x). \end{aligned}$$

Damit misst $g_y(x)$ eine Art Abstand zwischen x und y mittels e :



Diese Konstruktion funktioniert für jede konvexe Funktion \tilde{e} . Die zugehörige Distanz heißt dann Bregman-Distanz. Sie misst den Abstand zwischen dem Funktionswert $\tilde{e}(y)$ und dem Wert der Linearisierung $\tilde{e}(x) - \tilde{e}'(x)(y - x)$ an einer Stelle x .

4.7 Beispiel: Betrachte $\tilde{e} = \frac{1}{2}\|x\|^2$. Dann gilt $\tilde{e}'(x) = x$, womit die Bregman-Distanz gegeben ist durch:

$$z_{\tilde{e}}(y, x) = \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x\|^2 - \langle x, (y - x) \rangle = \frac{1}{2}(\|y\|^2 - 2\langle x, y \rangle + \|x\|^2) = \frac{1}{2}\|x - y\|^2.$$

4.2 Beispiele für Regularisierungsterme

Was ist ein potentieller Nachteil von Tikhonov oder ähnlichen Verfahren? Es gilt

$$K^*Ku + \alpha u = K^*f \Rightarrow u_\alpha = (K^*K + \alpha I)^{-1}K^*f = K^*[(K^*K + \alpha I)^{-1}f],$$

also $u_\alpha = K^*w_\alpha$. Ist K^* ein glättender Operator, so ist $u_\alpha \in \mathcal{R}(K^*)$ glatt, was in vielen Anwendungen ein großer Nachteil sein kann.

4.8 Beispiel: Sei $X = Y = L^2(\Omega)$ und $(Ku)(x) = \int_{\Omega} k(x, y)u(y) dy$. Es gilt dann:

$$(K^*w_\alpha)(x) = \int_{\Omega} k(y, x)w_\alpha(y) dy = u_\alpha(x).$$

Als a-priori Informationen bei Tikhonov haben wir $u \in L^2(\Omega)$ und dass $\|u\|_{L^2}$ möglichst klein sein soll (wegen äquivalenter Formulierung als Minimierungsproblem, siehe Satz 3.12). Damit wird u sehr glatt (abhängig von K bzw. K^*).

Wir wollen nun untersuchen, was bei Verwendung eines anderen Datenterms passiert, d.h. wir betrachten das Minimierungsproblem:

$$D(Ku, f) + \frac{\alpha}{2} \|u\|^2 \longrightarrow \min_u.$$

Sei zunächst $f_v(t) = D(K(u_\alpha + tv), f) + \frac{\alpha}{2} \|u_\alpha + tv\|^2$, wobei u_α den Minimierer des obigen Problems bezeichne. Dann ist $f_v(t)$ minimal bei $t = 0$, d.h. es gilt $f'_v(0) = 0$ für alle $v \in X$. Falls D differenzierbar ist, gilt:

$$f'_v(t) = \langle \partial_1 D(K(u_\alpha + tv), f), Kv \rangle + \alpha \langle u_\alpha + tv, v \rangle.$$

Bei $t = 0$ gilt dann für alle $v \in X$:

$$\begin{aligned} 0 = f'_v(0) &= \langle \partial_1 D(K(u_\alpha), f), Kv \rangle + \alpha \langle u_\alpha, v \rangle \\ &\Leftrightarrow 0 = \langle K^* \partial_1 D(Ku_\alpha, f) + \alpha u_\alpha, v \rangle \\ &\Leftrightarrow 0 = K^* \partial_1 D(Ku_\alpha, f) + \alpha u_\alpha \\ &\Leftrightarrow u_\alpha = K^* \left(-\frac{1}{\alpha} \partial_1 D(Ku_\alpha, f) \right) = K^* w_\alpha. \end{aligned}$$

Damit gilt $u_\alpha \in \mathcal{R}(K^*)$ unabhängig vom Datenterm.

4.9 Beispiel (Entrauschung): Wir wollen nun die a-priori Information, dass ∇u klein sein soll, betrachten. Sei dazu:

$$H_\diamond^2(\Omega) := \left\{ u \in L^2(\Omega) \mid \nabla u \in L^2(\Omega)^d, \int_\Omega u(x) dx = 0 \right\}, \quad \|u\|_{H_\diamond^2} := \sqrt{\int_\Omega |\nabla u(x)|^2 dx}.$$

Wir betrachten den Einbettungsoperator $K : H_\diamond^2(\Omega) \hookrightarrow L^2(\Omega)$; $u \mapsto u$. Wie beim Tikhonov-Verfahren setzen wir $J_\alpha(u) = \frac{1}{2} \|u - f\|_{L^2}^2 + \frac{\alpha}{2} \|u\|_{H_\diamond^2}^2$. Sei $f_v = J_\alpha(u_\alpha + tv)$ und u_α der Minimierer von J_α . Dann gilt:

$$\begin{aligned} 0 &= f'_v(0) \\ &= \langle u - f, v \rangle_{L^2} + \alpha \langle \nabla u, \nabla v \rangle_{L^2} \\ &= \int_\Omega (u - f)v + \alpha \nabla u \cdot \nabla v dx. \end{aligned}$$

Sei $v = 0$ auf $\partial\Omega$. Dann gilt weiter:

$$0 = \int_\Omega (u - f)v - \alpha \nabla \cdot \nabla u v dx,$$

also $u - f - \alpha \Delta u = 0 \Leftrightarrow \Delta u = \frac{1}{\alpha}(u - f) \in L^2(\Omega)$. Damit erhalten wir zuviel Glattheit im Vergleich zur a-priori Information, denn die zweite Ableitung liegt in $L^2(\Omega)$.

Wir halten fest: Klassische (lineare) Regularisierungsverfahren liefern $u_\alpha = K^* w_\alpha$, d.h. u_α ist abhängig von den Eigenschaften von K^* überglättet. Wir geben also eine alternative Konstruktion:

4.10 Sparsity-Regularisierung (Compressed Sensing): Wir haben eine Sammlung von typischen Lösungen (Dictionary). Die Regularisierung soll Lösungen, die sich aus wenigen Elementen dieser Sammlung zusammensetzen, bevorzugen, d.h. Linearkombinationen von wenigen

Elementen. Der Koeffizientenvektor soll also sparse sein. Zur Rekonstruktion:

Sei eine Sammlung w_1, \dots, w_N für groSSes $N \in \mathbb{N}$ gegeben. Dann besitzt eine Lösung die Darstellung $u = \sum_{i=1}^N z_i w_i$ für $z_i \in \mathbb{R}$, sodass wir eine neue Unbekannte $z \in \mathbb{R}^N$ erhalten. Es gilt dann $\sum_{i=1}^N z_i K w_i = Ku = f$. Setze $Az = \sum_{i=1}^N z_i K w_i$. Wir definieren einen Regularisierungsterm durch

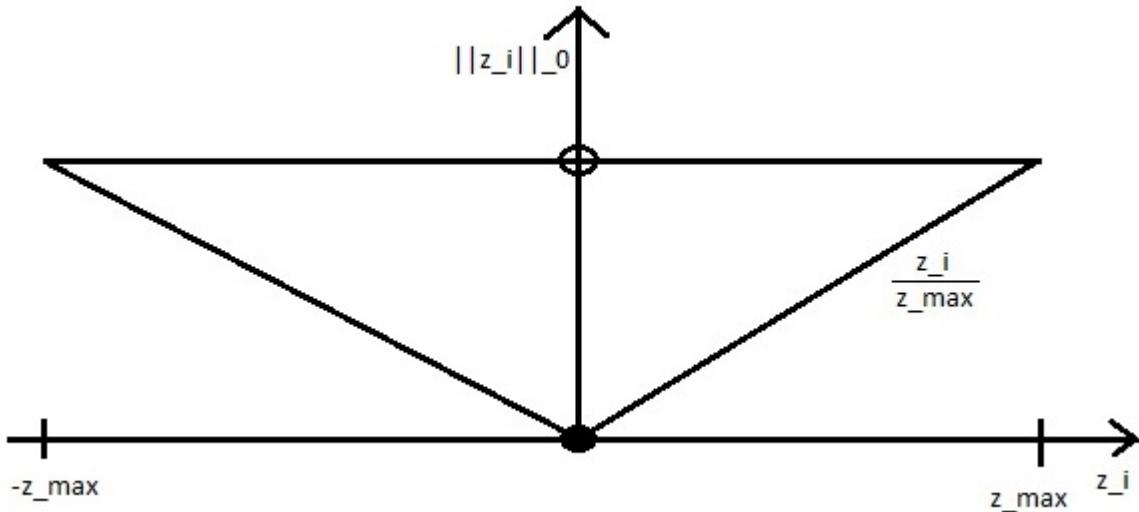
$$R^0(z) = \|z\|_0 = \#\{z_i \mid z_i \neq 0\}$$

Damit erhalten wir das regularisierte Problem:

$$D(Az, f) + \alpha R^0(z) \longrightarrow \min_{z \in \mathbb{R}^N} .$$

Es ergeben sich allerdings 2^N Möglichkeiten für Null- oder Nichtnulleinträge in z , womit ein Ausprobieren für groSSes $N \in \mathbb{N}$ unmöglich ist.

Um dieses Problem zu Lösen, betrachten wir als Ansatz konvexe Funktionen. Betrachte:



Dann ist $\frac{z_i}{z_{max}}$ die grösste konvexe Funktion, die unter der Nullnorm von z_i liegt.

Idee: Wir ersetzen $R^0(z) = \|z\|_0$ durch

$$R^1(z) = \|z\|_1 = \sum_{i=1}^N |z_i|.$$

Erhält die l^1 -Norm jedoch die a-priori Information der Sparsity?

Sei dazu $A \in \mathbb{R}^{M \times N}$ und $D(Az, f) = \frac{1}{2}\|Az - f\|^2$. Dann lautet das regularisierte Problem:

$$J_\alpha(z) = \frac{1}{2}\|Az - f\|^2 + \alpha\|z\|_1 \longrightarrow \min_{z \in \mathbb{R}^N} .$$

Sei $\tilde{z} \in \mathbb{R}^N$ ein Minimierer von J_α . Für $\tilde{z}_i \neq 0$ gilt $\frac{\partial}{\partial \tilde{z}_i} J_\alpha(\tilde{z}) = 0$ und weiter

$$\frac{\partial}{\partial z_i} J_\alpha(\tilde{z}) = (A^T(A\tilde{z} - f))_i + \alpha \operatorname{sgn}(\tilde{z}_i) = 0.$$

Damit folgt $A^T(-\frac{1}{\alpha}(A\tilde{z} - f)) = A^T w_\alpha = \operatorname{sgn}(\tilde{z}_i)$. Für ein kanonisches w_α kommen bei $A^T w_\alpha$ nur sehr wenige Einträge $\neq \pm 1$ heraus, also nur sehr wenige Einträge $\tilde{z}_i \neq 0$.

Alternative Erklärung (Beispiel): Sei $N = M = 1, A = 1$ und

$$J_\alpha(z) = \frac{1}{2}|z - f|^2 + \alpha|z| \longrightarrow \min_{z \in \mathbb{R}} .$$

Dann gilt $J_\alpha(0) = \frac{1}{2}|f|^2$. Ist dann der Minimierer $\tilde{z} \neq 0$, so gilt $0 = J_\alpha(\tilde{z}) = (\tilde{z} - f) + \alpha \operatorname{sgn}(\tilde{z})$, also $\tilde{z} \in \{f - \alpha, f + \alpha\}$. Dann gilt:

$$J_\alpha(f - \alpha) = \frac{1}{2}\alpha^2 + \alpha|f - \alpha|, \quad J_\alpha(f + \alpha) = \frac{1}{2}\alpha^2 + \alpha|f + \alpha|.$$

Es gilt $\tilde{z} = 0$, falls:

$$(1) \quad \frac{1}{2}f^2 \leq \frac{1}{2}\alpha^2 + \alpha|f - \alpha| \Leftrightarrow \frac{1}{2}(f - \alpha)(f - \alpha) \leq \alpha|f - \alpha| \Leftrightarrow \frac{f-\alpha}{|f-\alpha|} \frac{f+\alpha}{\alpha} \leq 2 \text{ (für } f \leq \alpha\text{);}$$

$$(2) \quad \frac{1}{2}f^2 \leq \frac{1}{2}\alpha^2 + \alpha|f + \alpha| \Leftrightarrow \frac{f+\alpha}{|f+\alpha|} \frac{f-\alpha}{\alpha} \leq 2 \text{ (für } f \geq \alpha\text{).}$$

Insgesamt erhalten wir den sogenannten Shrinkage-Operator:

$$\tilde{z} = S_\alpha(f) = \begin{cases} 0, & |f| \leq \alpha, \\ f - \alpha, & f > \alpha, \\ f + \alpha, & f < \alpha. \end{cases}$$

4.10 Beispiel (1D-Entrauschung): Sei $\Omega = (0, 1)$. Dann betrachten wir:

$$D(Ku, f) = \frac{1}{2} \int_{\Omega} (u(x) - f(x))^2 dx, \quad R(u) = \int_{\Omega} S(u'(x)) dx,$$

wobei S konvex, nichtnegativ und differenzierbar ist, z.B. $S(u') = |u'|^{p \frac{1}{p}}$. Wir erhalten dann das Funktional

$$J_\alpha(u) = \frac{1}{2} \int_0^1 (u(x) - f(x))^2 dx + \alpha \int_0^1 S(u'(x)) dx.$$

Für dessen Minimierer u_α gilt $\frac{d}{dt}|_{t=0} J_\alpha(u_\alpha + tv) = 0$ und

$$\frac{d}{dt} J_\alpha(u_\alpha + tv) = \int_0^1 (u_\alpha + tv - f)v dx + \alpha \int_0^1 S'(u' + tv') v' dx \stackrel{t=0, v(0)=v(1)=0}{=} \int_0^1 v(u_\alpha - f - \frac{d}{dt} \alpha S' \left(\frac{du_\alpha}{dx} \right)) dx.$$

Damit folgt $\frac{d}{dx} S' \left(\frac{du_\alpha}{dx} \right) = \frac{1}{\alpha}(u_\alpha - f) = w_\alpha \in L^2(\Omega)$ und weiter $S' \left(\frac{du_\alpha}{dx} \right) = c + \int_0^x w_\alpha(y) dy \in H^1$. Sei weiter S' invertierbar (z.B. $S'' > 0$). Dann folgt $\frac{du_\alpha}{dx} = (S')^{-1}(c + \int_0^1 w_\alpha(y) dy)$. Damit ist $\frac{du_\alpha}{dx}$ glatt und besitzt im Normalfall eine schwache zweite Ableitung.

