5 Analyse

In diesem Kapitel werden wir grundlegende Techniken zur Analyse von Variationsmethoden in Banach-Räumen kennen lernen, sowie die Anwendung auf die typische Form der Funktionale in der Bildverarbeitung.

5.1 Grundlegende Funktionalanalysis und Existenz

Die Lösung der Variationsprobleme aus den letzten Kapiteln ist ein Problem der Variationsrechnung, bei der man sich mit Optimierungsproblemen der Form

$$J(u) = \int_\Omega G(x, (Ku)(x), u(x), \nabla u(x))\; dx \to \min_u \tag{5.1}$$

beschäftigt. So erhält man z.B. das Variationsmodell zum Entrauschen mit additivem Gauss-Modell und gradientenbasierter Regularisierung als Spezialfall für

$$G(x, z, p, q) = \frac{1}{2}(p - f(x))^2 + F(x, q),$$

mit $F(x, q) = \alpha|q|$ für das ROF-Modell.

Bei genauerer Betrachtung des Optimierungsproblems (5.1) stellt sich die Frage über welche Klasse von Funktionen u minimiert werden soll, d.h. in welchem Funktionenraum. Bei der obigen Form würde sich auf den ersten Blick der Raum der stetig differenzierbaren Funktionen

$$C^1(\Omega) := \{u \in C(\Omega) \mid \nabla u \in C(\Omega)\}$$

anbieten, dann wären die Integrale auch ohne Schwierigkeiten zu definieren. Dies ist aber sowohl aus Sicht der Anwendung in der Bildverarbeitung (Bilder mit Kanten sind unstetig und daher sicher auch nicht C^1) als auch für die Analysis ungünstig, wie wir sehen werden.

Die grundlegenden Probleme bei der Analyse von (5.1) sind wiederum Existenz und Eindeutigkeit eines minimierenden Elements (das wir auch kurz als Minimum bezeichnen), im speziellen Fall von eq:kanonischesFunktional ist man dann auch an der Abhängigkeit der Lösung von den Daten f und vom Parameter α interessiert. Bei einem Optimierungsproblem lässt sich die Frage der Existenz letztlich immer auf den Fundamentalsatz der Optimierung zurückführen, den wir in einer recht allgemeinen Form benutzen:

Theorem 5.1.1. *Sei $J : (\mathcal{X}, \tau) \to \mathbb{R} \cup \{+\infty\}$ ein Funktional auf einem topologischen Raum $\mathcal{X}$ mit (lokalkonvexer) Topologie τ, das die beiden folgenden Bedingungen erfüllt:*

- **Folgenunterhalbstetigkeit:** *für $u_k \to u$ in der Topologie τ gelte*

$$J(u) \leq \liminf_k J(u_k). \tag{5.2}$$

- **Kompaktheit von Niveaumengen(Sub-Level-Sets):** *es existiert ein $\alpha \in \mathbb{R}$, sodass*

$$\mathcal{S}_\alpha := \{u \in \mathcal{X} \mid J(u) \leq \alpha\}$$

  nicht leer und kompakt in der Topologie τ ist.

Dann existiert ein Minimum $\hat{u} \in \mathcal{X}$, d.h.

$$J(\hat{u}) = \inf_{u \in \mathcal{X}} J(u).$$

Beweis. Sei $(u_k)_{k \in \mathbb{N}}$ eine minimierende Folge. Dann gilt für k hinreichend gross $u_k \in \mathcal{S}_\alpha$. Damit ist $(u_k)_{k \geq k_0}$ in einer kompakten Menge enthalten und hat also eine konvergente Teilfolge, die wir wieder mit (u_k) bezeichnen. Ihr Grenzwert sei $\hat{u}$. Wegen der Folgenunterhalbstetigkeit gilt dann

$$\inf_u J(u) \leq J(\hat{u}) \leq \liminf_k J(u_k) = \inf_u J(u),$$

also ist $\hat{u}$ ein globales Minimum von J. □

In der endlichdimensionalen Optimierung kann dieser Satz sehr einfach zum Existenzbeweis verwendet werden. Die Unterhalbstetigkeit ist dort noch recht anschaulich und die Kompaktheit folgt immer aus der Beschränktheit. In Funktionenräume gilt letzteres wegen der unendlichen Dimension nicht mehr. Um aus

Beschränktheit noch Kompaktheit folgern zu können, sind schwache bzw. schwach-* Topologien nötig. Sei $\mathcal{X}$ ein Banachraum und $\mathcal{X}^*$ sein Dualraum. Dann ist die schwache Topologie auf $\mathcal{X}$ definiert durch

$$u_k \rightharpoonup u \qquad :\Leftrightarrow \langle v, u_k \rangle \to \langle v, u \rangle \quad \forall\ v \in \mathcal{X}^*,$$

und die schwach-* Topologie auf $\mathcal{X}^*$ definiert durch

$$v_k \rightharpoonup *v \qquad :\Leftrightarrow \langle v_k, u \rangle \to \langle v, u \rangle \quad \forall\ u \in \mathcal{X}.$$

Die schwach-* Topologie auf $\mathcal{X}^*$ ist eine schwächere als die schwache Topologie auf diesem Raum, da $\mathcal{X} \subset \mathcal{X}^{**}$. Bei einem reflexiven Banachraum ($\mathcal{X} = \mathcal{X}^{**}$) fallen die schwache und die schwach-* Topologie zusammen.

Das zentrale Resultat zur Kompaktheit ist dann der Satz von Banach-Alaoglu, der zumindest in der schwach-* Topologie die Kompaktheit aus der Beschränktheit folgert:

Theorem 5.1.2 (Banach-Alaoglu)**.** *Sei $X = Z^*$ der Dualraum eines Banachraums Z und $\mathcal{M}$ eine beschränkte Menge in X. Dann ist $\mathcal{M}$ präkompakt in der schwach* Topologie.*

Beweis. Der allgemeine Beweis folgt aus dem Satz von Tikhonov, einem relativ allgemeinen topologischen Resultat. Wir beschränken uns auf den Fall, dass Z separabel ist, was für unsere Anwendung ausreichen wird. In diesem Fall ist die schwach* Konvergenz auf beschränkten Mengen auch metrisierbar, sodass wir uns auf Folgenkonvergenz einschränken können. Sei also x_k eine beschränkte Folge in X und $\{z_j\}_{j\in\mathbb{N}}$ sei dicht in Z (so eine Menge existiert wegen der Separabilität). Wir wählen nun induktiv Teilfolgen

$$\mathbb{N} \supset \Lambda_1 \supset \Lambda_2 \supset \ldots,$$

sodass

$$\langle x_k, z_j \rangle \to a_j \qquad \text{für } k \in \Lambda_j, k \to \infty$$

gilt. Dies ist möglich, da

$$|\langle x_k, z_j \rangle| \le ||x_k||_X ||z_j||_Z \le c||z_j||_Z$$

gilt, und damit $\langle x_k, z_j \rangle$ eine beschränkte Folge reeller Zahlen ist, die eine konvergente Teilfolge besitzt. Sei Λ die Diagonalfolge der Λ_k, dann gilt für alle j

$$\langle x_k, z_j \rangle \to a_j \qquad \text{für } k \in \Lambda, k \to \infty.$$

Wir definieren nun ein lineares Funktional ℓ zunächst auf $\{z_j\}$ durch $\ell(z_j) := a_j$, das wir automatisch auf die lineare Hülle der z_j durch

$$\ell(\sum c_j z_j) = \sum c_j a_j$$

fortsetzen. Man sieht daraus sofort, dass für $z = \sum c_j z_j$ gilt

$$\ell(z) = \sum c_j a_j = \sum c_j \lim_{k\in\Lambda} \langle x_k, z_j\rangle = \lim_{k\in\Lambda} \langle x_k, \sum c_j z_j\rangle = \lim_{k\in\Lambda} \langle x_k, z\rangle.$$

Damit ist ℓ wegen der gleichmässigen Beschränktheit der x_k ein beschränktes lineares Funktional auf der linearen Hülle der z_j, es gilt ja

$$|\ell(z)| \leq \limsup_{k\in\Lambda} |\langle x_k, z\rangle| \leq \limsup_k ||x_k||_X ||z||_Z \leq c||z||_Z.$$

Ein beschränktes lineares Funktional ist immer auf den Abschluss fortsetzbar, also in diesem Fall auf ganz Z. D.h. es existiert ein $x \in X$ mit $\langle x, z\rangle = \ell(x)$ für alle z aus der linearen Hülle der z_j.

