

Kapitel 2

Entrauschen von Bildern

In diesem Kapitel werden wir uns mit dem Entrauschen von Bildern und damit verwandten Problemen beschäftigen. Wir werden dieses Kapitel wieder in einzelne Unterabschnitte entsprechend verschiedener Methoden unterteilen, aber auch versuchen die Gemeinsamkeiten der verschiedenen Methoden herauszuarbeiten. Wir werden in allen Fällen im wesentlichen das kontinuierliche Modell zur Darstellung verwenden, aber auch die analogen Methoden für das diskrete Modell kurz erwähnen.

Bevor wir uns den verschiedenen Methoden zuwenden, sollten wir noch die mögliche Bewertung der Qualität einer Methode diskutieren. Eine einfache Bewertung wäre natürlich eine Kennzahl, die bei einer perfekten Rekonstruktion null wird bzw. mit schlechterer Qualität wächst (oder umgekehrt). Zum Test einer Methode mit künstlichen Daten kann man dabei einfach den Abstand zwischen der Rekonstruktion und dem sauberen Bild messen, und zwar in einer geeigneten Norm. Benutzen wir die Notation \hat{f} für das saubere Bild, f für die verrauschten Daten, und \hat{u} für die Rekonstruktion einer gewissen Methode, dann ist ein entsprechendes Fehlermaß gegeben durch

$$E_{\text{absolut}} = \|\hat{u} - \hat{f}\|. \quad (2.1)$$

Man kann natürlich auch den relativen Fehler in Bezug auf das Rauschen betrachten, d.h.

$$E_{\text{relativ}} = \frac{\|\hat{u} - \hat{f}\|}{\|f - \hat{f}\|} \quad (2.2)$$

oder den skalierten Fehler

$$E_{\text{skaliert}} = \frac{\|\hat{u} - \hat{f}\|}{\|\hat{f}\|}. \quad (2.3)$$

Eine Variante des skalierten Fehlers ist das sehr gebräuchliche Signal-to-Noise ratio (SNR, Signal-Rausch-Verhältnis). Die Namensgebung basiert auf der Idee, dass $\|\hat{u} - \hat{f}\|_2$ so etwas wie das Rauschen der Methode misst (beachte $\|f - \hat{f}\|_2$ ist das tatsächliche Rauschen) und $\|\hat{f}\|_2$ den Gehalt des Signals. Hier ist $\|\cdot\|_2$ im kontinuierlichen Fall die L^2 -Norm

$$\|u\|_2 = \sqrt{\int_{\Omega} u(x)^2 dx}$$

und im diskreten Fall die skalierte ℓ^2 -Norm

$$\|u\|_2 = \sqrt{\frac{1}{N^2} \sum_{i,j} u_{i,j}^2}$$

. Als Signal-to-Noise ratio erhält man dann

$$SNR = -\log \left(\frac{\|\hat{u} - \hat{f}\|_2}{\|\hat{f}\|_2} \right). \quad (2.4)$$

Man beachte dabei die Umkehrung der Monotonie, das SNR ist hoch bei guter Qualität und niedrig bei schlechter Qualität der Rekonstruktion. Eine ebenfalls weit verbreitete Variante des SNR ist das Peak-Signal-to-Noise ratio

$$PSNR = -\log \left(\frac{\|\hat{u} - \hat{f}\|_2}{\|\hat{f}\|_\infty} \right). \quad (2.5)$$

Hierbei vergleicht man das Rauschen mit dem Peak im Bild, d.h. $\|\hat{f}\|_\infty = \sup |\hat{f}|$.

In manchen Fällen wird es nötig sein, auch andere Fehlermaße als jene in einer Norm zu betrachten, z.B. wenn man primär an der Rekonstruktion von Kanten interessiert ist. In diesem Fall wird man eher geometrische Fehlermaße (z.B. in allgemeineren Metriken) für die Abstände der Kantenmengen verwenden.

2.1 Filtermethoden

Im folgenden werden wir einige Filtermethoden zum Entrauschen diskutieren. Die Bezeichnung Filter hat in der Ursprung in der Signalanalyse, als ein Verfahren das nur einen Teil des Signals (bzw. Bilds) liefert. Die Hoffnung beim Entrauschen ist natürlich, dass das saubere Bild aus dem Rauschen gefiltert wird.

2.1.1 Lokale Glättungsfiler

Lokale Glättungsfiler basieren auf der Idee, dass lokal meist ähnliche Grau- oder Farbwerte auftreten. Deshalb versucht man das Bild durch eine lokale Mittelung zu ersetzen. Dabei werden dann auch die zufälligen Störungen (das Rauschen) gemittelt. Geht man bei letzteren von unabhängigen gleichverteilten Zufallsvariablen aus, dann sollte dadurch auch die Varianz reduziert werden.

Ein linearer lokaler Glättungsfiler hat die allgemeine Form

$$\hat{u} = G_\epsilon * f, \quad (2.6)$$

wobei $G_\epsilon * f$ die Faltung des verrauschten Bilds mit einem Kern $G_\epsilon = \frac{1}{\epsilon^d} G(\frac{\cdot}{\epsilon})$ von spezieller Form ist. Um eine sinnvolle Konvexkombination zu erhalten sollte G nichtnegativ sein und Mittelwert eins haben. Um die Lokalität zu erhalten sollte G sein Maximum bei Null annehmen und für grosse Argumente gegen Null abfallen. Ein einfaches und oft verwendetes Beispiel einer solchen Funktion ist wieder die Gauss-Verteilung, aber auch Faltungskerne mit lokalem Träger sind von Interesse. Der Parameter ϵ misst die Skala auf der gemittelt wird. Die Faltung kann man dabei sowohl im kontinuierlichen als

$$G_\epsilon * f = \frac{1}{\epsilon^d} \int_{\Omega} G\left(\frac{x-y}{\epsilon}\right) f(y) dy \quad (2.7)$$

als auch im diskreten Modell als

$$(G_\epsilon * f)_{ij} = \frac{1}{\epsilon^d} \sum_{k,\ell} G\left(\frac{x_{ij} - x_{k\ell}}{\epsilon}\right) f_{k\ell} \quad (2.8)$$

definieren. Im diskreten Fall ist natürlich darauf zu achten, dass ϵ in geeignet in Abhängigkeit von der Gitterweite gewählt wird, bei zu kleinem ϵ tritt effektiv keine Glättung auf.

Lokale Glättungsverfahren reduzieren in der Praxis tatsächlich das Rauschen, allerdings gibt es auch ein potentielles Problem solcher Filter-Verfahren, nämlich Überglättung. Diese betrifft vor allem Kanten, da bei einer Kante am stärksten die Annahme lokal ähnlicher Grauwerte verletzt ist. Denken wir z.B. an eine schwarz-weiße Kante, so werden die lokalen Grauwerte an der Kante zu Grau gemittelt. Damit wird die Kante zu einem kontinuierlichen Übergang der Grauwerte gemittelt, und damit erscheint die Kante optisch verschmiert.

Zum näheren Verständnis betrachten wir den diskreten Fall und einen speziellen lokalen Filter der Form

$$\hat{u}_{ij} = (1 - 4\alpha)f_{ij} + \alpha(f_{i-1j} + f_{i+1j} + f_{ij-1} + f_{ij+1}), \quad (2.9)$$

d.h. der Filter wirkt nur auf die benachbarten Pixel. Dabei ist $\alpha \in (0, \frac{1}{5})$ der Parameter, der die Gewichtung misst. Der Skalenparameter ϵ kann mit der Grösse des Gitters identifiziert werden. Wir nehmen wieder an, dass die Beobachtung durch punktweise Gauss'sches Rauschen entstanden ist, d.h.

$$f_{ij} = \hat{f}_{ij} + n_{ij}^\sigma.$$

Durch den Filter erhalten wir einen systematischen Fehler, d.h.

$$\begin{aligned} \mathbb{E}(\hat{u}_{ij}) &= (1 - 4\alpha)\mathbb{E}(f_{ij}) + \alpha(\mathbb{E}(f_{i-1j}) + \mathbb{E}(f_{i+1j}) + \mathbb{E}(f_{ij-1}) + \mathbb{E}(f_{ij+1})) \\ &= (1 - 4\alpha)\hat{f}_{ij} + \alpha(\hat{f}_{i-1j} + \hat{f}_{i+1j} + \hat{f}_{ij-1} + \hat{f}_{ij+1}), \end{aligned}$$

und damit im allgemeinen $\mathbb{E}(\hat{u}_{ij}) \neq \hat{f}_{ij}$. Der Filter hat also einen gewissen Nachteil, den man nur dann in Kauf nehmen sollte, wenn zumindest der mittlere Fehler kleiner wird. Dazu betrachten wir als

$$\begin{aligned} \mathbb{E}((\hat{u}_{ij} - \hat{f}_{ij})^2) &= \mathbb{E}((\hat{u}_{ij} - \mathbb{E}(\hat{u}_{ij}) + \mathbb{E}(\hat{u}_{ij}) - \hat{f}_{ij})^2) \\ &= \mathbb{E}((\hat{u}_{ij} - \mathbb{E}(\hat{u}_{ij}))^2) + (\mathbb{E}(\hat{u}_{ij}) - \hat{f}_{ij})^2 \end{aligned}$$

Wir beginnen mit dem ersten Term und erhalten wegen der Unabhängigkeit der f_{ij} und $\mathbb{E}(f_{ij}) = \hat{f}_{ij}$:

$$\begin{aligned} \hat{u}_{ij} - \mathbb{E}(\hat{u}_{ij}) &= (1 - 4\alpha)(f_{ij} - \hat{f}_{ij}) + \alpha(f_{i-1j} - \hat{f}_{i-1j} + f_{i+1j} - \hat{f}_{i+1j} + \\ &\quad f_{ij-1} - \hat{f}_{ij-1} + f_{ij+1} - \hat{f}_{ij+1}) \\ &= (1 - 4\alpha)n_{ij}^\sigma + \alpha(n_{i-1j}^\sigma + n_{i+1j}^\sigma + n_{ij-1}^\sigma + n_{ij+1}^\sigma), \end{aligned}$$

wobei n_{ij}^σ unabhängige normalverteilte Zufallsvariablen mit Mittelwert 0 und Standardabweichung σ sind. Wir erhalten dann

$$\begin{aligned} \mathbb{E}((\hat{u}_{ij} - \mathbb{E}(\hat{u}_{ij}))^2) &= (1 - 4\alpha)^2\mathbb{E}((n_{ij}^\sigma)^2) + \alpha^2(\mathbb{E}((n_{i+1j}^\sigma)^2) + \mathbb{E}((n_{i-1j}^\sigma)^2) + \\ &\quad \mathbb{E}((n_{ij+1}^\sigma)^2) + \mathbb{E}((n_{ij-1}^\sigma)^2)) \\ &= (1 - 4\alpha)^2\sigma^2 + 4\alpha\sigma^2 = (1 - 8\alpha + 20\alpha^2)\sigma^2. \end{aligned}$$

Das Rauschen wird also durch den Filter verringert, da dieser Teil des Fehlers für $\alpha > 0$ kleiner als σ ist. Nun betrachten wir noch den zweiten Term in der obigen Abschätzung, die den systematischen Fehler beschreibt. Es gilt ja

$$\mathbb{E}(\hat{u}_{ij}) - \hat{f}_{ij} = \alpha(f_{i-1j} + f_{i+1j} + f_{ij-1} + f_{ij+1} - 4f_{ij}).$$

Der Einfachheit halber nehmen wir an, dass f_{ij} der Pixelwert zum Index (ih, jh) ist, wobei h die (kleine) Pixelgrösse ist. Wir denken also $f_{ij} = f(x_{ij})$ für eine geeignete Grauwertfunktion f . Ist f zweimal stetig differenzierbar, so folgt aus dem Mittelwertsatz die Existenz von $\xi_1 \in ((i-1)h, (i+1)h) \times \{jh\}$ und $\xi_2 \in \{ih\} \times ((j-1)h, (j+1)h)$, sodass

$$f_{i-1j} + f_{i+1j} - 2f_{ij} = \frac{\partial^2 f}{\partial x_1^2}(\xi_1)h^2$$

und

$$f_{ij-1} + f_{ij+1} - 2f_{ij} = \frac{\partial^2 f}{\partial x_2^2}(\xi_2)h^2.$$

Also können wir den zweiten Teil des Fehlers durch

$$\mathbb{E}(\hat{u}_{ij} - \hat{f}_{ij})^2 \leq 4 \max\left\{\left\|\frac{\partial^2 f}{\partial x_1^2}\right\|_\infty, \left\|\frac{\partial^2 f}{\partial x_2^2}\right\|_\infty\right\} \alpha^2 h^4$$

abschätzen.