Für $S(u') = |u'|^{p \frac{1}{p}}$ ist $S'(u') = |u'|^{p-1} \operatorname{sgn}(u')$ und $w_\alpha = c + \int_0^x w_\alpha(y) dy$. Weiter gelten dann:

$$\frac{du_\alpha}{dx} = (S')^{-1}(w_\alpha) = |w_\alpha|^{\frac{1}{p-1}} \operatorname{sgn}(w_\alpha), \quad \frac{du_\alpha}{dx^2} = \frac{1}{p-1} |w_\alpha|^{\frac{1}{p-2}} w_\alpha.$$

Für $p > 1$ findet eine Überglättung von $\frac{du_\alpha}{dx}$ statt, besser $p < 2$ als $p > 2$. Für $p = 1$ gilt formal:

$$S' \left(\frac{du_\alpha}{dx} \right) = \underbrace{\operatorname{sgn} \left(\frac{du_\alpha}{dx} \right)}_{\in \{\pm 1\}} = w_\alpha,$$

wobei $\operatorname{sgn} \left(\frac{du_\alpha}{dx} \right) \in \{\pm 1\}$ abhängig ist von der GröSSe der Ableitung.

Fazit: 1-Normen bzw. nicht-differenzierbare/nicht strikt-konvexe Regularisierungen können eine Überglättung vermeiden und effektiver a-priori Informationen einbringen.

5 Variationsmethoden zur Regularisierung

Wir haben bereits gesehen, dass ein Problem der Form $Ku = f$ meistens nicht stetig invertierbar ist. Deshalb muss das Problem durch eine stetig invertierbare Approximation ersetzt werden. Ein Hauptkonzept dazu ist die Konstruktion von Funktionalen, die idealerweise einen eindeutigen stationären Punkt besitzen, der der Lösung des ursprünglichen inversen Problems in einem gewissen FehlermaSS nahe ist. Diese Funktionale bestehen normalerweise aus zwei Teilen: einem Datenterm, der die Abweichung zwischen den gemessenen Daten f und dem Modell angewandt auf Input-Daten u misst, und einem Regularisierungsterm, der die Input-Daten u zwingt gewissen Anforderungen zu genügen.

Konkret untersuchen wir für einen Regularisierungsparameter $\alpha > 0$ das Minimierungsproblem

$$J_\alpha(u) = D(Ku, f) + \alpha R(u) \longrightarrow \min_{u \in X}$$

mit einem beschränkten linearen Operator $K : X \rightarrow Y$ zwischen separablen Banachräumen X, Y , einem konvexen Datenterm $D : Y \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ und einem konvexen Regularisierungsterm $R : X \rightarrow \mathbb{R} \cup \{+\infty\}$. Konvex heiSt in diesem Fall, dass für alle $u, v \in X, \lambda \in [0, 1]$ gilt:

$$R(\lambda u + (1 - \lambda)v) \leq \lambda R(u) + (1 - \lambda)R(v).$$

5.1 Lemma: Seien $F_1, F_2 : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex und $c_1, c_2 > 0$. Dann ist auch $c_1 F_1 + c_2 F_2$ konvex.

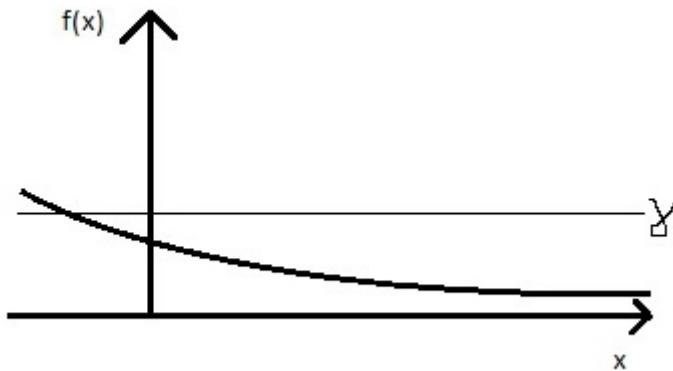
5.2 Lemma: Sei $F : Y \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex und $K : X \rightarrow Y$ linear. Dann ist $F \circ K)(u) = F(Ku)$ konvex.

Damit ist insbesondere J_α konvex.

5.1 Existenz von Minima

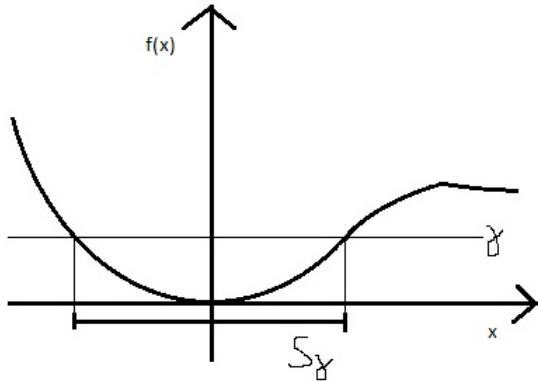
Anhand von Beispielen wollen wir wichtige Bedingungen für die Existenz von Minima herleiten. Zunächst benötigen wir die Beschränktheit von unten, d.h. ist $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$, so soll ein $c > 0$ mit $J(u) \geq c$ für alle $u \in X$ existieren. Betrachte nun folgende Beispiele:

- (i) Betrachte $f(x) = e^{-x}$:



Dann gilt $\inf f(x) = 0$, aber 0 wird nicht als Minimum angenommen. Wir definieren die sogenannten Sublevel-Sets als $S_\gamma := \{u \in X \mid J(u) \leq \gamma\}$. Im Beispiel haben wir dann $S_\gamma = [-\log(\gamma), +\infty)$ für $\gamma > 0$, d.h. S_γ ist unbeschränkt.

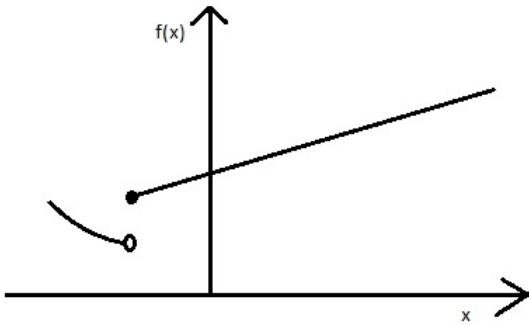
Betrachte nun:



In diesem Beispiel ist S_γ kompakt (beschränkt).

Der Unterschied zwischen den Beispielen ist nun, dass S_γ im ersten Fall unbeschränkt und im zweiten Fall beschränkt (präkompakt) für γ klein genug und S_γ nicht leer.

(2) Betrachte:



Dann wird das Infimum nicht angenommen. Wir benötigen die Unterhalbstetigkeit, d.h. für alle Folgen $u_k \rightarrow u$ soll $J(u) \leq \liminf_k J(u_k)$ gelten.

Folgender Satz besagt, dass die oben hergeleiteten Forderungen genügen:

5.3 Satz: Sei $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ für einen topologischen Raum (X, τ) mit metrisierbarer Topologie (charakterisiert durch Folgenkonvergenz). Weiter erfülle J folgende Bedingungen:

- (1) Koerzivität: Es existiert ein $\gamma > 0$, sodass $S_\gamma := \{u \in X \mid J(u) \leq \gamma\}$ nicht leer und präkompakt ist.
- (2) (Folgen-)Unterhalbstetigkeit: Ist $u_k \xrightarrow{\tau} u$ eine konvergente Folge in X , so gilt:

$$J(u) \leq \liminf_k J(u_k).$$

- (3) Beschränktheit von unten: Es existiert ein $c > 0$ mit $J(u) > c$ für alle $u \in X$.

Dann existiert ein Minimierer $\tilde{u} \in X$, d.h. es gilt

$$J(\tilde{u}) \leq J(u) \quad \forall u \in X.$$

Wir müssen also eine Topologie finden, sodass die Bedingungen aus diesem Satz erfüllt sind. Dazu:

5.4 Definitionen (Topologien in Banachräumen): Sei X ein Banachraum und $(u_k)_{k \in \mathbb{N}}$ eine Folge in X . Dann sagen wir:

(1) $(u_k)_{k \in \mathbb{N}}$ ist stark konvergent gegen $u \in X$, wenn $\|u_k - u\|_X \rightarrow 0$ für $n \rightarrow \infty$ gilt. Wir schreiben dann $u_k \rightarrow u$.

(2) $(u_k)_{k \in \mathbb{N}}$ heiSt schwach konvergent gegen $u \in X$, falls

$$\langle v, u_k \rangle \rightarrow \langle v, u \rangle \quad \forall v \in X^*$$

gilt. Wir schreiben dann $u_k \rightharpoonup u$.

(3) Für $X = Z^*$: $(u_k)_{k \in \mathbb{N}}$ heiSt schwach-*konvergent gegen $u \in X$, falls

$$\langle u_k, v \rangle \rightarrow \langle u, v \rangle \quad \forall v \in Z$$

gilt, d.h. u_k konvergiert punktweise gegen u . Wir schreiben dann $u_k \rightharpoonup^* u$.

Es gilt dann $(1) \Rightarrow (2) \Rightarrow (3)$, aber im Allgemeinen nicht die Umkehrung.

5.5 Satz (Banach-Alaoglu): Sei Z ein Banachraum, $X = Z^*$ und $(u_k)_{k \in \mathbb{N}}$ eine beschränkte Folge in X , d.h. $\|u_k\|_X$ ist eine beschränkte Folge in \mathbb{R}^+ . Dann besitzt $(u_k)_{k \in \mathbb{N}}$ eine schwach-*konvergente Teilfolge.

5.6 Bemerkungen:

(1) Ist $X = (X^*)^*$ reflexiv, so gilt $Z = X^*$, d.h. die schwache und die schwach-* Topologie stimmen überein.

(2) Ist X ein Hilbertraum, so gilt $X = X^*$, also $(X^*)^* = X$. Weiter gilt dann:

$$u_k \rightharpoonup u, \|u_k\| \rightarrow \|u\| \Rightarrow u_k \rightarrow u,$$

denn

$$\begin{aligned} \|u_k - u\|^2 &= \underbrace{\|u_k\|^2}_{\rightarrow \|u\|^2} + \|u\|^2 - 2 \underbrace{\langle u_k, u \rangle}_{\rightarrow \langle u, u \rangle = \|u\|^2} \longrightarrow 2\|u\|^2 - 2\|u\|^2 = 0. \end{aligned}$$

Eine typische Koerzivitätsbedingung ist

$$[\|u_k\|_X \rightarrow \infty \Rightarrow J(u_k) \rightarrow \infty] \Leftrightarrow [J(u) < \infty \Rightarrow \|u\|_X < \infty].$$

Für $J_\alpha(u) = D(Ku, f) + \alpha R(u)$ sollte gelten:

$$R(u_k) \rightarrow \infty \text{ für } \|u_k\|_X \rightarrow \infty.$$

Unter der Annahme $D \geq 0$ gilt dann

$$J_\alpha(u) \leq \gamma \Rightarrow R(u) \leq \frac{\gamma}{\alpha}.$$

Beispielsweise könnte $R(u) = \|u\|_X^p$ für $p \geq 1$ sein.

Mit geeigneter Wahl von R kann also erreicht werden, dass J_α koerziv ist.

Wir wollen uns nun der Unterhalbstetigkeit in der schwach-*Topologie zuwenden. Aus dem Beweis des Existenzresultats und dem Satz von Banach-Alaoglu folgt, dass es genügt, die Unterhalbstetigkeit auf beschränkten Teilmengen von X zu zeigen. Wir betrachten Datenterm und Regularisierungsterm getrennt:

Datenterm: Ist $K : X \rightarrow Y$ kompakt und $u_n \rightharpoonup u$ schwach-*konvergent in X , so folgt $Ku_n \rightarrow Ku$ in Y . Dann folgt aus der Unterhalbstetigkeit $D(Ku, f) \leq \liminf_n D(Ku_n, f)$ die starke Unterhalbstetigkeit, d.h.

$$D(v, f) \leq \liminf_n D(v_n, f) \text{ für } v_n \rightarrow v.$$

Wir betrachten dazu noch:

(1) Beispiel (für Abschätzung): Sei $D(v, f) = \frac{1}{2}\|v - f\|^2$. Dann gilt:

$$|D(v, f) - D(v_n, f)| = \frac{1}{2}|\langle v - v_n, v - f + v_n - f \rangle| \stackrel{\text{C.S.}}{\leq} \frac{1}{2} \underbrace{\|v - v_n\|}_{\rightarrow 0} \underbrace{\|v + v_n - 2f\|}_{\|2v - 2f\|} \rightarrow 0.$$

(2) Beispiel (Kullback-Leibler): Sei $D(v, f) = \int f \log\left(\frac{f}{v}\right) + v - f$. Dann ist D stetig, wenn v von Null wegbeschränkt ist. Aber sonst?

Regularisierungsterm (und offene Datenterme): Eine kanonische Annahme ist die Konvexität:

5.7 Lemma: Sei $J : X \rightarrow \mathbb{R}$ konvex und beschränkt, d.h. aus $\|u\|_X \leq C_1$ folgt $J(u) \leq C_2$. Dann ist J schwach unterhalbstetig.

5.8 Lemma: Sei $X = Z^*$ und $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$, sodass $J(u) = \sup_{z \in Z} (\langle u, z \rangle - H(z))$ für ein Funktional $H : Z \rightarrow \mathbb{R} \cup \{+\infty\}$ gilt. Dann ist J schwach-*-unterhalbstetig.

5.9 Beispiel: Sei $J(u) = \frac{1}{2}\|u\|_X^2$ für einen Hilbertraum X ($\Rightarrow Z = X$), sodass wie im Lemma $J(u) = \sup_z (\langle u, z \rangle - H(z))$ gilt. Falls $H(z) = \sup_v (\langle z, v \rangle - G(v))$, so folgt:

$$\begin{aligned} J(u) &= \sup_z (\langle u, z \rangle - \sup_v (\langle z, v \rangle - G(v))) \\ &= \sup_z \inf_v (\langle u, z \rangle - \langle v, z \rangle + G(v)) \\ &\stackrel{\text{Annahme}}{=} \inf_v \sup_z (\langle u - v, z \rangle + G(v)). \end{aligned}$$

Dann folgt

$$J(u) = \inf_v \begin{cases} G(u), & u = v, \\ +\infty, & v \neq u. \end{cases} = G(u).$$

SchlieSSlich erhalten wir:

$$\begin{aligned} H(z) &= \sup_v \langle z, v \rangle - J(v) \\ &= \sup_v \langle z, v \rangle - \frac{1}{2}\|v\|_X^2 \\ &= \sup_v \left(-\frac{1}{2}\|v - z\|^2 + \frac{1}{2}\|z\|_X^2 \right) \\ &= \frac{1}{2}\|z\|^2. \end{aligned}$$

5.10 Beispiel: Sei $J(u) = \|u\|_X$. Dann gilt:

$$J(u) = \|u\|_X = \sup_{z \neq 0} \frac{\langle u, z \rangle}{\|z\|} = \sup_{\|z\|=1} \langle u, z \rangle = \sup_z \langle u, z \rangle - H(z) \text{ mit } H(z) = \begin{cases} 0, & \|z\| \leq 1, \\ +\infty, & \|z\| > 1. \end{cases}$$

5.11 Folgerung: R sollte koerziv und konvex sein bzw. $R(u) = \sup_z \langle u, z \rangle - H(z)$. Dann erhalten wir die Existenz eines Minimierers von J_α für $\alpha > 0$.