Sei nun $z \in Z$ beliebig, dann gibt es eine Indexmenge $I \subset N$ mit $z_j \to z$, $j \in I$.

$$\begin{aligned}
|\langle x_k - x, z\rangle| &\leq |\langle x_k, z - z_j| + |\langle x, z - z_j| + |\langle x_k - x, z_j| \\
&\leq (||x_k||_X + ||x||_X)||z - z_j||_Z + |\langle x_k - x, z_j| \\
&\leq (\sup_k ||x_k||_X + ||x||_X)||z - z_j||_Z + |\langle x_k - x, z_j| \\
&\leq C||z - z_j|| + |\langle x_k - x, z_j|.
\end{aligned}$$

Für festes j erhalten wir nun im Grenzwert $k \to \infty$ (wegen $\langle x_k, z_j\rangle \to \langle x, z_j\rangle$)

$$\limsup_k |\langle x_k - x, z\rangle| \leq C||z - z_j||.$$

Da j beliebig ist, können wir für $j \in I$ zum Grenwert übergehen, damit gilt wegen $z_j \to z$

$$\limsup_k |\langle x_k - x, z\rangle| = 0,$$

d.h.

$$\langle x_k, z\rangle \to \langle x, z\rangle \quad \forall\, z \in Z.$$

Also konvergiert x_k schwach$*$ gegen x. □

Die schwach-* Topologie ist essentiell notwendig um die Kompaktheit zu erhalten, wie man leicht an Beispielen von Konvergenz gegen δ-Distributionen im Raum $L^1(\Omega)$ sieht. Für die Variationsrechnung ist dies interessant, falls

$$G(x, p, q) \geq c(|p|^r + |q|^r)$$

für ein $r \geq 1$. Dann folgt

$$\int_\Omega G(x, u(x), \nabla u(x))\ dx \geq c \int_\Omega (|u(x)|^r + |\nabla u(x)|^r)\ dx = c||u||^r_{W^{1,r}(\Omega)},$$

und damit gilt auf dem α-Sub Level Set

$$||u||_{W^{1,r}(\Omega)} \leq \left(\frac{\alpha}{c}\right)^{1/r}.$$

Konsequenterweise erhalten wir die Beschränktheit der Norm im Sobolevraum $W^{1,r}(\Omega)$ und können über den Satz von Banach-Alaoglu für $r > 1$ schwache Kompaktheit folgern. Komplizerter und Interessanter ist der Fall $r = 1$, den wir noch genauer diskutieren werden.

Umgekehrt ist die zu erzielende Kompaktheit natürlich auch eine Forderung an die Wahl der Regularisierung R. Wenn D nichtnegativ ist, gilt im Fall von (4.1) $J \geq \alpha R$, also sollte für positives α das Regularisierungsfunktional R die Kompaktheit liefern. Dies ist einer Regularisierung mit Potenzen Normen in Banachräumen mit schwach-* Topologie automatisch gegeben. Sind D und R unterhalbstetig und die Niveaumengen von R kompakt, dann kann man analog zum Existenzbeweis auch die Konvergenz der Lösung zeigen, wenn die Datenstörung gegen Null geht. Dazu verwenden wir folgendes Setup wie im letzten Kapitel: u_* sei die exakte Lösung zu Daten f_*, d.h. es gilt $D(u_*, f_*) = 0$. Konvergenz der Störung gegen Null bedeutet wir haben Daten f_k mit $D(u_*, f_k) \to 0$. Sei nun u_k das Minimum von

$$J_k(u) = D(u, f_k) + \alpha_k R(u),$$

dann folgt

$$D(u_k, f_k) + \alpha_k R(u_k) \leq D(u*, f_k) + \alpha_k R(u_*).$$

Insbesondere ist

$$R(u_k) \leq \frac{D(u_*, f_k)}{\alpha_k} + R(u_*).$$

Wir sehen also dass wir die Folge α_k so wählen müssen, dass der Quotient auf der rechten Seite beschränkt bleibt, d.h. α_k darf nicht schneller gegen Null konvergieren als die Störung $D(u_*, f_k)$. Ist dies der Fall dann liegt u_k in einer kompakten Menge und hat eine konvergente Teilfolge, die wir wiederum u_k nennen. Ist nun D unterhalbstetig in beiden Komponenten, so folgt für den Grenzwert $\hat{u}$

$$D(\hat{u}, f_*) \leq \liminf_k D(u_k, f_k) \leq \lim_k \left(D(u_*, f_k) + \alpha_k R(u_*)\right) = 0.$$

Ist die Lösung vn $D(u, f_*) = 0$ eindeutig, so folgt direkt $\hat{u} = u_*$. Ist die Lösung nicht eindeutig, dann konvergiert u_k gegen irgendeine Lösung. Gilt nun sogar $\frac{D(u_*, f_k)}{\alpha_k} \to 0$, dann kann man

$$R(\hat{u}) \leq \liminf_k R(u_k) \leq \lim_k \left(\frac{D(u_*, f_k)}{\alpha_k} + R(u_*)\right) = R(u_*)$$

benutzen um zu zeigen, dass $\hat{u}$ eine Lösung ist, die R minimiert unter allen u mit $D(u, f_*) = 0$.

Die Kompaktheit in der schwach-* Topologie macht dann auch die Untersuchung der Unterhalbstetigkeit in dieser Topologie nötig, was oft kein einfaches Unterfangen ist. In der Variationsrechnung benötigt man dabei meist die Konvexität der Funktion G bezüglich der letzten Variable (und noch weitere Bedingungen). Für das Funktional J ergibt sich aus dieser Bedingung eine konvexe Abhängigkeit von der zweiten Variable (dem Gradienten), eine analoge Bedingung wie bei Diffusionsfiltern, wo nichtkonvexe Energien auf instabile Rückwärtsdiffusion führen. durchführen.

5.2 Konvexität und Dualität

In diesem Kapitel werden wir uns mit Konvexität als eine der nützlichsten Eigenschaften von Funktionalen auseinandersetzen. Wir werden in diesem Rahmen auch Dualität kennenlernen, die besonders bei nichtdifferenzierbaren konvexen Funktionalen nützlich ist. Wir beginnen mit einer Definition: Ein Funktional $J : X \to \mathbb{R} \cup \{\infty\}$ heisst konvex, falls

$$J(\beta u + (1-\beta)v) \leq \beta J(u) + (1-\beta)J(v), \qquad \forall\, u, v \in \mathcal{X}, \forall\, \beta \in [0,1], \tag{5.3}$$

Für die Eindeutigkeit eines Minimums ist die strikte Konvexität eine essentielle Eigenschaft. Ein Funktional $J : X \to \mathbb{R} \cup \{\infty\}$ heisst strikt konvex, falls

$$J(\beta u + (1-\beta)v) < \beta J(u) + (1-\beta)J(v), \qquad \forall\, u \neq v \in \mathcal{X}, \forall\, \beta \in (0,1). \tag{5.4}$$

Theorem 5.2.1. *Sei $J : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ strikt konvex, dann existiert höchstens ein globales Minimum.*

Beweis. Seien $u \neq v$ zwei globale Minima von J. Dann folgt für $\beta \in (0,1)$

$$J(\beta u + (1-\beta)v) < \beta J(u) + (1-\beta)J(v) = \inf J,$$

ein Widerspruch. □

Ein wichtiges Konzept in der Analyse von Variationsproblemen ist die konvex Konjugierte bzw. Legendre-Transformation. Für ein Funktional $J : X \to \mathbb{R} \cup \{\infty\}$ ist die Konjugierte

$$J^*(v) = \sup_{v \in X^*} \langle v, u \rangle - J(u), \tag{5.5}$$

wobei $\langle v, u \rangle$ das Dualitätsprodukt bezeichnet, d.h. die Anwendung des linearen Funktionals v auf u. Auch wenn J beliebig ist, dann ist die Legendre-Transformation konvex:

Proposition 5.2.2. *Sei $J : X \to \mathbb{R} \cup \{\infty\}$ ein Funktional auf einem Banachraum. Dann ist $J^* : X^* \to \mathbb{R} \cup \{\infty\}$ konvex.*

Beweis. Seien $v, w \in X^*$ und $\beta \in (0,1)$, dann gilt

$$\begin{aligned} J(\beta v + (1-\beta)w) &= \sup_{u \in X} (\beta\langle v,u\rangle + (1-\beta)\langle w,u\rangle - J(u)) \\ &= \sup_{u \in X} (\beta(\langle v,u\rangle - J(u)) + (1-\beta)(\langle w,u\rangle - J(u))) \\ &\leq \beta \sup_{u \in X}(\langle v,u\rangle - J(u)) + (1-\beta)\sup_{u\in X}(\langle w,u\rangle - J(u)). \\ &= \beta J^*(v) + (1-\beta)J^*(w). \end{aligned}$$

□

5.2.1 Konvexe Einhüllende

Die Definition der konvex Konjugierten kann man natürlich wiederholen und auch ein bikonjugiertes Funktional $J^{**} : X^{**} \to \mathbb{R} \cup \{\infty\}$ definieren. Im Fall eines reflexiven Banachraums ist J^{**} auf X definiert, also kann man J^{**} mit J vergleichen. Wie oft bei zweifacher Dualisierung gelangt man zum Ausgang zurück, jedoch nur wenn J konvex ist (man beachte dass J^{**} die konvex Konjugierte von J^* ist und damit konvex). Genauer gilt folgendes Resultat:

Proposition 5.2.3. *Sei $J : X \to \mathbb{R} \cup \{\infty\}$ ein Funktional auf einem Banachraum. Dann ist $J^{**} : X \to \mathbb{R} \cup \{\infty\}$ die konvex einhüllende, d.h. das grösste konvexe Funktional das unter J liegt:*

$$J^{**}(u) \leq J(u) \qquad \forall\ u \in X,$$

und für alle Funktionale H mit dieser Eigenschaft:

$$J^{**}(u) \geq H(u) \qquad \forall\ u \in X.$$

Beweis. Wir beginnen mit der Berechnung von J^{**}

$$\begin{aligned} J^{**}(u) &= \sup_{v \in X^*} (\langle v,u\rangle - J^*(v)) \\ &= \sup_{v \in X^*} (\langle v,u\rangle - \sup_{w \in X}(\langle v,w\rangle - J(w))) \\ &= J(u) + \sup_{v \in X^*} \inf_{w \in X} (\langle v, u-w\rangle + J(w) - J(u)) \end{aligned}$$

Nun wissen wir für jedes $v \in X^*$

$$\inf_{w \in X}(\langle v, u-w\rangle + J(w) - J(u)) \leq \langle v, u-u\rangle + J(u) - J(u) = 0,$$

also gilt immer $J^{**}(u) \leq J(u)$.