Damit erhalten wir abschliessend die Abschätzung

$$\mathbb{E}((\hat{u}_{ij} - \hat{f}_{ij})^2) \leq (1 - 8\alpha + 20\alpha^2)\sigma^2 + 4 \max\left\{\left\|\frac{\partial^2 f}{\partial x_1^2}\right\|_\infty, \left\|\frac{\partial^2 f}{\partial x_2^2}\right\|_\infty\right\} \alpha^2 h^4. \quad (2.10)$$

Insbesondere ist für h hinreichend klein, der Fehler im quadratischen Mittel kleiner als σ^2 und somit erscheint es sinnvoll, den Filter zu benutzen. Man sieht aber auch, dass die zweite Ableitung von f wesentlich ist, also sollte man gerade bei weniger glatten Stellen wie Kanten Probleme erwarten. Diese treten in der Praxis auch auf und bewirken ein signifikantes Verschmieren, weshalb eine Vielfalt anderer Filter entwickelt wurden, die besser mit Kanten umgehen können.

Den obigen Filter und seine Abhängigkeit vom Parameter kann man in verschiedener Weise interpretieren: als variabel oder fix und klein. Die erste Sichtweise nimmt α als variablen Parameter, den man natürlich optimal wählen sollte. Die optimale Wahl des Parameters muss dabei die zwei konkurrierenden Effekte des Filters ausbalanzieren. Einerseits hat der Filter ja den positiven Effekt die Varianz des Rauschens durch die Mittelung auszuglätten, und die Varianz sinkt mit α (und erreicht sein Minimum bei $\alpha = \frac{1}{5}$, wie man aus dem ersten Teil des mittleren quadratischen Fehlers leicht nachrechnet). Andererseits wächst der systematische Fehler (siehe zweiter Teil) mit α , man erwartet also ein optimales α abhängig von den Daten irgendwo im Intervall $(0, \frac{1}{5})$. Dies sieht man auch durch eine andere Charakterisierung des Schätzers als Minimum des Funktionals (bei geeigneter Behandlung der Gitterpunkte am Rand)

$$\begin{aligned} J(u) &= \frac{1}{2} \sum_{i,j} (u_{ij} - f_{ij})^2 + \frac{\alpha}{2} \sum_{i,j} [(f_{i+1j} - f_{ij})^2 + (f_{ij+1} - f_{ij})^2] + \\ &\quad \alpha \sum_{i,j} [(f_{i+1j} - f_{ij})(u_{i+1j} - u_{ij}) + (f_{ij+1} - f_{ij})(u_{ij+1} - u_{ij})]. \end{aligned} \quad (2.11)$$

Das Regularisierungsfunktional, d.h.

$$E(u) = J(u) - \frac{1}{2} \sum_{i,j} (u_{ij} - f_{ij})^2$$

¹Bild erzeugt und bearbeitet von Bärbel Schlake und Melanie Schröter.



Abbildung 2.1: Entrauschen durch iteratives Anwenden eines lokalen Glättungsfilters. Sauberes Bild, verrauschtes Bild und Resultat¹.

kann als Taylor-Entwicklung erster Ordnung des Funktionals

$$\tilde{E}(u) = \frac{\alpha}{2} \sum_{i,j} [(u_{i+1j} - u_{ij})^2 + (u_{ij+1} - u_{ij})^2]$$

um f interpretieren. Da für kleines α das Minimum \hat{u} nahe bei f liegen wird, sollte der Unterschied zwischen $E(\hat{u})$ und $\tilde{E}(\hat{u})$ klein sein. Es liegt die Vermutung nahe, dass \hat{u} nahe beim Minimum von $J - E + \tilde{E}$ liegt, einem Funktional, das wir bereits in der Einleitung als MAP-Schätzer kennengelernt haben. Den kontinuierlichen Grenzwert ($h \rightarrow 0$) dieses Funktionals, nämlich

$$\hat{E}(u) = \frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx + \int_{\Omega} |\nabla u|^2 dx,$$

sowie Varianten davon werden wir unten im Abschnitt über Variationsmethoden noch näher diskutieren.

Alternativ kann man α auch als eine fixe gegebene Grösse ansehen und den Filter einfach mehrmals wiederholen um schrittweise die Varianz zu reduzieren. Daraus erhält man ein iteratives Verfahren, bei dem der Filter immer auf den Schätzer aus dem letzten Schritt angewendet wurde. Mit $u^0 := f$ erhalten wir dann

$$u_{ij}^{k+1} = (1 - 4\alpha)u_{ij}^k + \alpha(u_{i-1j}^k + u_{i+1j}^k + u_{ij-1}^k + u_{ij+1}^k).$$

Unter der Annahme, dass h klein und $\alpha \leq \frac{1}{5}$ ist, kann man auch den kontinuierlichen Grenzwert betrachten. es gilt ja

$$\frac{u_{ij}^{k+1} - u_{ij}^k}{h^2 \alpha} = \frac{1}{h^2} (u_{i-1j}^k + u_{i+1j}^k + u_{ij-1}^k + u_{ij+1}^k - 4u_{ij}^k) = \Delta u(x_{ij}) + \mathcal{O}(h^2),$$

wobei $\Delta = \nabla \cdot \nabla$ den Laplace-Operator bezeichnet. Mit $\tau = \alpha h^2 \lll 1$ können wir diese Gleichung auch als Vorwärts-Euler Diskretisierung der Wärmeleitungsgleichung (Diffusionsgleichung)

$$\frac{\partial u}{\partial t} = \Delta u$$

mit Anfangswert $u(0) = f$ interpretieren. Damit erhalten wir zum ersten Mal ein Verfahren basierend auf einer partiellen Differentialgleichung (PDE), einen Ansatz den wir ebenfalls noch genauer diskutieren werden. Bei Verfahren dieser Art gibt es ebenfalls die Frage nach einem optimal gewählten Parameter, nämlich den optimalen Abbruchindex bei der diskreten Iteration bzw. die optimale Endzeit bei der Diffusionsgleichung.



Abbildung 2.2: Entrauschen durch iteratives Anwenden eines lokalen Glättungsfilters. Sauberes Bild, verrauschtes Bild und Resultat².

²Bild erzeugt und bearbeitet von Bärbel Schlake und Melanie Schröter.

2.1.2 Frequenzraumfilter

Sehr analog zu Glättungsfiltern oder eigentlich nur eine andere Betrachtungsweise sind Frequenzraumfilter, die in der Form

$$\hat{u} = \mathcal{F}^{-1}(\psi_\epsilon \mathcal{F}(f)) \quad (2.12)$$

gegeben sind, wobei ψ_ϵ eine Funktion ist, die gewisse Frequenzen dämpft oder ganz eliminiert (wiederum abhängig von einem Skalenparameter ϵ). Den Zusammenhang mit der Filterdefinition (2.6) liefert wieder der Faltungssatz, denn mit $G_\epsilon := (2\pi)^{-d/2} \mathcal{F}^{-1}(\psi_\epsilon)$ folgt

$$G_\epsilon * f = (2\pi)^{-d/2} \mathcal{F}^{-1}(\mathcal{F}(G_\epsilon) \mathcal{F}(f)) = \mathcal{F}^{-1}(\psi_\epsilon \mathcal{F}(f)).$$

Die ursprüngliche Motivation der Konstruktion von Frequenzraumfiltern ist eine etwas andere als jene der lokalen Glättung. Im Frequenzraum ist die zentrale Idee, dass Rauschen meist hohen Frequenzen entspricht und man deshalb diese Frequenzen zumindest dämpfen sollte. Man spricht dann von einem *Lowpass-Filter*, da nur die niedrigen Frequenzen unverändert bleiben (analog einem Hardware-Filter für ein Signal, bei dem nur die Niedrigfrequenten Anteile durchkommen). Die einfachste Wahl für die Funktion $\psi_\epsilon(\omega) = \psi(\epsilon\omega)$ erhält man mit

$$\psi(\omega) = \begin{cases} 1 & \text{für } |\omega| \leq 1 \\ 0 & \text{sonst.} \end{cases} \quad (2.13)$$

Damit schneidet ψ_ϵ alle Frequenzanteile über $1/\epsilon$ ab (beachte die inverse Relation da ϵ ja der Wellenlänge entspricht), jene unter $1/\epsilon$ bleiben unverändert. Durch Berechnung der inversen Fouriertransformation sieht man, dass man in der Form (2.6) eine sinc-Funktion ($G(x) \sim \sin|x|/|x|$) als Faltungskern erhält. Wegen der lokalen Oszillation des Sinus ergibt dies eine etwas ungewöhnliche lokale Mittelung, da die Gewichtung nicht monoton mit dem Abstand fällt. Einen Filter dieser Form hätte man also wohl kaum ohne die Frequenzrauminterpretation konstruiert. Dennoch hängen die beiden Sichtweisen auch stark zusammen, da ja das Fehlen hoher Frequenzanteile automatisch bedingt, dass die lokale Variation des Grauwerts nicht extrem stark sein kann und umgekehrt, also basieren die ursprüngliche Sichtweise und die Frequenzrauminterpretation auf sehr ähnlichen Annahmen. Noch deutlicher wird der Zusammenhang wieder bei der Gauss-Verteilung als Faltungskern. Da die Fourier-Transformation dieser Funktion wieder (bis auf Konstante) wieder die Gauss-Verteilung ergibt, wird also im Frequenzraum wieder mit dieser schnell fallenden Funktion gedämpft.

Bisher haben wir nur die Anwendung eines gegebenen Filters auf die verrauschten Daten f betrachtet. Man kann sich aber auch die Frage nach der Konstruktion des optimalen Filters bei gegebenen Daten stellen. Eine weit verbreitete Möglichkeit für so einen Daten-getriebenen Filter sind sogenannte *Wiener³-Filter*. In diesem Fall setzt man G als Unbekannte an und versucht den mittleren quadratischen Fehler zwischen dem Schätzer $G * f$ und dem sauberen Bild \hat{f} zu minimieren, d.h.

$$G = \arg \min_g \mathbb{E}[(g * f(x) - \hat{f}(x))^2]. \quad (2.14)$$

Wegen der Definition von G als Minimum hat insbesondere die eindimensionale Funktion

$$e_g(t) := \mathbb{E}[(G + tg) * f(x) - \hat{f}(x)]^2$$

³Norbert Wiener (1894-1964), US-Amerikanischer Mathematiker. Gilt als Begründer der Kybernetik. Nach ihm wurde auch der Wiener Prozess in der Wahrscheinlichkeitstheorie benannt.

ihr Minimum in $t = 0$ für jede Variation g . Damit folgt

$$0 = e'_g(0) = 2\mathbb{E}[(G * f(x) - \hat{f}(x))g * f(x)].$$

Da g beliebig ist, können wir auch $g(x) = \delta(x - a)$ mit beliebigem a wählen (oder zumindest durch Funktionen beliebig annähern) und da dann $g * f(x) = f(x + a)$ gilt, folgt

$$0 = \mathbb{E}[(G * f(x) - \hat{f}(x))f(y)],$$

wobei $y = x + a$ beliebig ist. Diese Gleichung nennt man Orthogonalitätsbedingung des Wiener Filters. Wir betrachten nun die Korrelationsfunktionen, wobei allgemein die Definition

$$R_{uv}(x, y) := \mathbb{E}[u(x)v(y)].$$

Die Orthogonalitätsbedingung kann dann als

$$R_{G*ff}(x, y) = R_{\hat{f}f}(x, y) \quad \forall y$$

geschrieben werden. Diese Gleichung ist im allgemeinen schwierig zu lösen, leichter wird es aber wenn die Korrelation nur vom Abstand abhängt, d.h.

$$R_{fg}(x, y) = \tilde{R}_{fg}(x - y).$$

Solch eine Form ist natürlich für die meisten Rauschmodelle, insbesondere ist sie für das Gauss'sche Rauschmodell erfüllt, wo die punktwisen Störungen ja unabhängig sind. Berechnet man aber die Korrelationsfunktion von Bildern und nicht nur des Rauschens ist die Annahme allerdings schwieriger zu rechtfertigen. Unter der Vereinfachung wird die Orthogonalitätsbedingung zu

$$R_{G*ff}(z) = R_{\hat{f}f}(z), \quad z = x - y.$$

Wegen der Linearität des Erwartungswert können wir \mathbb{E} mit der Faltung vertauschen, d.h.

$$\mathbb{E}(G * f(x)f(y)) = \int G(x - z)\mathbb{E}(f(z)f(y)) dz = G * R_{ff}(x, y)$$

und damit folgt

$$G * R_{ff}(z) = R_{\hat{f}f}(z).$$

Nun beachten wir noch, dass das Rauschen $n = f - \hat{f}$ unabhängig vom Signal ist und Mittelwert Null hat, d.h.

$$E[\hat{f}(x)f(y)] = E[\hat{f}(x)\hat{f}(y)] + E[\hat{f}(x)]E[n(y)] = E[\hat{f}(x)\hat{f}(y)],$$

und damit ist $R_{\hat{f}f} = R_{\hat{f}\hat{f}}$ und wir erhalten die Gleichung

$$G * R_{ff}(z) = R_{\hat{f}\hat{f}}(z).$$

Sei nun S_{ff} die Fouriertransformation von R_{ff} , dann folgt aus dem Faltungssatz

$$\mathcal{F}G(\omega) = (2\pi)^{d/2} \frac{S_{\hat{f}\hat{f}}(\omega)}{S_{ff}(\omega)}.$$

Wegen der Unabhängigkeit des Rauschens vom Signal gilt $S_{\hat{f}\hat{f}} = S_{ff} - S_{nn}$ und damit können wir den Wiener Filter aus den Daten und dem Rauschmodell als

$$\mathcal{F}G(\omega) = (2\pi)^{d/2} \frac{S_{ff}(\omega) - S_{nn}(\omega)}{S_{ff}(\omega)}$$

berechnen.