5.12 Satz: Sei $R_X \rightarrow \mathbb{R} \cup \{+\infty\}$ wie oben und $X = Z^*$, $K : X \rightarrow Y$ kompakt und $D(\cdot, f)$ nicht-negativ und (stark) stetig auf Y . Dann existiert ein Minimum $u_\alpha \in X$ von J_α .

5.2 Eindeutigkeit von Minimierern

5.13 Definition: Ein Funktional $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ heiSt strikt konvex, wenn für alle $u, v \in X$ mit $u \neq v$ gilt:

$$J\left(\frac{1}{2}u + \frac{1}{2}v\right) < \frac{1}{2}J(u) + \frac{1}{2}J(v).$$

5.14 Satz: Sei $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ strikt konvex. Dann hat J höchstens einen Minimierer.

$J_\alpha(u) = D(Ku, f) + \alpha R(u)$ ist strikt konvex, falls eine der folgenden Bedingungen erfüllt ist:

- (1) D ist konvex, R ist strikt konvex.
- (2) R ist konvex, D ist strikt konvex und $\mathcal{N}(K) = \{0\}$.
- (3) Kombination: D ist strikt konvex, R soll sich gut genug an $\mathcal{N}(K)$ verhalten, z.B.:

$$\frac{1}{2}\|Ku - f\|_2^2 + \alpha\|u\|_1.$$

5.3 Stabilität von Minimierern

5.15 Definition: Sei f^{δ_n} eine Folge fehlerhafter Daten und $u_\alpha^{\delta_n}$ ein Minimierer von

$$J_\alpha^{\delta_n}(u) = D(Ku, f^{\delta_n}) + \alpha R(u).$$

Gilt dann $f^{\delta_n} \rightarrow f$ in Y , so heiSt das Minimierungsproblem stabil bezüglich des Datenfehlers, wenn $(u_\alpha^{\delta_n})_{n \in \mathbb{N}}$ eine schwach-*konvergente Teilfolge mit Grenzwert \tilde{u} besitzt, sodass \tilde{u} ein Minimierer von $J_\alpha(u) = D(Ku, f) + \alpha R(u)$ ist.

Um die Stabilität zu erreichen, benötigen wir Koerzivität (von R) und Unterhalbstetigkeit. Zusätzlich sollte D stetig in f sein. Sei dazu $u_\alpha^{\delta_n}$ ein Minimierer von $J_\alpha^{\delta_n}$. Dann gilt

$$\alpha R(u_\alpha^{\delta_n}) \stackrel{D \geq 0}{\leq} J_\alpha^{\delta_n}(u_\alpha^{\delta_n}) \leq J_\alpha^{\delta_n}(u)$$

für alle $u \in X$, also gilt insbesondere für einen Minimierer u_α von J_α :

$$\alpha R(u_\alpha^{\delta_n}) \stackrel{D \geq 0}{\leq} J_\alpha^{\delta_n}(u_\alpha^{\delta_n}) \leq J_\alpha^{\delta_n}(u_\alpha).$$

Damit folgt:

$$R(u_\alpha^{\delta_n}) \leq \frac{1}{\alpha} J_\alpha^{\delta_n}(u_\alpha) = \frac{1}{\alpha} D(Ku_\alpha, \underbrace{f^{\delta_n}}_{\rightarrow f}) + R(u_\alpha) \xrightarrow[D \text{ stetig}]{\alpha} \frac{1}{\alpha} D(Ku_\alpha, f) + R(u_\alpha).$$

Aus der Konvergenz von $\frac{1}{\alpha} J_\alpha^{\delta_n}$ folgt insbesondere die Beschränktheit von $R(u_\alpha^{\delta_n})$. Wegen der Koerzivität von R ist dann auch $\|u_\alpha^{\delta_n}\|_X$ beschränkt in \mathbb{R} , womit $(u_\alpha^{\delta_n})_{n \in \mathbb{N}}$ laut dem Satz von Banach-Alaoglu (für $X = Z^*$) eine konvergente Teilfolge bezüglich der schwach-*Topologie besitzt. Sei nun o.B.d.A. $(u_\alpha^{\delta_k})_{k \in \mathbb{N}}$ eine schwach-*konvergente Teilfolge, d.h. es existiert ein $\tilde{u} \in X$ mit $u_\alpha^{\delta_k} \rightharpoonup^* \tilde{u}$. Wir zeigen, dass \tilde{u} ein Minimierer von J_α ist. Zunächst gilt:

$$J_\alpha^{\delta_k}(u_\alpha) \geq J_\alpha^{\delta_k}(u_\alpha^{\delta_k}) = J_\alpha(u_\alpha^{\delta_k}) - D(Ku_\alpha^{\delta_k}, f) + D(Ku_\alpha^{\delta_k}, f^{\delta_k}).$$

Dann gilt auch

$$\liminf_k (J_\alpha(u_\alpha^{\delta_k}) - D(Ku_\alpha^{\delta_k}, f) + D(Ku_\alpha^{\delta_k}, f^{\delta_k})) \leq \liminf_k J_\alpha^{\delta_k}(u_\alpha) = \lim_k J_\alpha^{\delta_k}(u_\alpha),$$

und weiter

$$\lim_k (D(Ku_\alpha^{\delta_k}, f^{\delta_k}) - D(Ku_\alpha^{\delta_k}, f)) + \underbrace{\liminf_k J_\alpha(u_\alpha^{\delta_k})}_{\geq J_\alpha(\tilde{u}) \text{ (unterhalbstetig)}} \leq J_\alpha(u_\alpha).$$

Wenn $\lim_k (D(Ku_\alpha^{\delta_k}, f^{\delta_k}) - D(Ku_\alpha^{\delta_k}, f)) = 0$, so folgt $J_\alpha(\tilde{u}) \leq J_\alpha(u_\alpha)$, also ist \tilde{u} ein Minimierer von J_α .

Es gilt $D(Ku_n, f_n) - D(Ku, f) \rightarrow 0$, wenn $f_n \xrightarrow{\text{stark}} f$ in Y und $u_n \rightharpoonup^* u$ in X gilt. Verschiedene einfachere (d.h. stärkere) Forderungen können sein:

- (1) $D(v_n, f_n) - D(v_n, f) \rightarrow 0$ für $(v_n)_{n \in \mathbb{N}}$ beschränkt.
- (2) K kompakt: $D(v_n, f_n) - D(v, f_n) \rightarrow 0$ für $(v_n)_{n \in \mathbb{N}}$ stark konvergent.
- (3) $D : Y \times Y \rightarrow \mathbb{R}$ stetig, K kompakt, $v_n \rightarrow v$, $f_n \rightarrow f$:

$$D(v_n, f_n) - D(v_n, f) = \underbrace{D(v_n, f_n) - D(v, f)}_{\rightarrow 0} - \underbrace{(D(v_n, f) - D(v, f))}_{\rightarrow 0} \rightarrow 0.$$

Beispielsweise ist $D(v, f) = \|v - f\|_Y^p$ für $p \geq 1$ stetig.

5.16 Konvergenz der Regularisierung: Sei $u^* \in X$, sodass $f = Ku^*$ und sei $u_\alpha \in X$ ein Minimierer von $J_\alpha(u) = D(Ku, f) + \alpha R(u)$. Dann gilt $J_\alpha(u_\alpha) \leq J_\alpha(u)$ für alle $u \in X$, also insbesondere $J_\alpha(u_\alpha) \leq J_\alpha(u^*)$. Unter der Voraussetzung $D(v, f) \geq 0$, $D(f, f) = 0$ gilt dann

$$\underbrace{D(Ku_\alpha, f) + \alpha R(u_\alpha)}_{\geq 0} \leq \underbrace{D(Ku^*, f) + \alpha R(u^*)}_{=f=0},$$

also $R(u_\alpha) \leq R(u^*)$. Unter der Voraussetzung $R(u^*) < \infty$ gilt dann auch $R(u_\alpha) \leq R(u^*) < \infty$ für alle $\alpha > 0$. Wenn wir Koerzivität voraussetzen, ist $(u_\alpha)_{\alpha > 0}$ beschränkt, besitzt also eine schwach-*konvergente Teilfolge. Sei diese mit $(u_{\alpha_n})_{n \in \mathbb{N}}$ bezeichnet. Dann gibt es ein $\tilde{u} \in X$ mit $u_{\alpha_n} \rightharpoonup^* \tilde{u}$. Damit haben wir:

$$0 \leq \underbrace{\alpha_n}_{\rightarrow 0} \underbrace{R(u_{\alpha_n})}_{\text{beschr.}} + D(Ku_{\alpha_n}, f) = J_{\alpha_n}(u_{\alpha_n}) \leq J_{\alpha_n}(u^*) = \alpha_n R(u^*) \xrightarrow{\alpha_n \rightarrow 0} 0.$$

Damit folgt $D(Ku_{\alpha_n}, f) \rightarrow 0$. Setzen wir D als unterhalbstetig voraus, so folgt $D(K\tilde{u}, f) = 0$. Falls $v = f$ das eindeutige Minimum von $D(v, f)$ ist, so gilt $K\tilde{u} = f$.

Weiter wissen wir, dass $R(u_\alpha) \leq R(u^*)$. Wegen Unterhalbstetigkeit folgt dann:

$$R(\tilde{u}) \leq \liminf_n R(u_{\alpha_n}) \leq R(u^*),$$

also folgt $R(\tilde{u}) \leq R(u^*)$ für jede Lösung $u^* \in X$ von $Ku = f$.

Diese Überlegungen führen zu folgender Definition:

5.17 Definition: Ein $\tilde{u} \in X$ heiSt R -minimierende Lösung von $Ku = f$, wenn $K\tilde{u} = f$ und $R(\tilde{u}) \leq R(u^*)$ für jedes $u^* \in X$ mit $Ku^* = f$ gilt.

5.18 Konvergenz bei gestörten Daten: Sei u_α^δ ein Minimierer von

$$J_\alpha^\delta(u) = D(Ku, f^\delta) + \alpha R(u).$$

Dann gilt $J_\alpha^\delta(u_\alpha^\delta) \leq J_\alpha^\delta(u)$ für alle $u \in X$. Insbesondere gilt dann:

$$\alpha R(u_\alpha^\delta) \leq J_\alpha^\delta(u_\alpha^\delta) \leq J_\alpha^\delta(u^*) = D(\underbrace{Ku^*}_{=f}, f^\delta) + \alpha R(u^*) \Rightarrow R(u_\alpha^\delta) \leq R(u^*) + \frac{D(f, f^\delta)}{\alpha}.$$

Als zusätzliche Bedingung erhalten wir $\frac{D(f, f^\delta)}{\alpha} \rightarrow 0$ für $\delta \rightarrow 0$. Dann hat (u_α^δ) eine schwach-*konvergente Teilfolge und der Grenzwert jeder schwach-*konvergenten Teilfolge ist eine R -minimierende Lösung von $Ku = f$.

5.4 Ableitung von Funktionalen und Operatoren in Banachräumen

Seien X, Y Banachräume, $\phi : X \rightarrow \mathbb{R}$ ein Funktional und $F : X \rightarrow Y$ ein Operator.

5.19 Definition (Richtungsableitung von ϕ): Fixiere eine Richtung $v \in X$ und definiere:

$$\phi_v^u : \mathbb{R}^+ \rightarrow \mathbb{R}; \quad \phi_v^u(t) := \phi(u + tv).$$

Betrachte dann die Ableitung an der Stelle $t = 0$:

$$D_v\phi(u) = \frac{d}{dt} \phi_v^u(0_+) = \lim_{t \downarrow 0} \frac{\phi_v^u(t) - \phi_v^u(0)}{t} = \lim_{t \downarrow 0} \frac{\phi(u + tv) - \phi(u)}{t}.$$

Dann ist die die Richtungsableitung $D_v\phi(u)$ definiert über der Grenzwert (falls existent).

5.20 Definition (Richtungsableitung von F): Betrachte $u \mapsto \langle z, F(u) \rangle$ für ein $z \in Y^*$. Dann ist $\langle z, F(\cdot) \rangle : X \rightarrow \mathbb{R}$ ein Funktional und $D_v\langle z, F(u) \rangle$ definiert wie in Definition 5.19. Wenn nun ein $G_v : X \rightarrow Y$ existiert mit

$$D_v\langle z, F(u) \rangle = \langle z, G(u) \rangle \quad \forall z \in Y^*,$$

so heiSt F richtungsdifferenzierbar in Richtung $v \in X$ und wir setzen $D_v F(u) := G(u)$.

5.21 Beispiel:

5.22 Definition: Seien ϕ, F wie oben. Dann heiSt ϕ bzw. F Gateaux-differenzierbar, wenn $D_v\phi$ bzw. $D_v F$ für alle Richtungen $v \in X$ existiert.

