Lecture notes

An introduction to variational image processing

Benjamin Berkels

February 23rd, 2018

1st Spring School on Models and Data
DASIV Smart State Center, Columbia, University of South Carolina

typeset using KOMA-Script and LaTeX

§ 1 Calculus of variations

Real-life problem (denoising, registration, segmentation, ...)

$$\downarrow \text{Modeling}$$

$$\text{Minimization problem } J[y^*] = \min_{y \in M} J[y], M \subset X, \dim X = \infty \longrightarrow \textit{Calculus of variations}$$

Variational approaches

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Examples *Denoising* Given a noisy image/signal $f : \Omega \to \mathbb{R}$, i.e. $f = f_0 + n$, find f_0.

$$J[u] = \underbrace{\int_\Omega (u - f)^2 \,\mathrm{d}x}_{\text{data term}} + \lambda \underbrace{\int_\Omega \|\nabla u(x)\| \,\mathrm{d}x}_{\text{regularizer}} \quad \text{(Rudin-Osher-Fatemi)}$$

Deconvolution/Deblurring Given a blurry image/signal $f : \Omega \to \mathbb{R}$, i.e. $f = Af_0$, find f_0.

$$J[u] = \int_\Omega (Au - f)^2 \,\mathrm{d}x + \lambda \int_\Omega \|\nabla u(x)\| \,\mathrm{d}x$$

Segmentation Decompose an image $f : \Omega \to \mathbb{R}$ in foreground $\mathcal{O}$ (color c_1) and background $\Omega \setminus \mathcal{O}$ (color c_2).

$$J[\mathcal{O}] = \int_{\mathcal{O}} (f - c_1)^2 \,\mathrm{d}x + \int_{\Omega \setminus \mathcal{O}} (f - c_2)^2 \,\mathrm{d}x + \lambda \operatorname{Per}(\mathcal{O}) \quad \text{(binary Mumford–Shah functional)}$$

Registration Given two images $f, g : \Omega \to \mathbb{R}$, find a deformation $\phi : \Omega \to \Omega$, such that $f \approx g \circ \phi$ holds and ϕ is smooth.

$$J[\phi] = \int_\Omega |f(x) - g(\phi(x))|^2 \,\mathrm{d}x + \lambda \int_\Omega \|D(\phi(x) - x)\|^2 \,\mathrm{d}x$$

General task

$$\begin{aligned} \text{Given:} \quad & \text{Normed vector space } (X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}, \\ \text{Find:} \quad & y^* \in M \text{ such that } J[y^*] \le J[y] \text{ for all } y \in M. \end{aligned}$$

J is often called *objective functional*, M is the *admissible set*. In the following, a vector space is always a real vector space.

The structure is very similar to classical optimization, but in contrast to optimization we have $\dim(X) = \infty$. Usually X is a function space.

Central theoretical questions

- Existence of minimizers? Direct method in the calculus of variations

  sequentially compactness of $L_\gamma(J) := \{y \in M \,:\, J[y] \leq \gamma\}$

  $+$ lower semi-continuity of J ($\Leftarrow J$ convex)

  $\Rightarrow$ Existence

  $\dim X = \infty$! Choice of X and type of convergence are crucial (weak convergence)

  Suitable spaces? $C^k(\Omega)$, $L^p(\Omega)$, $W^{m,p}(\Omega)$, $BV(\Omega)$, ...

- Necessary conditions? "First Variation = 0", Euler–Lagrange equation $\rightarrow$ PDE

- Sufficient conditions? "Convexity $\Rightarrow$ Euler–Lagrange equation sufficient"

- Uniqueness of minimizers? $\Leftarrow$ Strict convexity

§ 2 Existence of minimizers

§ *2.1 Remark.* First, J needs to be bounded from below on the admissible set, i.e.

$$\underline{J} := \inf_{x \in X} J[x] > -\infty.$$

This condition is not sufficient (e.g. $J[x] = e^x$, this J is even strictly convex and analytic), but ensure the existence of a *minimizing sequence*, i.e. $(x_n)_n \in X^{\mathbb{N}}$ with $J[x_n] \to \underline{J}$ for $n \to \infty$.

The direct method in the calculus of variations consists of the following steps

(i) Selection of a minimizing sequence $(x_n)_n \in X^{\mathbb{N}}$

(ii) Getting a convergent subsequence $(x_{n_k})_k \in X^{\mathbb{N}}$ (denoting the limit by $x^* \in X$)

(iii) Proving *lower semi-continuity* of J, i.e.

$$J[y] \leq \liminf_{n \to \infty} J[y_n] \text{ for all } (y_n)_n \in X^{\mathbb{N}} \text{ with } y_n \to y \in X.$$

This means that function values do not "jump down".

Then, x^* is a minimizer, i.e. $J[x^*] = \underline{J}$, since

$$\underline{J} = \lim_{n \to \infty} J[x_n] = \lim_{k \to \infty} J[x_{n_k}] = \liminf_{k \to \infty} J[x_{n_k}] \geq J[x^*] \geq \underline{J}.$$

§ *2.2 Example.* Let us consider a simple finite dimensional example to study the existence of minimizers, denoising in 1D after discretization. Given a noisy, discrete signal $f \in \mathbb{R}^n =: X$, the objective function is

$$J[x] = \sum_{i=1}^{n} (x_i - f_i)^2 + \lambda \sum_{i=1}^{n-1} \frac{1}{h} |x_{i+1} - x_i| \,.$$

The admissible set is X. Here, we have $J[x] \geq 0$, so we can select a minimizing sequence $(x_n)_n \in X^{\mathbb{N}}$. For all $x \in \mathbb{R}^n$ we have

$$\|x\|_2 \leq \|x - f\|_2 + \|f\|_2 = \sqrt{\|x - f\|_2^2} + \|f\|_2 \leq \sqrt{2J[x]} + \|f\|_2 \,.$$

In particular, we have

$$\|x_n\|_2 \leq \sqrt{2J[x_n]} + \|f\|_2 \leq C,$$

since $J[x_n]$ is convergent and thus bounded. That means x_n is a bounded sequence in $\mathbb{R}^n$, thus there is a convergent subsequence $(x_{n_k})_k \in X^{\mathbb{N}}$. Moreover, J is continuous and thus also lower semi-continuous. Thus, the direct method in the calculus of variations can be applied.

To show that the minimizing sequence is norm-bounded, we have used that $\|x\|$ can be bounded in terms of $J[x]$. This property is a property of the objective J and called coercivity:

§ 2.3 Definition. Let X be a normed vector space and $M \subset X$. $J : X \to \mathbb{R}$ is called *coercive* on M, if there are constants $r, C > 0$ and $\beta \geq 0$, such that

$$J[y] \geq C \|y\|^r - \beta \text{ for all } y \in M.$$

A more important property that we used in the direct method is a property of the underlying space, i.e. that bounded sequences in $\mathbb{R}^n$ have a convergent subsequence. For this, the notion of boundedness and convergence is crucial.

§ *2.4 Remark.*

(i) In finite dimensional vector spaces all norms are equivalent, i.e. if X is a vector space with $\dim(X) < \infty$ and $\|\cdot\|_a$, $\|\cdot\|_b$ norms on X, there exists $c, C \in (0, \infty)$ with

$$c\,\|x\|_a \le \|x\|_b \le C\,\|x\|_a \ \text{ for all } x \in X.$$

For infinite dimensional vector spaces, this is not true!

We will show that $\|\cdot\|_{L^\infty}$ and $\|\cdot\|_{L^2}$ are not equivalent on $X = \mathrm{C}([0,1])$. For $n \in \mathbb{N}$, let

$$f_n(x) := \begin{cases} 2nx & x \in [0, \frac{1}{2n}) \\ 2n\left(\frac{1}{n} - x\right) & x \in [\frac{1}{2n}, \frac{1}{n}) \\ 0 & x \in [\frac{1}{n}, 1] \end{cases}.$$

Obviously, $f_n \in X$ for all $n \in \mathbb{N}$. Moreover, $\|f_n\|_{L^\infty} = 1$ and

$$\|f_n\|_{L^2} \le \left(\frac{1}{n}1^2\right)^{\frac{1}{2}} = \frac{1}{\sqrt{n}}.$$

Assuming there exists $c > 0$ with $c\,\|f\|_{L^\infty} \le \|f\|_{L^2}$ for all $f \in X$. Then,

$$c = c\,\|f_n\|_{L^\infty} \le \|f_n\|_{L^2} \le \frac{1}{\sqrt{n}} \ \text{for all } n \in \mathbb{N}.$$

The implies $c \le 0$ ↯. Thus, the two norms are not equivalent.

(ii) The choice of the norm defines the notion of a neighbourhood

$$B_r(x) := \{y \in X \;:\; \|x - y\| < r\}$$

and is essential for convergence and continuity.

(iii) In infinite dimensional vector spaces, norm-bounded sequences in general do not have a convergent subsequence. One can even show that $\overline{B_1(0)}$ compact $\Leftrightarrow \dim(X) < \infty$.

To show the former, consider the sequence $(f_n) \subset L^2[0, \pi]$ given by $f_n(x) = \sin(nx)$. A straightforward computation shows $\|f_n\|_{L^2} = \sqrt{\frac{\pi}{2}}$ for all $n \in \mathbb{N}$ and $\|f_n - f_m\|_{L^2} = \sqrt{\pi}$ for $n \ne m$. Thus, (f_n) is bounded in the L^2-norm but no subsequence converges in the L^2-norm. (f_n) also does not converge pointwise.

The problem is that norm-convergence is too restrictive. The notion of weak convergence can be used to fix this problem.

§ 2.5 Definition (Dual space)**.** Let X be a normed vector space with norm $\|\cdot\|$.

(i) The set

$$X' := \left\{x' : X \to \mathbb{R} \;:\; x' \text{ linear and continuous wrt. } \|\cdot\|\right\}$$

is called *(topological) dual space* of X. Without the continuity, the set is called *algebraic dual space.*

(ii) The so-called *dual pairing* of $x \in X$ and $x' \in X'$ is defined as $\langle x', x\rangle := x'(x)$.

(iii) The dual space of the dual space is called *double dual* $X'' := (X')'$.

§ *2.6 Remark.* With

$$
\begin{aligned}
+ &: X' \times X' \to X', (x', y') \mapsto ((x'+y') : x \mapsto x'(x) + y'(x)),\\
\cdot &: \mathbb{R} \times X' \to X', (\alpha, x') \mapsto ((\alpha x') : x \mapsto \alpha x'(x)),
\end{aligned}
$$

X' is a vector space. Moreover, the so-called *operator norm*

$$
|||\cdot||| : X' \to \mathbb{R}, x' \mapsto \left|\left|\left|x'\right|\right|\right| := \sup_{\|x\|\le 1} \left|x'(x)\right|
$$

is a norm on X'. With the completeness of $\mathbb{R}$ (Cauchy sequences converge), one can show the completeness of X' wrt. $|||\cdot|||$ (exercise). Thus, X' is a Banach space (complete, normed vector space).

Moreover, it holds that

$$
\left\langle x', x\right\rangle \le \left|\left\langle x', x\right\rangle\right| \le |||x'||| \, \|x\| \, .
$$

For $x = 0$ this is obviously true, for $x \neq 0$ it holds that

$$
\left|\left\langle x', x\right\rangle\right| = \left|x'(x)\right| = \|x\| \left|x'\left(\tfrac{x}{\|x\|}\right)\right| \le \|x\| \sup_{\|\tilde{x}\|\le 1} \left|x'(\tilde{x})\right| = |||x'||| \, \|x\| \, .
$$

§ **2.7 Definition.** Let X be a normed vector space.