2.1.3 Kantenkorrektur

Wie wir oben gesehen haben, liegt ein Hauptproblem der lokalen Glättungsfilter im Verschmieren von Kanten. Zur Verbesserung der Performance versucht man nun, die Mittelung nur entlang der Kanten und nicht in Normalenrichtung durchzuführen, um die Performance zu verbessern. Um eine Glättung über Kanten hinweg zu vermeiden, kann man versuchen die Mittelung nur entlang der Isokonturen des Bilds durchführen. Die Schwierigkeit dabei ist natürlich die Bestimmung der Isokonturen bzw. ihrer Tangentialen eben wegen der Verrauschtheit des Bilds.

Eine Möglichkeit besteht in einem zweistufigen vorgehen. Zunächst bestimmen wir wie oben eine linear gefilterte Version des Bilds $v = G_\epsilon * f$, eventuell sogar mit Überglättung (zu grosses ϵ). Damit werden zwar die Kanten verschmiert, aber das Rauschen ausgeglättet und die Isokonturen von v sollten immer noch eine gute Approximation der Isokonturen des Bilds sein. Da der Gradient von v in Normalenrichtung der Isokontur steht, ist ein normierter Tangentialvektor durch $\mathbf{t} = \frac{\nabla v^\perp}{|\nabla v|}$ gegeben, wobei ∇v^\perp den Orthogonalvektor zu ∇v bezeichnet. Als zweiten Schritt kann man nun nur entlang dieser Tangentiale \mathbf{t} glätten, um ein geglättetes Bild der Form

$$\hat{u}(x) = \int_{\mathbb{R}} \tilde{G}_\delta(s) u(x + s\mathbf{t}) ds, \quad \mathbf{t} = \frac{\nabla v^\perp}{|\nabla v|}, v = G_\epsilon * f \quad (2.15)$$

zu erhalten, jetzt mit einem eindimensionalen Kern \tilde{G} und Skalenparameter $\delta > 0$. Aufgrund der obigen Argumentation sollte ein Filter dieser Art Kanten besser erhalten als ein linearer Filter. In der Praxis ist (2.15) allerdings schwer zu implementieren, zumindest in der obigen Form. Hat man v berechnet (was wiederum diskret einfach mit einer Quadraturformel passieren kann), benötigt man zunächst den Gradienten ∇v . Dies kann mit einer Differenzenformel (zentral oder einseitig) geschehen, wegen der Vorglättung mit G_ϵ sind dabei keine Instabilitäten zu erwarten. Der letzte Schritt ist nun die Berechnung der eindimensionalen Faltung in Tangentialrichtung. Da im allgemeinen nicht zu erwarten ist, dass die Tangenten durch Gitterpunkte (Pixel) laufen, stellt sich die Frage, welche der Pixel-Werte man zur Berechnung des Integrals verwenden soll. Es ist naheliegend, die Pixelwerte abnehmend mit ihrem Normalabstand von der Tangente zu gewichten, allerdings sollte man dabei eigentlich auch den Gradienten von u (bzw. die Normalableitung) berücksichtigen, und dieser ist beim verrauschten Bild wieder schwer zu approximieren. Eine analoge Methode mit einfacherer Implementierung erhält man im Grenzwert der Skalenparameter $\epsilon \rightarrow 0$, $\delta \sim \epsilon \rightarrow 0$, dann gilt nämlich

$$\hat{u}(x) - f(x) = c\epsilon^2 |\nabla f| \kappa(f), \quad \kappa(f) = \nabla \cdot \left(\frac{\nabla f}{|\nabla f|} \right). \quad (2.16)$$

Hier ist $\kappa(f)$ die (mittlere) Krümmung der Isokontoren und (2.16) entspricht einem Zeitschritt (der Länge $c\epsilon^2$) der mittleren Krümmungsgleichung (Mean Curvature Flow),

$$\frac{\partial u}{\partial t} = |\nabla u| \kappa(u), \quad \kappa(u) = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right). \quad (2.17)$$

Wir werden uns im späteren Abschnitt über PDE-Methoden noch näher mit dieser Gleichung befassen.

Bei der obigen Konstruktion von (2.15) haben wir nur in Tangentialrichtung der Isokonturen geglättet. Dies ist eine gute Strategie bei Kanten, aber in glatten oder flachen Bereichen könnte man auch in andere Richtungen glätten und somit eine stärkere Varianzreduktion

erreichen. Dies schafft man durch eine geeignete Gewichtung abhängig von der Grösse des Gradienten,

$$\begin{aligned}\hat{u}(x) &= \frac{1}{C(x)} \int_{\mathbb{R}} \tilde{G}_\delta(s) u(x + \mathbf{st}) f_1(|\nabla v(x + \mathbf{st})|) ds + \\ &\quad \frac{1}{C(x)} \int_{\mathbb{R}} \tilde{G}_\delta(s) u(x + \mathbf{st}^\perp) f_2(|\nabla v(x + \mathbf{st}^\perp)|) ds.\end{aligned}\tag{2.18}$$

Die Konstante $C(x)$ soll dabei so gewählt werden, dass die Faltung volumserhaltend ist, d.h.

$$C(x) = \int_{\mathbb{R}} \tilde{G}_\delta(s) f_1(|\nabla v(x + \mathbf{st})|) ds + \frac{1}{C(x)} \int_{\mathbb{R}} \tilde{G}_\delta(s) f_2(|\nabla v(x + \mathbf{st}^\perp)|) ds.$$

Um die Glättung über Kanten zu vermeiden sollte $f_2 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ eine fallende Funktion sein mit $f_2(p) \rightarrow 0$ für $p \rightarrow \infty$.

2.1.4 Nachbarschaftsfilter

Nachbarschaftsfilter (cf. [9, 15]) basieren nicht nur auf einer örtlichen Nachbarschaft, sondern einer Nachbarschaft der Grau- oder Farbwerte. Ein wichtiges Beispiel ist der sogenannte *Sigma-Filter*

$$\hat{u}(x) = \frac{1}{C(x)} \int_{\mathbb{R}^d} G_\epsilon(x - y) f(y) e^{-\frac{(f(x) - f(y))^2}{\delta^2}} dy,\tag{2.19}$$

wieder mit einem Normalisierungsfaktor $C(x)$. Die meist verwendete Version von (2.19) benutzt dabei G als charakteristische Funktion des Einheitskreises. In diesem Fall hat man zwei Skalenparameter zur Verfügung. Für kleines ϵ und grosses δ verhält sich der Nachbarschaftsfilter (2.19) wie ein lokaler Glättungsfilter. Für grosses ϵ und kleines δ hingegen erhält man eine nichtlokale Mittelung über ähnliche Farbbereiche. Die Hoffnung dabei ist, dass gewisse selbstähnliche Strukturen in den Bildern auftreten und diese zur Varianzreduktion des Rauschens genutzt werden können.

Auch das asymptotische Verhalten des Filters hängt natürlich von der Relation der Skalenparameter ab. Im Fall eines Kerns G , der als charakteristische Funktion des Einheitskreises gewählt wird, wurde das Verhalten näher untersucht (cf. [1]). Die Relation zwischen den Skalenparametern wird dabei durch $\delta \sim \epsilon^\alpha$ ausgedrückt. Für $\alpha < 1$ geht δ also langsamer gegen Null als für ϵ und umgekehrt für $\alpha > 1$. Es ist also nicht überraschend, dass die Asymptotik im wesentlichen die Fälle $\alpha < 1$, $\alpha = 1$, und $\alpha > 1$ unterscheiden muss. Am einfachsten ist dabei wieder der Fall $\alpha < 1$, denn dort dominiert die lokale Glättung und es ist nicht überraschend, dass die Asymptotik wie im linearen Fall als (zeitdiskrete) Diffusionsgleichung

$$\hat{u} - f = c\epsilon^2 \Delta f + \mathcal{O}(\epsilon^{2+\alpha})$$

gegeben ist. Der Fall $\alpha = 1$ ist da schon wesentlich interessanter, hier erhält man asymptotisch eine nichtlineare Diffusionsgleichung der Form

$$\hat{u} - f = \epsilon^2 A\left(\frac{\epsilon}{\delta} |\nabla f|\right) \mathbf{n}^T (D^2 f) \mathbf{n} + \epsilon^2 B\left(\frac{\epsilon}{\delta} |\nabla f|\right) \mathbf{t}^T (D^2 f) \mathbf{t} + \mathcal{O}(\epsilon^3),$$

mit $\mathbf{n} = \frac{\nabla f}{|\nabla f|}$, $\mathbf{t} = \mathbf{n}^\perp$, und $D^2 f$ der Hesse-Matrix von f . Die Funktionen A und B sind gegeben durch

$$A(p) = 3B(p) + \frac{3}{p^2} B(p) - \frac{1}{2p^2}, \quad B(p) = \frac{pe^{-p^2}}{\int_0^p e^{-s^2} ds}.$$

Wie wir aus dem Plot der Funktionen A und B in Abbildung 2.3 sehen, ist B positiv und monoton fallend, während A für Argumente grösser 1 negativ wird. Dies bedeutet, dass analog zur Kantenkorrektur entlang der Kanten immer geglättet wird, in Normalenrichtung jedoch nur für kleine Gradienten, d.h. in glatten Teilen des Bilds. Für grosse Gradienten erhält man einen negativen Vorfaktor in Normalenrichtung, als Rückwärtsdiffusion. Wie wir im nächsten Kapitel noch sehen werden, bewirkt Rückwärtsdiffusion in Normalenrichtung sogar eine Schärfung der Kanten. Für $\alpha > 1$ erhält man wieder nur eine Gewichtung der Diffusion mit der Funktion B , also Vorwärtsdiffusion und damit Glättung in alle Richtungen.

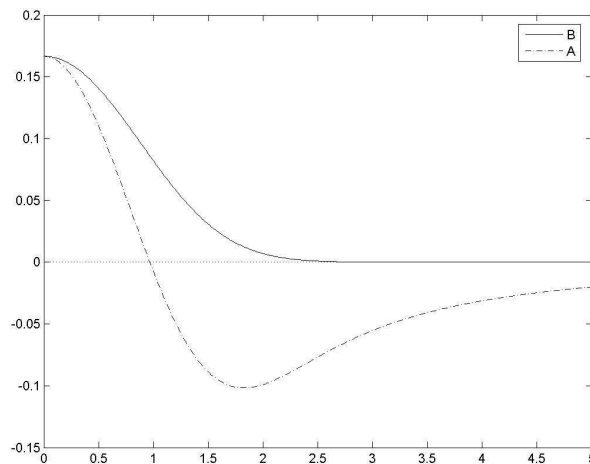


Abbildung 2.3: Funktionen A und B in der asymptotischen Darstellung des Sigma-Filters.

Um bezüglich der Nachbarschaft das Rauschen zu unterdrücken, kann man in der Exponentialfunktion wieder eine vorgeglättete Version v von f verwenden, d.h.

$$\hat{u}(x) = \frac{1}{C(x)} \int_{\mathbb{R}^d} G_\epsilon(x-y) f(y) e^{-\frac{(v(x)-v(y))^2}{\delta^2}} dy, \quad v = \tilde{G}_\delta * f. \quad (2.20)$$

Für genauere Studien des Verhaltens von Nachbarschaftsfiltern sowie Erweiterungen verweisen wir auf [1].