5.23 Definition: Seien ϕ, F wie oben. Dann heiSt ϕ bzw. F Frechet-differenzierbar an der Stelle $u \in X$, wenn ein beschränktes lineares Funktional $\nabla\phi(u) \in X^*$ bzw. ein beschränkter linearer Operator $F'(u) : X \rightarrow Y$ existiert, sodass

$$D_v\phi(u) = \langle \nabla\phi(u), v \rangle, \quad \frac{|\phi(u + w) - \phi(u) - \langle \nabla\phi(u), w \rangle|}{\|w\|_X} \rightarrow 0 \quad \forall \|w\|_X \rightarrow 0,$$

bzw.

$$D_v F(u) = F'(u)v, \quad \frac{\|F(u + w) - F(u) - F'(u)w\|_Y}{\|w\|_X} \rightarrow 0 \quad \forall \|w\|_X \rightarrow 0.$$

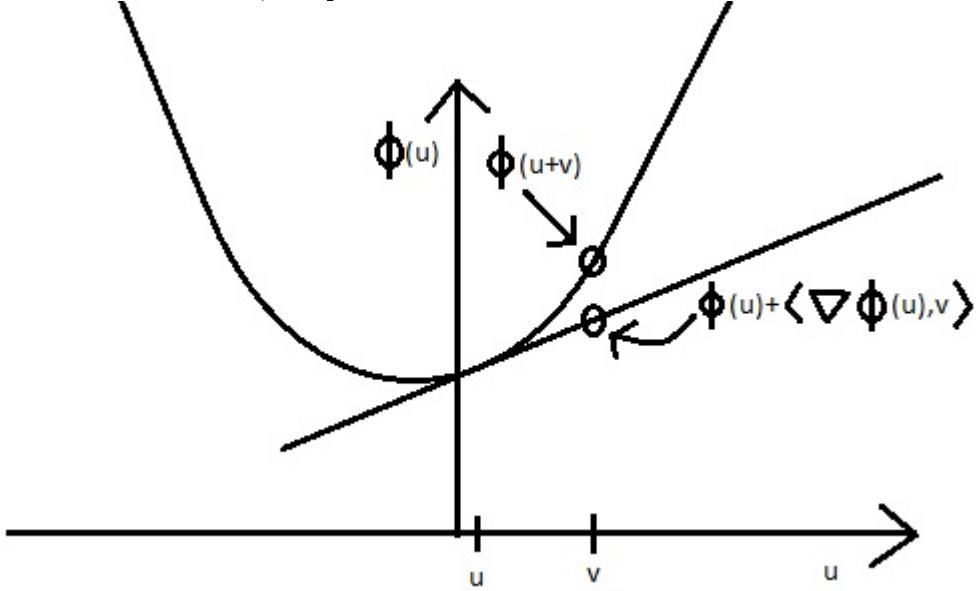
5.24 Beispiel:

5.25 Beispiel:

5.26 Beispiel:

5.5 Konvexe Analysis

Sei ϕ ein konvexe Funktion, beispielsweise:



Dann liegt ϕ über der Tangente, d.h.

$$\phi(u + v) \geq \phi(u) + \langle \nabla \phi(u), v \rangle \quad \forall v \in X.$$

Sei umgekehrt ϕ Frechet-differenzierbar und es gebe ein $p \in X^*$ mit

$$\phi(u + v) \geq \phi(u) + \langle p, v \rangle \quad \forall v \in X.$$

Dann gilt $p = \nabla \phi(u)$, denn es gilt

$$\langle p, v \rangle \leq \frac{\phi(u + tv) - \phi(u)}{t} \quad \forall t \geq 0 \Rightarrow \langle p, v \rangle \leq \langle \nabla \phi(u), v \rangle \quad \forall v \in X$$

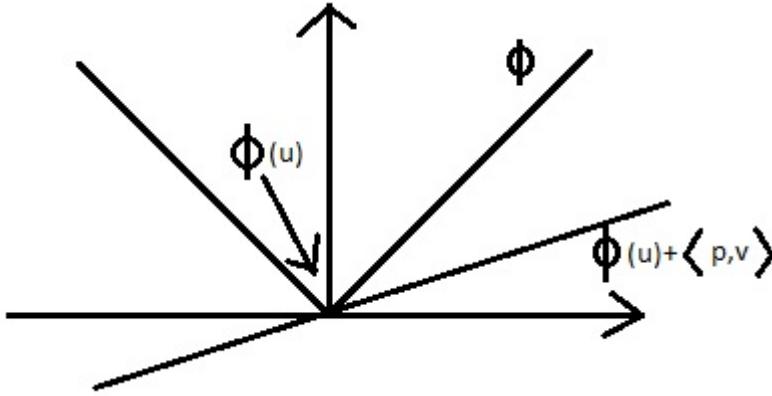
und analog $\langle p, -v \rangle \leq \langle \nabla \phi(u), -v \rangle$, also $\langle p, v \rangle = \langle \nabla \phi(u), v \rangle$ für alle $v \in X$. Damit folgt dann $p = \nabla \phi(u)$.

5.27 Definition: Sei $\phi : X \rightarrow \mathbb{R} \cup \{\infty\}$ konvex. Dann ist das Subdifferential von ϕ an der Stelle u definiert als

$$\partial\phi(u) := \{p \in X^* \mid \phi(u) + \langle p, v \rangle \leq \phi(u + v) \quad \forall v \in X\}.$$

Ein Element $p \in \partial\phi(u)$ heiSt Subgradient. Ist ϕ Frechet-differenzierbar, so besteht das Subdifferential nur aus der Frechet-Ableitung. Ist z.B. B ein stetiger linearer Operator auf einem Hilbertraum, so ist $\phi(u) = \frac{1}{2}\|Bu\|^2$ Frechet-differenzierbar, sodass $\partial\phi(v) = \{B^*Bv\}$.

Betrachte beispielsweise die Betragsfunktion:



Es gilt dann offensichtlich $\partial\phi(0) = [-1, 1]$.

5.28 Satz: Sei $\phi : X \rightarrow \mathbb{R}$ konvex und Frechet-differenzierbar. Dann gilt $\partial\phi(u) = \{\nabla\phi(u)\}$ für alle $u \in X$.

5.29 Satz: Sei $\infty \neq \phi : X \rightarrow \mathbb{R}$ konvex. Dann gilt

$$\tilde{u} \in \arg \min_{u \in X} \phi(u) \Leftrightarrow 0 \in \partial\phi(\tilde{u}).$$

Beweis: Es gilt:

$$\begin{aligned} 0 \in \partial\phi(\tilde{u}) &\Leftrightarrow \phi(\tilde{u}) \leq \phi(\tilde{u} + v) \quad \forall v \in X \\ &\Leftrightarrow \phi(\tilde{u}) \leq \phi(u) \quad \forall u \in X \\ &\Leftrightarrow \tilde{u} \text{ ist Minimierer von } \phi \\ &\Leftrightarrow \tilde{u} \in \arg \min_{u \in X} \phi(u). \end{aligned}$$

□

Wir wenden uns nun der Dualität zu:

5.30 Definition: Sei $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex. Dann ist die Legendre-Transformation definiert als

$$J^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}; \quad J^*(p) = \sup_{u \in X} \langle p, u \rangle - J(u).$$

J^* ist auch wohldefiniert, wenn J nicht konvex ist.

5.31 Satz: Sei $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$. Dann ist $F^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ wie oben konvex. Ebenso ist $F^{**} := (F^*)^* : X^{**} \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex. Ist X reflexiv, so haben wir $F^{**} : X \rightarrow \mathbb{R} \cup \{+\infty\}$.

5.32 Lemma: Sei X reflexiv und $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$. Dann gilt $F^{**}(u) \leq F(u)$ für alle $u \in X$. Ist $G : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex mit $G(u) \leq F(u)$ für alle $u \in X$, so folgt auch $G^{**}(u) \leq F^{**}(u) \leq F(u)$ für alle $u \in X$.

5.33 Satz: Sei X reflexiv und $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex. Dann gilt $J^{**} = J$. Ist weiter $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$ beliebig, so ist F^{**} die maximale konvexe Funktion, die unterhalb von F liegt (die konvexe Einhüllende).

5.34 Beispiel: Sei $J(u) = \langle q, u \rangle$ für ein $q \in X^*$. Dann gilt:

$$J^*(p) = \sup_{u \in X} \langle p, u \rangle - \langle q, u \rangle = \begin{cases} 0, & p = q, \\ +\infty, & p \neq q. \end{cases}$$

5.35 Rechenregeln für konvex konjugierte und Subdifferentiale:

- (1) Sind $F, G : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex, so gilt $\partial(F + G) = \partial F + \partial G$.
- (2) Ist $c > 0$, so gilt $\partial(cF) = c\partial F$.
- (3) Kettenregel: Sei $F(u) = G(Ku)$ für $K : X \rightarrow Y$, $G : Y \rightarrow \mathbb{R} \cup \{+\infty\}$. Ist $p \in \partial F(u)$, so gilt:

$$\langle p, v - u \rangle + F(u) \leq F(v), \quad \langle p, v - u \rangle + G(Ku) \leq G(Kv).$$

Ist $q \in \partial G(\tilde{u})$, so gilt:

$$\langle q, \tilde{v} - \tilde{u} \rangle + G(\tilde{u}) \leq G(\tilde{v}).$$

Mit $\tilde{u} = Ku$ und $\tilde{v} = Kv$ gilt dann:

$$\langle q, K(v - u) \rangle + G(Ku) \leq G(Kv), \quad \langle K^*q, v - u \rangle + G(Ku) \leq G(Kv).$$

Wir erhalten also:

$$q \in \partial_{\tilde{u}}G(Ku) \Rightarrow K^*q \in \partial_u(G(Ku)).$$

Die Umkehrung gilt nur, wenn $q = K^*\tilde{q}$, d.h. K^* muss surjektiv sein.

- (4) Im Allgemeinen gilt $(F + G)^*(p) \neq (F^* + G^*)(p)$.
- (5) Ist $c > 0$, so gilt $(cF)^*(p) = cF^*(\frac{p}{c})$.
- (6) Sei $G(u) = F(u - w)$ (oft in Datentermen). Dann gilt:

$$\begin{aligned} G^*(p) &= \sup_{u \in X} \langle p, u \rangle - F(u - w) \\ &= \sup_{\tilde{u} \in X} \langle p, \tilde{u} + w \rangle - F(\tilde{u}) \\ &= \langle p, w \rangle + \sup_{\tilde{u}} \langle p, \tilde{u} \rangle - F(\tilde{u}) \\ &= F^*(p) + \langle p, w \rangle. \end{aligned}$$

Ist $G(u) = F(u) + \langle q, u \rangle$, so gilt:

$$G^*(p) = \sup_{u \in X} \langle p, u \rangle - G(u) = \sup_{u \in X} \langle p - q, u \rangle - F(u) = F^*(p - q).$$

5.36 Zusammenhang von Dualität und Subgradienten:

Sei $F^*(p) = \sup_{u \in X} (\langle p, u \rangle - F(u))$. Dann gilt:

$$\begin{aligned} \text{Der Supremum wird angenommen bei } u &\Leftrightarrow \langle p, u \rangle - F(u) \geq \langle p, v \rangle - F(v) \quad \forall v \in X \\ &\Leftrightarrow F(v) \geq F(u) + \langle p, v - u \rangle \quad \forall v \in X \\ &\Leftrightarrow p \in \partial F(u). \end{aligned}$$

Sei nun $p \in \partial F(u)$. Ist F strikt konvex und differenzierbar, so gilt $p = F'(u)$, also $u = (F')^{-1}(p)$. Man kann zeigen, dass dann $(F^*)'(p) = (F')^{-1}(p) = u$ gilt.

Allgemeiner gilt unter geeigneten Regularitätsannahmen:

$$p \in \partial F(u) \Leftrightarrow u \in \partial F^*(p).$$

Wir haben bereits gezeigt: Ist \tilde{u} ein Minimum, so gilt $0 \in \partial F(u)$. Wir wissen nun, dass damit auch $\tilde{u} \in \partial F^*(0)$ gilt.

5.37 Definition: Sei $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ konvex und $p \in \partial J(u)$. Dann ist die Bregman-Distanz zum Subgradienten p definiert als

$$D_J^p(v, u) := J(v) - J(u) - \langle p, v - u \rangle,$$

d.h. sie misst den Abstand zwischen J , ausgewertet an der Stelle v , und der Linearisierung von J bei u , ausgewertet an der Stelle v . ($\Rightarrow v$ gibt Auswertungspunkt vor, u Linearisierungspunkt.) Die Bregman-Distanz ist keine Metrik im üblichen Sinne. Es gilt aber zumindest $D_J^p(u, u) = 0$ und $D_J^p(v, u) \geq 0$ für alle u, v . Ist J strikt konvex, so gilt auch $D_J^p(v, u) > 0$ für $u \neq v$. Allerdings haben wir keine Dreiecksungleichung oder eine Symmetrieeigenschaft.

Sind $p \in \partial J(u), q \in \partial J(v)$, so ist die symmetrische Bregman-Distanz definiert als

$$D_J^{p,q}(u, v) = D_J^p(v, u) + D_J^q(u, v) = \langle p - q, u - v \rangle.$$

5.6 Anwendung auf Variationsmethoden zur Regularisierung

5.38 Optimalitätsbedingungen: Zunächst wollen wir die Ergebnisse aus den beiden vorherigen Abschnitten anwenden, um Optimalitätsbedingungen herzuleiten. Wir nehmen an, dass D Fréchet-differenzierbar ist. Dann gilt:

$$\partial J_\alpha(u) \stackrel{5.35}{=} \partial_u(D(Ku, f)) + \alpha \partial_u R(u) = \{K^* D'(Ku, f)\} + \alpha \partial R(u).$$

Daraus erhalten wir:

$$u_\alpha \text{ ist Minimierer von } J_\alpha \Leftrightarrow 0 \in \partial J_\alpha(u_\alpha) \Leftrightarrow K^* D'(Ku_\alpha, f) + \alpha p_\alpha = 0, \quad p_\alpha \in \partial R(u_\alpha),$$

wobei D' die Ableitung nach Ku bezeichnet.

5.39 Beispiel: Sei $D(Ku, f) = \frac{1}{2} \|Ku - f\|^2$. Dann lautet die Optimalitätsbedingung:

$$K^*(Ku_\alpha - f) + \alpha p_\alpha = 0, \quad \text{also } p_\alpha = K^*(-\frac{1}{\alpha}(Ku_\alpha - f)).$$

Damit erhalten wir die Quellbedingung $p^* = K^* w \in \partial J(u^*)$.

5.40 Dualität (Fenchel-Dualität): Unter geeigneten Regularitätsannahmen gilt:

$$\underbrace{\inf_{u \in X} D(Ku, f) + \alpha R(u)}_{\text{primales Problem}} = \underbrace{\inf_{u \in X} \sup_{w \in Y} (\alpha R(u) + \overline{\langle K^* w, u \rangle} - D^*(w, f))}_{\text{primal-duales Problem}} = \underbrace{\sup_{w \in Y} (-\alpha R^*(-\frac{K^* w}{\alpha}) - D^*(w, f))}_{\text{duales Problem}}.$$

Dann definieren wir den Primal-Dual-Gap als

$$PDG(u, w) := D(Ku, f) + \alpha R(u) + \alpha R^*(-\frac{K^* w}{\alpha}) + D^*(w, f).$$

Wegen $\inf_u(\dots) = \sup_w(\dots)$ gilt $PDG(u, w) \geq 0$ für alle $u \in X, w \in Y$. Weiter folgt:

$PDG(\tilde{u}, \tilde{w}) = 0 \Rightarrow \tilde{u}$ löst primales Problem, \tilde{w} löst duales Problem, (\tilde{u}, \tilde{w}) löst primal-duales Problem.