(i) A sequence $(x_n) \subset X$ *converges weakly* to $x \in X$

$:\Leftrightarrow x_n \rightharpoonup x$

$:\Leftrightarrow \forall x' \in X' : \langle x', x_n\rangle \to \langle x', x\rangle$

(ii) A sequence $(x'_n) \subset X'$ *converges weakly-** to $x' \in X'$

$:\Leftrightarrow x'_n \overset{*}{\rightharpoonup} x'$

$:\Leftrightarrow \forall x \in X : \langle x'_n, x\rangle \to \langle x', x\rangle$

(iii) A set $M \subset X$ (or X') is called *weakly* (or *weakly-**) *sequentially compact*, if all sequences in M contain a subsequence that converges weakly (or weakly-*) to an element in M.

(iv) A mapping $J : X \to \mathbb{R}$ is called *weakly lower semi-continuous*, if for every weakly convergent sequence $(x_n) \subset X$ with $x_n \rightharpoonup x \in X$, it holds that

$$
J[x] \le \liminf_{n\to\infty} J[x_n].
$$

Weakly-* lower semi-continuity of $J : X' \to \mathbb{R}$ is defined analogously.

§ **2.8 Lemma.** *If a sequence $(x_n) \subset X$ (or $(x'_n) \subset X'$) converges strongly, i.e. in the norm, to $x \in X$ (or $x' \in X'$), then it also converges weakly to x (or weakly-* to x').*

Proof

$$x_n \to x \Rightarrow \left|\left\langle x', x_n - x\right\rangle\right| \le \left\|\left|x'\right|\right\| \underbrace{\|x_n - x\|}_{\to 0} \Rightarrow \left\langle x', x_n\right\rangle \to \left\langle x', x\right\rangle$$

$$x'_n \to x' \Rightarrow \left|\left\langle x' - x'_n, x\right\rangle\right| \le \underbrace{\left\|\left|x'_n - x'\right|\right\|}_{\to 0} \|x\| \Rightarrow \left\langle x'_n, x\right\rangle \to \left\langle x', x\right\rangle$$

□

§ *2.9 Remark.*

(i) Strong convergence implies weak convergence. In this sense, weak convergence is weaker than our usual notion of convergence.

Caution: For continuity, this is the other way around! Since more sequences converge weakly than strongly, weak (lower) continuity is stronger than strong (lower) continuity, since the corresponding condition has to be satisfied for more sequences in the weak case.

(ii) Weakly (or weakly-*) converging sequences are bounded. This can be shown with the Banach–Steinhaus theorem (also called uniform boundedness principle), see [1, Remark 8.3.(5)].

(iii) The weak and weak-* limits are unique (for a proof, see [1, Remark 8.3.(1)]).

§ *2.10 Example.* Again, we consider the sequence $(f_n) \subset L^2[0,\pi]$ given by $f_n(x) = \sin(nx)$. We already know that it is bounded in the L^2-norm. Now we will show that f_n converges weakly to 0. Since $L^2[0,\pi]$ is a Hilbert space, we know from Riesz representation theorem that $\left(L^2[0,\pi]\right)' \cong L^2[0,\pi]$. Thus, it is sufficient to show that

$$\int_0^\pi y(x)\sin(nx)\,\mathrm{d}x \underset{n\to\infty}{\longrightarrow} \int_0^\pi y(x)0\,\mathrm{d}x = 0 \text{ for all } y \in L^2[0,\pi]. \qquad (*)$$

For $a, b \in [0,\pi]$ and $c \in \mathbb{R}$, we get

$$\int_a^b c\sin(nx)\,\mathrm{d}x = -\frac{c}{n}\cos(nx)\bigg|_a^b = \frac{c}{n}\left(-\cos(nb) + \cos(na)\right) \underset{n\to\infty}{\longrightarrow} 0,$$

thus $(*)$ follows for step functions. For arbitrary but fixed $y \in L^2[0,\pi]$ and $\epsilon > 0$, there exists a step function $u \in L^2[0,\pi]$ with $\|y - u\|_{L^2} < \epsilon$. Thus,

$$\left|\int_0^\pi y(x)\sin(nx)\,\mathrm{d}x\right| \le \underbrace{\left|\int_0^\pi u(x)\sin(nx)\,\mathrm{d}x\right|}_{\longrightarrow 0 \text{ for } n\to\infty} + \underbrace{\left|\int_0^\pi (y(x) - u(x))\sin(nx)\,\mathrm{d}x\right|}_{\text{(Hölder's ineq.) } \le \|y-u\|_{L^2}\|\sin(n\cdot)\|_{L^2} < C\epsilon}$$

and $(*)$ follows.

§ 2.11 Proposition. *Let X be a normed vector space. The mapping $J : X \to X''$ given by*

$$\left\langle Jx, x'\right\rangle := \left\langle x', x\right\rangle \text{ for all } x \in X, x' \in X'$$

is linear and isometric (thus, in particular, injective and continuous).

For a proof, see [1, Proposition 8.2.(1)]. The proof uses $|\langle x', x\rangle| \leq |||x'||| \, \|x\|$ and that, for every $0 \neq x \in X$, there exists $x' \in X'$ with $|||x'||| = 1$ and $x'(x) = \|x\|$ (a consequence of the Hahn-Banach theorem).

§ 2.12 Definition (Reflexivity)**.** Let X be a normed vector space. X is called *reflexive*, if the mapping J from Proposition § 2.11 is surjective. Thus, J is bijective, and X and X'' are isomporphic, i.e. $X \cong X''$.

§ 2.13 Proposition. *Let X be a reflexive Banach space. Then, the closed unit ball*

$$\overline{B_1(0)} = \overline{\{x \in X \, : \, \|x\| < 1\}} \subset X$$

is weakly (or weakly-) sequentially compact.*

For a proof, see [1, Proposition 8.10].

§ *2.14 Remark.* For $1 < p < \infty$, $H^{m,p}(\Omega)$ is reflexive (see [1, 8.11.(3)]). In particular, Proposition § 2.13 implies that bounded sequences in $H^{m,p}(\Omega)$ have weakly convergent subsequences. For $m \geq 1$, such sequences even converge strongly in $H^{m-1,p}(\Omega)$ due to Rellich's theorem.

§ 2.15 Theorem. *Let X be a reflexive Banach space, $M \subset X$ and $J : X \to \mathbb{R}$. If M is nonempty weakly sequentially closed (if a sequence from M converges weakly, the limit is in M) and J coercive on M and weakly lower semi-continuous, then there exists $y^* \in M$ with*

$$J[y^*] \leq J[y] \text{ for all } y \in M.$$

Proof We use the direct method in the calculus of variations. Due to the coercivity of J on M, J is bounded from below. Thus, there is a minimizing sequence $(y_n)_n \in M^{\mathbb{N}}$ with

$$\lim_{n\to\infty} J[y_n] = \inf_{y\in M} J[y] =: \underline{J}.$$

In particular, the sequence $(J[y_n])_n$ is bounded and combined with the coercivity it follows that

$$\tilde{C} \geq J[y_n] \geq C \, \|y_n\|^r - \beta \Rightarrow \|y_n\| \leq \left(\frac{\tilde{C} + \beta}{C}\right)^{\frac{1}{r}}.$$

Thus, the sequence $(y_n)_n$ is bounded. Since X is reflexive, it follows from Proposition § 2.13 that there is a weakly convergent subsequence $(y_{n_k})_k$ with $y_{n_k} \rightharpoonup y^* \in X$. Since M is weakly sequentially closed, we get $y^* \in M$ and with the weak lower semi-continuity of J we have $J[y^*] \leq \underline{J}$ and thus $J[y^*] = \underline{J}$. □

§ *2.16 Example.* For a given $g \in L^2(\Omega)$ the simple denoising functional

$$J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \frac{\lambda}{2} \|\nabla y\|_{L^2}^2$$

is coercive on $H^{1,2}(\Omega)$. This can be shown as follows.

$$\|y\|_{L^2} \leq \|y - g\|_{L^2} + \|g\|_{L^2} = \sqrt{\|y - g\|_{L^2}^2} + \|g\|_{L^2} \leq \sqrt{2J[y]} + \|g\|_{L^2}$$

Combined with the inequality $(a + b)^2 \leq 2a^2 + 2b^2$

$$\big(0 \leq (a - b)^2 \Rightarrow 2ab \leq a^2 + b^2 \Rightarrow (a + b)^2 = a^2 + 2ab + b^2 \leq 2a^2 + 2b^2\big)$$

we get

$$
\begin{aligned}
\|y\|_{H^{1,2}}^2 &= \|y\|_{L^2}^2 + \|\nabla y\|_{L^2}^2 \le \left(\sqrt{2J[y]} + \|g\|_{L^2}\right)^2 + \frac{2}{\lambda} J[y] \\
&\le 2\sqrt{2J[y]}^2 + 2\,\|g\|_{L^2}^2 + \frac{2}{\lambda} J[y] = \left(4 + \frac{2}{\lambda}\right) J[y] + 2\,\|g\|_{L^2}^2
\end{aligned}
$$

$$
\Rightarrow J[y] \ge \frac{\lambda}{4\lambda + 2}\,\|y\|_{H^{1,2}}^2 - \frac{\lambda}{2\lambda+1}\,\|g\|_{L^2}^2
$$

In contrast to the assumption $g \in C(\overline{\Omega})$ in Example § 3.12, one only needs $g \in L^2(\Omega)$ here.

§ 2.17 Proposition. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain (i.e. open, nonempty and connected) with piecewise smooth boundary,*

$$
f : \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}, (x, y, \xi) \mapsto f(x, y, \xi)
$$

continuous, wrt. to the third variable continuously differentiable, bounded from below and convex in the third argument, i.e. for every $x \in \Omega$ and $y \in \mathbb{R}$, the function

$$
\mathbb{R}^d \to \mathbb{R}, \xi \mapsto f(x, y, \xi)
$$

is convex. For $1 < p < \infty$, then

$$
J : H^{1,p}(\Omega) \to \mathbb{R}, y \mapsto J[y] := \int_\Omega f(x, y(x), \nabla y(x))\,\mathrm{d}x
$$

is weakly lower semi-continuous.

Proof Let $(y_n) \subset H^{1,p}(\Omega)$ be a weakly convergent sequence with $y_n \rightharpoonup y \in H^{1,p}(\Omega)$ and $l := \liminf J[y_n]$. WLOG $l = \lim J[y_n]$ (else consider a suitable subsequence). We need to show $J[y] \le l$. Due to Rellich's theorem (cf. [1, A8.4], here the boundary regularity of Ω is needed), weak convergence in $H^{1,p}$ implies strong convergence in L^p, i.e. y_n converges strongly to y in L^p. Thus, there is another subsequence, again denoted with y_n, which converges pointwise a.e. in Ω to y (cf. [1, Lemma 3.22(1)]).