2.2 PDE-Methoden und Diffusionsfilter

Wie wir oben gesehen haben, führen lineare und nichtlineare Filter in sinnvollen Asymptotiken immer auf die Lösung parabolischer Differentialgleichungen der Form

$$\frac{\partial u}{\partial t} = F(\nabla u, D^2 u), \quad u(0) = f, \quad (2.21)$$

das entrauschte Bild entspricht dann der Lösung dieser Gleichung zu einem gewissen Zeitpunkt $T > 0$. Da man in einem solchen Fall Glättung durch Diffusion erreicht, spricht man von einem Diffusionsfilter. Auch der Ausdruck Scale Space Method ist dafür sehr gebräuchlich, denn wie wir noch sehen werden, eliminiert die Methode für wachsendes T die kleineren Skalen und die verbleibenden werden immer größer.

Um die Skaleneigenschaft zu verstehen, betrachten wir zunächst die örtlich eindimensionale Wärmeleitungsgleichung

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad u(0) = f, \quad (2.22)$$

auf $[0, \pi] \times \mathbb{R}_+$. Als Randbedingungen nehmen wir an

$$\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(\pi, t) = 0,$$

d.h. wir nehmen an, die Grauwerte ändern sich am Rand nicht in Normalenrichtung. In diesem Fall können wir die Lösung durch Fourier-Cosinus-Entwicklung berechnen. Für Funktionen mit den obigen Randbedingungen existiert eine konvergente Reihendarstellung der Form

$$u(x, t) = \frac{1}{\sqrt{\pi}} a_0(t) + \sum_{k=1}^{\infty} a_k(t) \sqrt{\frac{2}{\pi}} \cos(kx),$$

wobei $a_k(0)$ die Fourier-Cosinus-Koeffizienten von f sind. Setzen wir die Entwicklung in die Wärmeleitungsgleichung ein, so folgt durch Koeffizientenvergleich

$$a'_k(t) = -k^2 a_k(t)$$

und die Lösung dieser Gleichung erhält man als $a_k(t) = a_k(0)e^{-k^2 t}$. Dies bedeutet, dass höhere Frequenzen schneller gedämpft werden als niedrige. Die Lösung der Wärmeleitungsgleichung liefert also einen Frequenzraumfilter wie oben (in diesem Fall mit Fourierreihen statt -transformation). Die einzige Frequenz, die nicht gedämpft wird, ist die niedrigste bei $k = 0$, man erhält

$$\frac{1}{\pi} \int_0^{\pi} u(x, t) dx = \frac{1}{\sqrt{\pi}} a_0(t) = \frac{1}{\sqrt{\pi}} a_0(0) = \frac{1}{\pi} \int_0^{\pi} f(x) dx,$$

der Mittelwert bleibt also erhalten. Bei der mehrdimensionalen Wärmeleitungsgleichung ist eine analoges Vorgehen möglich, wenn man die Cosinus-Funktionen durch die jeweiligen Eigenfunktionen des Laplace-Operators ersetzt.

Wir sehen aus der obigen Reihenentwicklung, dass wir uns keine Sorgen um die Konvergenz der Reihenentwicklung für $u(x, t)$ machen müssen. Da die Cosinus-Funktionen mit der obigen Skalierung ein Orthonormalsystem bilden, folgt aus der Parseval'schen Identität

$$\int_{\Omega} u(x, t)^2 dx = \sum_{k=0}^{\infty} a_k(t)^2 \leq \sum_{k=0}^{\infty} a_k(0)^2 = \int_{\Omega} f(x)^2 dx.$$

Die L^2 -Norm der Funktion u zu jedem Zeitpunkt ist also durch jene von f beschränkt. Dies gilt nicht, wenn man statt der Wärmeleitungsgleichung (Vorwärtsdiffusion) die Rückwärtsdiffusions-Gleichung

$$\frac{\partial v}{\partial t} = -\frac{\partial^2 v}{\partial x^2}, \quad v(0) = f, \quad (2.23)$$

betrachtet. Mit der analogen Reihendarstellung

$$u(x, t) = \frac{1}{\sqrt{\pi}} b_0(t) + \sum_{k=1}^{\infty} b_k(t) \sqrt{\frac{2}{\pi}} \cos(kx),$$

würde $b'_k(t) = +k^2 b_k(t)$ folgen und damit $b_k(0)e^{+k^2 t}$. In diesem Fall werden hohe Frequenzen sogar verstärkt, wobei die Lösung im allgemeinen nach sehr kleiner Zeit gar nicht mehr existiert. Für $v(\cdot, t) \in L^2([0, 1])$ müsste ja gelten, dass

$$\int_{\Omega} v(x, t)^2 dx = \sum_{k=0}^{\infty} b_k(t)^2 = \sum_{k=0}^{\infty} b_k(0)^2 e^{2k^2 t} < \infty,$$

d.h. die Koeffizienten $b_k(0)$ müssen extrem schnell gegen 0 fallen. Dies ist eine extrem starke Einschränkung an den Anfangswert f , z.B. bedeutet es auch, dass f analytisch sein muss. In mathematischer Hinsicht ist Rückwärtsdiffusion im Gegensatz zu Vorwärtsdiffusion extrem problematisch, nicht nur wegen der möglichen Nicht-Existenz von Lösungen. Auch in praktischer Hinsicht kann Rückwärtsdiffusion zu Problemen führen, vor allem da Fehler in hohen Frequenzen extrem verstärkt werden. Dennoch kann ein wenig Rückwärtsdiffusion richtig eingesetzt in der Bildverarbeitung nützlich sein, wie wir noch sehen werden.

Nach diesen ersten Überlegungen zur linearen Diffusion werden wir uns im Folgenden vor allem mit nichtlinearen partiellen Differentialgleichungen beschäftigen. Wie wir gesehen haben, führen die interessanteren und erfolgreicher Filtermethoden asymptotisch auf nichtlineare parabolische Gleichungen, weshalb eine weitere Untersuchung vor allem im nichtlinearen Fall von Interesse ist.

2.2.1 Einschub - Nichtlineare Parabolische Gleichungen

Wir beginnen mit einer Diskussion der Eigenschaften nichtlinearer parabolischer Gleichungen bzw. einiger Klassen davon. Die Lösungstheorie (Existenz und Eindeutigkeit) für Probleme der Form (2.21) ist ein schwieriger und immer noch recht aktueller Forschungszweig der angewandten Analysis. Unter relativ allgemeinen Voraussetzungen, die auch eine degenerierte Diffusion erlauben (z.B. nur in Tangentialrichtung wie beim mittleren Krümmungsfluss), benötigt es eigene Lösungskonzepte wie *Viskositätslösungen*, auf die wir hier aber nicht näher eingehen wollen.

Wir wollen hier zwei Unterklassen von (2.21) untersuchen, nämlich *nichtlineare Diffusionsfilter* der Form

$$\frac{\partial u}{\partial t} = \nabla \cdot (a(|\nabla u|^2) \nabla u), \quad u(t=0) = f, \quad (2.24)$$

und *morphologische Methoden der Form*

$$\frac{\partial u}{\partial t} = |\nabla u| B\left(\frac{\nabla u}{|\nabla u|}, \kappa(u)\right), \quad \kappa(u) = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right), \quad u(t=0) = f. \quad (2.25)$$

Die Diffusionsfilter betrachten wir auf einem Gebiet Ω mit natürlichen Randbedingungen

$$a(|\nabla u|^2) \frac{\partial u}{\partial n} = 0 \quad \text{auf } \partial\Omega \times \mathbb{R}_+, \quad (2.26)$$

die morphologischen Methoden der Einfachheit halber auf $\mathbb{R}^d \times \mathbb{R}_+$ ohne Randbedingungen. Wir werden später noch sehen, dass Randbedingungen im morphologischen Fall entweder keine Rolle spielen oder aber sehr speziell gewählt werden sollten.

Zur einfacheren Schreibweise definieren wir den Vorwärtsflussoperator

$$S_t : f \mapsto u(\cdot, t), \quad (2.27)$$

mit u der Lösung von (2.24) oder (2.26). Dahinter liegt natürlich wieder diese Annahme, dass der Fluss auch wohldefiniert ist, was wieder aus der Existenz und Eindeutigkeit der Lösung folgen würde. Wie wir schon oben bei der Wärmeleitungsgleichung gesehen haben, können wir die Zeit nicht sinnvoll umkehren, d.h. wir müssen damit rechnen, dass S_t keine stetige Inverse hat.

Die wichtigsten Eigenschaften dieser Klassen von nichtlinearen Gleichungen sind:

Maximumprinzip: Die allgemeine nichtlineare parabolische Gleichung (2.21) erfüllt ein Maximumprinzip, wenn F monoton von der zweiten Ableitung abhängt, d.h. $F(p, X) \geq F(p, Y)$ für $X - Y$ positiv semidefinit, und $F(0, 0) = 0$ erfüllt. Das parabolische Maximumprinzip sagt für solche Gleichungen aus, dass die Funktion ihr Maximum am parabolischen Rand, d.h. in $\Omega \times \{0\} \cup \partial\Omega \times (0, T)$ annimmt. Die starke Version des Maximumprinzips (unter stärkeren Annahmen) sagt sogar aus, dass das Maximum und Minimum nur dort angewendet werden können. Die Beweisidee ist ein Widerspruchsbeweis. Hätte man eine glatte Lösung (zumindest C^2 im Raum, C^1 in der Zeit), dann würde für ein Maximum (x, t) in $\Omega \times (0, T]$ gelten, dass $\frac{\partial u}{\partial t} \geq 0$ und $\nabla u = 0$ ist. Weiter ist für ein Maximum die Hesse-Matrix negativ semidefinit, d.h. $D^2u \leq 0$. Damit folgt aus der Gleichung

$$0 \leq \frac{\partial u}{\partial t}(x, t) = F(\nabla u(x, t), D^2u(x, t)) = F(0, D^2u(x, t)) \leq F(0, 0) = 0.$$

Dies ist noch kein Widerspruch, durch richtige Approximation von F und u kann man aber erreichen, dass $F_\epsilon(0, D^2u_\epsilon(x, t)) < 0$ gilt und erhält für die Approximation dann das starke Maximumprinzip. Im Grenzwert der Approximation bleibt dann nur das schwache Maximumprinzip erhalten.

Die vielleicht wichtigste Folgerung aus dem Maximumprinzip ist das *Vergleichsprinzip* (Comparison Principle). Hierbei betrachtet man zwei Lösungen u_1 und u_2 der Gleichung mit den selben Randwerten, aber verschiedenen Anfangswerten f_1 und f_2 . Das Vergleichsprinzip (oder auch Monotonieeigenschaft der Evolution) garantiert dann, dass für $f_1 \geq f_2$ in Ω auch $u_1 \geq u_2$ in $\Omega \times R_+$ gilt.

Das Vergleichsprinzip kann in der Bildverarbeitung z.B. bedeuten, dass der Grauwert im vernünftigen positiven Bereich bleibt. Solange $F(0, 0) = 0$ gilt, ist ja offensichtlich jede konstante Funktion eine Lösung von (2.21). Also kann man z.B. $f_2 = u_2$ konstant gleich dem minimalen Grauwert wählen, und somit ist für jede Lösung mit sinnvollem Anfangswert auch für spätere Zeit der Grauwert grösser diesem Minimalwert.

Erhaltungsgrößen und Invarianz: Diffusionsfilter und morphologische Methoden unterscheiden sich vor allem in Bezug auf ihre Invarianzeigenschaften bzw. mögliche Erhaltungsgrößen. Unter Invarianz bezüglich einer speziellen (invertierbaren) Operation Φ auf den Bildern verstehen wir

$$\Phi(S_t(f)) = S_t(\Phi(f)). \tag{2.28}$$

Dies bedeutet, dass es egal ist, ob zuerst die Transformation oder die Glättung mit der parabolischen Gleichung durchgeführt wird. Interessante Invarianzeigenschaften sind:

- *Grauwert-Verschiebungsinvarianz* wird beschrieben durch $\Phi(f) = f + c$ für eine Konstante $c \in \mathbb{R}$. Da die Funktionen selbst nicht in der Gleichung (2.21) und etwaigen Randbedingungen auftreten, verschiebt sich auch die Lösung u um die Konstante c bei Anfangswert $f + c$. Die PDE-Methoden sind damit in jedem Fall invariant bezüglich Verschiebung des Grauwerts.