Nehmen wir nun an, das J konvex ist, dann folgt dass auch die Niveaumengen von J konvex sind. Insbesondere gilt dies für die Menge ($\epsilon > 0$)

$$M_\epsilon = \{(w, J(w)) \in X \mid J(w) \leq J(u) - \epsilon\}. \tag{5.6}$$

Da $(u, J(u)) \notin M_\epsilon$ gibt es nach dem konvexen Trennungssatz (einer Folgerung aus dem Satz von Hahn-Banach) ein lineares Funktional, das M_ϵ und den Punkt $(u, J(u))$ trennt. D.h. es existiert $\gamma > 0$ und $v \in X^*$ sodass

$$\gamma J(u) - \langle v, u\rangle \leq \gamma J(w) - \langle v, w\rangle.$$

Dividieren wir durch γ und definieren v zu $\frac{v}{\gamma}$ um, so folgt

$$\langle v, u - w\rangle + J(w) - J(u) \geq 0.$$

Ist nun $J(w) \geq J(u)$, so folgt mit $v = 0$, dass

$$\langle 0, u - w\rangle + J(w) - J(u) \geq 0$$

ist. Insbesondere existiert dann für jedes w ein v sodass

$$\langle v, u - w\rangle + J(w) - J(u) \geq 0$$

und damit ist

$$\sup_{v \in X^*} \inf_{w \in X} (\langle v, u - w\rangle + J(w) - J(u)) \geq 0,$$

also

Nun kann man leicht zeigen, dass für $H \leq J$ immer auch $H^{**} \leq J^{**}$ gilt. Damit folgt für jedes konvexe $H \leq J$ auch

$$H(u) = H^{**}(u) \leq J^{**}(u).$$

□

5.2.2 Unterhalbstetigkeit

Ein grosser Vorteil von konvex konjugierten ist die Unterhalbstetigkeit in der schwach-* Topologie. Wir beweisen dazu folgendes wichtige Resultat:

Theorem 5.2.4. *Sei Z ein Banachraum und $H : Z \to \mathbb{R} \cup \infty$. Dann ist $J = H^*$ ein schwach-* folgenunterhalbstetiges Funktional auf $X = Z^*$.*

Beweis. Sei $u_k \rightharpoonup^* u$. Wir wählen $z^\epsilon \in Z$ mit

$$J(u) - \epsilon \leq \langle u, z^\epsilon \rangle - H(z^\epsilon),$$

was wegen der Definition von J als Supremum für jedes $\epsilon > 0$ möglich ist. Damit folgt für ϵ beliebig aber fix aus der schwach-* Konvergenz

$$\begin{aligned} J(u) - \epsilon &\leq \langle u, z^\epsilon \rangle - H(z^\epsilon) \\ &= \lim_k \left(\langle u_k, z^\epsilon \rangle - H(z^\epsilon) \right) \\ &\leq \liminf_k \sup_z \left(\langle u_k, z \rangle - H(z) \right) \\ &= \liminf_k J(u_k). \end{aligned}$$

In der Ungleichung

$$J(u) - \epsilon \leq \liminf_k J(u_k)$$

kann man nun ϵ gegen Null gehen lassen und erhält die Folgenunterhalbstetigkeit. □

5.2.3 Fenchel-Dualität

Im Folgenden werden wir ein wichtiges Dualitätskonzept, nämlich jenes von Fenchel kennenlernen, das besonders interessant für Funktionale ist, die aus eine Summe der Form

$$J(u) = F(Au) + G(u) \tag{5.7}$$

bestehen, wobei A ein beschränkter linearer Operator ist. Grundlage der Dualitätsbetrachtung ist die Beobachtung

$$H(u) + H^*(v) \geq \langle v, u \rangle, \tag{5.8}$$

die ja sofort aus der Definition des dualen Funktionals folgt, Benutzt man nun $H = F$ mit Argument Au bzw. $H = G$ mit Argument $-A^*v$ für das duale Funktional, so folgt

$$\begin{aligned} F(Au) + F^*(v) &\geq \langle v, Au \rangle, \\ G(u) + G^*(-A^*v) &\geq -\langle A^*v, u \rangle. \end{aligned}$$

Addiert man diese beiden Ungleichungen und verwendet die Definition des adjungierten Operators, so folgt

$$F(Au) + G(u) + F^*(v) + G^*(-Av) \geq 0 \tag{5.9}$$

für alle u und v und damit auch

$$\inf_u\{F(Au)+G(u)\}+\inf_v\{F^*(v)+G^*(-Av)\}\geq 0 \tag{5.10}$$

Findet man also ein $\hat{u}$ und ein $\hat{v}$, sodass die obige Summe Null ergibt, dann weiss man sofort dass man eine Lösung des primalen Problems

$$J(u)=F(Au)+G(u)\to\min_u \tag{5.11}$$

als auch des sogenannten dualen Problems

$$F^*(v)+G^*(-A^*v)\to\min_v \tag{5.12}$$

gefunden hat. Für konvexe Funktionale F und G kann man unter gewissen Regularitätsbedingungen tatsächlich beweisen, dass die Summe der Infima (Minima) Null ergibt. Wie wir später sehen werden gibt es auch einen direkten Zusammenhang zwischen der Lösung des primalen und des dualen Problems. Dies kann man teilweise ausnutzen, wenn das direkte Problem leichter zu lösen ist (etwa auch in der Numerik) oder um gewisse Eigenschaften von Lösungen nachzuweisen.

5.3 Ableitungen / Subgradienten

Neben der Existenz und Eindeutigkeit sind auch Optimalitätsbedingungen für Minima basierend auf Ableitungen interessant. Da wir hier nur konvexe Funktionale betrachten werden, bei denen stationäre Punkte immer globale Minima sind, genügt die Betrachtung erster Ableitungen. Das einfachste Konzept um eine Ableitung eines Funktionals J auf einem linearen normierten Raum zu definieren ist die *Richtungsableitung*

$$dJ(u;v):=\lim_{t\downarrow 0}\frac{J(u+tv)-J(u)}{t}. \tag{5.13}$$

Man nennt das Funktional J differenzierbar in die Richtung v, falls der obige Grenzwert existiert und endlich ist. Die Sammlung dieser Richtungsableitungen $dJ(u;.)$ nennt man Gateaux-Ableitung. Man beachte, dass durch die einseitige Definition der Richtungsableitung nicht notwendigerweise $dJ(u;v)=dJ(u;-v)$ gilt. Damit ist z.B. auch die Betragsfunktion im Punkt 0 Gateaux-differenzierbar.

Für gegebenes u ist $dJ(u;\cdot)$ ein Funktional der zweiten Variable, bei klassischer Differenzierbarkeit sollte es ein lineares sein, das die lokale Variation von J beschreibt. Existiert so ein lineares und beschränktes Funktional $J'(u)\in\mathcal{X}^*$, sodass

$$J'(u)v=dJ(u;v)\qquad\forall v\in\mathcal{X}$$

und
$$\frac{||J(u+v) - J(u) - J'(u)v||}{||v||} \to 0 \qquad \text{für } ||v|| \to 0,$$
dann heisst J Frechet-differenzierbar in u und $J'(u)$ bezeichnet man als die Frechet-Ableitung. In vielen Fällen (z.b. für $\mathcal{X} = L^p(\Omega)$) ist man auch an einer Darstellung des Funktionals $J'(u)$ interessiert (z.B. in $L^{p'}(\Omega)$), man nennt diese Darstellung dann den Gradienten.

Für differenzierbare Funktionale J lassen sich sehr einfach analoge Optimalitätsbedingungen wie in $\mathbb{R}^n$ herleiten. Ist J Gateaux-differenzierbar, so muss für ein Minimum $\hat{u}$ gelten:
$$dJ(\hat{u}; v) \geq 0 \qquad \forall\, v \in \mathcal{X}.$$
Für ein Frechet-differenzierbares Funktional folgt (wegen $dJ(u;v) = dJ(u;-v)$) die klassische Optimalitätsbedingung erster Ordnung $J'(\hat{u}) = 0$.