5.41 Beispiel: Sei $X = \mathbb{R}^n, Y = \mathbb{R}^m, K \in \mathbb{R}^{m \times n}$. Wir untersuchen das Problem

$$\frac{1}{2} \|Ku - f\|_2^2 + \alpha \|u\|_1 \longrightarrow \min_{u \in X}.$$

Die Optimalitätsbedingung lautet dann

$$K^T(Ku_\alpha - f) + \alpha p_\alpha = 0, \quad p_\alpha \in \partial \|u_\alpha\|_1.$$

Es gilt $p \in \partial\|u\|_1$, wenn

$$\|u\|_1 + \langle p, v - u \rangle \leq \|y\|_1 \quad \forall v \in \mathbb{R}^n.$$

Schreiben wir $u = (u^i)_{i=1,\dots,N}$, so ist dies äquivalent zu

$$\sum_{i=1}^N |u^i| + \sum_{i=1}^N p^i(v^i - u^i) \leq \sum_{i=1}^N |v^i| \Leftrightarrow \sum_{i=1}^N (|u^i| - |v^i| + p^i(v^i - u^i)) \leq 0 \quad \forall v \in \mathbb{R}^n.$$

Wählen wir $v = (u^1, \dots, u^{i-1}, v^i, u^{i+1}, \dots, u^N)$, so ist dies äquivalent zu

$$|u^i| - |v^i| + p^i(v^i - u^i) \leq 0 \quad \forall v \in \mathbb{R}^n, i = 1, \dots, N,$$

womit wir das eindimensionale Problem

$$|u^i| - p^i u^i \leq |v^i| - p^i v^i \quad \forall v^i \in \mathbb{R}$$

erhalten. Betrachte nun die Spezialfälle

- $v^i = 0 : |u^i| \leq p^i u^i \Leftrightarrow \begin{cases} p^i \geq 1, & u^i > 0, \\ p^i \leq -1, & u^i < 0, \\ p^i \text{ irgendwie,} & u^i = 0. \end{cases}$
- $v^i = 2u^i : |u^i| - p^i u^i \leq 2(|u^i| - p^i u^i) \Leftrightarrow |u^i| \geq p^i u^i \Leftrightarrow \begin{cases} p^i \leq 1, & u^i > 0, \\ p^i \geq -1, & u^i < 0, \\ p^i \text{ irgendwie,} & u^i = 0. \end{cases}$

Damit wissen wir $|u^i| - p^i u^i = 0$. (Tatsächlich gilt $J(u) = \langle p, u \rangle$ für $p \in \partial J(u)$, falls $J(tu) = |t|J(u)$, d.h. J ist ein 1-homogenes Funktional.) Damit vereinfacht sich das eindimensionale Problem zu $p^i v^i \leq |v^i|$ für alle $v^i \in \mathbb{R}$. Es gilt $p^i = 1$ für $u^i > 0$ und $p^i = -1$ für $u^i < 0$. Weiter ist $p^i v^i \leq |p^i||v^i| \leq |v^i|$ für $|p^i| \leq 1$ erfüllt. Also folgern wir:

$$p \in \partial\|u\|_1 \Leftrightarrow \begin{cases} p^i = 1, & u^i > 0, \\ p^i = -1, & u^i < 0, \\ p^i \in [-1, 1], & u^i = 0. \end{cases}$$

Die Optimalitäts- bzw. Quellbedingung lautet dann $p = K^T w$. Wegen $\|p\|_\infty \leq 1$ muss w klein genug sein, damit $\|K^T w\|_\infty \leq 1$ gilt. Für kanonisches w und typisches K mit $\max_i |(K^T w)_i| = 1$ gilt $|p_j| = |(K^T w)_j| < 1$ für die meisten j . Also folgt auch $u_j = 0$ für die meisten j .

Wir wollen nun das duale Problem betrachten. Dazu berechnen wir D^* und R^* :

$$D^*(w, f) = \sup_v \langle w, v \rangle - \frac{1}{2} \|v - f\|^2 = \langle w, f + w \rangle - \frac{1}{2} \|w\|_2^2 = \frac{1}{2} \|w\|^2 + \langle f, w \rangle = \frac{1}{2} \|w + f\|^2 - \frac{1}{2} \|f\|^2,$$

$$R^*(z) = \sup_u \underbrace{\langle z, u \rangle - \|u\|_1}_{=\sum_{i=1}^N z^i u^i - |u^i|} = \begin{cases} 0, & \|z\|_\infty \leq 1, \\ +\infty, & \text{sonst.} \end{cases}$$

Damit ergibt sich das duale Problem

$$\frac{1}{2} \|w + f\|^2 \longrightarrow \min_w \quad \text{unter} \quad \left\| -\frac{K^T w}{\alpha} \right\|_\infty \leq 1.$$

Mit $\tilde{w} = -w$ ergibt sich das äquivalente Problem

$$\frac{1}{2} \|\tilde{w} - f\|_2^2 \longrightarrow \min_{\tilde{w}} \quad \text{unter} \quad \|K^T \tilde{w}\|_\infty \leq \alpha \Leftrightarrow (K^T \tilde{w})^i \leq \alpha \quad \forall i = 1, \dots, N.$$

Damit ist das optimale \tilde{w} gegeben als Projektion von f auf den Simplex, der durch $K^T \tilde{w} \leq 1$ beschrieben wird.

5.7 Fehlerabschätzungen

Bisher haben wir Fehlerabschätzungen für $\|u_\alpha^\delta - u^*\|$ unter Quellbedingungen wie $u^* = K^*w$ über die Singulärwertzerlegung hergeleitet.

Als Zwischenschritt berechnen wir Fehlerabschätzungen für die (lineare) Tikhonov-Regularisierung ohne Singulärwertzerlegung, d.h.:

$$u_\alpha \text{ minimiert } J_\alpha(u) = \frac{1}{2} \|Ku - f\|_Y^2 + \frac{\alpha}{2} \|u\|_X^2 \Leftrightarrow K^*(Ku_\alpha - f) + \alpha u_\alpha = 0.$$

Sei $u^* = K^*w$ erfüllt. Dann gibt es ein f_α , sodass $K^*(Ku^* - f_\alpha) + \alpha u^* = 0$ gilt.

Wir untersuchen also ein Abschätzung zwischen u_α und \tilde{u}_α , wobei $u_\alpha, \tilde{u}_\alpha$ Lösungen der Tikhonov-Regularisierung mit Daten f, \tilde{f} sind, d.h. $u_\alpha = R_\alpha f, \tilde{u}_\alpha = R_\alpha \tilde{f}$ ($\Rightarrow \|u_\alpha - \tilde{u}_\alpha\| \leq \|R_\alpha\| \|f - \tilde{f}\|$). Wir haben dann:

$$\begin{aligned} K^* Ku_\alpha + \alpha u_\alpha &= K^* f \\ K^* K \tilde{u}_\alpha + \alpha \tilde{u}_\alpha &= K^* \tilde{f}. \end{aligned}$$

Subtrahieren ergibt $K^* K(u_\alpha - \tilde{u}_\alpha) + \alpha(u_\alpha - \tilde{u}_\alpha) = K^*(f - \tilde{f})$, also

$$\|K^* K(u_\alpha - \tilde{u}_\alpha) + \alpha(u_\alpha - \tilde{u}_\alpha)\|^2 = \|K^*(f - \tilde{f})\|^2.$$

Damit ergibt sich weiter:

$$\|K^* K(u_\alpha - \tilde{u}_\alpha)\|^2 + \alpha^2 \|u_\alpha - \tilde{u}_\alpha\|^2 + 2\alpha \underbrace{\langle K^* K(u_\alpha - \tilde{u}_\alpha), u_\alpha - \tilde{u}_\alpha \rangle}_{=\|K(u_\alpha - \tilde{u}_\alpha)\|^2} \leq \|K^*\|^2 \|f - \tilde{f}\|^2.$$

Alle drei Terme links sind nicht negativ, also ist jeder durch die rechte Seite beschränkt. Im Einzelnen erhalten wir:

$$(1) \quad \alpha^2 \|u_\alpha - \tilde{u}_\alpha\|^2 \leq \|K^*\|^2 \|f - \tilde{f}\|^2, \text{ also } \|u_\alpha - \tilde{u}_\alpha\| \leq \frac{\|K^*\|}{\alpha} \|f - \tilde{f}\|. \text{ Es gilt } C_\alpha = \frac{2}{\sqrt{\alpha}} \text{ und } \|f - f_\alpha\| = \mathcal{O}(\alpha), \text{ d.h. wir würden nur } \|u_\alpha - u^*\| = \mathcal{O}(1) \text{ haben.}$$

$$(2) \quad 2\alpha \|K(u_\alpha - \tilde{u}_\alpha)\|^2 \leq \|K^*\|^2 \|f - \tilde{f}\|^2, \text{ also } \|K(u_\alpha - \tilde{u}_\alpha)\| \leq \frac{c}{\sqrt{\alpha}} \|f - \tilde{f}\|, \text{ wobei vor } \|f - \tilde{f}\| \text{ eigentlich } \mathcal{O}(1) \text{ sein sollte.}$$

Insgesamt erhalten wir keine gute Abschätzung.

Wir wissen $u_\alpha = K^* w_\alpha, \tilde{u}_\alpha = K^* \tilde{w}_\alpha$. Für $w_\alpha = \frac{1}{\alpha}(f - Ku_\alpha)$ und $\tilde{w}_\alpha = \frac{1}{\alpha}(\tilde{f} - K\tilde{u}_\alpha)$ erhalten wir $Ku_\alpha + \alpha w_\alpha = f$ und $K\tilde{u}_\alpha + \alpha \tilde{w}_\alpha = \tilde{f}$. Durch Subtraktion erhalten wir

$$K(u_\alpha - \tilde{u}_\alpha) + \alpha(w_\alpha - \tilde{w}_\alpha) = f - \tilde{f},$$

und damit analog zu oben:

$$\|K(u_\alpha - \tilde{u}_\alpha)\|^2 + \alpha^2 \|w_\alpha - \tilde{w}_\alpha\|^2 + 2\alpha \langle K(u_\alpha - \tilde{u}_\alpha), w_\alpha - \tilde{w}_\alpha \rangle = \|f - \tilde{f}\|^2,$$

wobei $\langle K(u_\alpha - \tilde{u}_\alpha), w_\alpha - \tilde{w}_\alpha \rangle = \langle (u_\alpha - \tilde{u}_\alpha), K^* w_\alpha - K^* \tilde{w}_\alpha \rangle = \|u_\alpha - \tilde{u}_\alpha\|^2$. Wegen Nichtnegativität folgt dann:

$$\|K(u_\alpha - \tilde{u}_\alpha)\| \leq \|f - \tilde{f}\|, \quad \|w_\alpha - \tilde{w}_\alpha\| \leq \frac{1}{\alpha} \|f - \tilde{f}\|, \quad \|u_\alpha - \tilde{u}_\alpha\| \leq \frac{1}{\sqrt{2\alpha}} \|f - \tilde{f}\|.$$

Wir erhalten eine Abschätzung für $u_\alpha^\delta - u^*$ als Lösungen des Tikhonov-Verfahrens für f^δ bzw. $f_\alpha = f + \alpha w$:

$$\|u_\alpha^\delta - u^*\| \leq \frac{1}{\sqrt{2\alpha}} \|f^\delta - (f + \alpha w)\| \leq \frac{1}{\sqrt{2\alpha}} (\|f^\delta - f\| + \alpha \|w\|) \leq \frac{\delta}{\sqrt{2\alpha}} + \sqrt{\frac{\alpha}{2}} \|w\|.$$

Nun wollen wir den allgemeinen Fall $J_\alpha(u) = \frac{1}{2}\|Ku - f\|^2 + \alpha R(u)$ für R konvex untersuchen. Die Optimalitätsbedingung ist erneut $K^*(Ku_\alpha - f) + \alpha p_\alpha = 0$ für $p_\alpha \in \partial R(u_\alpha)$. Ist u^* die Lösung von $R(u) \rightarrow \min_{K(u)=f}$ (vergleiche auch Definition 5.17), so minimiert u^* auch

$$\sup_w R(u) + \langle f - Ku, w \rangle.$$

Falls ein Sattelpunkt (u^*, w^*) von $R(u) + \langle f - Ku, w \rangle$ existiert, d.h. es gilt

$$\sup_w R(u^*) + \langle f - Ku^*, w \rangle = R(u^*) + \langle f - Ku, w^* \rangle,$$

so ist u^* ein Minimierer von $R(u) + \langle f - Ku, w^* \rangle$. Ableiten nach u liefert dann $p^* - K^*w^* = 0$. Die Quellbedingung ist somit die Existenz von w^* und wir brauchen die Existenz von $p^* \in \partial R(u^*)$ mit $p^* = K^*w$. (Dies ist konsistent mit dem Tikhonov-Verfahren: Dort haben wir $R(u) = \frac{1}{2}\|u\|^2$ und $\partial R(u) = \{u\}$, also $u^* = K^*w^*$.) Wenn die Quellbedingung erfüllt ist, gilt $K^*(Ku^* - f) = 0$, also auch

$$K^*(Ku^* - f - \alpha w^*) + \alpha p^* = \underbrace{K^*(Ku^* - f)}_{=0} - \underbrace{\alpha K^*w^*}_{=0} + \alpha p^* = 0.$$

Damit erfüllt u^* die Optimalitätsbedingung für $f_\alpha = f + \alpha w^*$. Wir können also wieder den Fehler zwischen u_α und \tilde{u}_α (als Lösungen mit f bzw. \tilde{f}) abschätzen:

$$\begin{aligned} K^*Ku_\alpha + \alpha p_\alpha &= K^*f, \quad p_\alpha \in \partial R(u_\alpha), \\ K^*K\tilde{u}_\alpha + \alpha \tilde{p}_\alpha &= K^*\tilde{f}, \quad \tilde{p}_\alpha \in \partial R(\tilde{u}_\alpha). \end{aligned}$$

Gibt es w_α und \tilde{w}_α mit $p_\alpha = K^*w_\alpha$ und $\tilde{p}_\alpha = K^*\tilde{w}_\alpha$, so folgt

$$\begin{aligned} Ku_\alpha + \alpha w_\alpha &= f, \\ K\tilde{u}_\alpha + \alpha \tilde{w}_\alpha &= \tilde{f}. \end{aligned}$$

Damit ergibt sich wieder:

$$\|Ku_\alpha - K\tilde{u}_\alpha\|^2 + \alpha^2\|w_\alpha - \tilde{w}_\alpha\|^2 + 2\alpha \langle K(u_\alpha - \tilde{u}_\alpha), w_\alpha + \tilde{w}_\alpha \rangle = \|f - \tilde{f}\|^2.$$

Wegen $\langle K(u_\alpha - \tilde{u}_\alpha), w_\alpha + \tilde{w}_\alpha \rangle = \langle K^*(w_\alpha - \tilde{w}_\alpha), u_\alpha - \tilde{u}_\alpha \rangle = \langle p_\alpha - \tilde{p}_\alpha, u_\alpha - \tilde{u}_\alpha \rangle = D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) \geq 0$ und Nichtnegativität der beiden anderen Terme folgt dann:

$$\begin{aligned} \|K(u_\alpha - \tilde{u}_\alpha)\| &\leq \|f - \tilde{f}\|, \\ \|w_\alpha - \tilde{w}_\alpha\| &\leq \frac{1}{\alpha} \|f - \tilde{f}\|, \\ D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) &\leq \frac{1}{2\alpha} \|f - \tilde{f}\|^2. \end{aligned}$$

Als Abschätzung zwischen u_α^δ und u^* mit $f = f^\delta$ und $\tilde{f} = f_\alpha = f + \alpha w^*$ ergibt sich dann:

$$D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha^\delta, u^*) \leq \frac{1}{2\alpha} \|f^\delta - f_\alpha\|^2 = \frac{1}{2\alpha} \|f^\delta - f - \alpha w^*\|^2 \leq \frac{1}{2\alpha} (\delta + \alpha \|w^*\|)^2.$$

Für $\alpha \sim \delta$ ergibt sich daraus:

$$D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha^\delta, u^*) = \mathcal{O}(\delta).$$

Beachte: Die Abschätzung gilt für jedes $p^* \in \partial R(u^*)$, welches $p^* = K^*w^*$ erfüllt, d.h. wir erhalten eine mehrwertige Abschätzung und damit auch mehr Informationen durch geeignete Wahl von p^* bzw. w^* .