- *Translationsinvarianz* wird beschrieben durch $\Phi(f(x)) = f(x + x_0)$ für ein $x_0 \in \mathbb{R}^d$. Da in der Gleichung (2.21) x nicht explizit auftritt, sondern nur Ableitungen nach x , die bei der Translation unverändert bleiben, erhält man die Translationsinvarianz von (2.21)
- *Rotationsinvarianz* wird beschrieben durch $\Phi(f(x)) = f(Qx)$ mit einer Rotationsmatrix Q (orthogonal mit Determinante eins). Aus der Kettenregel erhält man nach der Rotation mit Q den Transformierten Gradienten $\bar{\nabla} = Q^T \nabla$ und analog die transformierte Divergenz $\bar{\nabla} \cdot = \nabla \cdot Q$. Bei einem Diffusionsfilter erhält man für skalares a

$$\bar{\nabla} \cdot (a(|\bar{\nabla}|^2) \bar{\nabla} u) = \nabla \cdot (Qa(|Q^T \nabla u|^2) Q^T \nabla u) = \nabla \cdot (a(|\nabla u|^2) \nabla u)$$

wobei wir benutzt haben, dass Q normerhaltend ($|Q^T \nabla u| = |\nabla u|$) ist und $QQ^T = I$. Bei einer morphologischen Methode sieht man analog, dass die Krümmung $\kappa(u)$ und der Vorfaktor $|\nabla u|$ unter Rotation invariant sind. Nicht notwendigerweise invariant ist eine morphologische Methode bei expliziter Abhängigkeit von der Normalen an die Isokontur $\frac{\nabla u}{|\nabla u|}$.

- *Morphologische Invarianz* wird beschrieben durch $\Phi(f) = G(f)$ mit einer monoton wachsenden Funktion G . Wie schon der Name andeutet sind morphologische Methoden unter einer solchen Grauwerttransformation invariant. Sei $v = G(u)$, dann gilt

$$\frac{\partial v}{\partial t} = G'(u) \frac{\partial u}{\partial t} = G'(u) |\nabla u| B\left(\frac{G'(u) \nabla u}{G'(u) |\nabla u|}, \kappa(u)\right) = |\nabla v| B\left(\frac{\nabla v}{|\nabla v|}, \kappa(v)\right)$$

und da $v(t=0) = G(f)$ folgt $v = \Psi(S_t(f)) = S_t(\Psi(f))$. Bei Diffusionsfiltern gilt die morphologische Invarianz hingegen nicht, wie man meist schon bei einfachen linearen Transformationen $G(s) = cs$ sieht.

Neben Invarianzeigenschaften sind auch Erhaltungseigenschaften interessant. Als Erhaltungsgrösse bezeichnet man ein Funktional $\Psi(u) \in \mathbb{R}$, wenn

$$\Psi(u(\cdot, t)) = \Psi(f) \quad \forall t \geq 0$$

gilt. Ein klassisches Beispiel für eine Erhaltungsgrösse ist der mittlere Grauwert

$$\Psi(u) = \frac{1}{|\Omega|} \int_{\Omega} u(x, t) \, dx,$$

der z.B. bei Diffusionsfiltern erhalten wird. Mit dem Gauss'schen Integralsatz folgt

$$\frac{d}{dt} \Psi(u(\cdot, t)) = \frac{1}{|\Omega|} \int_{\Omega} \frac{\partial u}{\partial t} \, dx = \int_{\Omega} \nabla \cdot (a \nabla u) \, dx = \int_{\partial \Omega} a \nabla u \cdot n \, d\sigma = 0,$$

wobei wir die homogenen Neumann-Randbedingungen für u verwendet haben. Weitere Erhaltungsgrößen sind eher spezielle Eigenschaften diverser Gleichungen.

Dissipation und Glättung: Wichtig für das Entrauschen sind natürlich vor allem Glättungseigenschaften, die man auch über Dissipation von Energiefunktionalen E . Eine Dissipationsgleichung hat die Form

$$\frac{d}{dt} E[u(\cdot, t)] = -D[u(\cdot, t)], \quad (2.29)$$

wobei D ein nichtnegatives Dissipationsfunktional. Insbesondere ist E dann eine Lyapunov-Funktional für die Evolution, d.h. $E[u(\cdot, t)]$ ist monoton nicht-ansteigend in der Zeit t . Das Energiefunktional sowie das Dissipationsfunktional können einiges über die Glättung des Verfahrens aussagen. Dies sieht man z.B. bei Diffusionsfiltern bei Funktionalen der Form

$$E[u] = \int_{\Omega} A(|\nabla u|^2) dx,$$

die für eine positive Funktion F ein Maß für die Grösse der Gradienten und damit die Glattheit darstellen. Es gilt ja, falls $A' = a$

$$\begin{aligned} \frac{d}{dt} E[u] &= 2 \int_{\Omega} A'(|\nabla u|^2) \nabla u \nabla \frac{\partial u}{\partial t} dx \\ &= 2 \int_{\Omega} a(|\nabla u|^2) \nabla u \nabla (\nabla \cdot (a(|\nabla u|^2) \nabla u)) dx \\ &= -2 \int_{\Omega} (\nabla \cdot (a(|\nabla u|^2) \nabla u))^2 dx \leq 0. \end{aligned}$$

Durch die Abnahme des Funktionals E wird also die Lösung im Verlauf der Zeit glatter.

Abgesehen vom obigen Energiefunktional erhält man auch Dissipation für Funktionale der Form

$$E[u] = \int_{\Omega} B(u) dx$$

für eine konvexe Funktion B . Es gilt ja

$$\begin{aligned} \frac{d}{dt} E[u] &= \int_{\Omega} B'(u) \frac{\partial u}{\partial t} dx = \int_{\Omega} B'(u) \nabla \cdot (a(|\nabla u|^2) \nabla u) dx \\ &= - \int_{\Omega} a(|\nabla u|^2) B''(u) |\nabla u|^2 dx \leq 0. \end{aligned}$$

Im Fall morphologischer Methoden erhält man meist Dissipation für geometrische Funktionale wie z.B. die Länge der Isokonturen. Dies werden wir noch für den mittleren Krümmungsfluss näher sehen.

Langzeitverhalten: Das Langzeitverhalten beschreibt den Grenzwert der Funktion für $t \rightarrow \infty$. Die meisten Methoden zur Bildverarbeitung wie Diffusionsfilter liefern im Langzeitverhalten eine homogene Grauwertverteilung. Dies sieht man auch aus den Dissipationseigenschaften, durch Integration in der Zeit folgt

$$E[u(\cdot, t)] + \int_0^t D[u(\cdot, s)] ds \leq E[f].$$

Da das Integral über $D[u(\cdot, s)]$ gleichmässig in der Zeit beschränkt ist, folgt sofort die Existenz einer Teilfolge mit $D[u(\cdot, t_k)] \rightarrow 0$. Man sieht aus der ersten Wahl von Energie- und Dissipationsfunktional, dass dann im Grenzwert $\nabla \cdot (a(|\nabla u|^2) \nabla u) = 0$ erfüllt ist, d.h. das Langzeitverhalten wird durch die stationäre Version der Gleichung erfüllt.

2.2.2 Numerische Methoden

Die numerische Lösung partieller Differentialgleichungen ist eigentlich ein Thema für eine eigene Vorlesungsreihe, und eine detaillierte Betrachtung und Analyse würde den Rahmen

dieser Vorlesung sprengen. Wir werden an dieser Stelle nur ein paar spezifische Aspekte bei der Bildverarbeitung ansprechen.

Da digitale Bilder in Pixeln, d.h. auf einem regelmässigen kartesischen Gitter gegeben sind, werden üblicherweise finite Differenzenmethoden in diesem Zusammenhang verwendet, d.h. man ersetzt die Differentialoperatoren durch Differenzenoperatoren auf dem Gitter. Analog kann man die Zeitableitung durch einen Differenzenquotienten approximieren.

In den meisten Fällen ist man bei der Bildverarbeitung nur an einer Verbesserung des Bildes (Entrauschen) interessiert, und die klassischen Fragen der numerischen Analysis wie die Konvergenz der diskreten Methoden gegen die Lösung der Differentialgleichung spielen eine eher untergeordnete Methoden. Wichtiger ist, dass die diskreten Methoden die Eigenschaften der Differentialgleichung, wie Glättung oder Kantenschärfung, gut erhalten. Deshalb verwendet man meist nicht Verfahren höherer Ordnung (in Bezug auf Konsistenz und Konvergenz), sondern sehr einfache Verfahren meist erster Ordnung. Da man auch meist die Gleichungen nur für kurze Zeitintervalle lösen möchte, ist es weniger wichtig grosse Zeitschritte zu machen. Deshalb sind in der Bildverarbeitung explizite Zeitdiskretisierungen wie vor allem das explizite Euler-Verfahren der Form

$$u(x_{ij}, t_{k+1}) = u(x_{ij}, t_k) + \tau F_h(g(x_{ij}, t_k), H(x_{ij}, t_k)) \quad (2.30)$$

weit verbreitet. Hier sind $g(x_{ij}, t_k)$ und $H(x_{ij}, t_k)$ geeignete Approximationen des Gradienten bzw. der Hessematrix von $u(\cdot, t_k)$ im Gitterpunkt x_{ij} , und F_h ist eine mögliche Approximation der Nichtlinearität F . Da die Differenzenquotienten von den Werten von $\mathbf{u}(t_k) := (u(x_{ij}, t_k))$ and den umliegenden Gitterpunkten abhängen, kann man für jede spezifische Wahl der Differenzenquotienten das Verfahren in der Form

$$u(x_{ij}, t_{k+1}) = u(x_{ij}, t_k) + \tau G_{ij}(\mathbf{u}(t_k)) \quad (2.31)$$

schreiben.

Die Konsistenz eines solchen Verfahrens lässt bei vernünftiger Wahl von F_h durch Taylor-Entwicklung zeigen. Noch wichtiger als die Konsistenz ist die Stabilität der diskreten Verfahren. Man ist ja nicht unbedingt an der exakten Approximation der Differentialgleichung, sondern eher an den analogen Glättungseigenschaften interessiert, und daher wäre es besonders ungünstig, wenn die Glättung durch Instabilitäten zerstört würde. Eine Möglichkeit die Stabilität einer Diskretisierung mit finiten Differenzen zu analysieren, sind Monotonietechniken. Ein Verfahren ist dann monoton, wenn das Vergleichsprinzip auch nach der Diskretisierung gilt. Monotonie des Verfahrens kann unter einer geeigneten Monotonieeigenschaft von G_{ij} erreicht werden, nämlich

$$\frac{\partial G_{ij}}{\partial x_{ij}}(p) \leq 0 \quad \text{für } i, j = 1, \dots, N, p \in \mathbb{R}^{N \times N} \quad (2.32)$$

$$\frac{\partial G_{ij}}{\partial x_{\ell m}}(p) \geq 0 \quad \text{für } i, j, \ell, m = 1, \dots, N, (i, j) \neq (\ell, m), p \in \mathbb{R}^{N \times N}. \quad (2.33)$$

Diese Eigenschaft ist für parabolische Probleme sehr natürlich, wie man wieder im Fall der Wärmeleitung sofort sieht. Verwendet man einen Standard Fünf-Punkt Differenzenquotienten, so führt dies auf den oben näher untersuchten Filter

$$u_{ij}^{k+1} = (1 - 4\alpha)u_{ij}^k + \alpha(u_{i-1j}^k + u_{i+1j}^k + u_{ij-1}^k + u_{ij+1}^k)$$

wobei $\alpha = \frac{\tau}{h^2}$. Es gilt also in der Notation von (2.31) $\frac{\partial G_{ij}}{\partial x_{ij}} \equiv -\frac{4}{h^2} < 0$ und $\frac{\partial G_{ij}}{\partial x_{\ell m}} \equiv \frac{1}{h^2} > 0$. Mit der Monotonieeigenschaft (2.32), (2.33) erhält man ein stabiles Verfahren, wenn man den Zeitschritt klein genug wählt:

Satz 2.1. *Die Funktion $G_{ij} : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}$ sei stetig differenzierbar und (2.32), (2.33) seien erfüllt. Dann ist das Verfahren (2.31) monoton und folglich stabil in der Maximums-Norm, falls der Zeitschritt τ die Bedingung*

$$\tau \max_{i,j} \sup_V \left| \frac{\partial G_{ij}}{\partial V_{ij}}(V) \right| \leq 1 \quad (2.34)$$

erfüllt.