Wir werden im folgenden häufig auch konvexe Funktionale betrachten, die nicht Frechet-differenzierbar sind. Besonders nützlich sind dabei sogenannte Subgradienten, die als Tangenten, die unter dem Funktional liegen interpretiert werden können. Das Subdifferential eines konvexen Funktionals J ist gegeben durch
$$\partial J(u) := \{p \in \mathcal{X}^* \mid J(u) + \langle p, v-u \rangle \leq J(v)\} \qquad \forall\, v \in \mathcal{X}. \tag{5.14}$$
Die einzelnen Elemente des Subdifferentials nennt man Subgradienten. Ist J Frechet-differenzierbar, so erhält man eine "Ubereinstimmung mit dem klassischen Begriff, da $\partial J(u) = \{J'(u)\}$. Für nicht differenzierbare Funktionale ist das Subdifferential im allgemeinen wirklich mehrwertig, z.B. gilt für die Betragsfunktion $J(u) = |u|$ im nichtdifferenzierbaren Punkt $J'(u) = [-1, 1]$.

Um eine Optimalitätsbedingung zu erhalten, können wir Subgradienten verwenden. Es gilt folgendes Resultat

Theorem 5.3.1. *Ein Element $u \in \mathcal{X}$ ist genau dann Minimum des konvexen Funktionals J, falls $0 \in \partial J(u)$.*

Beweis. Sei $0 \in \partial J(u)$, dann gilt nach Definition des Subgradienten
$$0 = \langle 0, v-u \rangle \leq J(v) - J(u) \qquad \forall\, v \in \mathcal{X}$$
und damit ist u ein globales Minimum von v. Sei umgekehrt $0 \notin \partial J(u)$, dann existiert zumindest ein $v \in \mathcal{X}$ mit
$$J(v) - J(u) < \langle 0, v-u \rangle = 0$$

und damit kann u kein Minimum sein. □

Zum Abschluss geben wir auch noch die Konsistenz von Subdifferentialen mit Frechet-Ableitungen an

Theorem 5.3.2. *Sei J konvex und Frechet-differentierbar in u. Dann gilt*

$$\partial J(u) = \{J'(u)\}.$$

5.4 Funktionen beschränkter Variation und das ROF Modell

In diesem Abschnitt werden wir uns zunächst näher mit der Frage beschäftigen, welche Art von Regularisierungsterm (F) bzw. welchen Funktionenraum wir für das Entrauschen von Bildern sinnvollerweise benutzen können. Die Lebesgue-Räume $L^p(\Omega)$ sollten alle sinnvollen Bilder enthalten, da aber auch Rauschen in diesen Räumen liegen kann (insbesondere das Gauss'sche Rauschen in unserem Modell in $L^2(\Omega)$) sind sie aber zu gross gewählt - man kann in $L^p(\Omega)$ nicht zwischen Signal und Rauschen unterscheiden. Besonders naheliegend ist deshalb die Betrachtung der Sobolev-Räume $W^{1,p}(\Omega)$, $p \geq 1$. Diese sind Teilräume der Lebesgue-Räume $L^p(\Omega)$, die auch noch einen p-integrierbaren Gradienten haben sollen und damit Oszillationen (Rauschen) zu einem signifikant höheren Wert der Norm führen. Andererseits könnte die Wahl eines solchen Sobolev-Raums zu einschränkend sein, um interessante Bilder (mit Kanten) zuzulassen. Dies ist für $p > 1$ auch der Fall, wie wir aus den folgenden Resultaten sehen.

Lemma 5.4.1. *Sei $u \in W^{1,p}(\Omega)$, $p > 1$ wobei $\Omega \subset \mathbb{R}^1$ ein Intervall sei. Dann ist u stetig.*

Beweis. Sei zunächst $u \in C^1(\Omega) \subset W^{1,p}(\Omega)$. Dann folgt

$$|u(y) - u(x)| = |\int_x^y u'(z)\ dz| \leq \left(\int_x^y |u'(z)|^p\ dz\right)^{1/p} |x-y|^{1/p'},$$

wobei $1/p + 1/p' = 1$ gilt. Also folgt

$$|u(y) - u(x)| \leq ||u||_{W^{1,p}(\Omega)} |x-y|^{1/p'}.$$

Für beliebieges $u \in W^{1,p}(\Omega)$ können wir nun die Dichtheit von $C^1(\Omega)$ in $W^{1,p}(\Omega)$ verwenden, d.h. es existiert eine Folge $u_n \in C^1(\Omega)$ die gegen u in der Norm von $W^{1,p}(\Omega)$ und auch punktweise konvergiert. Damit folgt fast überall

$$|u(y)-u(x)| = \lim |u_n(x)-u_n(y)| \leq \lim ||u_n||_{W^{1,p}(\Omega)}|x-y|^{1/p'} = ||u||_{W^{1,p}(\Omega)}|x-y|^{1/p'}.$$

Damit ist u sogar Hölder-stetig mit Exponent $1/p'$. □

Lemma 5.4.2. *Sei $D \subset \Omega$ ein Gebiet mit C^1-Rand. Dann ist die Funktion*

$$u(x) = \begin{matrix} 1 & \text{für } x \in D \\ 0 & \text{sonst} \end{matrix}$$

nicht in $W^{1,p}(\Omega)$ für $p \geq 1$.

Beweis. Der distributionelle Gradient einer Funktion u ist definiert durch das lineare Funktional

$$\langle \nabla u, \varphi \rangle = -\int_\Omega u(\nabla \cdot \varphi)\, dx \qquad \forall \varphi \in C_0^\infty(\Omega; \mathbb{R}^d).$$

Für eine Funktion der obigen Gestalt folgt dann mit dem Gauss'schen Satz

$$\langle \nabla u, \varphi \rangle = -\int_D \nabla \cdot \varphi\, dx = -\int_{\partial D} \varphi \cdot n\, d\sigma.$$

Da dieses lineare Funktional auf ∂D, d.h. einer Nullmenge des Lebesgue-Maßes konzentriert ist, kann es für kein $p' \in [1, \infty]$ zu einem linearen Funktional in $L^{p'}(\Omega, \mathbb{R}^d)$ erweitert werden. Damit liegt der distributionelle Gradient auch nicht in $L^p(\Omega, \mathbb{R}^d) \subset L^{p'}(\Omega, \mathbb{R}^d)^*$. □

Als Folgerung sehen wir also, dass $W^{1,p}$ für $p > 1$ sicher nicht in Frage kommt, da Unstetigkeiten (Kanten) nicht zugelassen werden. Auch der Fall $p = 1$ würde nach dem obigen Argument stückweise konstante Funktionen nicht zulassen. Darüber hinaus ergibt sich in $W^{1,1}(\Omega)$ eine funktionalanalytische Schwierigkeit, da dieser Raum (analog zu $L^1(\Omega)$) nicht der Dualraum eines Banachraumes ist. Man kann also keine schwach-* Topologie verwenden und damit auch den Satz von Banach-Alaoglu nicht anwenden. In beiden Fällen liegt das Problem darin, dass gewisse Maße wie zum Beispiel Dirac δ-Distributionen nicht als Gradient einer Funktion $W^{1,1}(\Omega)$ zugelassen werden, aber durch $W^{1,1}$-Funktionen mit beschränkter Norm approximiert werden können.

Um einen sinnvolleren Funktionenraum für Bilder zu erhalten, muss $W^{1,1}(\Omega)$ noch einmal vergrössert werden. Dabei geht man analog vor wie beim "Ubergang von $L^1(\Omega)$ zum Raum der Radon-Maße. Wir definieren deshalb den Raum $BV(\Omega)$ von Funktionen beschränkter (totaler) Variation

$$BV(\Omega) = \{u \in L^1(\Omega) \mid |u|_{BV} < \infty\}, \tag{5.15}$$

wobei die totale Variation gegeben ist durch

$$|u|_{BV} := \sup_{\varphi \in C_0^\infty(\Omega;\mathbb{R}^d), ||\varphi||_\infty \leq 1} \int_\Omega u \nabla \cdot \varphi\, dx. \tag{5.16}$$

Die Norm in $BV(\Omega)$ ist gegeben durch

$$||u||_{BV} := ||u||_{L^1} + |u|_{BV}. \tag{5.17}$$

Die Wahl des Raums $C_0^\infty(\Omega;\mathbb{R}^d)$ in der Definition ist nicht essentiell. Wegen der Dichte dieses Raums in anderen Räumen erhalten wir das selbe Supremum in jedem $C_0^k(\Omega;\mathbb{R}^d)$ für $k \geq 0$. Wir werden im folgenden auch diese unterschiedlichen Räume in der Definition je nach Notwendigkeit einsetzen. Für Funktionen in $u \in W^{1,1}(\Omega)$ rechnet man leicht nach, dass $||u||_{BV} = ||u||_{W^{1,1}}$ gilt. Damit enthält $BV(\Omega)$ auf jeden Fall $W^{1,1}$. Man sieht auch leicht, dass $BV(\Omega)$ auch unstetige Funktionen enthält. Für die stückweise konstante Funktion u aus Lemma 5.4.2 gilt

$$\begin{aligned} |u|_{BV} &= \sup_{\varphi \in C_0^\infty(\Omega;\mathbb{R}^d), ||\varphi||_\infty \leq 1} \int_D \nabla \cdot \varphi \, dx \\ &= \sup_{\varphi \in C_0^\infty(\Omega;\mathbb{R}^d), ||\varphi||_\infty \leq 1} \int_{\partial D} \varphi \cdot n \, d\sigma \\ &= \int_{\partial D} d\sigma < \infty. \end{aligned}$$

Sobald also ∂D ein endliches $d-1$-dimensionales Hausdorff-Maß (bei Kurven also einfach endliche Länge) hat, ist die totale Variation der stückweise konstanten Funktion endlich. Wir sehen auch dass die totale Variation in diesem Fall gleich der Länge der Kurve ist, d.h. eine Minimierung (bzw. zumindest Verkleinerung) der totalen Variation sollte auch die Kurven glätten.