Let $\epsilon > 0$ be arbitrary but fixed. Due to Egorov's theorem (cf. [1, A3.18]), the pointwise convergence implies that there exists a set $E_\epsilon \subset \Omega$, such that

$$
y_n \to y \text{ uniformly one } E_\epsilon, \text{ where } |\Omega \setminus E_\epsilon| \le \epsilon.
$$

Let $F_\epsilon := \left\{x \in \Omega \,:\, |y(x)| + \|\nabla y(x)\| \le \frac{1}{\epsilon}\right\}$. Due to

$$
\infty > \|y\|_{H^{1,p}}^p \ge \int_{\Omega \setminus F_\epsilon} y^p + \|\nabla y\|^p\,\mathrm{d}x \overset{(|a|+|b|)^p \le C_p(|a|^p+|b|^p)}{\ge} \int_{\Omega\setminus F_\epsilon} \frac{1}{C_p \epsilon^p}\,\mathrm{d}x = \frac{1}{C_p\epsilon^p}\,|\Omega \setminus F_\epsilon|\,,
$$

we get $|\Omega \setminus F_\epsilon| \to 0$ for $\epsilon \to 0$. For $G_\epsilon := E_\epsilon \cap F_\epsilon$, it follows that

$$
0 \le |\Omega \setminus G_\epsilon| = |\Omega \setminus (E_\epsilon \cap F_\epsilon)| = |(\Omega \setminus E_\epsilon) \cup (\Omega \setminus F_\epsilon)| \le |\Omega \setminus E_\epsilon| + |\Omega \setminus F_\epsilon| \to 0 \text{ for } \epsilon \to 0.
$$

WLOG $f \ge 0$ (else consider $f + C$ for a lower bound C of f). With the convexity of f in ξ, we get

$$
\begin{aligned}
J[y_n] &= \int_\Omega f(x, y_n(x), \nabla y_n(x))\,\mathrm{d}x \ge \int_{G_\epsilon} f(x, y_n(x), \nabla y_n(x))\,\mathrm{d}x \\
(\S\ 3.16)\ &\ge \int_{G_\epsilon} f(x, y_n(x), \nabla y(x)) + \nabla_\xi f(x, y_n(x), \nabla y(x)) \cdot [\nabla y_n(x) - \nabla y(x)]\,\mathrm{d}x\,.
\end{aligned}
$$

Due to the continuity of f on $\overline{\Omega} \times \mathbb{R} \times \mathbb{R}^d$, f is bounded on $D := \overline{\Omega} \times \left[-\frac{2}{\epsilon}, \frac{2}{\epsilon}\right] \times \overline{B_{\frac{1}{\epsilon}}(0)}$. Moreover, due to the uniform convergence of y_n on E_ϵ, it follows that for sufficiently large n, $y_n(x) \in \left[-\frac{2}{\epsilon}, \frac{2}{\epsilon}\right]$ for all $x \in G_\epsilon$. The above combined with the dominated convergence theorem implies

$$\lim_{n\to\infty} \int_{G_\epsilon} f(x, y_n(x), \nabla y(x))\,\mathrm{d}x = \int_{G_\epsilon} f(x, y(x), \nabla y(x))\,\mathrm{d}x\,.$$

$\nabla_\xi f$ is uniformly continuous on D and $\nabla_\xi f(\cdot, y_n(\cdot), \nabla y(\cdot))$ converges uniformly on G_ϵ to $\nabla_\xi f(\cdot, y(\cdot), \nabla y(\cdot))$. Moreover, ∇y_n converges weakly to ∇y in L^p due to the weak convergence of y_n in $H^{1,p}$. Thus,

$$\begin{aligned}
&\left|\int_{G_\epsilon} \nabla_\xi f(x, y_n, \nabla y) \cdot [\nabla y_n - \nabla y]\,\mathrm{d}x\right| \\
&\le \underbrace{\left|\int_{G_\epsilon} (\nabla_\xi f(x, y_n, \nabla y) - \nabla_\xi f(x, y, \nabla y)) \cdot [\nabla y_n - \nabla y]\,\mathrm{d}x\right|}_{\to 0 \text{ due to the uniform convergence of } \nabla_\xi f(\cdot, y_n(\cdot), \nabla y(\cdot)) \text{ and } \nabla y_n \rightharpoonup \nabla y} \\
&\quad + \underbrace{\left|\int_{G_\epsilon} \nabla_\xi f(x, y, \nabla y) \cdot [\nabla y_n - \nabla y]\,\mathrm{d}x\right|}_{\to 0 \text{ since } \nabla y_n \rightharpoonup \nabla y}
\end{aligned}$$

In total, we get

$$l = \lim_{n\to\infty} J[y_n] \ge \int_{G_\epsilon} f(x, y(x), \nabla y(x))\,\mathrm{d}x\,.$$

This holds for an arbitrary $\epsilon > 0$. Going to the limit $\epsilon \to 0$, we get from the monotone convergence theorem

$$l \ge \int_\Omega f(x, y(x), \nabla y(x))\,\mathrm{d}x = J[y].$$

Here, we used $f \ge 0$ and, that one can construct E_ϵ such that it increases for $\epsilon \to 0$. □

§ *2.18 Example.* Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with piecewise smooth boundary and $g \in C(\Omega)$. Then $f(x, y, \xi) = \frac{1}{2}(y - g(x))^2 + \frac{\lambda}{2}\xi^2$ fulfills the assumptions in Proposition § 2.17. Thus,

$$J[y] = \int_\Omega \frac{1}{2}(y(x) - g(x))^2 + \frac{\lambda}{2}(\nabla y(x))^2\,\mathrm{d}x$$

is weakly lower semi-continuous on $H^{1,2}(\Omega)$. Moreover, J is coercive on $H^{1,2}(\Omega)$ (Example § 2.16). Since $H^{1,2}(\Omega)$ is reflexive (since $M = X = H^{1,2}(\Omega)$, M is weakly sequentially closed), Theorem § 2.15 ensures the existence of a minimizer of J. Thus, the simple denoising problem can be solved, in fact in $H^{1,2}(\Omega)$.

The assumption $g \in C(\Omega)$ is too strong for real images (in particular, noisy images are not continuous), but can be weakened for the denoising problem easily:

§ 2.19 Lemma. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with piecewise smooth boundary and $g \in L^2(\Omega)$. If $(y_n) \subset H^{1,2}(\Omega)$ converges weakly to $y \in H^{1,2}(\Omega)$, then*

$$\|y - g\|_{L^2}^2 \le \liminf_{n\to\infty} \|y_n - g\|_{L^2}^2\,.$$

Proof WLOG $\liminf \|y_n - g\|_{L^2}^2 = \lim \|y_n - g\|_{L^2}^2$ (else consider a suitable subsequence). Like in the proof of Proposition § 2.17, one shows that there is a subsequence, again denoted by y_n, which converges pointwise a.e. in Ω to y. Thus, $(y_n - g)^2$ converges pointwise a.e. to $(y - g)^2$ and we get

$$\|y - g\|_{L^2}^2 = \int_\Omega (y-g)^2 \,\mathrm{d}x = \int_\Omega \lim_{n\to\infty} (y_n - g)^2 \,\mathrm{d}x \overset{\text{Fatou}}{\leq} \liminf_{n\to\infty} \int_\Omega (y_n - g)^2 \,\mathrm{d}x$$
$$= \liminf_{n\to\infty} \|y_n - g\|_{L^2}^2 .$$

□

§ 2.20 Lemma. *Let X be a normed vector space and $J, K : X \to \mathbb{R}$ weakly lower semi-continuous. Then, $J + K$ is weakly lower semi-continuous. The analogous statement holds for weakly-* lower semi-continuity mappings $X' \to \mathbb{R}$.*

Proof Let $(x_n) \subset X$ be a weakly convergent sequence with $x_n \rightharpoonup x \in X$. Then,

$$J[x] + K[x] \leq \liminf_{n\to\infty} J[x_n] + \liminf_{n\to\infty} K[x_n] \leq \liminf_{n\to\infty} (J+K)[x_n].$$

□

§ *2.21 Remark.* We have shown that minimizers of our simple denoising model

$$J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \frac{\lambda}{2} \|\nabla y\|_{L^2}^2$$

are in $H^{1,2}(\Omega)$. What can we infer from this about the suitability of this model to denoise images? One important property of images is that they can have edges (jumps in image intensity). Is it possible to have edges in $H^{1,2}(\Omega)$?

§ 2.22 Lemma. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and $D \subset \mathbb{R}^d$ a bounded domain with piecewise smooth boundary with $\overline{D} \subset \Omega$. Then, the* characteristic function χ_D *of D, given by*

$$\chi_D(x) := \begin{cases} 1 & x \in D \\ 0 & \textit{else} \end{cases},$$

is not in $H^{1,p}(\Omega)$ for $1 \leq p \leq \infty$.

Proof Exercise. □

§ *2.23 Remark.* Even though $H^{1,1}$ does not allow for jumps, the $H^{1,1}$-norm is a good starting point. Let $y \in C^1[0,1]$ be increasing. Then, we have

$$|y|_{H^{1,1}} = \int_0^1 |y'(t)| \,\mathrm{d}t = \int_0^1 y'(t) \,\mathrm{d}t = y(1) - y(0).$$

Thus, for increasing functions in 1D, the $H^{1,1}$-seminorm is independent of the size of the derivative, just the difference of the function values at the interval boundary matters. In particular, a function with jump like

$$(0,1) \to \mathbb{R}, t \mapsto \begin{cases} 0 & t < \frac{1}{2} \\ 1 & t \geq \frac{1}{2} \end{cases}$$

can be approximated with a sequence that is bounded in the $H^{1,1}$-norm. To extend the $H^{1,1}$-norm to such functions, we need a more general concept than weak derivatives. For $x \in \mathbb{R}^d$ with $x \neq 0$, we have

$$\|x\|_2 = x \cdot \frac{x}{\|x\|_2} \leq \sup_{\|p\|_2 \leq 1} -x \cdot p \leq \sup_{\|p\|_2 \leq 1} \|x\|_2 \|p\|_2 = \|x\|_2 \Rightarrow \|x\|_2 = \sup_{\|p\|_2 \leq 1} -x \cdot p.$$

Thus, for $y \in H^{1,1}(\Omega)$

$$\begin{aligned} \int_\Omega \|\nabla y\|_2 \,\mathrm{d}x &= \sup_{p \in C_c^\infty(\Omega, \mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega -\nabla y \cdot p \,\mathrm{d}x = \sup_{\left\| \|p\|_2 \right\|_{L^\infty} \leq 1} - \int_\Omega \sum_{i=1}^d \partial_i y p_i \,\mathrm{d}x \\ &= \sup_{\left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega \sum_{i=1}^d y \partial_i p_i \,\mathrm{d}x = \sup_{\left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \operatorname{div} p \,\mathrm{d}x, \end{aligned}$$

which motivates the following definition:

§ 2.24 Definition. For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega, \mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \operatorname{div} p \,\mathrm{d}x \,.$$

The space of functions of bounded variation is

$$BV(\Omega) := \left\{ y \in L^1(\Omega) \,:\, |y|_{BV(\Omega)} < \infty \right\}.$$

The BV-norm of $y \in BV(\Omega)$ is defined as

$$\|y\|_{BV(\Omega)} := \|y\|_{L^1(\Omega)} + |y|_{BV(\Omega)} \,.$$

§ *2.25 Remark.* Let $y \in H^{1,1}(\Omega)$. As shown above, we have $\int_\Omega \|\nabla y\|_2 \,\mathrm{d}x = |y|_{BV}$. From

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{d}\, \|x\|_2 \ \text{ and } \ |y|_{H^{1,1}} = \int_\Omega \|\nabla y\|_1 \,\mathrm{d}x,$$

it follows that

$$|y|_{BV} \leq |y|_{H^{1,1}} \leq \sqrt{d}\, |y|_{BV} \Rightarrow \|y\|_{BV} \leq \|y\|_{H^{1,1}} \leq \sqrt{d}\, \|y\|_{BV} \,.$$

In particular, we have $H^{1,1}(\Omega) \subset BV(\Omega) \subset L^1(\Omega)$.

§ *2.26 Example.*

(i) For the *Heaviside Function*

$$H : \mathbb{R} \to \mathbb{R}, t \mapsto \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}$$

it holds that $|H|_{BV(\mathbb{R})} = 1$. Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$. Then,

$$\begin{aligned} \int_\mathbb{R} H(t) \operatorname{div} p(t) \,\mathrm{d}t &= \int_0^\infty p'(t) \,\mathrm{d}t = \lim_{n \to \infty} \int_0^n p'(t) \,\mathrm{d}t = \lim_{n \to \infty} (p(n) - p(0)) \\ &= -p(0) \leq 1. \end{aligned}$$

Since there is an admissible p with $p(0) = -1$, we get $|H|_{BV(\mathbb{R})} = 1$.

(ii) Let $a, b \in \mathbb{R}$ with $a < b$. Then, for the characteristic function of the interval $[a, b]$, we have

$$\left|\chi_{[a,b]}\right|_{BV(\mathbb{R})} = 2.$$

Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$. Then,

$$\int_{\mathbb{R}} \chi_{[a,b]}(t)\,\mathrm{div}p(t)\,\mathrm{d}t = \int_a^b p'(t)\,\mathrm{d}t = p(b) - p(a) \leq 2.$$

Since there is an admissible p with $p(a) = -1$ and $p(b) = 1$, the statement follows.