Beweis. Seien \mathbf{u} und \mathbf{v} zwei Lösungen, so dass $u_{ij}(t_k) \leq v_{ij}(t_k)$ für alle i, j gilt. Dann folgt aus der Differenzierbarkeit von G

$$\begin{aligned} u_{ij}(t_{k+1}) - v_{ij}(t_{k+1}) &= u_{ij}(t_k) - v_{ij}(t_k) + \tau(G_{ij}(\mathbf{u}(t_k)) - G_{ij}(\mathbf{v}(t_k))) \\ &= \int_0^1 [(u_{ij}(t_k) - v_{ij}(t_k)) + \tau \nabla G_{ij}((1 - \sigma)\mathbf{v}(t_k) + \sigma\mathbf{u}(t_k)) \cdot (\mathbf{u}(t_k) - \mathbf{v}(t_k))] d\sigma \\ &= \int_0^1 (1 + \tau \frac{\partial G_{ij}}{\partial V_{ij}}((1 - \sigma)\mathbf{v}(t_k) + \sigma\mathbf{u}(t_k))) d\sigma (u_{ij}(t_k) - v_{ij}(t_k)) \\ &\quad + \tau \sum_{(\ell, m) \neq (i, j)} \int_0^1 \frac{\partial G_{ij}}{\partial V_{\ell m}}((1 - \sigma)\mathbf{v}(t_k) + \sigma\mathbf{u}(t_k)) d\sigma (u_{\ell m}(t_k) - v_{\ell m}(t_k)). \end{aligned}$$

Wegen (2.34) folgt

$$1 + \tau \frac{\partial G_{ij}}{\partial V_{ij}}((1 - \sigma)\mathbf{v}(t_k) + \sigma\mathbf{u}(t_k)) \geq 0$$

und wegen $u_{ij}(t_k) \leq v_{ij}(t_k)$ ist der erste Term nichtpositiv. Aus (2.33) und $u_{\ell m}(t_k) \leq v_{\ell m}(t_k)$ folgt auch die Nichtpositivität des zweiten Terms und damit $u_{ij}(t_{k+1}) \leq v_{ij}(t_{k+1})$. Also ist das Verfahren monoton. Die Stabilität in der Supremumsnorm folgt dann sofort durch den Vergleich mit konstanten Lösungen. \square

Bei den parabolischen Gleichungen ist die Bedingung (2.34) sehr restriktiv und impliziert $\tau = \mathcal{O}(h^2)$. Da man in der Bildverarbeitung aber meist nicht zu weit in der Zeit gehen muss, um die nötige Glättung zu erreichen, ist dies akzeptabel und effizienter zu realisieren als ein implizites Verfahren mit grösseren Zeitschritten.

2.2.3 Fallstudie - Perona-Malik Modell

Als wichtiges Beispiel einer PDE-Methode betrachten wir das Perona-Malik Modell

$$\frac{\partial u}{\partial t} = \nabla \cdot (a(|\nabla u|^2) \nabla u), \quad u(t=0) = f, \quad a(p) = \frac{1}{1 + bp} \quad (2.35)$$

wobei $b > 0$ ein Gewichtungsparmeter ist. Die Hauptmotivation des Perona-Malik Modells ist es, eine Glättung in den Regionen mit kleiner Grauwertvariation durch Diffusion zu erreichen (dort ist $\nabla u \approx 1$), aber die Kanten nicht zu verschmieren ($a(|\nabla u|^2) \rightarrow 0$ für $|\nabla u| \rightarrow \infty$). Noch

besser lässt sich das Perona-Malik Modell verstehen, wenn wir (analog zu den Nachbarschaftsfiltern) die Diffusion wieder in einen Tangential- und Normalanteil entlang der Isokonturen aufspalten. Die Tangente und Normale sind dabei wie oben durch

$$\mathbf{t} = \frac{\nabla u^\perp}{|\nabla u|}, \quad \mathbf{n} = \frac{\nabla u}{|\nabla u|}$$

gegeben. Rechnen wir die Divergenz in (2.35) aus, so folgt

$$\begin{aligned} \nabla \cdot (a(|\nabla u|^2)\nabla u) &= a(|\nabla u|^2)\Delta u + 2a'(|\nabla u|^2)\nabla u \cdot (D^2u)\nabla u \\ &= A(|\nabla u|^2)\mathbf{n} \cdot (D^2u\mathbf{n}) + B(|\nabla u|^2)\mathbf{t} \cdot (D^2u\mathbf{t}), \end{aligned}$$

mit

$$A(p) = a(p) + 2a'(p)p = \frac{1 - bp}{(1 + bp)^2}, \quad B(p) = a(p) = \frac{1}{1 + bp}.$$

Man sieht sofort, dass B immer positiv ist, während A bei grossen Werten des Gradienten negativ wird. Damit hat man ein analoges Verhalten wie bei der Asymptotik des Nachbarschaftsfilters, nämlich Vorwärtsdiffusion in tangentialer Richtung und Rückwärtsdiffusion in normaler Richtung bei grossen Gradienten. Durch die Rückwärtsdiffusion in normaler Richtung kommt es wiederum zu einer Schärfung der Kanten.

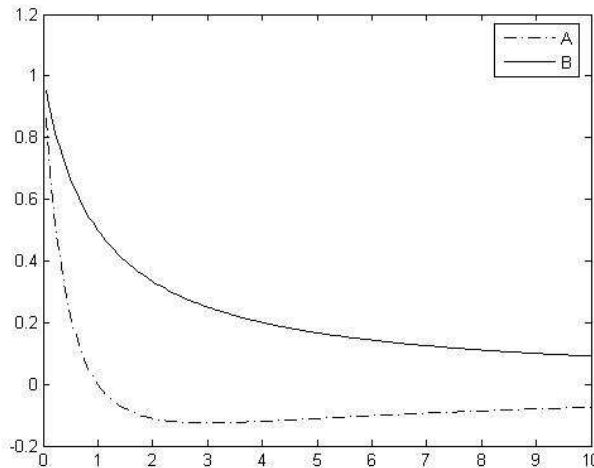


Abbildung 2.4: Funktionen A und B beim Perona-Malik Modell für $b = 1$.

Aus mathematischer Sicht ist die Rückwärtsdiffusion ein grosses Problem, wie wir schon im linearen Fall gesehen haben, kann dann die Existenz einer Lösung nicht mehr garantiert werden. Auch wenn die Lösung existiert, können immer noch kleine Oszillationen verstärkt werden, was gerade bei Rauschen ein grosses Problem ist. Um dies zu vermeiden, kann man eine analoge Modifikation des Perona-Malik Modells wie bei den obigen Filter betrachten,

$$\frac{\partial u}{\partial t} = \nabla \cdot (a(|\nabla G_\epsilon * u|^2)\nabla u), \quad u(t=0) = f, \quad a(p) = \frac{1}{1 + bp}. \quad (2.36)$$

Dieser Ansatz wurde von Catte, Coll, Lions und Morel [2] untersucht, und die Existenz einer Lösung kann bei sinnvoller Wahl des Faltungskerns G garantiert werden. Damit kann auch das

Rauschen besser unterdrückt werden. Die Kanten werden für sehr kleine Zeiten dann eher ein bisschen ausgeglättet, für etwas längere Zeit werden dann die Kanten geschärft. In diesem Fall ist es auch interessant das Langzeitverhalten zu betrachten, da man für grosse Zeiten einen Cartoon des Bildes erhält, d.h. die Texturen und kleinen Oszillationen sind verschwunden und nur die Kanten grösserer Objekte bleiben übrig.

2.2.4 Fallstudie - Mittlerer Krümmungsfluss

Der mittlere Krümmungsfluss basiert auf einem morphologischen Ansatz, man betrachtet zunächst die Glättung der Isokonturen. Bei zweidimensionalen Bildern sind dies Kurven und man erreicht eine Glättung in dem man diese Kurven in einem Geschwindigkeitsfeld der Form $V = -\kappa \mathbf{n}$ bewegt, wobei κ die Krümmung und \mathbf{n} die Normale ist. Dadurch werden konkave Bereiche ($\kappa < 0$) nach aussen bewegt und konvexe ($\kappa > 0$) nach innen. Dadurch wird natürlich die Oszillation dieser Kurven geringer. Wie wir schon gesehen haben, können wir die Normale und die Krümmung als

$$\mathbf{n} = \frac{\nabla u}{|\nabla u|}, \quad \kappa = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right)$$

schreiben und somit direkt aus dem Grauwert berechnen. Dabei benutzt man für Krümmung die Formel $\kappa = \nabla \cdot \mathbf{n}$ benutzt, die aus der Frenet'schen Formel und einer Ausdehnung der Normalen weg von der Kurve folgen. Nun wäre es natürlich ideal, die gesamte Änderung des Bildes direkt über den Grauwert berechnen zu können, ohne die Isokonturen zu rekonstruieren und diese explizit zu bewegen. Dazu betrachten wir einfach den Übergang von der Lagrange'schen Betrachtungsweise (den einzelnen Punkten auf den Isokonturen folgend) zur Euler'schen Betrachtungsweise. Bei letzterer fixieren wir einen Punkt (Pixel) und betrachten die Änderung des Grauwerts in diesem Punkt während der Evolution durch den Krümmungsfluss. Sei also der Grauwert im Punkt x zur Zeit t gegeben durch $u(x, t)$, dann gilt (aus der Lagrange'schen Sicht) für $s \geq t$

$$\frac{d}{ds} u(\xi(s), s) = 0, \quad \xi(t) = x, \quad \frac{d\xi}{ds}(s) = V(\xi(s), s),$$

wobei V die durch Krümmung und Normale definierte Geschwindigkeit ist. Man berechnet mit der Kettenregel

$$\frac{d}{ds} u(\xi(s), s) = \frac{\partial}{\partial t} u(\xi(s), s) + \nabla u(\xi(s), s) \cdot \frac{d\xi}{ds}(s).$$

Verwenden wir diese Identitäten für $t = s$ und setzen auch noch die Form des Geschwindigkeitsfelds als $V = -\kappa \mathbf{n}$ ein, so folgt

$$0 = \frac{\partial u}{\partial t}(x, t) + V(x, t) \cdot \nabla u(x, t) = \frac{\partial u}{\partial t} - \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \frac{\nabla u}{|\nabla u|} \cdot \nabla u.$$

Durch weiteres Ausmultiplizieren erhält man daraus die Gleichung für den mittleren Krümmungsfluss als

$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right). \quad (2.37)$$

Wegen seiner morphologischen Konstruktion hat der mittlere Krümmungsfluss eine Menge wichtiger Invarianzen, unter anderem hat der Fluss alle oben erwähnten Invarianzen.

Der mittlere Krümmungsfluss ist kein Diffusionsfilter in der obigen Form, da er nicht in Divergenzform geschrieben werden kann. Dennoch hat er ähnliche Dissipationseigenschaften, vor allem wenn man die Form

$$\frac{1}{|\nabla u|} \frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right)$$

interpretiert. Man kann dann (unter geeigneten Randbedingungen) zeigen, dass das Energiefunktional

$$E[u] = \int |\nabla u| \, dx$$

in der Zeit abnimmt. Wegen der morphologischen Invarianz erhält man auch, dass

$$E_\phi[u] = \int |\nabla \phi(u)| \, dx = \int \phi'(u) |\nabla u| \, dx$$

in der Zeit abnimmt, für jede monotone Funktion ϕ . Aus der Coarea-Formel folgt die äquivalente Darstellung

$$\int \phi'(u) |\nabla u| \, dx = \int_{\mathbb{R}} \phi'(s) \int_{\partial\{u < s\}} d\sigma \, ds.$$

Insbesondere kann man nun ϕ als Approximation der Heaviside-Funktion $H(\cdot - \alpha)$ wählen bzw. ϕ' als Approximation der bei α zentrierten Dirac-delta Distribution $\delta(\cdot - \alpha)$. Da alle diese Funktionale in der Zeit abnehmen, gilt dies auch für den jeweiligen Limes

$$\int_{\mathbb{R}} \delta(s - \alpha) \int_{\partial\{u < s\}} d\sigma \, ds = \int_{\partial\{u < \alpha\}} d\sigma.$$

Damit wird die Länge aller Isokonturen während des Flusses kleiner, eine spezielle morphologische Glättungseigenschaft.