Nach dieser Argumentation ist es naheliegend geglättete Funktionen im Raum beschränkter Variation zu suchen, bzw. die totale Variation sogar als Regularisierungsfunktional zu verwenden. Damit erhält man das Rudin-Osher-Fatemi (ROF) Modell zum Entrauschen (cf. [**?**]), das in der Minimierung

$$J(u) = \frac{\lambda}{2} \int_\Omega (u-f)^2 \, dx + |u|_{BV} \to \min_{u \in BV(\Omega)} \tag{5.18}$$

besteht. Dieses Modell und Variationen davon sind heute eines der populärsten Forschungsgebiete in der Bildverarbeitung. Wie wir sehen werden, erhält man daraus (im Gegensatz zu vielen anderen Variationsmethoden und Filtern) scharfe Kanten bei der Rekonstruktion und auch recht klare geometrische Interpretationen bezüglich der Isokonturen.

Wir beginnen die Analyse des Modells mit der Existenz eines Minimums. Wir werden dazu im weiteren immer annehmen, dass Ω ein beschränktes Gebiet mit stückweise glattem Rand ist. Durch die Beschränktheit gilt dann folgende Abschätzung zwischen der L^2 und der L^1 Norm (als Folgerung aus Cauchy-Schwarz Ungleichung)

$$||u||_{L^1} \leq \sqrt{|\Omega|}||u||_{L^2}. \tag{5.19}$$

Diese Abschätzung ist nützlich um eine Schranke für die Norm in $BV(\Omega)$ auf den Sub-Level Sets von J zu erhalten:

Lemma 5.4.3. *Sei $\lambda > 0$ und $J(u) \leq \alpha < \infty$, dann gilt*

$$||u||_{BV(\Omega)} \leq \alpha + \sqrt{|\Omega|}\left(\frac{2\alpha}{\lambda} + ||f||_{L^2(\Omega)}\right).$$

Beweis. Wegen der Nichtnegativität des ersten Terms in J folgt sofort $|u|_{BV} \leq \alpha$. Wegen der Nichtnegativität der totalen Variation folgt

$$||u-f||_{L^2} \leq \frac{2\alpha}{\lambda},$$

und mit der Dreiecksungleichung sowie der obigen Abschätzung für die L^1-Norm

$$||u||_{L^1} \leq \sqrt{|\Omega|}\left(\frac{2\alpha}{\lambda} + ||f||_{L^2(\Omega)}\right).$$

Durch Addition der Abschätzungen für $||u||_{L^1}$ und $|u|_{BV}$ folgt die Aussage. □.

Die Abschätzung der Norm ist der erste Schritt zur Kompaktheit, wir benötigen nun nur noch eine schwach-* Topologie (um den Satz von Banach-Alaoglu anwenden zu können). Dazu betrachten wir den Raum

$$\mathcal{Y} := \{(c, \nabla \cdot \varphi) \mid c \in \mathbb{R}, \varphi \in C(\overline{\Omega}; \mathbb{R}^d), \varphi|_{\partial\Omega} \cdot n = 0\}$$

mit der Norm

$$||(c,\psi)|| = \max\{|c|, \inf_{\varphi, \nabla\cdot\varphi=\psi} ||\varphi||_\infty\}.$$

Der Dualraum dieses Raums sind alle stetigen linearen Funktionale der Form

$$\ell_1(c) + \ell_2(\psi)$$

und da ℓ_1 ein stetiges lineares Funktional in $\mathbb{R}$ ist folgt $\ell_1(c) = \gamma c$ für ein $\gamma \in \mathbb{R}$. Für p hinreichend gross (abhängig von der Dimension) lässt sich

$$W_0^{1,p}(\Omega; \mathbb{R}^d) = \{v \in L^p(\Omega; \mathbb{R}^d) \mid \nabla v \in L^p(\Omega; \mathbb{R}^{d\times d}, v|_{\partial\Omega} \cdot n = 0\}$$

stetig in $C_0(\overline{\Omega}; \mathbb{R}^d)$ einbetten. Da für $v \in W_0^{1,p}(\Omega; \mathbb{R}^d)$ die Funktion $\psi = \nabla \cdot v$ im Lebesgue-Raum

$$L^p_\diamond(\Omega) := \{u \in L^p(\Omega) \mid \int_\Omega u \, dx = 0\}$$

liegt, ist ℓ_2 ein stetiges lineares Funktional insbesondere auf diesem Teilraum. Wegen des Darstellungssatzes für lineare Funktionale in Lebesgue-Räumen folgt dann

$$\ell_2(\psi) = \int_\Omega w\psi \, dx$$

für ein $w \in L_{\diamond}^{p'}(\Omega)$, wobei $p' > 1$ gilt. Definieren wir nun $u = \gamma + w$, dann ist u eine Funktion in $L^1(\Omega)$. Wegen der notwendigen Stetigkeit des linearen Funktionals $\ell_1 + \ell_2$ folgt die Bedingung

$$\sup_{c \in \mathbb{R}, \nabla\cdot\varphi, \max\{c, ||\varphi||_\infty\}=1} \left(\overline{u}c + \int_\Omega u \nabla \cdot \varphi \; dx \right) < \infty,$$

wobei $\overline{u} = \frac{1}{|\Omega|} \int_\Omega u \; dx$. Diese Bedingung kann leicht zu

$$|\overline{u}| + |u|_{BV} < \infty$$

umgerechnet werden. Da man zeigen kann, dass $|\overline{u}| + |u|_{BV}$ eine äquivalente Norm auf $BV(\Omega)$ definiert, können wir $BV(\Omega)$ also mit dem Dualraum von $\mathcal{Y}$ identifizieren. Als direkte Konsequenz daraus erhalten wir eine schwach-* Topologie auf $BV(\Omega)$.

Nachdem wir die Kompaktheit geklärt haben, untersuchen wir uns nun noch die schwach-*-Folgenunterhalbstetigkeit. Dabei ist wieder die duale Definition der totalen Variation sehr nützlich. Sei $\varphi_k \in C_0^\infty(\Omega; \mathbb{R}^d)$ eine Folge mit $||\varphi||_\infty \leq 1$, sodass

$$|u|_{BV} = \lim_k \int_\Omega u \nabla \cdot \varphi_k \; dx.$$

Dann gilt für eine schwach-* konvergente Folge $u_n \rightharpoonup^* u$

$$\begin{aligned}
\int_\Omega u \nabla \cdot \varphi_k \; dx &= \lim_n \int_\Omega u_n \nabla \cdot \varphi_k \; dx \\
&= \liminf_n \int_\Omega u_n \nabla \cdot \varphi_k \; dx \\
&\leq \liminf_n \sup_{\varphi \in C_0^\infty(\Omega;\mathbb{R}^d), ||\varphi||_\infty \leq 1} \int_\Omega u_n \nabla \cdot \varphi \; dx \\
&= \liminf_n |u_n|_{TV}.
\end{aligned}$$

Damit können wir auch den Limes über k bilden und erhalten

$$|u|_{BV} \leq \liminf_n |u_n|_{BV}.$$

Analog kann die Folgenunterhalbstetigkeit des ersten Terms in J nachgewiesen, und damit sind die Bedingungen des Fundamentalsatzes der Optimierung erfüllt. Wir erhalten also:

Theorem 5.4.1. *Sei $f \in L^2(\Omega)$ und $\lambda > 0$. Dann existiert ein Minimum von* (**??**).

Die totale Variation ist ein konvexes Funktional, da

$$\sup_{\varphi \in C_0^\infty(\Omega;\mathbb{R}^d), ||\varphi||_\infty \leq 1} \int_\Omega (\alpha u + (1-\alpha)v) \nabla \cdot \varphi \; dx \leq \alpha \sup_{\cdots} \int_\Omega u \nabla \cdot \varphi \; dx + (1-\alpha) \sup_{\cdots} \int_\Omega v \nabla \cdot \varphi \; dx.$$

Das quadratische Funktional $\frac{\lambda}{2} \int_\Omega (u - f)^2$ ist sogar strikt konvex, wie man leicht nachrechnet. Daher ist auch J strikt konvex und wir erhalten damit:

Theorem 5.4.2. *Sei $f \in L^2(\Omega)$ und $\lambda > 0$. Dann ist das Minimum von (**??**) eindeutig.*

Das Subdifferential des Funktionals J in (**??**) kann als Summe der Ableitung des ersten (differenzierbaren) Terms und des Subdifferentials der totalen Variation berechnet werden (diese Formel ist nicht trivial, siehe [**?**]), d.h.

$$\partial J(u) = \{\lambda(u - f) + p\}, \qquad p \in \partial |u|_{BV}.$$

Die Subgradienten der totalen Variation sind schwieriger zu charakterisieren, man kann aber zeigen, dass für $p \in \partial |u|_{BV}$ als notwendige Bedingung gilt

$$p = \nabla \cdot g, \quad g \in L^\infty(\Omega; \mathbb{R}^d), ||g||_\infty \leq 1$$

und $\langle p, u \rangle = |u|_{TV}$. Die letzte Formel ist eine einfache Folgerung aus der positiven Homogenität der totalen Variation ($|cu|_{BV} = |c|\ |u|_{BV}$ für alle $c \in \mathbb{R}$): wählen wir in der Definition des Subgradienten $v = 2u$ so gilt

$$\langle p, u \rangle = \langle p, v - u \rangle \leq |v|_{BV} - |u|_{BV} = |2u|_{BV} - |u|_{BV} = |u|_{BV}.$$

Man kann auch leicht zeigen, dass in Gebieten mit $|\nabla u| \neq 0$ lokal

$$p(x) = \nabla \cdot \left(\frac{\nabla u(x)}{|\nabla u(x)|} \right)$$

gelten muss, d.h. der Subgradient ist gleich der (mittleren) Krümmung der Isokonturen. Auf Kanten kann man ebenfalls zeigen, dass p mit der (mittleren) Krümmung der Kante übereinstimmen muss.