(iii) Consider the characteristic function of the circle $B_r(0) \subset \mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\|\|p\|_2\|_{L^\infty} \leq 1$, we get

$$\begin{aligned}\int_{\mathbb{R}^2} \chi_{B_r(0)}(x)\,\mathrm{div}p(x)\,\mathrm{d}x &= \int_{B_r(0)} \mathrm{div}p(x)\,\mathrm{d}x = \int_{\partial B_r(0)} p(x)\cdot\nu(x)\,\mathrm{d}A(x)\\ &\leq \int_{\partial B_r(0)} \|p(x)\|_2\,\|\nu(x)\|_2\,\mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.\end{aligned}$$

Since there is an admissible p with $p = \nu$ on $\partial B_r(0)$, we get $\left|\chi_{B_r(0)}\right|_{BV(\mathbb{R}^2)} = 2\pi r$.

In general, for a bounded domain $D \subset \mathbb{R}^d$ with piecewise smooth boundary and $\overline{D} \subset \Omega$, $|\chi_D|_{BV(\Omega)}$ is the length of the boundary of D. Thus, as regularizer, $|\cdot|_{BV}$ smoothens the boundary of the sublevel sets $\{x \in \Omega \,:\, y(x) < c\}$.

§ *2.27 Remark.* As shown above, $BV(\Omega)$ allows for edges. Thus, the space suggests itself to treat images and motivates a variant of our simple denoising functional (cf. Example § 3.12):

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of the *ROF-functional*

$$J : BV(\Omega) \to \mathbb{R}_\infty, y \mapsto J[y] = \frac{1}{2}\|y - g\|_{L^2}^2 + \lambda\,|y|_{BV} \quad \text{(Rudin, Osher, Fatemi, 1992).}$$

Existence of minimizers can be shown again with the direct method, but not using reflexivity since $BV(\Omega)$ is not reflexive. Instead, one can show that $BV(\Omega)$ is the dual space of a separable Banach space (cf. [2, 3.12]). Moreover, it is known that if X is a separable Banach space, $\overline{B_1(0)} \subset X'$ is weakly-* sequentially compact.

Thus, bounded sequences in $BV(\Omega)$ have weakly-* converging subsequences. Moreover, one can show that weakly-* converging sequences in $BV(\Omega)$ converge strongly in L^1 (to the same limit), cf. [2, 3.11 + 3.12]. With this, one can show that J is weakly-* lower semi-continuous. Also, J is coercive and we can finally apply the direct methods to guarantee existence of minimizers.

§ 3 Characterization of minimizers

§ *3.1 Remark.* Idea from classical optimization:

Let x^* be a minimizer of $J : \mathbb{R}^d \to \mathbb{R}$, i.e. $J(x^*) \leq J(x)$ for all $x \in \mathbb{R}^d$.

Choose a perturbation direction $s \neq 0 \in \mathbb{R}^d$ and consider $\varphi : \mathbb{R} \to \mathbb{R}, \epsilon \mapsto J(x^* + \epsilon s)$. 0 is a minimizer of φ and if J is differentiable in x^*, we get $\varphi'(0) = 0$. Since s was arbitrary and

$$\partial_s J(x^*) = \varphi'(0) = 0,$$

every directional derivative of J vanishes. Since there are only finitely many linearly independent directions in $\mathbb{R}^d$, it follows that

$$\nabla J(x^*) \cdot s = \partial_s J(x^*) = 0,$$

hence $\nabla J(x^*) = 0$. This is the first order necessary condition known from analysis: A minimizer (the same holds for maximizers) is a zero of the gradient.

This idea can be applied analogously to general vector spaces: Let X be a vector space and $M \subset X$. Let $y^* \in M$ be a global minimizer of $J : M \to \mathbb{R}$. Choose $z \in X$ as *test function* (or *perturbation function*) and $\epsilon \in \mathbb{R}$ as *perturbation parameter*. Since $y = y^* + \epsilon z$ has to compete with y^*, we need $y \in M$. In case $M = X$, this results in no further restrictions. If M, for instance, imposes boundary values ($M = \left\{y \in C^1([a,b]) \,:\, y(a) = y_a \wedge y(b) = y_b\right\}$), then it must hold that

$$y_a = y(a) = y^*(a) + \epsilon z(a),$$

$$y_b = y(b) = y^*(b) + \epsilon z(b).$$

Here, the test function z has to fulfill $z(a) = z(b) = 0$, to ensure $y \in M$, i.e. to make y admissible.

For $\epsilon_0 > 0$ such that $y^* + \epsilon z \in M$ for all $\epsilon \in (-\epsilon_0, \epsilon_0)$, consider

$$\varphi : (-\epsilon_0, \epsilon_0) \to \mathbb{R}, \epsilon \mapsto J[y^* + \epsilon z].$$

Since y^* is a global minimizer of J, 0 is a global minimizer of φ. If J is sufficiently regular, φ is differentiable and we get

$$0 = \frac{d}{d\epsilon}\varphi(\epsilon)\Big|_{\epsilon=0}.$$

§ 3.2 Definition. Let X be a vector space, $D \subset X$ open and $J : D \to \mathbb{R}$. Moreover, let $z \in X$ such that $y + \epsilon z \in D$ for all ϵ with sufficiently small absolute value. If the limit

$$\lim_{\epsilon \to 0} \frac{J[y + \epsilon z] - J[y]}{\epsilon} = \frac{d}{d\epsilon}(J[y + \epsilon z])\Big|_{\epsilon=0} =: \left\langle J'[y], z\right\rangle,$$

exists, $\langle J'[y], z\rangle$ is called *Gâteaux-differential* of J in direction z at the position y. If X is normed and the mapping

$$J'[y] : X \to \mathbb{R}, z \mapsto \left\langle J'[y], z\right\rangle$$

linear and continuous, J is called *Gâteaux-differentiable* at the position y and $J'[y]$ is called *Gâteaux-derivative* of J at the position y. $J'[y]$ is also called *first variation* of J.

§ *3.3 Remark.* If $X = \mathbb{R}^d$, $D \subset X$ open and $J \in C^1(D)$, it follows from the chain rule that

$$\big\langle J'[y], z \big\rangle = \nabla J(y) \cdot z \text{ for all } z \in \mathbb{R}^d.$$

Thus, the Gâteaux-derivative and the classical derivative coincide for differentiable functions in $\mathbb{R}^d$.

§ 3.4 Proposition (Necessary condition)**.** *Let X be a vector space, $y^* \in M \subset X$ a global extremum of $J : M \to \mathbb{R}$ and $z \in X$ such that $y^* + \epsilon z \in M$ for all ϵ with sufficiently small absolute value. If $\langle J'[y^*], z\rangle$ exists, it holds that*

$$\big\langle J'[y^*], z \big\rangle = 0.$$

Proof WLOG y^* is a minimizer (else consider $-J$). Since $\langle J'[y^*], z\rangle$ exists, $\varphi(\epsilon) = J[y^* + \epsilon z]$ is differentiable in 0, and the proposition follows as in Remark § 3.1. □

§ 3.5 Corollary. *Let X be a normed vector space, $y^* \in \mathring{M} \subset X$ a local extremum of $J : M \to \mathbb{R}$ and $z \in X$. If $\langle J'[y^*], z\rangle$ exists, it holds that $\langle J'[y^*], z\rangle = 0$.*

Proof Since y^* is from the interior of M and a lokal extremum, there is $r > 0$ with $B_r(y^*) \subset M$ such that y^* is a global extremum of J on $B_r(y^*)$. Then the proposition follows from Proposition § 3.4. □

§ *3.6 Example.* Let $\Omega \subset \mathbb{R}^d$ be a domain,

$$X = C^1(\overline{\Omega}) := \big\{ y \in C^1(\Omega) \cap C(\overline{\Omega}) \,:\, \|y\|_{C^1} := \max\{\|y\|_{L^\infty}, \|\|\nabla y\|_2\|_{L^\infty}\} < \infty \big\}$$

and $J : X \to \mathbb{R}$.

(i) For $g \in C(\overline{\Omega})$, let $J[y] = \frac{1}{2}\|y - g\|_{L^2}^2$. We get

$$\begin{aligned}
&\frac{1}{\epsilon}\left(J[y + \epsilon z] - J[y]\right) = \frac{1}{2\epsilon}\int_\Omega (y(x) + \epsilon z(x) - g(x))^2 - (y(x) - g(x))^2 \,\mathrm{d}x \\
&= \frac{1}{2\epsilon}\int_\Omega (y(x) - g(x))^2 + 2\epsilon(y(x) - g(x))z(x) + \epsilon^2 z(x)^2 - (y(x) - g(x))^2 \,\mathrm{d}x \\
&= \frac{1}{2}\int_\Omega 2(y(x) - g(x))z(x) + \epsilon z(x)^2 \,\mathrm{d}x \\
&\xrightarrow[\epsilon \to 0]{} \big\langle J'[y], z \big\rangle = \int_\Omega (y(x) - g(x))z(x)\,\mathrm{d}x = (y - g, z)_{L^2}
\end{aligned}$$

$\langle J'[y], z\rangle$ is linear in z. Does this hold for all J?

(ii) Let $X = \mathbb{R}^2$, so $\dim(X) < \infty$. For

$$J(x) = \begin{cases} x_1^2\left(1 + \frac{1}{x_2}\right) & x_2 \neq 0 \\ 0 & x_2 = 0 \end{cases}$$

and $z \in \mathbb{R}^2$ with $z_2 \neq 0$ it holds that

$$\frac{1}{\epsilon}\left(J(0 + \epsilon z) - J(0)\right) = \frac{1}{\epsilon}J(\epsilon z) = \frac{1}{\epsilon}\epsilon^2 z_1^2\left(1 + \frac{1}{\epsilon z_2}\right) = \epsilon z_1^2 + \frac{z_1^2}{z_2}$$

$$\xrightarrow[\epsilon \to 0]{} \big\langle J'(0), z \big\rangle = \frac{z_1^2}{z_2}$$

If $z_2 = 0$, then $\langle J'(0), z\rangle = 0$. Thus, the Gâteaux-differential $\langle J'(0), z\rangle$ exists for all $z \in \mathbb{R}^2$, but it is neither continuous nor linear in z.

(iii) For $J[y] = \frac{1}{2} \| \|\nabla y\|_2 \|_{L^2}^2 = \frac{1}{2} \int_\Omega \|\nabla y(x)\|_2^2 \,\mathrm{d}x$ we get

$$\frac{1}{\epsilon}\left(J[y+\epsilon z] - J[y]\right) = \frac{1}{2\epsilon}\int_\Omega \|\nabla(y+\epsilon z)\|^2 - \|\nabla y\|^2 \,\mathrm{d}x$$

$$= \frac{1}{2\epsilon}\int_\Omega \|\nabla y + \epsilon \nabla z\|^2 - \|\nabla y\|^2 \,\mathrm{d}x$$

$$= \frac{1}{2\epsilon}\int_\Omega \|\nabla y\|^2 + 2\epsilon \nabla y \cdot \nabla z + \epsilon^2 \|\nabla z\|^2 - \|\nabla y\|^2 \,\mathrm{d}x$$

$$= \frac{1}{2}\int_\Omega 2\nabla y \cdot \nabla z + \epsilon \|\nabla z\|^2 \,\mathrm{d}x$$

$$\xrightarrow[\epsilon \to 0]{} \left\langle J'[y], z \right\rangle = \int_\Omega \nabla y \cdot \nabla z \,\mathrm{d}x = (\nabla y, \nabla z)_{L^2}$$

Note: The perturbation z leads to a response in ∇z.