5.42 Beispiel: Wir knüpfen an Beispiel 5.41 an: Sei $R : \mathbb{R}^N \rightarrow \mathbb{R}$; $R(u) = \|u\|_1$. Dann gilt:

$$p \in \partial \|u\|_1 \Leftrightarrow \begin{cases} p^i = 1, & u^i > 0, \\ p^i = -1, & u^i < 0, \\ p^i \in [-1, 1], & u^i = 0. \end{cases}$$

Wir berechnen die zugehörige Bregman-Distanz:

$$\begin{aligned} D_{\|\cdot\|_1}^{p^*}(u, u^*) &= \|u\|_1 - \|u^*\|_1 - \langle p^*, u - u^* \rangle \\ &= \|u\|_1 - \langle p^*, u \rangle \\ &= \sum_{i=1}^N (|u^i| - (p^*)^i u^i) \\ &= \sum_{i: u^i (u^*)^i > 0} (|u^i| - |u^i|) + \sum_{i: u^i (u^*)^i < 0} 2|u^i| + \sum_{i: (u^*)^i = 0} |u^i| - (p^*)^i u^i \\ &= \sum_{i: u^i (u^*)^i < 0} 2|u^i| + \sum_{i: (u^*)^i = 0, u^i \neq 0} \underbrace{u^i (\text{sgn}(u^i) - (p^+)^i)}_{\in [0, 2|u^i|]}. \end{aligned}$$

Damit der zweite Term sinnvoll ist, ist es vorteilhaft, mehrere p^* auszuwählen und nicht nur $(p^*)^i = \text{sgn}(u^i)$. Der beste Fall ist $(p^*)^i = -\text{sgn}(u^i)$. Es gilt:

$$\sup_{p^* \in \partial \|u^*\|_1} D_{\|\cdot\|_1}^{p^*}(u, u^*) = 2 \sum_{i: \text{sgn}(u^i) \neq \text{sgn}((u^*)^i)} |u^i|.$$

5.43 Zur Fehlerabschätzung: Wir geben noch eine alternative Herleitung der Fehlerabschätzung an. Betrachte erneut:

$$\begin{aligned} K^*(Ku_\alpha - f) + \alpha p_\alpha &= 0, \\ K^*(K\tilde{u}_\alpha - \tilde{f}) + \alpha \tilde{p}_\alpha &= 0. \end{aligned}$$

Subtraktion liefert $K^*K(u_\alpha - \tilde{u}_\alpha) + \alpha(p_\alpha - \tilde{p}_\alpha) = K^*(f - \tilde{f})$. Bilden wir das Dualitätsprodukt mit $(u_\alpha - \tilde{u}_\alpha)$, so ergibt sich:

$$\begin{aligned} \underbrace{\langle K^*K(u_\alpha - \tilde{u}_\alpha), u_\alpha - \tilde{u}_\alpha \rangle}_{=\|Ku_\alpha - K\tilde{u}_\alpha\|^2} + \alpha \underbrace{\langle p_\alpha - \tilde{p}_\alpha, u_\alpha - \tilde{u}_\alpha \rangle}_{=D} &= \langle K^*(f - \tilde{f}), u_\alpha - \tilde{u}_\alpha \rangle \\ &= \langle f - \tilde{f}, K(u_\alpha - \tilde{u}_\alpha) \rangle \\ &\stackrel{\text{Young}}{\leq} \frac{1}{2} \|f - \tilde{f}\|^2 + \frac{1}{2} \|Ku_\alpha - K\tilde{u}_\alpha\|^2. \end{aligned}$$

Durch Subtraktion ergibt sich $\frac{1}{2} \|Ku_\alpha - K\tilde{u}_\alpha\|^2 + \alpha D \leq \frac{1}{2} \|f - \tilde{f}\|^2$.

Bemerkung: Einsetzen von $p_\alpha = K^*w_\alpha$, $\tilde{p}_\alpha = K^*\tilde{w}_\alpha$ und bilden des Dualitätsprodukts mit $(KK^*)^{-1}K^*(w_\alpha - \tilde{w}_\alpha)$ liefert den zweiten Teil der ursprünglichen Abschätzung.

5.44 Verallgemeinerte Quellbedingung: Wir suchen eine geeignete Verallgemeinerung für die Quellbedingung $u = (K^*K)^\mu w$ im linearen Fall.

Sei die Optimalitätsbedingung $K^*(Ku_\alpha^\delta - f^\delta) + \alpha p_\alpha^\delta = 0$ erfüllt und sei u^* die Lösung von $R(u) \rightarrow \min_u$ unter der Nebenbedingung $Ku = f$, $\|f - f^\delta\| < \delta$. Dann gilt auch:

$$K^*K(u_\alpha^\delta - u^*) + \alpha(p_\alpha^\delta - p^*) = K^*(f^\delta - f) - \alpha p^*, \quad p^* \in \partial R(u^*).$$

Durch Bilden des Dualitätsprodukts mit $(u_\alpha^\delta - u^*)$ ergibt sich:

$$\begin{aligned} \|K(u_\alpha^\delta - u^*)\|^2 + \alpha \langle p_\alpha^\delta - p^*, u_\alpha^\delta - u^* \rangle &= \langle f^\delta - f, K(u_\alpha^\delta - u^*) \rangle - \alpha \langle p^*, u_\alpha^\delta - u^* \rangle \\ &\stackrel{\text{Young}}{\leq} \frac{1}{2} \|f^\delta - f\|^2 + \frac{1}{2} \|K(u_\alpha^\delta - u^*)\|^2 \\ &\quad - \alpha \langle w, K(u_\alpha^\delta - u^*) \rangle + \alpha \langle K^*w - p^*, u_\alpha^\delta - u^* \rangle \\ &\leq \frac{\delta^2}{2} + \frac{\alpha^2}{2} \|w\|^2 + \alpha \langle K^*w - p^*, u_\alpha^\delta - u^* \rangle. \end{aligned}$$

Wir schränken uns nun auf den einfachsten Fall ein, nämlich $R(u) = \|u\|_X$. Dann gilt:

$$\langle K^*w - p^*, u_\alpha^\delta - u^* \rangle \leq \|K^*w - p^*\|_{X^*} \|u_\alpha^\delta - u^*\|_X \leq \|K^*w - p^*\|_{X^*} (\|u_\alpha^\delta\|_X + \|u^*\|_X).$$

Wegen

$$\underbrace{\frac{1}{2} \|Ku_\alpha^\delta - f^\delta\|^2 + \alpha R(u_\alpha^\delta)}_{\geq 0} \leq \frac{1}{2} \underbrace{\|Ku^* - f^\delta\|^2}_{\leq \delta^2} + \alpha R(u^*)$$

folgt $\|u_\alpha^\delta\|_X = R(u_\alpha^\delta) \leq R(u^*) + \frac{\delta^2}{2\alpha}$, also:

$$\|K^*w - p^*\|_{X^*} (\|u_\alpha^\delta\|_X + \|u^*\|_X) \leq \|K^*w - p^*\|_{X^*} (2R(u^*) + \frac{\delta^2}{2\alpha}).$$

Insgesamt erhalten wir die Abschätzung:

$$\langle p_\alpha^\delta - p^*, u_\alpha^\delta - u^* \rangle \leq \frac{\delta^2}{2\alpha} + \frac{\alpha\|w\|^2}{2} + \|K^*w - p^*\|_{X^*} \left(2R(u^*) + \frac{\delta}{2\alpha} \right).$$

Diese Herleitung motiviert folgende Definition:

5.45 Definition (Approximative Quellbedingung): Wir setzen

$$e(c) = \inf \{ \|K^*w - p^*\|_{X^*} \mid \|w\| \leq c \}$$

Dann ist die Glattheit formuliert über die Schnelligkeit der Konvergenz $e(c) \rightarrow 0$ für $c \rightarrow \infty$.

Damit ergibt sich für obige Abschätzung:

$$\langle p_\alpha^\delta - p^*, u_\alpha^\delta - u^* \rangle \leq \frac{\delta^2}{2\alpha} + \frac{\alpha c^2}{2} + e(c) \left(2R(u^*) + \frac{\delta}{2\alpha} \right).$$

5.46 Allgemeine Datenterme: Wir wollen allgemeine, konvexe Datenterme D betrachten und eine Abschätzung zwischen

$$K^* \partial D(Ku, f) + \alpha p = 0 \quad \text{und} \quad K^* \partial D(K\tilde{u}, \tilde{f}) + \alpha \tilde{p} = 0$$

erhalten. Subtraktion liefert $K^*(\partial D(Ku, f) - \partial D(K\tilde{u}, \tilde{f})) - \alpha(p - \tilde{p}) = 0$. Durch Bildung des Dualitätsprodukts mit $(u - \tilde{u})$ erhalten wir

$$\begin{aligned} \alpha \langle p - \tilde{p}, u - \tilde{u} \rangle &= -\langle \partial D(Ku, f) - \partial D(K\tilde{u}, \tilde{f}), K(u - \tilde{u}) \rangle \\ &= \langle \partial D(Ku, f), Ku - K\tilde{u} \rangle + \langle \partial D(K\tilde{u}, \tilde{f}), Ku - K\tilde{u} \rangle \\ &\stackrel{\text{Def.}}{\leq} D(K\tilde{u}, f) - D(Ku, f) + D(Ku, \tilde{f}) - D(K\tilde{u}, \tilde{f}). \end{aligned}$$

Damit ergibt sich schließlich:

$$\alpha \langle p - \tilde{p}, u - \tilde{u} \rangle + D(Ku, f) + D(K\tilde{u}, \tilde{f}) \leq D(K\tilde{u}, f) + D(Ku, \tilde{f}).$$

5.47 Zu Fenchel: Wir betrachten das Problem

$$D(v, f) + \alpha R(u) \rightarrow \min_{u, v: Ku = v}.$$

Es gilt dann:

$$\begin{aligned} \inf_{u, v} \sup_w D(v, f) + \alpha R(u) - \langle w, Ku - v \rangle &\stackrel{v \text{ elim.}}{=} \inf_u \sup_w \alpha R(u) - \langle w, Ku \rangle - D^*(-w, f) \\ &= \alpha \inf_u \sup_w R(u) - \langle \underbrace{\frac{w}{\alpha}}_{=: \tilde{w}}, Ku \rangle - \frac{1}{2} D^*(-w, f) \\ &= \alpha \inf_u \sup_{\tilde{w}} R(u) - \langle \tilde{w}, Ku \rangle - \frac{1}{2} D^*(-\alpha \tilde{w}, f) \\ &= \alpha \inf_u \sup_w R(u) - \langle w, Ku - f \rangle - \frac{1}{2} D^*(-\alpha, f) + \langle w, f \rangle. \end{aligned}$$

Dann ist eine Lösung u_α von $D(Ku, f) + \alpha R(u) \rightarrow \min_u$ ein Sattelpunkt (zusammen mit w_α) des obigen Problems.

Die Optimalitätsbedingung lautet $p_\alpha - K^* w_\alpha = 0$ für $p_\alpha \in \partial R(u_\alpha)$ und ist somit äquivalent zur Quellbedingung. Leiten wir nach der ersten Variablen ab, so ergibt sich:

$$-Ku_\alpha - \frac{1}{\alpha} \partial D^*(-\alpha w_\alpha, f)(-\alpha) = 0 \Leftrightarrow Ku_\alpha = \partial D^*(-\alpha w_\alpha, f), \quad K^* w_\alpha \in \partial R(u_\alpha).$$

Für Daten \tilde{f} gilt dann entsprechend:

$$K\tilde{u}_\alpha = \partial D^*(-\alpha \tilde{w}_\alpha, \tilde{f}), \quad K^* \tilde{w}_\alpha \in \partial R(\tilde{u}_\alpha).$$

Als symmetrische Bregman-Distanz ergibt sich dann entsprechend:

$$\begin{aligned} \langle p_\alpha - \tilde{p}_\alpha, u_\alpha - \tilde{u}_\alpha \rangle &= \langle K^* w_\alpha - K^* \tilde{w}_\alpha, u_\alpha - \tilde{u}_\alpha \rangle \\ &= \langle w_\alpha - \tilde{w}_\alpha, Ku_\alpha - K\tilde{u}_\alpha \rangle \\ &= \frac{1}{\alpha} \langle Ku_\alpha, -\alpha \tilde{w}_\alpha - (-\alpha \tilde{w}_\alpha) \rangle + \frac{1}{\alpha} \langle K\tilde{u}_\alpha, -\alpha w_\alpha - (-\alpha \tilde{w}_\alpha) \rangle \\ &\leq \frac{1}{\alpha} (D^*(-\alpha \tilde{w}_\alpha, f) - D^*(-\alpha w_\alpha, f)) + \frac{1}{\alpha} (D^*(-\alpha w_\alpha, \tilde{f}) - D^*(-\alpha \tilde{w}_\alpha, \tilde{f})). \end{aligned}$$

5.48 Beispiel: Wir betrachten $D(Ku, f) = \|Ku - f\|_{L^1} = \int |(Ku)(x) - f(x)| dx$, was dem Modell für einen Laplace-verteilten Fehler entspricht. Wir nehmen $\|f - f^\delta\|_{L^1} \leq \delta$ an. Wir benötigen in unserem Fall also eine Abschätzung in Bezug auf $\|f - f^\delta\|_{L^1}$.