2.2.5 Variationsmethoden und PDE

Zum Abschluss unserer Betrachtung von PDE-Methoden wollen wir noch die Verbindung mit bzw. den Übergang zu Variationsmethoden diskutieren. Dazu betrachten wir zunächst voll implizite Diskretisierungen für Diffusionsfilter, d.h. der Zeitschritt t_k wird aus dem letzten Zeitschritt durch Lösung der Gleichung

$$\frac{1}{\tau} (u(t_k) - u(t_{k-1})) = \nabla \cdot (a(|\nabla u(t_k)|^2) \nabla u(t_k)). \quad (2.38)$$

Der Einfachheit halber benutzen wir die Notation u für $u(t_k)$ und v für $u(t_{k-1})$. Wir betrachten zunächst den Differentialoperator auf der rechten Seite. In schwacher Form (d.h. nach Multiplikation mit einer Testfunktion und Integration) gilt

$$- \int_{\Omega} \nabla \cdot (a(|\nabla u|^2) \nabla u) \varphi \, dx = \int_{\Omega} a(|\nabla u|^2) \nabla u \cdot \nabla \varphi \, dx,$$

für jede Testfunktion φ . Wie man leicht nachrechnet, kann die rechte Seite mit der Richtungsableitung des Funktionals

$$E[u] := \frac{1}{2} \int_{\Omega} R(|\nabla u|^2) \, dx$$

identifiziert werden, wobei R die Stammfunktion von a ist. Um auch den rechten Term zu berücksichtigen addieren wir zum Energiefunktional E noch einen kinetischen Term und definieren ein Funktional

$$J_\tau[u] := E[u] + \frac{1}{2\tau} \int_{\Omega} (u - v)^2 dx.$$

Da die rechte Seite in (2.38) mit der Ableitung des kinetischen Terms identifiziert werden kann, ist (2.38) die Optimalitätsbedingung erster Ordnung $J'_\tau[u] = 0$. Da der kinetische Term offensichtlich konvex ist charakterisiert die Optimalitätsbedingung für konvexes R ein Minimum des Funktionals J_τ . Also kann der Zeitschritt $u(t_k)$ als Minimum von J_τ berechnet werden, bei gegebenem Zeitschritt $v = u(t_{k-1})$. Man beachte, dass die Minimierung von J_τ eine analoge Form wie bei den MAP-Schätzern in der Einleitung hat, man iteriert also eine Variationsmethode. Dies wird noch deutlicher beim ersten Zeitschritt (wo $v = f$ ist), mit der Notation $\lambda = \frac{1}{\tau}$. Das Minimierungsproblem, das $u = u(\tau)$ liefert ist dann

$$\frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx + \frac{1}{2} \int_{\Omega} R(|\nabla u|^2) dx \rightarrow \min_u,$$

u ist also durch ein Variationsproblem definiert, in dem ein zusammengesetztes Funktional aus Datenfitting und Regularisierung minimiert wird. Da eine implizite Zeitdiskretisierung unbedingt stabil ist und man die PDE nur für relativ kleine Zeitintervalle $(0, T)$ lösen will, könnte man versuchen $\tau = T$ zu wählen und nur einen Zeitschritt durchzuführen, was direkt auf eine Variationsmethode führt. Diese werden wir im nächsten Kapitel genauer betrachten.

2.3 Variationsmethoden

Im folgenden werden wir uns mit Variationsmethoden zum Entrauschen befassen und auch die dabei benötigten mathematischen Grundlagen diskutieren. Der Fokus liegt dabei wieder auf einem Gauss'schen Modell für das Rauschen. Wir nehmen an es ist eine verrauschte Version f eines Bilds u_0 gegeben, wobei die Varianz durch

$$\int_{\Omega} (u - f)^2 dx \leq \sigma^2 \tag{2.39}$$

Im Sinne eines MAP-Schätzers (bzw. der zugehörigen Log-Likelihood Funktion) addiert man bei einer Variationsmethode eine regularisierende Energie zum quadratischen Fitting-Funktional. Um Glättung durch die Regularisierung zu erreichen, versucht man normalerweise eine Gradienten-Energie zur Regularisierung zu verwenden, d.h. man minimiert ein Funktional der Form

$$J(u) = \frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx + \int_{\Omega} F(x, \nabla u) dx \tag{2.40}$$

mit einer nichtlinearen Funktion $F : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$. Der positive Parameter λ misst dabei die Gewichtung zwischen dem Fitting-Term und der Regularisierung, er sollte abhängig von der Rauschvarianz σ gewählt werden ($\lambda \rightarrow 0$ für $\sigma \rightarrow 0$).

Ein einfaches Beispiel für die Funktion F ist wiederum $F(x, p) = \frac{1}{2}|p|^2$. In diesem Fall kann man zeigen, dass das Minimum die lineare Gleichung

$$\lambda(u - f) - \Delta u = 0 \tag{2.41}$$

erfüllt. Auch die natürlichen Randbedingungen für die Minimierung lassen sich leicht herleiten. Dazu berechnen wir wieder Richtungsableitungen, denn für das Minimum gilt ja (als Minimum und damit stationärer Punkt in jeder Richtung)

$$0 = \frac{d}{dt} J(u + tv)|_{t=0} = \lambda \int_{\Omega} (u - f)v \, dx + \int_{\Omega} \int_{\Omega} \nabla u \cdot \nabla v \, dx.$$

Aus dem Gauss'schen Satz folgt

$$0 = \int_{\Omega} (\lambda(u - f) - \Delta u)v \, dx + \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, d\sigma.$$

Da v ja beliebig ist überlegt man sich leicht, dass sowohl der Term im Volumsintegral als auch der Randterm bei einem Minimum verschwinden müssen. Damit erhalten wir die obige Differentialgleichung und automatisch die homogene Neumann Randbedingung $\frac{\partial u}{\partial n} = 0$. Aus der Theorie der linearen partiellen Differentialgleichungen sieht man sofort, dass dann eine eindeutige Lösung, d.h. ein eindeutiges Minimum des Funktionals J existiert. Für allgemeinere Funktionale F ist eine solche Analyse wesentlich komplizierter und erfordert Methoden aus der Variationsrechnung, die wir im Folgenden kurz diskutieren.

2.3.1 Einschub - Methoden der Variationsrechnung

Die Minimierung des Funktionals J aus (2.40) ist ein Problem der Variationsrechnung, bei der man sich mit Optimierungsproblemen der Form

$$\int_{\Omega} G(x, u(x), \nabla u(x)) \, dx \rightarrow \min_u \tag{2.42}$$

beschäftigt. Das Funktional J erhält man dann als Spezialfall für $G(x, p, q) = \lambda(p - f(x))^2 + F(x, q)$.

Bei genauerer Betrachtung des Optimierungsproblems (2.42) 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 (2.42) sind wiederum Existenz und Eindeutigkeit eines minimierenden Elements (das wir auch kurz als Minimum bezeichnen), im speziellen Fall von (2.40) ist man dann auch an der Abhängigkeit der Lösung von den Daten f und vom Parameter λ 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:

Satz 2.2. *Sei $J : (\mathcal{X}, \tau) \rightarrow \mathbb{R} \cup \{+\infty\}$ ein Funktional auf einem topologischen Raum \mathcal{X} mit Topologie τ , das die beiden folgenden Bedingungen erfüllt:*

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

$$J(u) \leq \liminf_k J(u_k). \tag{2.43}$$

- **Kompaktheit von 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 . \square

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

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

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

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

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:

Satz 2.3. *Sei \mathcal{X}^* der Dualraum eines Banachraums \mathcal{X} und $C > 0$. Dann ist die Menge*

$$\{ v \in \mathcal{X}^* \mid \|v\|_{\mathcal{X}^*} \leq C \}$$

kompakt in der schwach- Topologie.*

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

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

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. Komplizierter und Interessanter ist der Fall $r = 1$, den wir unten noch genauer diskutieren werden.

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. Wir werden diese Analyse im nächsten Abschnitt exemplarisch für ein wichtiges Modell detailliert durchführen.

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

$$J(\alpha u + (1 - \alpha)v) \leq \alpha J(u) + (1 - \alpha)J(v), \quad \forall u, v \in \mathcal{X}, \forall \alpha \in [0, 1], \quad (2.44)$$

mit Gleichheit nur für $u = v$ oder $\alpha \in \{0, 1\}$.

Satz 2.4. *Sei $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$ strikt konvex, dann existiert höchstens ein globales Minimum.*

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

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

ein Widerspruch. \square

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*

$$dJ(u; v) := \lim_{t \downarrow 0} \frac{J(u + tv) - J(u)}{t}. \quad (2.45)$$

Man nennt das Funktional J differenzierbar in die Richtung v , falls der obige Grenzwert existiert und endlich ist. Die Sammlung dieser Richtungsableitungen $dJ(u; \cdot)$ 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 Funktional $J'(u) \in \mathcal{X}^*$, sodass

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

und

$$\frac{\|J(u+v) - J(u) - J'(u)v\|}{\|v\|} \rightarrow 0 \quad \text{für } \|v\| \rightarrow 0,$$

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:

$$dJ(\hat{u}; v) \geq 0 \quad \forall v \in \mathcal{X}.$$

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

$$\partial J(u) := \{ p \in \mathcal{X}^* \mid J(u) + \langle p, v - u \rangle \leq J(v) \} \quad \forall v \in \mathcal{X}. \quad (2.46)$$

Die einzelnen Elemente des Subdifferentials nennt man Subgradienten. Ist J Frechet-differenzierbar, so erhält man eine Übereinstimmung 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 $J'(u) = [-1, 1]$.

2.3.2 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 2.5. *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

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

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

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

Für beliebiges $u \in W^{1,p}(\Omega)$ 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 $1/p'$. \square

Lemma 2.6. *Sei $D \subset \Omega$ ein Gebiet mit C^1 -Rand. Dann ist die Funktion*

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

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

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

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

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 $L^p(\Omega, \mathbb{R}^d) \subset L^{p'}(\Omega, \mathbb{R}^d)^*$. \square

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 Übergang von $L^1(\Omega)$ zum Raum der Radon-Maße. Wir definieren deshalb den Raum $BV(\Omega)$ von Funktionen beschränkter (totaler) Variation

$$BV(\Omega) = \{ u \in L^1(\Omega) \mid |u|_{BV} < \infty \}, \quad (2.47)$$

wobei die totale Variation gegeben ist durch

$$|u|_{BV} := \sup_{\varphi \in C_0^\infty(\Omega; \mathbb{R}^d), \|\varphi\|_\infty \leq 1} \int_{\Omega} u \nabla \cdot \varphi \, dx. \quad (2.48)$$

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

$$\|u\|_{BV} := \|u\|_{L^1} + |u|_{BV}. \quad (2.49)$$

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 2.6 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. [13]), das in der Minimierung

$$J(u) = \frac{\lambda}{2} \int_{\Omega} (u - f)^2 \, dx + |u|_{BV} \rightarrow \min_{u \in BV(\Omega)} \quad (2.50)$$

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)

$$\|u\|_{L^1} \leq \sqrt{|\Omega|} \|u\|_{L^2}. \quad (2.51)$$

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

Lemma 2.7. *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. \square .

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

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

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

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

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

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

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

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

umgerechnet werden. Da man zeigen kann, dass $|\bar{u}| + |u|_{BV}$ 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:

Satz 2.8. *Sei $f \in L^2(\Omega)$ und $\lambda > 0$. Dann existiert ein Minimum von (2.50).*

Die totale Variation ist ein konvexes Funktional, da

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

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:

Satz 2.9. *Sei $f \in L^2(\Omega)$ und $\lambda > 0$. Dann ist das Minimum von (2.50) eindeutig.*

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

Satz 2.10. *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

$$0 = \langle 0, v - u \rangle \leq J(v) - J(u) \quad \forall v \in \mathcal{X}$$

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. \square

Das Subdifferential des Funktionals J in (2.50) kann als Summe der Ableitung des ersten (differenzierbaren) Terms und des Subdifferentials der totalen Variation berechnet werden (diese Formel ist nicht trivial, siehe [8]), d.h.

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

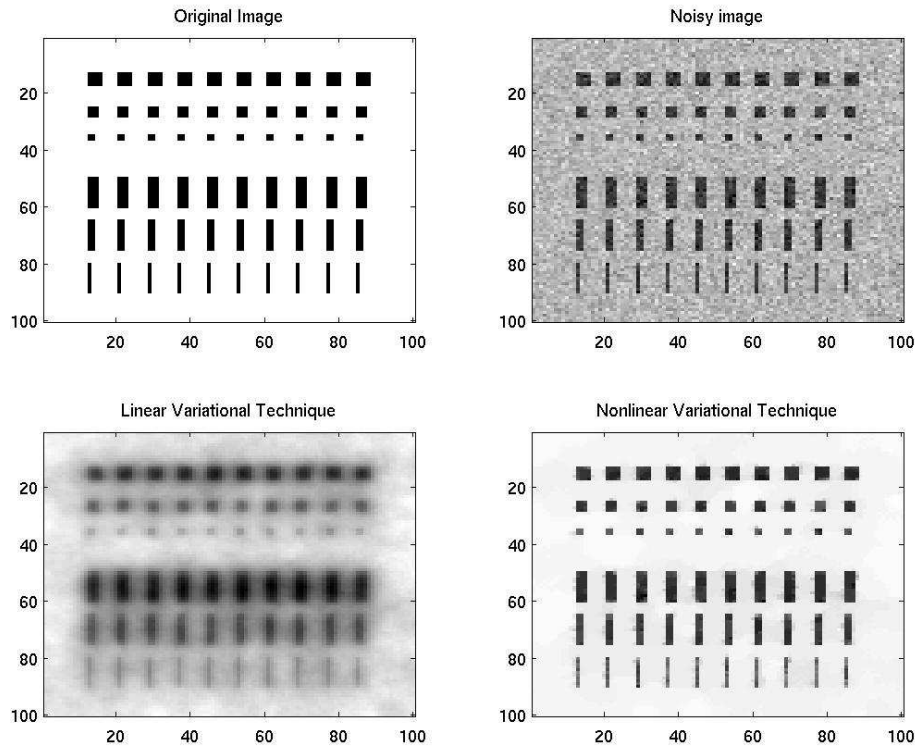


Abbildung 2.5: Entrauschen durch Variationsmethoden: Lineare Methode (2.41) im Vergleich mit nichtlinearer Methode (2.50). Sauberes Bild, verrauschtes Bild und Resultate.