For the sake of simplicity, we consider the special case $d = 1$ and $\Omega = [a, b]$. For $y \in C^2([a, b])$ and $z \in C_0^1([a, b])$, it holds that

$$\left\langle J'[y], z \right\rangle = \int_a^b y' z' \,\mathrm{d}t = y' z \Big|_a^b - \int_a^b y'' z \,\mathrm{d}t = (-y'', z)_{L^2}$$

Thus, $\langle J'[y], z\rangle = 0$ for all $z \in C_0^1([a, b])$ if $y'' = 0$ ($\Rightarrow y' = c_1 \Rightarrow y = c_1 t + c_2$). Does $\langle J'[y], z\rangle = 0$ for all $z \in C_0^1([a, b])$ also imply $y'' = 0$?

(iv) For a given $g \in C(\overline{\Omega})$ consider the sum of (i) and (iii), i.e.

$$J[y] = \frac{1}{2}\|y - g\|_{L^2}^2 + \frac{\lambda}{2} \| \|\nabla y\|_2 \|_{L^2}^2 = \int_\Omega \frac{1}{2}(y(x) - g(x))^2 + \frac{\lambda}{2}\|\nabla y(x)\|^2 \,\mathrm{d}x\,.$$

This is a very simple denoising model. While the square in the second term prevents edges from being preserved (more on this later), this model can be solved with classical methods: With $f(x, y, \xi) = \frac{1}{2}(y - g(x))^2 + \frac{\lambda}{2}\|\xi\|^2$, the model belongs to the following general class of variational models:

§ 3.7 Problem (Classical variation problem (VP))**.** Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with piecewise smooth boundary,

$$f : \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}, (x, y, \xi) \mapsto f(x, y, \xi)$$

continuous and with respect to the second and third argument continuously differentiable, $M := C^1(\overline{\Omega})$ and

$$J : M \to \mathbb{R}, y \mapsto J[y] := \int_\Omega f(x, y(x), \nabla y(x)) \,\mathrm{d}x\,.$$

Find: A global minimizer y^* of J on M, i.e $y^* \in \operatorname*{argmin}_{y \in M} J[y]$.

§ 3.8 Proposition. *Let Ω, f and J be as in (VP). Then, J is Gâteaux-differentiable with respect to the C^1-Norm and for $y, z \in C^1(\overline{\Omega})$ it holds that*

$$\left\langle J'[y], z \right\rangle = \int_\Omega \partial_y f(x, y(x), \nabla y(x)) z(x) + \nabla_\xi f(x, y(x), \nabla y(x)) \cdot \nabla z(x) \,\mathrm{d}x\,.$$

Proof For $y, z \in C^1(\overline{\Omega})$, we get

$$\begin{aligned}
\left\langle J'[y], z \right\rangle &= \frac{d}{d\epsilon} \int_\Omega f(x, y(x) + \epsilon z(x), \nabla y(x) + \epsilon \nabla z(x)) \,\mathrm{d}x \,\bigg|_{\epsilon=0} \\
&= \int_\Omega \frac{d}{d\epsilon} f(x, y(x) + \epsilon z(x), \nabla y(x) + \epsilon \nabla z(x)) \bigg|_{\epsilon=0} \,\mathrm{d}x \qquad \text{(Leibniz's rule)} \\
&= \int_\Omega \nabla_{(y,\xi)} f(x, y(x) + \epsilon z(x), \nabla y(x) + \epsilon \nabla z(x)) \cdot (z(x), \nabla z(x)) \bigg|_{\epsilon=0} \,\mathrm{d}x \\
&= \int_\Omega \partial_y f(x, y(x), \nabla y(x)) z(x) + \nabla_\xi f(x, y(x), \nabla y(x)) \cdot \nabla z(x) \,\mathrm{d}x \,.
\end{aligned}$$

For $z_1, z_2 \in C^1(\overline{\Omega})$, we get

$$\begin{aligned}
&\left| \left\langle J'[y], z_1 \right\rangle - \left\langle J'[y], z_2 \right\rangle \right| \\
&= \left| \int_\Omega \partial_y f(x, y(x), \nabla y(x))(z_1(x) - z_2(x)) + \nabla_\xi f(x, y(x), \nabla y(x)) \cdot (\nabla z_1(x) - \nabla z_2(x)) \,\mathrm{d}x \right| \\
&\le \int_\Omega |\partial_y f(x, y(x), \nabla y(x))(z_1(x) - z_2(x))| + |\nabla_\xi f(x, y(x), \nabla y(x)) \cdot (\nabla z_1(x) - \nabla z_2(x))| \,\mathrm{d}x \\
&\le \underbrace{\int_\Omega |\partial_y f(x, y(x), \nabla y(x))| + \|\nabla_\xi f(x, y(x), \nabla y(x))\| \,\mathrm{d}x}_{=\text{const, independent of } z_1, z_2} \|z_1 - z_2\|_{C^1} \,.
\end{aligned}$$

This shows the continuity with respect to the C^1-norm. □

§ 3.9 Corollary (Integration by parts)**.** *Let Ω be as in (VP), ν the outer normal to $\partial\Omega$, $\varphi \in C^1(\overline{\Omega})$ and $v \in C^1(\overline{\Omega})^d$. Then, it holds that*

$$\int_\Omega \nabla\varphi(x) \cdot v(x) \,\mathrm{d}x = -\int_\Omega \varphi(x) \,\mathrm{div} v(x) \,\mathrm{d}x + \int_{\partial\Omega} \varphi(x) v(x) \cdot \nu(x) \,\mathrm{d}A(x).$$

If $\psi \in C^2(\overline{\Omega})$, it holds that

$$\int_\Omega \nabla\varphi(x) \cdot \nabla\psi(x) \,\mathrm{d}x = -\int_\Omega \varphi(x) \Delta\psi(x) \,\mathrm{d}x + \int_{\partial\Omega} \varphi(x) \nabla\psi(x) \cdot \nu(x) \,\mathrm{d}A(x).$$

Proof The first statement follows from the divergence theorem applied to φv and the product rule:

$$\int_{\partial\Omega} \varphi(x) v(x) \cdot \nu(x) \,\mathrm{d}A(x) = \int_\Omega \mathrm{div}\,(\varphi(x) v(x)) \,\mathrm{d}x = \int_\Omega \nabla\varphi(x) \cdot v(x) \,\mathrm{d}x + \int_\Omega \varphi(x) \,\mathrm{div} v(x) \,\mathrm{d}x$$

The second statement follows from the first with $v = \nabla\psi$ since $\mathrm{div}\nabla\psi(x) = \Delta\psi(x)$. □

§ 3.10 Lemma (Fundamental lemma of calculus of variations)**.** *Let $\Omega \subset \mathbb{R}^d$ be a domain. For $y \in C(\Omega)$, the following statements are equivalent:*

(i) $y \equiv 0$, *i.e.* $y(x) = 0$ *for all* $x \in \Omega$

(ii) $\int_\Omega y(x) z(x) \,\mathrm{d}x = 0$ *for all* $z \in C_c^\infty(\Omega)$

Proof Exercise.

§ 3.11 Proposition (Euler–Lagrange equation (ELE) of calculus of variations)**.** *Let Ω, f and J be as in (VP) and y^* a solution of (VP). If $y^* \in C^2(\overline{\Omega})$ and $f \in C^2$, then y^* solves the boundary value problem*

$$\begin{aligned}\partial_y f(x, y^*(x), \nabla y^*(x)) - \operatorname{div}_x \left(\nabla_\xi f(x, y^*(x), \nabla y^*(x))\right) &= 0 \ \textit{in } \Omega,\\ \nabla_\xi f(x, y^*(x), \nabla y^*(x)) \cdot \nu(x) &= 0 \ \textit{on } \partial\Omega.\end{aligned}$$

Short notation: $\partial_y f - \operatorname{div}_x(\nabla_\xi f) = 0$. The boundary values that appear here are called natural boundary conditions*, since they are automatically satisfied if no explicit boundary conditions are prescribed.*

Proof For $y, z \in C^1(\overline{\Omega})$, Proposition § 3.8 implies

$$\langle J'[y], z\rangle = \int_\Omega \partial_y f(x, y(x), \nabla y(x)) z(x) + \nabla_\xi f(x, y(x), \nabla y(x)) \cdot \nabla z(x)\, \mathrm{d}x\,.$$

If $y^* \in C^2(\overline{\Omega})$ solves (VP), we get for $z \in C^1(\overline{\Omega})$ using Proposition § 3.4 (necessary condition) and Corollary § 3.9 (integration by parts)

$$\begin{aligned}0 = \langle J'[y^*], z\rangle = &\int_\Omega \left(\partial_y f(x, y^*(x), \nabla y^*(x)) - \operatorname{div}_x\left(\nabla_\xi f(x, y^*(x), \nabla y^*(x))\right)\right) z(x)\, \mathrm{d}x\\ &+ \int_{\partial\Omega} z(x) \nabla_\xi f(x, y^*(x), \nabla y^*(x)) \cdot \nu(x)\, \mathrm{d}A(x).\end{aligned}$$

For $z \in C_0^1(\overline{\Omega})$ the second term vanishes and the claimed equality in Ω follows with the fundamental lemma. Thus, the first term vanishes for all $z \in C^1(\overline{\Omega})$. The fundamental lemma also holds for integrals of the form

$$\int_{\partial\Omega} g(x) z(x)\, \mathrm{d}A(x),$$

which proves the claimed boundary values. □

§ *3.12 Example* (Denosing in $\mathbb{R}^d$). For a given $g \in C(\overline{\Omega})$, we consider again

$$J[y] = \int_\Omega \frac{1}{2}(y(x) - g(x))^2 + \frac{\lambda}{2} \|\nabla y(x)\|^2\, \mathrm{d}x$$

on $C^1(\overline{\Omega})$. It holds that $f(x, y, \xi) = \frac{1}{2}(y - g(x))^2 + \frac{\lambda}{2}\|\xi\|^2$. Thus, if $y^* \in C^2(\overline{\Omega})$ solves (VP), then, according to Proposition § 3.11, y^* fulfills the ELE

$$0 = \partial_y f - \operatorname{div}_x(\nabla_\xi f) = y^* - g - \operatorname{div}(\lambda \nabla y^*) = y^* - g - \lambda \Delta y^* \text{ in } \Omega$$

and the boundary values

$$0 = \nabla_\xi f \cdot \nu = \lambda \nabla y^* \cdot \nu = \lambda \partial_\nu y^* \Rightarrow \partial_\nu y^* = 0 \text{ on } \partial\Omega.$$

In general, the assumptions in (VP) are not strong enough to guarantee that solutions of the ELE solve (VP). In other words, solutions of the ELE fulfill the necessary condition for minimizers, but this condition is not sufficient. In this example, f is also convex in (y, ξ). This is sufficient to prove that every solution of the ELE is a minimizer. Here, f is even strictly convex in (y, ξ), which also implies the uniqueness of the minimizer.

§ 3.13 Definition. Let X be a vector space, $M \subset X$ a set and $J : M \to \mathbb{R}$ a mapping.

(i) M is called *convex* $:\Leftrightarrow \forall x, y \in M\ \forall \lambda \in [0,1] : \lambda x + (1-\lambda)y \in M$

(ii) J is called *convex* on M $:\Leftrightarrow \forall x, y \in M\ \forall \lambda \in [0,1]$:

$$[\lambda x + (1-\lambda)y \in M] \Rightarrow J[\lambda x + (1-\lambda)y] \leq \lambda J[x] + (1-\lambda)J[y].$$

If for $x \neq y$ and $\lambda \in (0,1)$ even "$<$" holds, J is called *strictly convex*.

§ 3.14 Proposition. *Let X be a normed vector space, $M \subset X$ convex, U an open neighborhood of M and $J : U \to \mathbb{R}$ Gâteaux-differentiable. Then,*

(i) J convex on M $\Leftrightarrow \forall x, y \in M : J[y] \geq J[x] + \langle J'[x], y - x\rangle$.