Ist $D(v, f) = \|v - f\|_{L^1}$, so wissen wir, dass $D^*(w, f) \stackrel{3.35}{=} D^*(w, 0) + \langle w, f \rangle$ gilt. Ist weiter $J(u) = \|u\|_X$, so gilt:

$$J^*(w) = \begin{cases} 0, & \|w\|_{X^*} \leq 1, \\ +\infty, & \text{sonst} \end{cases} = \chi_1^{X^*}(w).$$

Damit folgt:

$$D^*(w, 0) = \chi_1^{L^\infty}(w) = \begin{cases} 0, & \|w\|_{L^\infty} \leq 1, \\ +\infty, & \text{sonst.} \end{cases}$$

Wir erhalten dann:

$$D^*(-\alpha w, f) = \begin{cases} -\alpha \langle w, f \rangle, & \alpha \|w\|_{L^\infty} \leq 1, \\ +\infty, & \text{sonst.} \end{cases}$$

Für α hinreichend klein gilt dann $\alpha\|w_\alpha\| \leq 1, \alpha\|\tilde{w}_\alpha\| \leq 1$. Einsetzen in obige Fehlerabschätzung liefert dann:

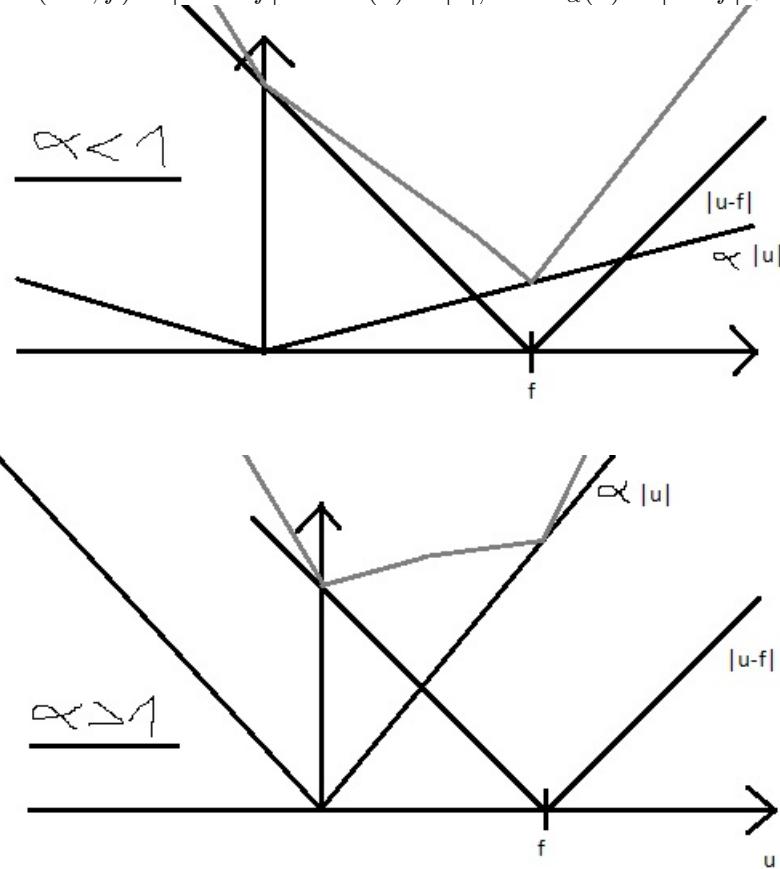
$$\begin{aligned}
 \langle p_\alpha - \tilde{p}_\alpha, u_\alpha - \tilde{u}_\alpha \rangle &\leq -\langle \tilde{w}_\alpha, f \rangle + \langle w_\alpha, f \rangle - \langle w_\alpha, \tilde{f} \rangle + \langle \tilde{w}_\alpha, \tilde{f} \rangle \\
 &= \langle w_\alpha - \tilde{w}_\alpha, f - \tilde{f} \rangle \\
 &\leq \frac{1}{\alpha} \underbrace{\|\alpha w_\alpha - \alpha \tilde{w}_\alpha\|_{L^\infty}}_{\leq \|\alpha w_\alpha\|_{L^\infty} + \|\alpha \tilde{w}_\alpha\|_{L^\infty} \leq 2} \|f - \tilde{f}\|_{L^1} \\
 &\leq \frac{2}{\alpha} \|f - \tilde{f}\|_{L^1}.
 \end{aligned}$$

Damit kann man ein fixes α wählen, nämlich so, dass $\alpha\|w_\alpha\| = 1$.

5.49 Bemerkung: Bei der L^1 -Norm ergibt sich kein zusätzlicher Term wie $\alpha\|w\|^2$ oder ähnliches für kleines α . Damit gibt es keinen systematischen Fehler der Regularisierung.

5.50 Beispiel: Sei $X = Y = \mathbb{R}$. Wir betrachten den Operator $Ku = u$.

(a) Sei $D(Ku, f) = |Ku - f|$ und $R(u) = |u|$, also $J_\alpha(u) = |u - f| + \alpha|u|$. Betrachte zwei Fälle:



Damit ergibt sich:

$$u_\alpha = \begin{cases} 0, & \alpha > 1, \\ f, & \alpha \leq 1. \end{cases}$$

Insbesondere haben wir $J(0) = |f|, J(f) = \alpha|f|$.

(b) Sei $D(Ku, f) = \frac{1}{2}|Ku - f|^2$ und $R(u) = |u|$, also $J_\alpha(u) = \frac{1}{2}|u - f|^2 + \alpha|u|$. Dann gilt:

$$u_\alpha = S_\alpha(f) = \begin{cases} 0, & |f| < \alpha, \\ f - \alpha, & f > \alpha, \\ f + \alpha, & f < -\alpha. \end{cases}$$

Zum systematischen Fehler: Auch groSSe f werden im Betrag um α reduziert.

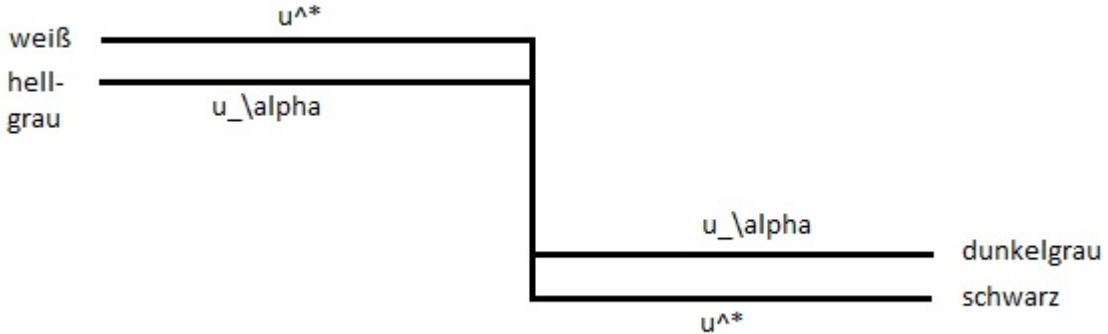
5.51 Abschlussüberlegung: Es gilt

$$\underbrace{D(Ku_\alpha, f)}_{\geq 0} + \alpha R(u_\alpha) \leq D(Ku^*, f) + \alpha R(u^*),$$

also

$$R(u_\alpha) \leq R(u^*) + \frac{1}{\alpha} D(Ku^*, f).$$

Oft gilt dann sogar $R(u_\alpha) < R(u^*)$. Ist beispielsweise $R(u) = \int |\nabla u|$, so ergibt sich bei der Bildrekonstruktion folgende Situation:



Der Minimierer von J_α weist also schwächere Kontraste auf.

6 Iterative Regularisierung basierend auf Variationsmethoden

6.1 Bregman-Iteration

Wir betrachten wieder

$$D(Ku, f) + \alpha R(u) \rightarrow \min_u.$$

Ein typischer Fall ist $R \geq 0$ mit einem Minimum bei $u = 0$, d.h. es gilt $0 \in \partial R(0)$, $R(0) = 0$. Dann ist das ursprüngliche Problem äquivalent zu:

$$D(Ku, f) + \alpha(R(u) - R(0) - \langle \underbrace{0}_{\in \partial R(0)}, u - 0 \rangle) \rightarrow \min_u \Leftrightarrow D(Ku, f) + \alpha D_R^0(u, 0) \rightarrow \min_u$$

Dies führt zur Bregman-Interation:

6.1 Definition: Initialisiere:

$$u^0 := 0, p^0 := 0 \in \partial R(u^0) \quad (\text{allgemeiner: } u^0 \text{ als Minimierer von } R).$$

Iteriere:

$$u^{k+1}, k = 0, 1, 2, \dots \text{ als L\"osung von } D(Ku, f) + \alpha D_R^{p_k}(u, u^k) \rightarrow \min_u.$$

Berechne: p^{k+1} aus der Optimalitätsbedingung

$$0 = K^* \partial D(Ku^{k+1}, f) + \alpha(p^{k+1} - p^k), \quad p^{k+1} \in \partial R(u^{k+1}), \quad p^{k+1} = p^k - \frac{1}{\alpha} K^*(\partial D(Ku^{k+1}, f)).$$

Für $D(Ku, f) = \frac{1}{2}\|Ku - f\|^2$ ergibt sich dann die Iteration:

$$u_0 = 0, \quad p_0 = 0 \in \partial R(u_0);$$

Für $k = 0, 1, 2, \dots$:

$$\begin{aligned} u_{k+1} &\in \arg \min_u [\frac{1}{2}\|Ku - f\|^2 + \alpha D_R^{p_k}(u, u_k)]; \\ p_{k+1} &= p_k - \frac{1}{\alpha} K^*(Ku_{k+1} - f) \in \partial R(u_{k+1}). \end{aligned}$$

Wir ersetzen also den Regularisierungsterm beim Variationsproblem durch die zugehörige Bregman-Distanz.

6.2 Satz: Seien K, R so, dass

$$\frac{1}{2}\|Ku - g\|^2 + \alpha R(u) \rightarrow \min_{u \in X}$$

für alle $g \in Y, \alpha > 0$ eine Lösung hat. Dann ist auch die Bregman-Iteration wohldefiniert und es gilt $p_k = K^*w_k$ für eine Folge $(w_k) \subseteq Y$, d.h. die p_k erfüllen die Quellbedingung.

6.3 Alternative Formulierung: Berechne

$$u_{k+1} \in \arg \min_{u \in X} (\frac{1}{2}\|Ku - k_k\|^2 + \alpha R(u))$$

mit $g_k = f + v_k, v_{k+1} = v_k + Ku_{k+1} + f$, also

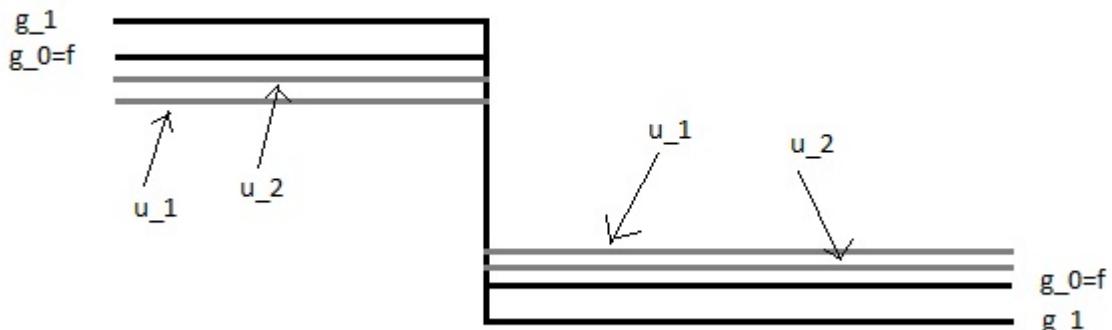
$$g_{k+1} = g_k + f - Ku_{k+1}, \quad g_0 = f.$$

Das Residuum $f - Ku_{k+1}$ wird also wieder zu den Daten addiert.

6.4 Beispiel (ROF-Modell): Betrachte

$$\frac{1}{2}\|u - f\|_{L^2}^2 + \alpha|u|_{BV} \longrightarrow \min_u .$$

Dann beobachten wir folgendes Phänomen:



Zur Interpretation: g_k addiert den systematischen Fehler zu den Daten; die Regularisierung macht

einen systematischen Fehler zu g_k , damit näher bei f .

6.5 Beispiel (Eindimensionales Shrinkage): Wir wollen $\frac{1}{2}|u - f|^2 + \alpha|u|$ minimieren. Setze $u_0 = 0$, $g_0 = f$. Dann gilt:

$$u_1 = S_\alpha(g_0) = \begin{cases} 0, & |f| \leq \alpha, \\ (|f| - \alpha)\operatorname{sgn}(f), & |f| \geq \alpha. \end{cases}$$

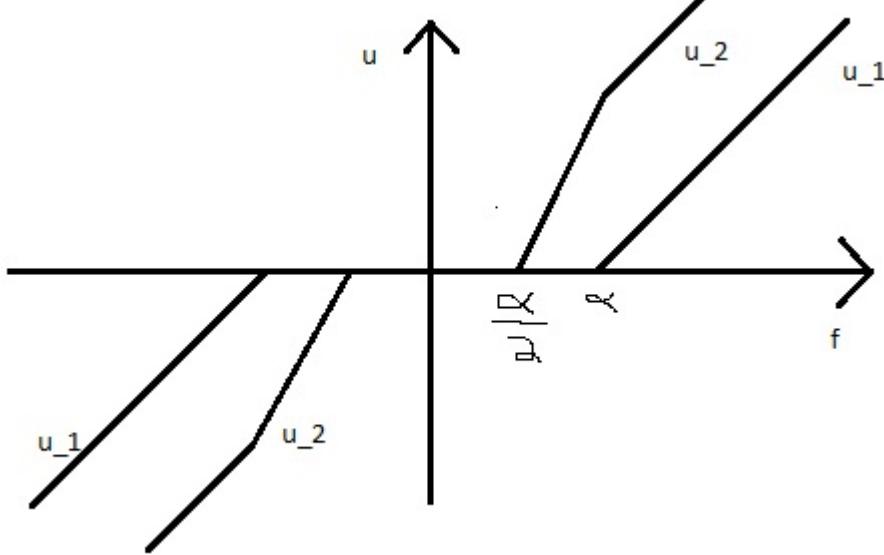
Berechne g_1 :

$$g_1 = g_0 + f - u_1 = 2f - u_1 = \begin{cases} 2f, & |f| < \alpha, \\ f + \alpha\operatorname{sgn}(f), & |f| \geq \alpha. \end{cases}$$

Berechne u_2 :

$$u_2 = S_\alpha(g_1) = \begin{cases} 0, & |g_1| < \alpha, \\ (|g_1| - \alpha)\operatorname{sgn}(g_1), & |g_1| \geq \alpha \end{cases} = \begin{cases} 0, & |f| < \frac{\alpha}{2}, \\ (|2f| - \alpha)\operatorname{sgn}(f), & \frac{\alpha}{2} < |f| < \alpha, \\ f, & |f| \geq \alpha. \end{cases}$$

Wir beobachten dann folgende Minimierer:



6.6 Folgerung: k Iterationen der Bregman-Iteration sind eher vergleichbar zur Variationsmethode mit Parameter $\frac{\alpha}{k}$. Ist umgekehrt α der Parameter der Variationsmethode, so entspricht dies dem Parameter αk bei k Bregman-Iterationen.