Die Optimalitätsbedingung für das Minimum ist dann ja $0 \in \partial J(u)$, d.h. es existiert ein Subgradient $p \in \partial |u|_{BV}$ mit

$$\lambda(u - f) + p = 0.$$

Die Optimalitätsbedingung kann als Glättung des Subgradienten interpretiert werden. Normalerweise wäre ja $p \in BV(\Omega)^*$, einem Raum von Distributionen, und damit nicht unbedingt eine Funktion. Aus der Optimalitätsbedingung folgt aber

$$p = \lambda(f - u) \in L^2(\Omega),$$

d.h. p ist eine quadratisch integrierbare Funktion. Dies bedeutet also, dass die Isokonturen und Kanten des entrauschten Bildes eine quadratisch integrierbare (mittlere) Krümmung haben, eine recht natürliche geometrische Eigenschaft. Die Glättung des Subgradienten impliziert unter anderem, dass die Kantenmengen keine Ecken haben dürfen, da diese ja starke Singularitätenmengen der Krümmung sind.

Interessant ist auch ein Vergleich mit der linearen Methode (**??**). Dort ist der Gradient des quadratischen Regularisierungsfunktionals gegeben durch $p = -\Delta u$, die Optimalitätsbedingung impliziert dort direkt eine Glättung der Funktion und eliminiert die Kanten, während das ROF-Modell die Kanten erhält.

Wie jede Glättungsmethode hat auch das ROF-Modell einen systematischen Fehler, wie wir aus dem folgenden Resultat sehen:

Theorem 5.4.3. *Sei u das Minimum von* (**??**) *für $\lambda > 0$ dann gilt $u = f$ nur für konstante Funktionen f.*

Beweis. Für konstantes f sieht man sofort, dass $u = f$ den Wert $J(u) = 0$ liefert, und damit ein Minimum des Funktionals sein muss. Falls umgekehrt $u = f$ gelten soll, folgt aus der Optimalitätsbedingung $p = 0$ und wegen $0 = \langle p, u \rangle = |u|_{BV}$ muss u konstant sein. □

Das ROF-Modell kann also nur konstante Bilder exakt rekonstruieren, sobald kleinere Skalen dazu kommen macht es einen Fehler. In jedem Fall wird aber die gröbste Skala richtig aufgelöst, d.h. der Mittelwert von u ist immer gleich dem von f:

Theorem 5.4.4. *Sei u das Minimum von* (**??**). *Dann gilt*

$$\int_\Omega u \; dx = \int_\Omega v \; dx.$$

Beweis. Sei u das Minimum von (**??**) und $v = u - \overline{u} + \overline{f}$. Dann gilt $|v|_{BV} = |u|_{BV}$ und

$$\begin{aligned} \int_\Omega (v-f)^2 \; dx &= \int_\Omega (u-f)^2 \; dx - 2\int_\Omega (\overline{u} - \overline{f})(u-f) \; dx + \int_\Omega (\overline{u} - \overline{f})^2 \; dx \\ &= \int_\Omega (u-f)^2 \; dx - |\Omega|(\overline{u} - \overline{f})^2. \end{aligned}$$

Damit erhalten wir

$$J(u) \leq J(v) \leq J(u) - \frac{\lambda}{2}|\Omega|(\overline{u} - \overline{f})^2.$$

Diese Ungleichung kann nur für $\overline{u} - \overline{f}$ erfüllt sein. □

Um weitere Informationen über die Struktur der Lösung zu erhalten, betrachten wir nun ein duales Problem, das wir mit formalen Argumenten herleiten. Es gilt ja

$$J(u) = \frac{\lambda}{2} \int_\Omega (u-f)^2 \; dx + \sup_{g, ||g||_\infty \leq 1} \int_\Omega u \nabla \cdot g \; dx.$$

Damit können wir die Minimierung von J als Sattelpunktsproblem in u und g schreiben

$$\inf_u J(u) = \inf_u \sup_{g,||g||_\infty \leq 1} \left(\frac{\lambda}{2} \int_\Omega (u-f)^2 \, dx + \int_\Omega u \nabla \cdot g \, dx \right).$$

Unter der Annahme, dass sich inf und sup vertauschen lassen (man kann dies tatsächlich mit fortgeschrittenen Methoden der konvexen Analysis zeigen) gilt

$$inf_u J(u) = \sup_{g,||g||_\infty \leq 1} \inf_u \left(\frac{\lambda}{2} \int_\Omega (u-f)^2 \, dx + \int_\Omega u \nabla \cdot g \, dx \right)$$

und das Infimum über u für fixes g können wir leicht berechnen. Bei fixem g hat man ja ein quadratisch-lineares Funktional in u, das insbesondere differenzierbar ist. Setzen wir die Ableitung gleich Null, so folgt die Optimalitätsbedingung

$$\lambda(u-f) + \nabla \cdot g = 0 \quad \Rightarrow u = f - \frac{1}{\lambda} \nabla \cdot g. \tag{5.20}$$

Wir können nun u mit dieser Formel ausrechnen und im Sattelpunktsproblem eliminieren. Dann erhalten wir das duale Problem, das nur mehr von g abhängt

$$\inf_u J(u) = \sup_{g,||g||_\infty \leq 1} \left(-\frac{1}{2\lambda} \int_\Omega (\nabla \cdot g - \lambda f)^2 \, dx + \frac{\lambda}{2} \int_\Omega f^2 \, dx \right).$$

Da wir konstante Terme ($\frac{\lambda}{2} \int_\Omega f^2 \, dx$) vernachlässigen können und statt einem Funktional zu maximieren auch das negative Funktional minimieren können, erhalten wir als duales Problem zunächst

$$\frac{1}{2} \int_\Omega (\frac{1}{\lambda} \nabla \cdot g - f)^2 \, dx \to \min_{g,||g||_\infty \leq 1}.$$

Nun können wir noch die Umskalierung $g = \lambda q$ einsetzen und erhalten dann

$$\frac{1}{2} \int_\Omega (\nabla \cdot q - f)^2 \, dx \to \min_q, \quad \text{mit Nebenbedingung } ||q||\infty \leq \frac{1}{\lambda}. \tag{5.21}$$

Mit (5.20) erhalten wir dann das entrauschte Bild wieder als $u = f - \nabla \cdot q$.

Im eindimensionalen Fall kann man aus dem dualen Problem sofort einiges über die Struktur von u folgern. Betrachten wir also $\Omega = (0,1)$, dann ist das duale Problem

$$\frac{1}{2} \int_0^1 (q' - f)^2 \, dx \to \min_q,$$

mit der Nebenbedingung

$$-\frac{1}{\lambda} \leq q \leq \frac{1}{\lambda}.$$

Für das Minimum q muss dann gelten, dass die Ableitung in eine zulässige Richtung nichtnegativ ist, d.h.

$$\int_0^1 (q' - f)(r' - q')\ dx \geq 0 \qquad \forall r, ||r||_\infty \leq \frac{1}{\lambda}.$$

Nach partieller Integration erhalten wir

$$\int_0^1 (q'' - f')(r - q)\ dx \leq 0$$

Sei nun $I \subset (0,1)$ ein offenes Intervall auf dem die Nebenbedingung nicht aktiv ist, d.h. $-\frac{1}{\lambda} < q < \frac{1}{\lambda}$. Dann können wir eine beliebige Funktion ψ mit Träger in I geben und $r = q \pm \epsilon\psi$ erfüllt dann die Nebenbedingung für ϵ hinreichend klein. Damit folgern wir

$$\int_0^1 (q'' - f')\psi\ dx = 0.$$

Wegen der Beliebigkeit von ψ folgt $u' = (f - q')' = 0$ auf I, d.h. u ist konstant. Auf anderen Teilintervallen gilt konstant $q = \frac{1}{\lambda}$ oder $q = -\frac{1}{\lambda}$, und damit folgt $q' = 0$ und damit $u = f - q' = f$. Damit ist auf verschiedenen Teilintervall das Bild gleich den Daten f oder stückweise konstant. Der letzte Fall tritt nach dem obigen Resultat sicher auf, da $u = f$ auf ganz Ω ja unmöglich ist. Bei starkem Rauschen ist $u = f$ fast nirgends zu erwarten und man erhält typischerweise eine stückweise konstante Rekonstruktion, damit aber auch in jedem Fall scharfe Kanten. Die stückweise konstante Rekonstruktion (auch als stair-casing bezeichnet) ist vor allem in Bild (**??**) gut sichtbar.