(ii) J strictly convex on M $\Leftrightarrow \forall x, y \in M, x \neq y : J[y] > J[x] + \langle J'[x], y - x\rangle$

Proof

(i) "$\Rightarrow$": Let $x, y \in M$ and $\lambda \in (0,1]$. Since J and M are convex, we get

$$J[x + \lambda(y-x)] = J[\lambda y + (1-\lambda)x] \leq \lambda J[y] + (1-\lambda)J[x] = J[x] + \lambda(J[y] - J[x]).$$

Reorganizing the terms and division by λ leads to

$$\tfrac{1}{\lambda}\left(J[x + \lambda(y-x)] - J[x]\right) \leq J[y] - J[x].$$

Since J is Gâteaux-differentiable, the limit $\lambda \to 0$ exists and it follows that

$$\left\langle J'[x], y - x \right\rangle \leq J[y] - J[x].$$

"$\Leftarrow$": Let $x, y \in M$ and $\lambda \in [0,1]$. Since M is convex, we get $\hat{x} = \lambda x + (1-\lambda)y \in M$ and due to the assumptions, it holds that

$$J[x] \geq J[\hat{x}] + \left\langle J'[\hat{x}], x - \hat{x} \right\rangle \tag{$*^1$}$$

$$J[y] \geq J[\hat{x}] + \left\langle J'[\hat{x}], y - \hat{x} \right\rangle \tag{$*^2$}$$

With $\lambda(*^1) + (1-\lambda)(*^2)$ it follows that

$$\lambda J[x] + (1-\lambda)J[y] \geq J[\hat{x}] + 0 = J[\lambda x + (1-\lambda)y].$$

Thus, J is convex.

(ii) "$\Rightarrow$": Let $x, y \in M$ with $x \neq y$. Then, $z = \frac{1}{2}(x+y) \in M$ and with the strict convexity it follows that

$$J[z] < \frac{1}{2}\left(J[x] + J[y]\right) \Rightarrow J[z] - J[x] < \frac{1}{2}\left(J[y] - J[x]\right).$$

Combined with (i), this leads to

$$J[y] - J[x] > 2(J[z] - J[x]) \overset{\text{(i)}}{\geq} 2\left\langle J'[x], z - x \right\rangle = \left\langle J'[x], y - x \right\rangle.$$

"$\Leftarrow$" follows analogously to (i) "$\Leftarrow$" using the strict inequality in $(*^1)$ and $(*^2)$. $\square$

§ 3.15 Corollary. *Let X, M and J be as in Proposition § 3.14 and J additionally convex. Moreover, let $y^* \in M$ with $J'[y^*] = 0$ (i.e. $\langle J'[y^*], z\rangle = 0$ for all $z \in X$). Then, $y^* \in \operatorname{argmin}_{y\in M} J[y]$.*

Proof Let $y \in M$ be arbitrary but fixed. Using Proposition § 3.14, it immediately follows that

$$J[y] \geq J[y^*] + \left\langle J'[y^*], y - y^*\right\rangle = J[y^*]. \qquad \square$$

§ 3.16 Corollary. *Let $M \subset \mathbb{R}^d$ be convex, U an open neighborhood of M and $f \in C^1(U)$. Then,*

(i) f convex on $M \Leftrightarrow \forall x, y \in M : f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$

(ii) f strictly convex on $M \Leftrightarrow \forall x, y \in M, x \neq y : f(y) > f(x) + \nabla f(x) \cdot (y - x)$

Proof Immediately follows from Remark § 3.3 and Proposition § 3.14. $\square$

§ 3.17 Proposition. *Let f be as in (VP) and additionally convex in the second and third argument, i.e. for all $x \in \Omega$ let the function*

$$\mathbb{R} \times \mathbb{R}^d \to \mathbb{R}, (y, \xi) \mapsto f(x, y, \xi)$$

be convex and $f \in C^2$. Then, every solution $y \in C^2(\overline{\Omega})$ of the ELE $\partial_y f - \operatorname{div}_x(\nabla_\xi f) = 0$ from Proposition § 3.11 with natural boundary conditions solves (VP). If f is strictly convex in the second and third argument, the solution of (VP) is unique.

Proof Let $y^* \in C^2(\overline{\Omega})$ be a solution of the ELE and $y \in C^1(\overline{\Omega})$ arbitrary. Then $z := y - y^* \in C^1(\overline{\Omega})$ and from the convexity of f and Corollary § 3.16 it follows that

$$\begin{aligned}
J[y] &= \int_\Omega f(x, y(x), \nabla y(x))\,\mathrm{d}x \\
(\S\ 3.16)\quad &\geq \int_\Omega f(x, y^*, \nabla y^*) + (\partial_y, \nabla_\xi) f(x, y^*, \nabla y^*) \cdot [(y, \nabla y) - (y^*, \nabla y^*)]\,\mathrm{d}x \\
&= \int_\Omega f(x, y^*, \nabla y^*) + \partial_y f(x, y^*, \nabla y^*) z + \nabla_\xi f(x, y^*, \nabla y^*) \cdot \nabla z\,\mathrm{d}x \\
&= \int_\Omega f(x, y^*, \nabla y^*) + \underbrace{\left(\partial_y f(x, y^*, \nabla y^*) - \operatorname{div}_x\left(\nabla_\xi f(x, y^*, \nabla y^*)\right)\right)}_{=0 \text{ (ELE)}} z\,\mathrm{d}x \\
&\qquad + \int_{\partial\Omega} z \underbrace{\nabla_\xi f(x, y^*, \nabla y^*) \cdot \nu}_{=0 \text{ (boundary conditions)}}\,\mathrm{d}A(x) = J[y^*].
\end{aligned}$$

Thus, y^* is a global minimizer of J on $C^1(\overline{\Omega})$ and, thus, solves (VP). Using the strict convexity, it follows analogously that for $y \neq y^*$ it holds that $J[y] > J[y^*]$ (here the continuity of the integrand and the positive volume of Ω is needed). Thus, the solution of (VP) is unique. $\square$

§ 3.18 Proposition. *Let X be a vector space, $M \subset X$ convex and $J : M \to \mathbb{R}$ strictly convex. Then, at most one global minimizer of J on M exists.*

Proof Assume there are two global minimizers $u, v \in M$ of J on M with $u \neq v$. Then, $\frac{1}{2}u + \frac{1}{2}v \in M$, since M is convex and it follows that

$$J\left[\tfrac{1}{2}u + \tfrac{1}{2}v\right] < \tfrac{1}{2}J[u] + \tfrac{1}{2}J[v] = J[u] \quad \lightning \text{ to } u \text{ global minimizer of } J \text{ on } M. \qquad \square$$

§ 3.19 Definition. Let X be a normed vector space and $J : X \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$ convex. $u \in X'$ is called *subgradient* of J at position $y \in X$, if

$$J[y] + \langle u, x - y\rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J at position y.

§ 3.20 Proposition. *Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. Then,*

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \;\Leftrightarrow\; 0 \in \partial J[y^*].$$

Proof y^* is a minimizer, if and only if $J[y^*] \leq J[y]$ for all $y \in X$. Due to

$$J[y^*] = J[y^*] + \langle 0, y - y^*\rangle\,,$$

this is equivalent to $0 \in \partial J[y^*]$. □

§ 3.21 Proposition. *Let X be a normed vector space and $J : X \to \mathbb{R}$ Gâteaux-differentiable and convex. Then, $\partial J[y] = \{J'[y]\}$ for all $y \in X$.*

Proof From Proposition § 3.14(i) it immediately follows that $J'[y] \in \partial J[y]$, thus $\{J'[y]\} \subset \partial J[y]$. Now let $u \in \partial J[y]$. Let $\epsilon > 0$ and $z \in X$ be arbitrary. Then,

$$J[y] + \langle u, (y + \epsilon z) - y\rangle \leq J[y + \epsilon z] \Rightarrow \langle u, z\rangle \leq \frac{1}{\epsilon}\left(J[y + \epsilon z] - J[y]\right).$$

Since J is Gâteaux-differentiable, going to the limit $\epsilon \to 0$ implies

$$\langle u, z\rangle \leq \left\langle J'[y], z\right\rangle \text{ for all } z \in X.$$

For $z \in X$, we have $-z \in X$ and it follows that

$$\langle u, -z\rangle \leq \left\langle J'[y], -z\right\rangle \Rightarrow -\langle u, z\rangle \leq -\left\langle J'[y], z\right\rangle \Rightarrow \langle u, z\rangle \geq \left\langle J'[y], z\right\rangle.$$

Altogether, it follows that $\langle u, z\rangle = \langle J'[y], z\rangle$ for all $z \in X$, i.e. $u = J'[y]$, i.e. $\partial J[y] \subset \{J'[y]\}$. □
It is possible to characterize convexity completely without differentiability:

§ 3.22 Proposition. *Let X be a normed vector space, $M \subset X$ convex and $J : M \to \mathbb{R}_\infty$. Then,*

(i) J convex $\Leftarrow \forall x \in M\ \exists u \in X'\ \forall y \in M : J[y] \geq J[x] + \langle u, y - x\rangle$.

(ii) J strictly convex $\Leftarrow \forall x \in M\ \exists u \in X'\ \forall y \in M \setminus \{x\} : J[y] > J[x] + \langle u, y - x\rangle$.

Proof The proof follows immediately from the proof of "$\Leftarrow$" in Proposition § 3.14. □

§ *3.23 Remark.* The definition of the subdifferential $\partial J[y]$ can also be considered, if J is not convex. With the proposition above, it follows immediately that a function is convex, in case $\partial J[y] \neq \emptyset$ for all $y \in X$.

The direction "$\Rightarrow$" in Proposition § 3.22 does not hold without further assumption, e.g. consider $X = \mathbb{R}$, $M = [0, 1]$ and $J = \chi_{\{0,1\}}$. If M is open and $J : M \to \mathbb{R}$ continuous, then "$\Rightarrow$" holds (exercise). If $\dim(X) < \infty$, M open and $J : M \to \mathbb{R}$ convex already imply the continuity of J.

§ 4 Minimization using the proximal operator

Convex functions have a lot of structure that can be exploited to numerically compute minimizers. It is even possible to create efficient minimization algorithms for non-smooth, but convex functions.

§ 4.1 Definition. Let X be a vector space and $J : X \to \mathbb{R}_\infty$. Then

$$\operatorname{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$$

is called *effective domain* of J. Moreover,

$$\operatorname{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$$

is called *epigraph* of J. A convex J is called *proper*, if $\operatorname{epi}(J) \neq \emptyset$, i.e. if there exists $x \in X$ with $J[x] < \infty$. If X is a normed vector space, a proper convex functional J is called *closed*, if $\operatorname{epi}(J)$ is closed.

The set of closed proper convex functionals on X is denoted by $\Gamma_0(X)$.

§ 4.2 Proposition. *Let X be a Banach space and $J \in \Gamma_0(X)$. Then, J is weakly lower semi-continuous and $\{J \leq t\} := \{y \in X \,:\, J[y] \leq t\}$ is weakly sequentially closed for all $t \in \mathbb{R}$.*

Proof Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$. Then, [3, Bem. 6.26] implies

- J convex $\Leftrightarrow$ $\operatorname{epi}(J)$ convex
- J convex and lower semi-continuous $\Leftrightarrow$ $\operatorname{epi}(J)$ convex and closed
- J convex and lower semi-continuous $\Rightarrow$ $\{J \leq t\}$ closed for all $t \in \mathbb{R}$

If X is also complete and J convex, then [3, Korollar 6.28] implies that J is weakly lower semi-continuous, if and only of it is lower semi-continuous.

Let $J \in \Gamma_0(X)$. Since J is convex, $\operatorname{epi}(J)$ is convex. Since $\operatorname{epi}(J)$ is convex and closed, J is lower semi-continuous. Then, [3, Korollar 6.28] implies the weak lower semi-continuity. Since $\{J \leq t\}$ is convex and closed, [1, Proposition 8.13] implies that it is weakly sequentially closed. □

§ 4.3 Proposition. *Let X be a reflexive Banach space and $J \in \Gamma_0(X)$. Then, the mapping*

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called proximal mapping / proximal operator.