6.7 Inverse Scale Space Method: Wir wollen nun die Bregman-Iteration für den Fall $\alpha \rightarrow \infty$ untersuchen. Wir haben

$$\alpha w_{k+1} = \alpha w_k + f - Ku_{k+1}, \quad K^*w_{k+1} \in \partial R(u_{k+1}).$$

Wir ersetzen nun $\alpha \rightarrow \infty$ durch $\frac{1}{\Delta t} \rightarrow \infty$ für $\Delta t \rightarrow 0$. Dann gilt:

$$\frac{w_{k+1} - w_k}{\Delta t} = f - Ku_{k+1}.$$

Mit $w(k\Delta t) = w_k$ und $u(k\Delta t) = u_k$ gilt dann:

$$\frac{w(k\Delta t + \Delta t) - w(k\Delta t)}{\Delta t} = f - Ku(k\Delta t + \Delta t).$$

Durch Grenzwertbildung $\Delta t \rightarrow 0$ ergibt sich dann:

$$\frac{\partial w}{\partial t}(t) = f - Ku(t), \quad K^*w(t) \in \partial R,$$

und analog:

$$\frac{\partial p}{\partial t}(t) = K^*(f - Ku(t)), \quad p(t) \in \partial R(u(t)).$$

Dies ist die Inverse Scale Space Method.

6.8 Beispiel (Shrinkage): Sei $R(u) = |u|$ und $K : \mathbb{R} \rightarrow \mathbb{R}$; $u \mapsto u$. Dann gilt:

$$p'(t) = f - u(t), \quad p(t) \in \partial|u(t)| = \operatorname{sgn}(u(t)).$$

Für $p(0) = 0$ folgt dann:

$$p(t) = \int_0^t (f - u(s)) \, ds.$$

Dann gibt es ein $t_0 > 0$, sodass $|p(t)| < 1$ für alle $t \in [0, t_0]$. Wegen $p(t) \in \operatorname{sgn}(u(t))$ folgt $u(t) = 0$ in $[0, t_0]$. Damit folgt $p(t) = tf$ in $[0, t_0]$, also $t_0 = \frac{1}{|f|}$. Für $t > t_0$ gilt $p' = f - u$, also

$$p(t) = p(t_0) + \int_{t_0}^t (f - u(s)) \, ds = \operatorname{sgn}(f) + \int_{t_0}^t (f - u(s)) \, ds.$$

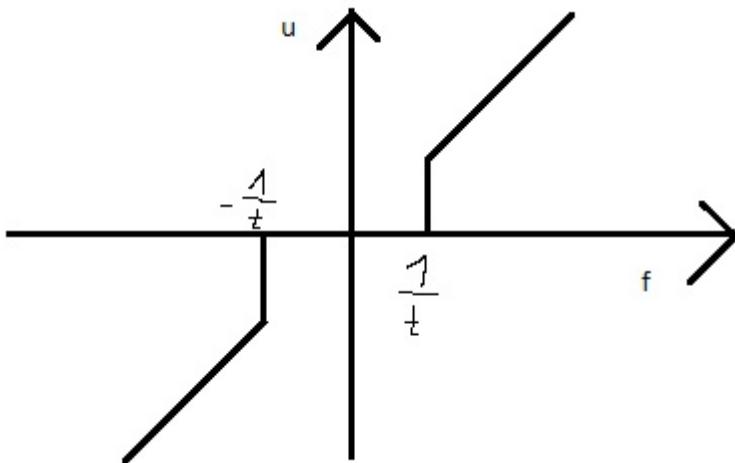
Die Lösung ist dann $u(t) = f$ für $t > t_0$, $p(t) = \operatorname{sgn}(f) = \operatorname{sgn}(u(t))$. Insgesamt ergibt sich:

$$u(t) = \begin{cases} 0, & t < t_0 = \frac{1}{|f|}, \\ f, & t > t_0. \end{cases}$$

Für fixes t gilt:

$$u(t) = \begin{cases} 0, & |f| < \frac{1}{t}, \\ f, & |f| \geq \frac{1}{t}. \end{cases}$$

Damit gibt es keinen systematischen Fehler bei grossen Werten.



6.9 Analyse der Iteration: Es liegt Semikonvergenz vor, d.h.:

- Bei exakten Daten $f = Ku^*$ gilt $u_k \rightarrow u^*$ für $k \rightarrow \infty$.

- Bei verrauschten Daten f^δ mit $\|f - f^\delta\| \leq \delta$ haben wir:

$$p_{k+1}^\delta = p_k^\delta - K^*(Ku_{k+1}^\delta - f^\delta).$$

Wir stoppen die Iteration dann für $k = k(\delta, f^\delta)$. Für die Rekonstruktion gilt:

$$u_{k(\delta, f^\delta)}^\delta \longrightarrow u^*, \quad \delta \rightarrow 0, k \rightarrow \infty.$$

6.10 Proposition: Sei R konvex und u_k^δ die Folge der Iterierten aus der Bregman-Iteration. Dann gilt:

$$\|Ku_{k+1}^\delta - f^\delta\| \leq \|Ku_k^\delta - f^\delta\|,$$

mit Gleichheit genau dann, wenn $D_R^{p_k^\delta}(u_{k+1}, u_k) = 0$. Die Iteration hört also nur auf, wenn die Lösung erreicht ist.

6.11 Zu ISS: Es gilt $\frac{d}{dt}\|Ku(t) - f\|^2 = 2\langle Ku - f, K\partial_t u \rangle$ und $[p \in \partial R(u) \Rightarrow u \in \partial R^*(p)]$. Falls R^* zweimal differenzierbar ist, so gilt $u(t) = \partial R^*(p(t))$ und

$$\partial_t u = \partial^2 R^*(p)\partial_t p = -\partial^2 R^*(p)K^*(Ku - f)$$

und damit

$$\begin{aligned} \frac{d}{dt}\|Ku(t) - f\|^2 &= -2\langle \underbrace{K^*(Ku - f)}_{=q}, \underbrace{\partial^2 R^*(p)(K^*(Ku - f))}_{=q} \rangle \\ &= -2\langle q, \underbrace{\partial^2 R^*(p)}_{\text{pos.semidef. } (R^* \text{ konv.})} q \rangle \leq 0, \end{aligned}$$

der Datenterm wird also kleiner.

Wir wollen nun untersuchen, was mit u_k^δ bzw. $u^\delta(t)$ für $k \rightarrow \infty$ bzw $t \rightarrow \infty$ passiert. Dazu betrachten wir den Abstand von u_k^δ bzw. $u^\delta(t)$ zu u^* in der Bregman-Distanz. Gilt

$$D_R^{p_{k+1}^\delta}(u^*, u_{k+1}^\delta) - D_R^{p_k^\delta}(u^*, u_k^\delta) \leq 0 \quad \text{bzw.} \quad \frac{d}{dt}D_R^{p^\delta(t)}(u^*, u^\delta(t)) \leq 0 \quad ?$$

Wir schränken uns auf ISS ein. Zunächst gilt:

$$\begin{aligned} \frac{d}{dt}D_R^{p^\delta(t)}(u^*, u(t)) &= \frac{d}{dt}(R(u^*) - R(u^\delta) - \langle p(t), u^* - u^\delta \rangle) \\ &= -\langle \partial R(u^\delta), \partial_t u^\delta \rangle - \langle \partial_t p^\delta, u^* - u^\delta \rangle + p^\delta, \partial_t u^\delta \rangle \\ &\stackrel{p^\delta = \partial R(u^\delta)}{=} -\langle \partial_t p^\delta, u^* - u^\delta \rangle \\ &= -\langle K^*(f^\delta - Ku^\delta), u^* - u^\delta \rangle \\ &= -\langle f^\delta - Ku^\delta, Ku^* - Ku^\delta \rangle \\ &= -\langle f^\delta - Ku^\delta, f^\delta - Ku^\delta \rangle + \langle f^\delta - Ku^\delta, f^\delta - Ku^* \rangle \\ &\leq -\|f^\delta - Ku^\delta\|^2 + \|f^\delta - Ku^\delta\| \underbrace{\|f^\delta - Ku^*\|}_{\leq \delta} \\ &\leq -\|f^\delta - Ku^\delta\|(\|f^\delta - Ku^\delta\| - \delta) \\ &< 0, \text{ solange } \|f^\delta - Ku^\delta\| > \delta. \end{aligned}$$

Es ergibt sich erneut eine Art Diskrepanzprinzip, das die Wahl der Abbruchzeit vorgibt:

$$T(\delta, f^\delta) := \inf\{t > 0 \mid \|Ku^\delta - f^\delta\| \leq \delta\}.$$

Wir wissen damit $\|Ku^\delta(T(\delta, f^\delta)) - f^\delta\| \leq \delta$, also $Ku^\delta(T(\delta, f^\delta)) \rightarrow f$ für $\delta \rightarrow 0$. Wegen

$$\frac{d}{dt} D_R^{p^\delta}(u^*, u^\delta) \leq -\|Ku^\delta - f^\delta\| (\|Ku^\delta - f^\delta\| - \delta)$$

folgt durch Integration:

$$D_R^{p^\delta(T)}(u^*, u^\delta(T)) + \int_0^T \|Ku^\delta - f^\delta\|^2 dt \leq \underbrace{D_R^{p^\delta(0)}(u^*, u^\delta(0))}_{=R(u^*)} + \delta \int_0^T \|Ku^\delta - f^\delta\| dt$$

bzw. $D_R^{p^\delta(T)}(u^*, u^\delta(T)) \leq R(u^*)$. Es gilt dann

$$\delta \int_0^T \|Ku^\delta - f^\delta\|^2 dt \leq \frac{1}{2} \int_0^T \delta^2 + \|Ku^\delta - f^\delta\|^2 dt,$$

also

$$\int_0^T \underbrace{\|Ku^\delta(t) - f^\delta\|^2}_{\geq \|Ku^\delta(T) - f^\delta\|^2} dt \leq \frac{\delta^2}{2} T + R(u^*)$$

und weiter

$$\frac{1}{2} \|Ku^\delta(T) - f^\delta\|^2 T \leq \frac{\delta^2 T}{2} + R(u^*).$$

Dies gilt für beliebiges T . Durch Multiplikation mit $\frac{2}{T}$ ergibt sich:

$$\|Ku^\delta - f^\delta\|^2 \leq \delta^2 + \frac{2}{T} R(u^*).$$

Insbesondere folgt dann für $\delta = 0$:

$$\|Ku - f\|^2 \leq \frac{2R(u^*)}{T} \rightarrow 0 \quad \text{für } T \rightarrow \infty.$$

6.12 Allgemeines Prinzip: Betracht folgende Differentialgleichungen:

$$E : \mathbb{R}^+ \rightarrow \mathbb{R}^+, \quad \frac{dE}{dt}(t) \leq 0 \quad \forall t > 0;$$

$$F : \mathbb{R}^+ \rightarrow \mathbb{R}^+, \quad \frac{dF}{dt}(t) \leq -cE(t), \quad c > 0.$$

Dann gilt:

$$F(t) + c \int_0^t E(s) ds \leq F(0).$$

Für $s \leq t$ gilt $E(s) \geq E(t)$, also $F(t) + ctE(t) \leq F(0)$. Da $F \geq 0$ folgt $E(t) \leq \frac{F(0)}{ct} = \mathcal{O}(\frac{1}{t})$. Vorher hatten wir

$$E(t) = \frac{1}{2} \|Ku(t) - f\|^2, \quad F(t) = D_R^{p(t)}(u^*, u(t)),$$

und jetzt:

$$E(t) = D_R^{p(t)}(u^*, u(t)), \quad F(t) = ?.$$

Der allgemeinere Fall kann nur unter zusätzlichen Annahmen an u^* funktionieren:
Es sei die Quellbedingung $K^*w^* \in \partial R(u^*)$ erfüllt. Beachte dann, dass

$$p(t) = K^*w(t), \quad \partial_t w(t) = f - Ku(t), \quad \partial_t w^*(t) = 0 = f - Ku^*, \quad \partial_t(w - w^*) = Ku^* - Ku$$

gelten. Wir haben dann:

$$\begin{aligned}
\langle \partial_t(w - w^*), w - w^* \rangle &= -\langle K(u - u^*), w - w^* \rangle \\
&= -\langle u - u^*, \underbrace{K^*w - K^*w^*}_{=p} \rangle \\
&= -(D_R^p(u^*, u) + D_R^{K^*w^*}(u, u^*)).
\end{aligned}$$

Weiter gilt dann:

$$-\underbrace{D_R^p(u^*, u(t))}_{=E(t)} \geq \frac{d}{dt} \underbrace{\left(\frac{1}{2} \|w - w^*\|^2 \right)}_{F(t)},$$

also $E(t) \leq \frac{F(0)}{t}$. Mit $u(0) = 0$ folgt dann schlieSSlich $D_R^p(u^*, u(t)) \leq \frac{\|w^*\|^2}{t}$. Damit erhalten wir analoge Abschätzungen zu den Variationsmethoden. Analog ergeben sich auch Abschätzungen für die diskrete Iteration:

$$E_{k+1} \leq E_k \text{ und } F_{k+1} - F_k \leq -cE_{k+1} \Rightarrow E_k \leq \frac{F_0}{ck}.$$

6.2 Konstruktion weiterer Iterationsverfahren

Betrachte $\partial_t p = K(f - Ku)$. Dann ist die Bregman-Iteration ein Rückwärts-Euler-Verfahren für diese Differentialgleichung. Eine Alternative wäre ein Vorwärts-Euler-Verfahren:

$$p_{k+1} = p_k + \tau K^*(f - Ku_k) \quad (\tau = \frac{1}{\alpha}).$$

Für $R(u) = \frac{1}{2}\|x\|^2$ folgt dann $p_k = u_k$, also $u_{k+1} = u_k + \tau K^*(f - Ku_k)$ (Landweber-Iteration). Für Konvergenz beim Landweber-Verfahren wähle $\tau < \frac{2}{\|K\|^2}$. Das Landweber-Verfahren ist dann äquivalent zur Minimierung von

$$u \in \arg \min_u \frac{1}{2} \|Ku_k - f\|^2 + \langle K(u - u_k), Ku_k - f \rangle + \alpha D_R^{p_k}(u, u_k).$$

Dies entspricht der Taylor-Entwicklung erster Ordnung für $\frac{1}{2}\|Ku - f\|^2$ um den Punkt u_k . R muss dann passend gewählt sein, damit das Funktional nach unten beschränkt ist. Dies ist insbesondere problematisch für $R = \|\cdot\|_{l^1}$ oder $R = |\cdot|_{TV}$. Eventuell ist u_{k+1} nicht wohldefiniert. Dabei ist die totale Variation definiert durch:

$$|u|_{TV} := \sup \left\{ \int_{\Omega} u \operatorname{div}(\varphi) dx \mid \varphi \in \mathcal{C}_0^{\infty}(\Omega; \mathbb{R}^n), \|\varphi\|_{\infty} \leq 1 \right\}.$$

Die Seminorm $|\cdot|_{TV}$ ist insofern relevant, als die L^p -Normen zu viele Funktionen zulassen, die in der Bildgebung nicht gewollt sind (z.B. stark oszillierende Funktionen), und die Sobolev-Normen zu restriktiv sind (beispielsweise sind stückweise konstante Funktionen nicht in Sobolev-Räumen enthalten).