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

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

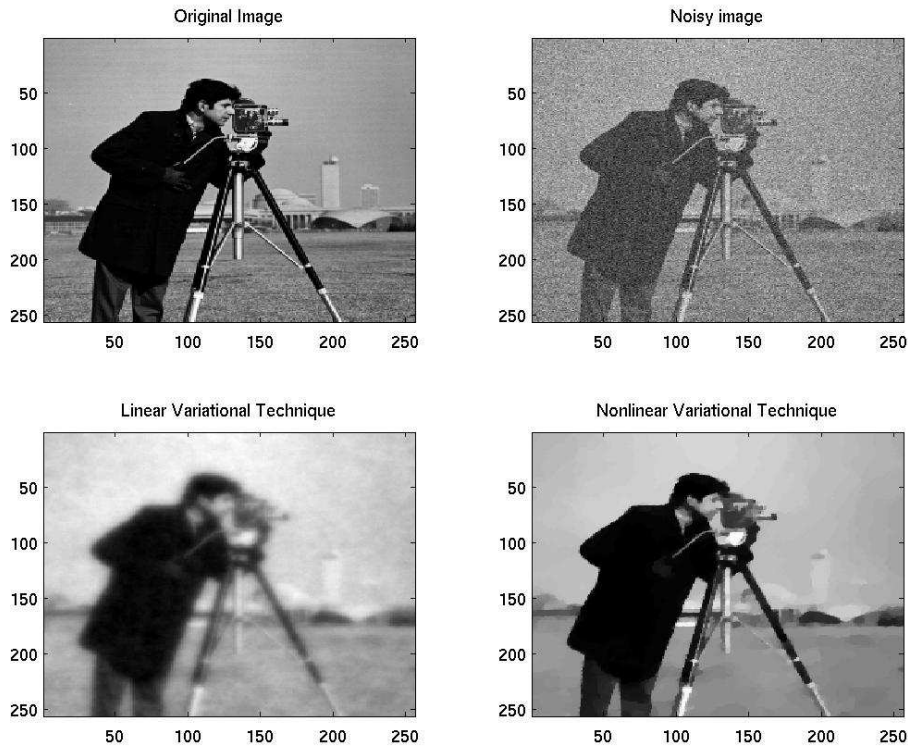


Abbildung 2.6: Entrauschen durch Variationsmethoden: Lineare Methode (2.41) im Vergleich mit nichtlinearer Methode (2.50) . Sauberes Bild, verrauschtes Bild und Resultate.

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 (2.41). 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. Dies sieht man auch aus den Vergleichen in Bildern (2.5), (2.6), (2.7).

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

Satz 2.11. *Sei u das Minimum von (2.50) für $\lambda > 0$ dann gilt $u = f$ nur für konstante Funktionen f .*

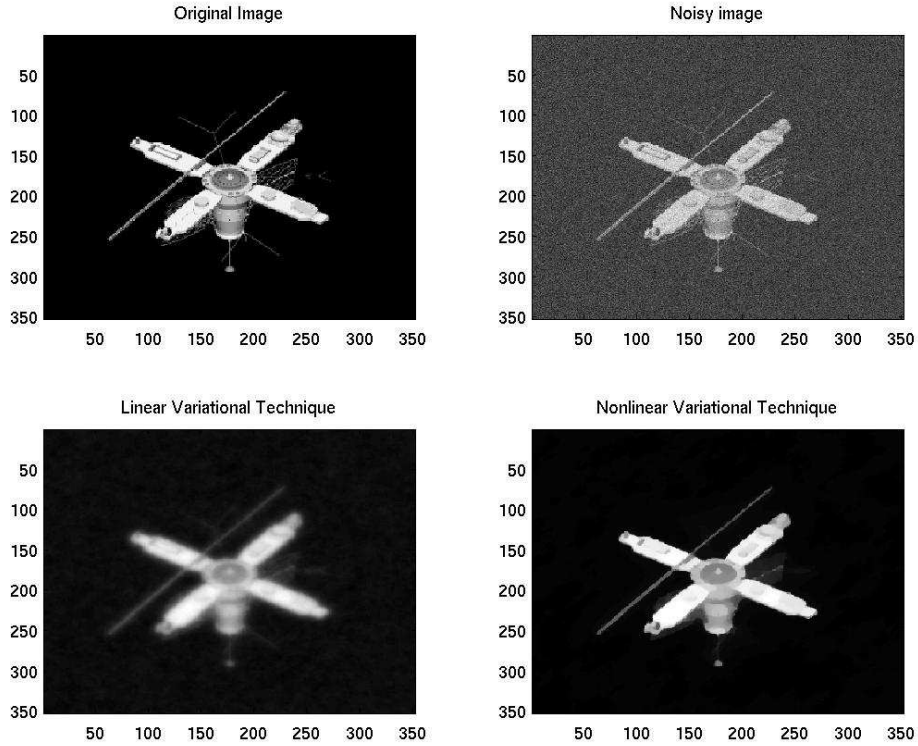


Abbildung 2.7: Entrauschen durch Variationsmethoden: Lineare Methode (2.41) im Vergleich mit nichtlinearer Methode (2.50) . Sauberes Bild, verrauschtes Bild und Resultate.

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. \square

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 größte Skala richtig aufgelöst, d.h. der Mittelwert von u ist immer gleich dem von f :

Satz 2.12. Sei u das Minimum von (2.50). Dann gilt

$$\int_{\Omega} u \, dx = \int_{\Omega} v \, dx.$$

Beweis. Sei u das Minimum von (2.50) und $v = u - \bar{u} + \bar{f}$. Dann gilt $|v|_{BV} = |u|_{BV}$ und

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

Damit erhalten wir

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

Diese Ungleichung kann nur für $\bar{u} - \bar{f}$ erfüllt sein. \square

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

$$\lambda(u - f) + \nabla \cdot g = 0 \quad \Rightarrow \quad u = f - \frac{1}{\lambda} \nabla \cdot g. \quad (2.52)$$

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

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

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

$$\frac{1}{2} \int_{\Omega} (\nabla \cdot q - f)^2 dx \rightarrow \min_q, \quad \text{mit Nebenbedingung } \|q\|_{\infty} \leq \frac{1}{\lambda}. \quad (2.53)$$

Mit (2.52) erhalten wir dann das entauschte 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

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

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 nicht-negativ ist, d.h.

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

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 (2.6) gut sichtbar.

Eine interessante Frage bei Variationsmethoden ist immer die Asymptotik $\lambda \rightarrow \infty$ bzw. $\sigma \rightarrow 0$. 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 $\lambda \rightarrow \infty$ 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

Satz 2.13. *Sei $\hat{f} \in BV(\Omega) \cap L^2(\Omega)$ und $f_n \in L^2(\Omega)$ ein verrauschtes Bild, das (2.39) 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 $\lambda_n \rightarrow \infty$ und $\lambda_n \sigma_n^2 \rightarrow 0$. Weiter bezeichne u_n das eindeutige Minimum von (2.50) mit $f = f_n$ und $\lambda = \lambda_n$. Dann gilt*

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

Beweis. Da u_n ein Minimum von (2.50) 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 (2.39) 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

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

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

Aus der obigen Abschätzung folgt weiter

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

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. \square

Zum Abschluss wollen wir noch kurz numerische Methoden zur Minimierung des ROF-Funktional (2.50) diskutieren. Dazu unterscheiden wir nach primalen, primal-dualen, oder nur dualen Ansätzen.

Primale Lösungsmethoden zielen auf die direkte Lösung und Diskretisierung des Optimierungsproblem (2.50). Hat man die Grauwerte u_{ij} im Pixel i, j gegeben, so könnte man die Integrale als Summe über die Pixel und die Ableitung durch einen finiten Differenzenquotienten approximieren, d.h.

$$J_N(U) = \frac{\lambda h}{2} \sum_{i,j=1}^N (u_{ij} - f_{ij})^2 + \sum_{i,j=1}^{N-1} \sqrt{(u_{i+1j} - u_{ij})^2 + (u_{ij+1} - u_{ij})^2 + \epsilon_N^2} \quad (2.54)$$

approximieren. Wählt man einen positiven Term ϵ_N bei der Approximation, so kann man auch Differenzierbarkeit erreichen und die Standard-Algorithmen der nichtlinearen Optimierung anwenden wie z.B. das Gradienten-Verfahren

$$U^{k+1} = U^k - \tau_k \nabla J_N(U^k),$$

mit geeigneter Schrittweite τ_k oder auch Varianten des Newton-Verfahrens. Verfahren dieser Art sind einfach zu implementieren, oft aber ineffizient vor allem für kleines ϵ_N , wo sich die Nähe zu einem nichtdifferenzierbaren Problem bemerkbar macht. Eine andere Alternative ist eine Reformulierung des Problems für $\epsilon_N = 0$, in dem ein weiteres Feld V einführt sowie die Nebenbedingungen

$$(u_{i+1j} - u_{ij})^2 + (u_{ij+1} - u_{ij})^2 \leq v_{ij}^2, \quad v_{ij} \geq 0.$$

Das Zielfunktional kann dann zu einem quadratisch-linearen (und damit wieder differenzierbaren) Funktional

$$\tilde{J}_N(U, V) = \frac{\lambda h}{2} \sum_{i,j=1}^N (u_{ij} - f_{ij})^2 + \sum_{i,j=1}^{N_1} v_{ij}$$

geändert werden. Die Tatsache, dass die Summe der v_{ij} minimiert werden soll, führt natürlich zu möglichst kleinen Werten von v_{ij} andererseits beschränken die Nebenbedingungen v_{ij} nach unten, und man erhält im Minimum tatsächlich

$$v_{ij} = \sqrt{(u_{i+1j} - u_{ij})^2 + (u_{ij+1} - u_{ij})^2}.$$

Während der Optimierung sind aber auch abweichende Werte möglich, man arbeitet also mit einer Relaxierung des ursprünglichen Problems. Zur Minimierung können dann effiziente Algorithmen der beschränkten Optimierung, wie Innere-Punkt Methoden angewandt werden.

Duale Lösungsmethoden versuchen das duale Problem (2.53) für q zu lösen. Ein Vorteil dabei ist das einfache quadratische (und insbesondere differenzierbare) Zielfunktional, ein Nachteil die Nebenbedingung. Im Gegensatz zu primalen Methoden kann aber das duale Problem einfach direkt diskretisiert werden, z.B. in dem man die Freiheitsgrade für q in die Mittelpunkte zwischen den Pixel-Werten von u gibt. Man erhält dann in 2D zwei Felder $q_{\ell k}^1$ und $q_{\ell k}^2$, die die Werte des Vektorfelds q in den Gitterpunkten beschreiben. Die Divergenz von q kann dann durch einseitige Differenzenquotienten berechnet werden, und damit effektiv als Wert im Pixel interpretiert werden. Dies erlaubt dann sinnvoll den Term $\nabla \cdot q - f$ im Zielfunktional und in der Formel für als Pixel-Wert zu berechnen. Für die Minimierung des Funktionals können dann wieder Methoden der beschränkten Optimierung verwendet werden.

Primal-Duale Lösungsmethoden versuchen die Optimalitätsbedingung zu lösen, basierend auf der Charakterisierung des Subgradienten. Das primal-duale System sucht u und g mit $\|g\|_\infty \leq 1$

$$\begin{aligned} \lambda(u - f) + \nabla \cdot g &= 0 \\ \int_{\Omega} u \nabla \cdot (g - \varphi) dx &\geq 0 \end{aligned}$$

für alle φ mit $\|\varphi\|_\infty \leq 1$, löst also das Sattelpunktsproblem für u und g . Durch Approximationen der Nebenbedingung an g und der daraus resultierenden Variationsungleichung erhält man ein Gleichungssystem, das mit Newton-artigen Verfahren gelöst werden kann.