Proof From Proposition § 4.2, it follows that J is weakly lower semi-continuous. Due to the weak lower semi-continuity of norms (cd. [3, Korollar 6.15]) $p_y[\cdot] := J[\cdot] + \frac{1}{2} \|\cdot - y\|^2$ is weakly lower semi-continuous. Moreover, p_y is coercive, since convex functions decay at most linearly (the Brøndsted-Rockafellar theorem implies that ∂J is dense in $\operatorname{dom}(J)$, in particular, there exists a subgradient that is a linear lower bound).

Since $\operatorname{dom}(J) \neq \emptyset$, there exists $z \in \operatorname{dom}(J)$. It holds that $p_y[z] < \infty$, thus $M := \{u \in X \,:\, p_y[u] \leq p_y[z]\}$ is weakly sequentially closed. From Theorem § 2.15, the existence of minimizers of p_y on M follows and thus also on all of X. This minimizer is unique according to Proposition § 3.18, since p_y is strictly convex on $\operatorname{dom}(J)$ and $\operatorname{dom}(J)$ is convex. □

In the following, X always denotes a Hilbert space and we identify X' with X.

§ 4.4 Lemma. *Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then,*

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \Leftrightarrow y^* = \operatorname{prox}_{\tau J}[y^*].$$

Proof This can be shown elementally, it is even sufficient if X is just a reflexive Banach space (exercise). The statement for Hilbert spaces later follows from a more general statement. □

§ 4.5 Lemma. *Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y, y^* \in X$*

$$y^* = \operatorname{prox}_{\tau J}[y] \Leftrightarrow y \in y^* + \tau \partial J[y^*].$$

In particular, $y \in \operatorname{prox}_{\tau J}[y] + \tau \partial J[\operatorname{prox}_{\tau J}[y]]$ and $\partial J[\operatorname{prox}_{\tau J}[y]] \neq \emptyset$.

Proof Let $y, y^* \in X$ be arbitrary but fixed and $G_{\tau,y}[z] := \frac{1}{2\tau} \|z - y\|^2$.

"⇒" Let $y^* = \operatorname{prox}_{\tau J}[y]$. Then, y^* is a minimizer of the convex function $J + G_{\tau,y}$. Due to Proposition § 3.20, we have $0 \in \partial(J + G_{\tau,y})[y^*]$. Moreover, $y^* \in \operatorname{dom}(J) \cap \operatorname{dom}(G_{\tau,y})$ and $G_{\tau,y}$ is continuous. Then, due to [3, Satz 6.51 3.], we get

$$\partial(J + G_{\tau,y})[y^*] = \partial J[y^*] + \partial G_{\tau,y}[y^*] = \tfrac{1}{\tau}(y^* - y) + \partial J[y^*]$$

It follows that $0 \in \frac{1}{\tau}(y^* - y) + \partial J[y^*] \Rightarrow y \in y^* + \tau \partial J[y^*]$.

"⇐" Let $y \in y^* + \tau \partial J[y^*]$. Analogously to "⇒", we get

$$0 \in \tfrac{1}{\tau}(y^* - y) + \partial J[y^*] = \partial J[y^*] + \partial G_{\tau,y}[y^*].$$

Since $\partial J[z] = \emptyset$ for $z \in X \setminus \operatorname{dom}(J)$, we get $y^* \in \operatorname{dom}(J) = \operatorname{dom}(J) \cap \operatorname{dom}(G_{\tau,y})$. Combined with the continuity of $G_{\tau,y}$, it follows from [3, Satz 6.51 3.] that

$$0 \in \partial J[y^*] + \partial G_{\tau,y}[y^*] = \partial(J + G_{\tau,y})[y^*].$$

Due to Proposition § 3.20, we get $y^* \in \operatorname{argmin}(J + G_{\tau,y})$. Due to the uniqueness of this minimization problem (Proposition § 4.3), it follows that $y^* = \operatorname{prox}_{\tau J}[y]$.

□

§ 4.6 Corollary. *Let $J \in \Gamma_0(X)$ and $\tau > 0$. The proximal operator coincides with the so-called* resolvent *of the subdifferential, i.e. for $y \in X$, we have*

$$\{\operatorname{prox}_{\tau J}[y]\} = (\operatorname{id} + \tau \partial J)^{-1}[y].$$

Here, for a set-valued mapping $A : X \to \mathcal{P}(Y)$, the inversion is defined by

$$A^{-1} : Y \to \mathcal{P}(X), y \mapsto A^{-1}[y] := \{z \in X \,:\, y \in A[z]\}.$$

Proof

"⊂" Let $z = \operatorname{prox}_{\tau J}[y]$. Due to Lemma § 4.5, we have

$$y \in z + \tau \partial J[z] = (\operatorname{id} + \tau \partial J)[z].$$

Thus, with the definition of the inversion, it holds that $z \in (\operatorname{id} + \tau \partial J)^{-1}[y]$.

"$\supset$" Let $z \in (\mathrm{id} + \tau\partial J)^{-1}[y]$. It immediately follows that $y \in (\mathrm{id} + \tau\partial J)[z]$ and Lemma § 4.5 implies $z = \mathrm{prox}_{\tau J}[y]$.

$\square$

§ 4.7 Corollary. *Let $J \in \Gamma_0(X)$, $\tau > 0$ and $y, y^* \in X$. Then,*

$$y^* \in \partial J[y] \Leftrightarrow y = \mathrm{prox}_{\tau J}[y + \tau y^*].$$

Proof With the definition of the inversion and Corollary § 4.6, we get

$$\begin{aligned} y^* \in \partial J[y] &\Leftrightarrow y + \tau y^* \in (\mathrm{id} + \tau\partial J)[y] \\ &\Leftrightarrow y \in (\mathrm{id} + \tau\partial J)^{-1}[y + \tau y^*] \\ &\Leftrightarrow y = \mathrm{prox}_{\tau J}[y + \tau y^*]. \end{aligned}$$

$\square$

§ *4.8 Remark.* With $y^* = 0$ and the corollary above together with Proposition § 3.20, it directly follows that Lemma § 4.4 holds.

For the sake of simplicity, in the following, we confine to the case $X = \mathbb{R}^n$ and $\|\cdot\| = \|\cdot\|_2$, i.e. we consider minimization problems after discretization (Discretize Then Optimize).

§ *4.9 Remark.* Due to Lemma § 4.4, finding a minimizer of $J \in \Gamma_0(\mathbb{R}^n)$ is equivalent to finding a fixed point of $\mathrm{prox}_{\tau J}$. This motivates the *proximal point algorithm*

$$y^{k+1} = \mathrm{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value $y^0 \in \mathbb{R}^n$. If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value (proof will be given later).

§ *4.10 Example.* Let $\tau > 0$.

(i) For $J \equiv c \in \mathbb{R}$, we have $\mathrm{prox}_{\tau J}(y) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\tau c + \frac{1}{2}\|u - y\|_2^2\right) = y$.

(ii) Let $g \in \mathbb{R}^n$ and $J(y) = \frac{1}{2}\sum_{i=1}^{n}(y_i - g_i)^2 = \frac{1}{2}\|y - g\|_2^2$. It holds that

$$u^* := \mathrm{prox}_{\tau J}(y) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\frac{\tau}{2}\|u - g\|_2^2 + \frac{1}{2}\|u - y\|_2^2\right)$$

$$\Rightarrow 0 = \tau(u^* - g) + (u^* - y) \Rightarrow \mathrm{prox}_{\tau J}(y) = \frac{y + \tau g}{1 + \tau}$$

(iii) For $J(y) = \sum_{i=1}^{n} J_i(y_i)$ with $J_i \in \Gamma_0(\mathbb{R})$, we have $\mathrm{prox}_{\tau J}(y) = (\mathrm{prox}_{\tau J_1}(y_1), \ldots, \mathrm{prox}_{\tau J_n}(y_n))$.

(iv) For $J(y) := \|y\|_1$, $\mathrm{prox}_{\tau J}(y)$ is the so-called *soft threshold* operator (exercise), i.e.

$$(\mathrm{prox}_{\tau J}(y))_i = \begin{cases} y_i - \tau & y_i \geq \tau \\ 0 & |y_i| < \tau \\ y_i + \tau & y_i \leq -\tau \end{cases}.$$

(v) If $C \subset \mathbb{R}^n$ is a nonempty, closed, convex set, then

$$I_C : \mathbb{R}^n \to \mathbb{R}_\infty, y \mapsto \begin{cases} 0 & y \in C \\ \infty & y \notin C, \end{cases}$$

is called *indicator function* of C in $\Gamma_0(\mathbb{R}^n)$ and we have $\operatorname{prox}_{\tau I_C}(y) = \Pi_C(y)$, where Π_C is the Euclidean projection to C, i.e. $\Pi_C(y) = \operatorname*{argmin}_{z \in C} \|z - y\|_2$.

§ *4.11 Remark.* If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$ is determined by the necessary condition:

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \Rightarrow y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

This is the same as the backward Euler discretization of the gradient descent of J. Thus, for differentiable J, the proximal point algorithm is equivalent to the fully implicit gradient descent.

With additional assumptions on the structure of J, one can construct algorithms that also work in case $\operatorname{prox}_{\tau J}$ cannot be computed with sufficient efficiency. Very widespread are so-called *operator splitting methods*.

§ *4.12 Remark.* For $J = G + H$, we consider the optimization problem

$$\min_{y \in \mathbb{R}^n} (G(y) + H(y)),$$

where $G \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$ and $H \in \Gamma_0(\mathbb{R}^n)$, i.e. a part of the objective function is differentiable. Then, the *proximal gradient algorithm* is given by

$$y^{k+1} = \operatorname{prox}_{\tau_k H} \left(y^k - \tau_k \nabla G(y^k) \right)$$

for step sizes $\tau_k > 0$ and an initial value $y^0 \in \mathbb{R}^n$. Using

$$\mathcal{F}_\tau(y) = \frac{1}{\tau} \left(y - \operatorname{prox}_{\tau H} \left(y - \tau \nabla G(y) \right) \right),$$

we get

$$y^{k+1} = y^k - \tau_k \mathcal{F}_{\tau_k}(y^k).$$

This method is also called *forward-backward splitting*, since it combines a forward Euler gradient descent step in G with a proximal point algorithm step in H, which, in the smooth case, is equivalent to a backward Euler gradient descent step in H.