Eine interessante Frage bei Variationsmethoden ist immer die Asymptotik $\lambda \to \infty$ bzw. $\sigma \to 0$. In diesem Fall sollte natürlich das Minimum des Funktionals gegen $\hat{f}$ konvergieren. Dies ist unter sinnvollen Annahmen an $\hat{f}$ tatsächlich der Fall, solange λ vernünftig in Abhängigkeit von σ gewählt wird. Grob sehen wir dies schon aus dem dualen Problem, da im Grenzwert $\lambda \to \infty$ die obere und untere Schranke an q gegen Null gehen. Deshalb bleibt als Grenzwert nur $q = 0$ und damit $u = f$ übrig. Für einen Beweis sind wieder die richtigen

Theorem 5.4.5. *Sei $\hat{f} \in BV(\Omega) \cap L^2(\Omega)$ und $f_n \in L^2(\Omega)$ ein verrauschtes Bild, das (**??**) mit $f = f_n$ und $\sigma = \sigma_n$ erfüllt, wobei σ_n eine gegen Null konvergierende Folge positiver Zahlen ist. Sei $\lambda_n = \lambda(\sigma_n)$ so gewählt, dass $\lambda_n \to \infty$ und $\lambda_n \sigma_n^2 \to 0$. Weiter bezeichne u_n das eindeutige Minimum von (**??**) mit $f = f_n$ und $\lambda = \lambda_n$. Dann gilt*

$$\begin{aligned} u_n &\to \hat{f} &&\text{in } L^2(\Omega) \\ u_n &\rightharpoonup^* \hat{f} &&\text{in } BV(\Omega). \end{aligned}$$

Beweis. Da u_n ein Minimum von (**??**) ist, folgt $J(u_n) \leq J(\hat{f})$, d.h.

$$\frac{\lambda_n}{2} \int_\Omega (u_n - f_n)^2 \; dx + |u_n|_{BV} \leq \frac{\lambda_n}{2} \int_\Omega (\hat{f} - f_n)^2 \; dx + |\hat{f}|_{BV}.$$

Setzen wir nun noch (**??**) ein, so folgt

$$\frac{\lambda_n}{2} \int_\Omega (u_n - f_n)^2 \; dx + |u_n|_{BV} \leq \frac{\lambda_n}{2} \sigma_n^2 + |\hat{f}|_{BV}.$$

Insbesondere ist dann

$$\int_\Omega (u_n - f_n)^2 \; dx \leq \sigma_n^2 + \frac{|\hat{f}|_{BV}}{\lambda_n},$$

was mit der Dreiecksungleichung die L^2-Abschätzung

$$||u_n - \hat{f}||_{L^2} \leq ||f_n - \hat{f}||_L^2 + \sqrt{\sigma_n^2 + \frac{|\hat{f}|_{BV}}{\lambda_n}} \leq \sigma_n + \sqrt{\sigma_n^2 + \frac{|\hat{f}|_{BV}}{\lambda_n}}$$

impliziert. Da die rechte Seite für $n \to \infty$ gegen Null konvergiert, folgt die Konvergenz von u_n gegen $\hat{f}$ in L^2.

Aus der obigen Abschätzung folgt weiter

$$u_n|_{BV} \leq \frac{\lambda_n}{2} \sigma_n^2 + |\hat{f}|_{BV} \to |\hat{f}|_{BV}.$$

Damit ist u_n gleichmässig in $BV(\Omega)$ beschränkt, und es folgt die schwach-* Konvergenz einer Teilfolge, wegen der Eindeutigkeit des Grenzwerts ebenfalls gegen $\hat{f}$. Aus der Eindeutigkeit des Grenzwerts kann man auch folgern, dass jede Teilfolge und somit die Folge selbst schwach-* gegen $\hat{f}$ konvergiert.

5.5 Entzerren und Rekonstruktion

Variationsmethoden sind sehr einfach vom Entrauschen zum Entzerren und zur Rekonstruktion zu verallgemeinern. Genauso wie in (**??**) muss nur im ersten Term im Zielfunktional u durch Ku ersetzt werden, um die Verzerrung oder die Bildmodalität, die zu den Daten führt, korrekt wiederzugeben. Bei einem Gradienten-Term in der Energie erhalten wir also

$$J(u) = \frac{\lambda}{2} \int_\Omega |Ku - f|^2 \; dx + \int_\Omega F(x, \nabla u) \; dx \tag{5.22}$$

als zu minimierendes Funktional. Damit erhält man auch eine Verallgemeinerung des ROF-Funktionals für die Entzerrung und Rekonstruktion, nämlich

$$J(u) = \frac{\lambda}{2} \int_\Omega |Ku - f|^2 \; dx + |u|_{BV}. \tag{5.23}$$

Um die Optimalitätsbedingung zu erhalten, muss die Ableitung des Funktionals

$$R(u) = \frac{\lambda}{2} \int_\Omega |Ku - f|^2 \; dx$$

berechnet werden. Die Richtungsableitung ergibt sich direkt als

$$dR(u;v) = \lambda \int_\Omega (Ku - f) Kv \; dx = \lambda \langle Ku - f, Kv \rangle_{L^2}.$$

Durch Verwendung des adjungierten Operators K^* erhalten wir

$$dR(u;v) = \lambda \langle K^*(Ku - f), v \rangle_{L^2}$$

und damit

$$R'(u) = \lambda K^*(Ku - f).$$

Es fehlt damit nur noch die Berechnung des adjungierten Operators K^*. Seien also $u, v \in L^2(\Omega)$, dann gilt

$$\begin{aligned} \langle Ku, v \rangle_{L^2} &= \int_\Omega \int_\Omega k(x,y) u(y) \; dy v(x) \; dx \\ &= \int_\Omega \int_\Omega k(x,y) v(x) \; dx u(y) \; dy \\ &= \langle u, K^* v \rangle_{L^2}, \end{aligned}$$

mit

$$(K^* v)(x) = \int_\Omega k(y,x) v(y) \; dy.$$

Die Optimalitätsbedingung ist dann

$$\lambda K^*(Ku - f) + p = 0, \qquad p \in \partial E(u), \tag{5.24}$$

wobei E das Regularisierungsfunktional ist, wie z.B. die totale Variation.

Es kann bei der Entzerrung durchaus sinnvoll sein auch Regularisierungsfunktionale zu verwenden, die nicht vom Gradienten sondern nur von u abhängen. Ein Beispiel dafür haben wir ja schon bei der Entfaltung mit Fourier-Transformation gesehen, nämlich

$$E(u) = \frac{1}{2} \int_\Omega u^2 \; dx.$$

Man berechnet leicht $E'(u) = u$ und damit ist die Optimalitätsbedingung

$$u = -\lambda K^*(Ku - f) = -\int_\Omega k(y,.)((Ku)(y) - f(y)) \; dy.$$

Die Lösung ist also proportional zu einem Integral des Residuums und damit automatisch glatt, wobei die exakte Glättung jetzt wieder vom Kern k abhängt (umso glätter k ist, umso mehr wird das Bild geglättet). Da sich dies beim obigen Fall direkter auswirkt als bei der totalen Variation (dort wird nur der Subgradient geglättet, das Bild kann unstetig bleiben !), erhält man eine fast paradoxe Situation: die Regularisierung mit einem Funktional, das nur von u abhängt kann mehr glätten als die Regularisierung mit einem Funktional, das vom Gradienten abhängt.

Auch andere Funktionale in Abhängigkeit von u sind bei der Regularisierung interessant. So verwendet man bei der Rekonstruktion von Dichtebildern gerne die logarithmische Entropie

$$E(u) = \int_\Omega (u \log u - u)\ dx,$$

die natürlich nur für positive u definiert ist. Die Optimalitätsbedingung wird in diesem Fall

$$\lambda K^*(Ku - f) + \log u = 0,$$

oder äquivalent

$$u = \exp(-\lambda K^*(Ku - f)),$$

und damit ist insbesondere die Lösung u immer positiv.