§ 4.13 Lemma. *Assumptions as in Remark § 4.12. Moreover, let ∇G be Lipschitz continuous with constant $L > 0$. Then, for all $y, z \in \mathbb{R}^n$ and $\tau \in [0, \frac{1}{L}]$, it holds that*

$$J(y - \tau \mathcal{F}_\tau(y)) \le J(z) + \mathcal{F}_\tau(y) \cdot (y - z) - \frac{\tau}{2} \|\mathcal{F}_\tau(y)\|_2^2 .$$

Proof Let $y, z \in \mathbb{R}^n$ be arbitrary but fixed. For an arbitrary $w \in \mathbb{R}^n$ and $v = w - y$, we get

$$\begin{aligned} G(w) &= G(y) + \nabla G(y) \cdot v + \int_0^1 (\nabla G(y + tv) - \nabla G(y)) \cdot v \, \mathrm{d}t \\ &\le G(y) + \nabla G(y) \cdot v + \int_0^1 \|\nabla G(y + tv) - \nabla G(y)\|_2 \|v\|_2 \, \mathrm{d}t \\ &\le G(y) + \nabla G(y) \cdot v + \int_0^1 Lt \|v\|_2^2 \, \mathrm{d}t = G(y) + \nabla G(y) \cdot v + \frac{L}{2} \|v\|_2^2 . \end{aligned}$$

For $w = y - \tau \mathcal{F}_\tau(y)$, it follows that $v = -\tau \mathcal{F}_\tau(y)$ and

$$G(y - \tau \mathcal{F}_\tau(y)) \leq G(y) - \tau \nabla G(y) \cdot \mathcal{F}_\tau(y) + \frac{\tau^2 L}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 .$$

For $\tau \in [0, \frac{1}{L}]$, we get

$$G(y - \tau \mathcal{F}_\tau(y)) \leq G(y) - \tau \nabla G(y) \cdot \mathcal{F}_\tau(y) + \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 . \tag{$*^1$}$$

Due to Lemma § 4.5, for $w \in \mathbb{R}^n$, it holds that

$$w \in \operatorname{prox}_{\tau H}(w) + \tau \partial H(\operatorname{prox}_{\tau H}(w)).$$

From the definition of $\mathcal{F}_\tau$, we get $\operatorname{prox}_{\tau H}(y - \tau \nabla G(y)) = y - \tau \mathcal{F}_\tau(y)$, thus, for $w = y - \tau \nabla G(y)$, it follows that

$$\begin{aligned} & y - \tau \nabla G(y) \in y - \tau \mathcal{F}_\tau(y) + \tau \partial H(y - \tau \mathcal{F}_\tau(y)) \\ & \Rightarrow \mathcal{F}_\tau(y) - \nabla G(y) \in \partial H(y - \tau \mathcal{F}_\tau(y)). \end{aligned}$$

For $w \in \mathbb{R}^n$ and $u \in \partial H(w)$, we get

$$H(w) + u \cdot (z - w) \leq H(z).$$

With $w = y - \tau \mathcal{F}_\tau(y)$ and $u = \mathcal{F}_\tau(y) - \nabla G(y)$, it holds that $u \in \partial H(w)$ and we get

$$H(y - \tau \mathcal{F}_\tau(y)) \leq H(z) - (\mathcal{F}_\tau(y) - \nabla G(y)) \cdot (z - y + \tau \mathcal{F}_\tau(y)) . \tag{$*^2$}$$

Combined with $G(y) \leq G(z) + \nabla G(y) \cdot (y - z)$ (Corollary § 3.16), $(*^1)$ and $(*^2)$ lead to

$$\begin{aligned} J(y - \tau \mathcal{F}_\tau(y)) &\leq G(y) - \tau \nabla G(y) \cdot \mathcal{F}_\tau(y) + \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 + H(y - \tau \mathcal{F}_\tau(y)) \\ &\leq G(z) + \nabla G(y) \cdot (y - z) - \tau \nabla G(y) \cdot \mathcal{F}_\tau(y) + \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 \\ &\quad + H(z) - (\mathcal{F}_\tau(y) - \nabla G(y)) \cdot (z - y + \tau \mathcal{F}_\tau(y)) \\ &\leq G(z) + H(z) + \mathcal{F}_\tau(y) \cdot (y - z) - \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 . \end{aligned}$$

□

§ 4.14 Theorem. *Assumptions as in Remark § 4.12. Moreover, let ∇G be Lipschitz continuous with constant $L > 0$, $\tau_k \in [\tau_{min}, \frac{1}{L}]$, where $\tau_{min} \in (0, \frac{1}{L}]$, and let minimizer y^* of J exist. Then, the proximal gradient algorithm converges. More precisely, it holds that*

$$0 \leq J(y^k) - J(y^*) \leq \frac{1}{2k\tau_{min}} \left\| y^0 - y^* \right\|_2^2 = O\left(\tfrac{1}{k}\right) .$$

Proof Let $y^+ = y - \tau \mathcal{F}_\tau(y)$. For $z = y$, it follows from Lemma § 4.13 that

$$J(y^+) \leq J(y) - \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 \leq J(y).$$

Thus, we have $J(y^{i+1}) \leq J(y^i)$ for $i \in \mathbb{N}_0$. For $z = y^*$ and $\underline{J} = J(y^*)$, Lemma § 4.13 gives

$$\begin{aligned} J(y^+) - \underline{J} &\leq \mathcal{F}_\tau(y) \cdot (y - y^*) - \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 = \frac{1}{2\tau} \left(\left\| y - y^* \right\|_2^2 - \left\| y - y^* - \tau \mathcal{F}_\tau(y) \right\|_2^2 \right) \\ &= \frac{1}{2\tau} \left(\left\| y - y^* \right\|_2^2 - \left\| y^+ - y^* \right\|_2^2 \right) . \end{aligned}$$

In particular, $\|y^+ - y^*\|_2^2 \leq \|y - y^*\|_2^2$, i.e. the distance to the minimizer decreases. If y^+ is not already a minimizer, we have $J(y^+) \neq \underline{J}$ and thus the strict inqeuality $\|y^+ - y^*\|_2^2 < \|y - y^*\|_2^2$ holds.

Summing the inequality for $y = y^{i-1}$ and $y^+ = y^i$ with $\tau = \tau_{i-1} \geq \tau_{\min}$ gives

$$\sum_{i=1}^{k} \left(J(y^i) - \underline{J}\right) \leq \sum_{i=1}^{k} \frac{1}{2\tau_{i-1}} \left(\left\|y^{i-1} - y^*\right\|_2^2 - \left\|y^i - y^*\right\|_2^2\right)$$
$$\leq \frac{1}{2\tau_{\min}} \left(\left\|y^0 - y^*\right\|_2^2 - \left\|y^k - y^*\right\|_2^2\right) \leq \frac{1}{2\tau_{\min}} \left\|y^0 - y^*\right\|_2^2 .$$

Since $J(y^{i+1}) \leq J(y^i)$, we get

$$J(y^k) - \underline{J} \leq \frac{1}{k} \sum_{i=1}^{k} \left(J(y^i) - \underline{J}\right) \leq \frac{1}{2k\tau_{\min}} \left\|y^0 - y^*\right\|_2^2 .$$

Thus, $O(1/\epsilon)$ iterations are necessary to get $J(y^k) - \underline{J} \leq \epsilon$. □

§ *4.15 Remark.* Since the proximal gradient algorithm is a generalization of several other methods, Theorem § 4.14 proves also their convergence.

- With $G = 0$ and $H = J$, one gets the proximal point algorithm and since $\nabla 0$ is Lipschitz continuous with constant 0, it follows (as long as J has a minimizer) the convergence for arbitrary, bounded step sizes.
- With $G = J$ and $H = 0$, one gets the fully explicit gradient descent. If ∇J is Lipschitz continuous and has a minimizer, we get convergence for suitable τ_n.
- If $C \subset \mathbb{R}^n$ is nonempty, convex and closed, $G = J$ and $H = I_C$ lead to the so-called *projected gradient descent*, which minimizes $J(y)$ under the constraint $y \in C$. If ∇J is again Lipschitz continuous and there exists a minimizer of J under the above constraint, we get convergence for suitable τ_n.

§ 4.16 Definition. Let $J : X \to \mathbb{R}_\infty$ be proper. Then,

$$J^* : X' :\to \mathbb{R}_\infty, x' \mapsto \sup_{x \in X} \left(\langle x', x \rangle - J[x]\right)$$

is called *Fenchel conjugate* of J.

§ *4.17 Remark.* Particularly relevant in image processing are problems of the type

$$\min_{y \in \mathbb{R}^n} (G(y) + H(Ay)),$$

where $G \in \Gamma_0(\mathbb{R}^n)$, $H \in \Gamma_0(\mathbb{R}^m)$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Here, the second term is decomposed in H and A to simplify the computation of the proximal operator. Since $H \in \Gamma_0(\mathbb{R}^m)$, it follows from [3, Lemma 6.63] that $H = H^{**}$. Thus, we get

$$\inf_{y \in \mathbb{R}^n} (G(y) + H(Ay)) = \inf_{y \in \mathbb{R}^n} (G(y) + H^{**}(Ay))$$
$$= \inf_{y \in \mathbb{R}^n} (G(y) + \sup_{z \in \mathbb{R}^m} (Ay \cdot z - H^*(z)))$$
$$= \inf_{y \in \mathbb{R}^n} \sup_{z \in \mathbb{R}^m} (Ay \cdot z + G(y) - H^*(z)) .$$

The necessary conditions for z and y are

$$0 \in \partial_z \left(Ay \cdot z - H^*(z)\right) = Ay - \partial H^*(z) \Rightarrow Ay \in \partial H^*(z),$$
$$0 \in \partial_y \left(A^T z \cdot y + G(y)\right) = A^T z + \partial G(y) \Rightarrow -A^T z \in \partial G(y).$$

Due to Corollary § 4.7, this is equivalent to

$$z = \operatorname{prox}_{\sigma H^*}[z + \sigma Ay],$$
$$y = \operatorname{prox}_{\tau G}[y - \tau A^T z]$$

for $\tau, \sigma > 0$ and motivates the algorithm

$$z^{k+1} = \operatorname{prox}_{\sigma H^*}[z^k + \sigma A\overline{y}^k]$$
$$y^{k+1} = \operatorname{prox}_{\tau G}[y^k - \tau A^T z^{k+1}]$$
$$\overline{y}^{k+1} = y^{k+1} + \theta(y^{k+1} - y^k)$$

for $\theta \in [0, 1]$, $\overline{y}^0 = y^0 \in \mathbb{R}^n$, $z^0 \in \mathbb{R}^m$. The third step of this is an extrapolation step. The algorithm shows the effect of the decomposition of the second term in H and A: One just has to compute $\operatorname{prox}_{\sigma H^*}$ and typically A is chosen such that $\operatorname{prox}_{\sigma H^*}$ can be computed pointwise, i.e. as in § 4.10 (iii)).

This algorithm was proposed by Chambolle and Pock in 2010, is currently very popular in image processing (2000+ citations) and belongs to the class of primal-dual methods. In particular, the algorithm is well suited for models that use the total variation as regularizer.

§ *4.18 Remark.* For the primal-dual method we consider discrete images $y = (y_{i,j}) \in X := \mathbb{R}^{M\times N}$, i.e. Ω is a rectangle and discretized with a cartesian grid with M nodes in x-direction, N nodes in y-direction and grid width h. The gradient of y is computed with forward difference quotients, i.e. $(\nabla^h y)_{i,j} = ((\partial_1^{h+} y)_{i,j}, (\partial_2^{h+} y)_{i,j}) \in X \times X$, where

$$(\partial_1^{h+} y)_{i,j} = \begin{cases} \frac{y_{i+1,j} - y_{i,j}}{h} & i < M; \\ 0 & i = M; \end{cases}, \; j = 1, \ldots, N,$$
$$(\partial_2^{h+} y)_{i,j} = \begin{cases} \frac{y_{i,j+1} - y_{i,j}}{h} & j < N; \\ 0 & j = N; \end{cases}, \; i = 1, \ldots, M.$$

Then, the discretized total variation of y is

$$H(\nabla^h y) := \left\|\nabla^h y\right\|_1 := \sum_{i,j} \left|(\nabla^h y)_{i,j}\right|.$$

For $g \in X$, the discretized data term is

$$G(y) = \frac{1}{2\lambda} \|y - g\|_2^2 \overset{\S\, 4.10\, \text{(ii)}}{\Longrightarrow} \operatorname{prox}_{\tau G}(y) = \frac{y + \frac{\tau}{\lambda} g}{1 + \frac{\tau}{\lambda}}.$$

With $\mathbb{R}^n \simeq X$, $\mathbb{R}^m \simeq X \times X$ and $A \simeq \nabla^h$, $G(y) + H(Ay)$ is a discretization of the ROF-model that fits to Remark § 4.17. One can show (exercise), that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \;:\; \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, it follows from Example § 4.10 (v) that $\mathrm{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

Thus, the full algorithm for the ROF model is

$$\begin{aligned} z^{k+1} &= \Pi_P(z^k + \sigma A\overline{y}^k) \\ y^{k+1} &= \frac{y^k - \tau A^T z^{k+1} + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}} \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{aligned}$$

There are variants of the primal-dual method, which exploit the strict convexity of the data term G for an even faster convergence.

Bibliography

[1] Hans Wilhelm Alt. *Linear Functional Analysis: An Application-Oriented Introduction.* Springer, 2016.

[2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems.* New York: Oxford University Press, 2000.

[3] Kristian Bredies and Dirk Lorenz. "Mathematische Bildverarbeitung." In: 1. Auflage. Vieweg+Teubner Verlag, 2011. Chap. 6.