5.6 Variationsrechnung

Im Folgenden betrachten wir nun die Minimierung von Funktionalen der Form

$$J(u) = \int_\Omega F(x, u, \nabla u)\ dx. \tag{5.25}$$

Wir nehmen an, dass die messbare Funktion $F : \Omega \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ eine Bedingung der Form

$$C_1|w|^p + C_2|s|^p + C_3 \geq F(x, s, w) \geq c_1|w|^p + c_2|s|^p + c_3 \tag{5.26}$$

erfüllt, mit Konstanten $C_1, C_2, C_3, c_1, c_2 \in \mathbb{R}^+$ und $c_3 \in \mathbb{R}$. Weiters nehmen wir an, dass gilt:

$$w \mapsto F(x, s, w) \quad \text{ist konvex für alle } x \in \Omega, s \in \mathbb{R} \tag{5.27}$$

und

$$|F(x, s_1, w) - F(x, s_2, w)| \leq G(x, s_1, s_2, w)|s_1 - s_2| \quad \forall\, x \in \Omega, w \in \mathbb{R}^n, s_1 \in \mathbb{R}, s_2 \in \mathbb{R} \tag{5.28}$$

für eine positive Funktion g mit

$$G(x, s_1, s_2, w) \leq \gamma_1|w|^{p-1} + \gamma_2(|s_1|^{p-1} + |s_2|^{p-1}). \tag{5.29}$$

Diese Form der Minimierung ist etwa durch die Probleme bei der Bildregistrierung motiviert, z.B. mit

$$F(x,s,w) = \frac{1}{2}\|f_T(s) - f_R(x)|^2 + \frac{\alpha}{2}|w|^2.$$

Da hier die Abhängigkeit vom Argument s über das

Wir sehen dann, dass

$$J(u) \geq \int_\Omega (c_1|\nabla u|^p + c_2|u|^p + c_3)\ dx \geq \min\{c_1, c_2\}||u||^p_{W^{1,p}(\Omega)} - |c_3|\ |\Omega| \tag{5.30}$$

gilt. Wir erhalten also, dass für C hinreichend gross, das sub-Level Set $\mathcal{M}_C$ nicht leer und beschränkt in $W^{1,p}(\Omega)$ ist. Also ist $W^{1,p}(\Omega)$ der richtige Raum für die Analyse von (5.25), mit (5.26) sind dann die ersten beiden Bedingungen von Satz **??** erfüllt und es bleibt uns nur mehr die schwach-* Unterhalbstetigkeit nachzuprüfen.

Lemma 5.6.1. *Sei J definiert durch (5.25) und F erfülle (5.26), (5.27), (5.29) für ein $p \in (1,\infty)$.Dann ist J schwach-unterhalbstetig in $W^{1,p}(\Omega)$.*

Beweis. Es gelte $u_k \rightharpoonup *u$ in $W^{1,p}(\Omega)$. Wegen der kompakten Einbettung von $W^{1,p}(\Omega)$ in $L^p(\Omega)$ konvergiert damit $u_k \to u$ stark in $L^p(\Omega)$. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass u_k eine beschränkte Folge ist, da für Teilfolgen mit $||u_{k_\ell}||_{W^{1,p}} \to \infty$ auch $J(u_{k_\ell}) \to \infty$ folgt (wegen (5.26)). Damit ist entweder der limes inferior gleich unendlich, was die Unterhalbstetigkeit trivial macht, oder wir können diese Teilfolgen eliminieren, da der limes inferior auch ohne diese Teilfolge noch derselbe ist.

Wir erhalten nun

$$\begin{aligned}
\liminf_k J(u_k) &= \liminf_k \int_\Omega F(x, u_k, \nabla u_k)\ dx \\
&\geq \liminf_k \int_\Omega F(x, u, \nabla u_k)\ dx + \\
&\quad \liminf_k \left[\int_\Omega F(x, u_k, \nabla u_k)\ dx - \int_\Omega F(x, u, \nabla u_k)\ dx\right].
\end{aligned}$$

Wir betrachten zunächst den zweiten Term und werden sehen, dass dieser gegen Null konvergiert. Es gilt wegen (5.28) und (5.29)

$$\begin{aligned}
&\left|\int_\Omega F(x, u_k, \nabla u_k)\ dx - \int_\Omega F(x, u, \nabla u_k)\ dx\right| \\
&\leq \int_\Omega |F(x, u_k, \nabla u_k) - F(x, u, \nabla u_k)|\ dx \\
&\leq \int_\Omega G(x, u_k, u, \nabla u_k)|u - u_k|\ dx \\
&\leq \int_\Omega (\gamma_1|\nabla u_k|^{p-1} + \gamma_2|u_k|^{p-1} + \gamma_2|u|^{p-1})|u - u_k|\ dx.
\end{aligned}$$

Sei $p_* = \frac{p}{p-1}$, dann folgt aus der Hölder-Ungleichung

$$\begin{aligned}
&\left| \int_\Omega F(x, u_k, \nabla u_k)\; dx - \int_\Omega F(x, u, \nabla u_k)\; dx \right| \\
&\leq \left[\gamma_1 \left(\int_\Omega |\nabla u_k|^p \; dx \right)^{1/p_*} + \gamma_2 \left(\int_\Omega |u_k|^p \; dx \right)^{1/p_*} + \gamma_2 \left(\int_\Omega |u|^p \; dx \right)^{1/p_*} \right] \left(\int_\Omega |u - u_k|^p \; dx \right)^{1/p} \\
&= \left[\gamma_1 ||\nabla u_k||_{L^p}^{1/p_*} + \gamma_2 ||u_k||_{L^p}^{1/p_*} + \gamma_2 ||u||_{L^p}^{1/p_*} \right] ||u - u_k||_{L^p}
\end{aligned}$$

Der erste Term ist beschränkt und wegen der starken Konvergenz in $L^p(\Omega)$ konvergiert $||u - u_k||_{L^p} = 0$. Also ist

$$\liminf_k \left[\int_\Omega F(x, u_k, \nabla u_k)\; dx - \int_\Omega F(x, u, \nabla u_k)\; dx \right] = 0.$$

Aus (5.27) sehen wir sofort, dass das Funktional

$$J_u(v) := \int_\Omega F(x, u, \nabla v)\; dx$$

konvex ist, und wegen (5.26) ist es auch endlich. Damit können wir Lemma **??** anwenden und erhalten die Unterhalbstetigkeit von J, insbesondere

$$J(u) = J_u(u) \leq \liminf_k J_u(u_k) = \liminf_k \int_\Omega F(x, u, \nabla u_k)\; dx.$$

Damit folgt die Unterhalbstetigkeit von J. □

Nun haben wir alle Voraussetzungen des Existenzresultats in Satz (**??**) nachgeprüft und erhalten:

Theorem 5.6.1. *Sei J definiert durch (5.25) und F erfülle (5.26), (5.27), (5.29) für ein $p \in (1, \infty)$. Dann existiert ein Minimierer von J in $W^{1,p}(\Omega)$.*

Ist die Abbildung $(s, w) \mapsto F(x, s, w)$ strikt konvex für fast alle $x \in \Omega$, dann ist auch E strikt konvex als Funktional in $W^{1,p}(\Omega)$ und wir erhalten unter dieser Bedingung die Eindeutigkeit des Minimierers.

Um nun die Verbindung mit elliptischen partiellen Differentialgleichungen herzustellen betrachten wir den Fall einer differenzierbaren Funktion F, genauer nehmen wir an, dass gilt

$$F(x, \cdot, \cdot) \in C^1(\mathbb{R} \times \mathbb{R}^n) \qquad \text{für fast alle } x \in \Omega, \tag{5.31}$$

sowie für alle partiellen Ableitungen nach s oder w_i

$$|\partial_\eta F(x, \cdot, \cdot)| \leq D_1 |w|^{p-1} + D_2 |s|^{p-1} + D_3 \qquad \text{für } \eta = s, \eta = w_i, \tag{5.32}$$

für positive Konstante D_1, D_2, D_3. In diesem Fall können wir Richtungsableitungen berechnen, die für ein Minimum offensichtlich verschwinden sollten und erhalten daraus eine Gleichung. Die genaue Argumentation ist die folgende: sei $\varphi \in W^{1,p}(\Omega)$ beliebig aber fix. Dann gilt für einen Minimierer u von E:

$$J(u) \leq J(u + t\varphi) \qquad \forall\, t \in \mathbb{R}. \tag{5.33}$$

Damit hat die eindimensionale Funktion $j_\varphi(t) := J(u + t\varphi)$ ein Minimum in $t = 0$, und unter den obigen Bedingungen an F ist j_φ auch wirklich differenzierbar. Damit folgt

$$0 = \frac{de_\varphi}{dt}(0) = \int_\Omega (\partial_s F(x, u, \nabla u)\varphi + \nabla_w F(x, u, \nabla u) \cdot \nabla\varphi)\; dx.$$

Es gilt also die Variationsgleichung

$$\int_\Omega (\partial_s F(x, u, \nabla u)\varphi + \nabla_w F(x, u, \nabla u) \cdot \nabla\varphi)\; dx = 0 \qquad \forall\, \varphi \in W^{1,p}(\Omega). \tag{5.34}$$

Mit entsprechender zusätzlicher Regularität von F und von u (z.B. $u \in W^{2,p}(\Omega)$) können wir den Gauss'schen Satz anwenden und erhalten daraus

$$\int_\Omega (\partial_s F(x, u, \nabla u) - \nabla \cdot (\nabla_w F(x, u, \nabla u)))\varphi\; dx + \int_{\partial\Omega} \nabla_w F(x, u, \nabla u) \cdot n\varphi\; d\sigma = 0.$$

Dies ist für beliebige Testfunktionen $\varphi \in W^{1,p}(\Omega)$ nur möglich, wenn u eine (starke) Lösung der partiellen Differentialgleichung

$$-\nabla \cdot (\nabla_w F(x, u, \nabla u)) + \partial_s F(x, u, \nabla u) = 0 \qquad \text{fast überall in } \Omega \tag{5.35}$$

erfüllt und dazu auch noch die Neumann-Randbedingung

$$\nabla_w F(x, u, \nabla u) \cdot n = 0 \qquad \text{auf } \partial\Omega. \tag{5.36}$$