Assignment 4 — Total Variation Denoising

Image Processing and Pattern Recognition

Deadline: January 31, 2025

1 Goal

In this assignment we will implement an image-denoising algorithm using a total variation regularizer. Specifically, we discuss primal-dual optimization algorithms, and their application to total-variation denoising. This assignment heavily relies on [1], which is recommended literature for anyone interested in this topic.

2 Methods

2.1 Primal-Dual Algorithm

Consider the optimization problem

$$\min_{x \in X} F(Kx) + G(x), \tag{1}$$

where $K : X \to Y$ is a linear operator, $F : Y \to [0, \infty]$ and $G : X \to [0, \infty]$ are proper convex lower-semicontinuous extended real-valued functions. See [2] for a definition of these (and the following) concepts. X and Y are finite-dimensional vector spaces equipped with inner product $\langle \cdot\,, \cdot \rangle$ and norm $|| \cdot || = \langle \cdot\,, \cdot \rangle^{\frac{1}{2}}$. Problems of this form can be transformed into the saddle-point problem

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y), \tag{2}$$

where y is the dual variable and F^* is the convex conjugate of F. The primal-dual hybrid gradient algorithm

$$\begin{cases} x^{k+1} = \operatorname{prox}_{\tau G}(x^k - \tau K^* y^k), \\ y^{k+1} = \operatorname{prox}_{\sigma F^*}(y^k + \sigma K(2x^{k+1} - x^k)), \end{cases} \tag{3}$$

can solve such problems efficiently. Here, the proximal operator $\operatorname{prox}_{\alpha H} : Z \to Z$ of a proper extended real-valued function $H : Z \to (-\infty, \infty]$ (Z is a Hilbert space) is the map

$$\operatorname{prox}_{\alpha H}(v) = \arg\min_{w} \left\{ \alpha H(w) + \frac{1}{2} \|v - w\|^2 \right\}. \tag{4}$$

The positive scalars τ and σ are the step-sizes of the primal and dual variable respectively. They have to be chosen such that $\tau\sigma \|K\|^2 < 1$, where $\|K\| = \max\{\|Kx\| : x \in X, \|x\| \leq 1\}$ is the induced operator norm of K.

2.2 Total Variation

In the continuous setting, the total variation reads

$$\mathrm{TV}(u) = \int_\Omega |Du|, \tag{5}$$

with the d-dimensional image domain $\Omega \subset \mathbb{R}^d$ and the distributional derivative Du, which reduces to ∇u for sufficiently smooth functions. In the discrete setting, we consider a Cartesian grid of size $M{\times}N$, i.e.

$$\{(ih, jh)\colon 1 \leq i \leq M, 1 \leq j \leq n\}, \tag{6}$$

where h is the distance between neighbouring grid points and we denote the indices of the discrete locations (ih, jh) in the image domain with (i, j). In this assignment we let $h{=}1$. The image u is assumed to be an element of a vector space $X = \mathbb{R}^{MN}$, with the standard inner product

$$\langle u, v\rangle_X = \sum_{i,j} u_{i,j} v_{i,j}, \quad u, v \in X. \tag{7}$$

The gradient ∇u is a vector in $Y{=}X{\times}X$ and we define $\nabla\colon X \to Y$, $u \mapsto \nabla u$ as

$$(\nabla u)_{i,j} = \begin{pmatrix} (\nabla u)^1_{i,j} \\ (\nabla u)^2_{i,j} \end{pmatrix} \tag{8}$$

where

$$(\nabla u)^1_{i,j} = \begin{cases} \frac{u_{i+1,j} - u_{i,j}}{h} & \text{if } i < M \\ 0 & \text{if } i = M \end{cases}, \quad (\nabla u)^2_{i,j} = \begin{cases} \frac{u_{i,j+1} - u_{i,j}}{h} & \text{if } j < N \\ 0 & \text{if } i = N \end{cases}. \tag{9}$$

We also define the scalar product in Y as

$$\langle p, q\rangle_Y = \sum_{i,j} p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j}, \quad p = (p^1, p^2),\ q = (q^1, q^2) \in Y. \tag{10}$$

The discrete total variation is then

$$\mathrm{TV}(u) = \|\nabla u\|_{2,1} \tag{11}$$

with the discrete total variation norm

$$\|\nabla u\|_{2,1} = \sum_{i,j} \left\|(\nabla u)_{i,j}\right\|_2, \qquad \left\|(\nabla u)_{i,j}\right\|_2 = \sqrt{\left((\nabla u)^1_{i,j}\right)^2 + \left((\nabla u)^2_{i,j}\right)^2}. \tag{12}$$

In addition, we choose the discrete divergence operator $\operatorname{div} : Y \to X$ to be adjoint to the gradient operator. In particular, by the identity $\langle \nabla u, p\rangle_Y = -\langle u, \operatorname{div} p\rangle_X$, we find $-\operatorname{div} = \nabla^*$, where $(\,\cdot\,)^*$ denotes adjointness.

Combining the total variation prior with a squared ℓ_2 data term yields the famous TV-ℓ_2 model

$$\min_u \|\nabla u\|_{2,1} + \frac{\lambda}{2}\|u-g\|^2 \tag{13}$$

where $g \in X$ is the noisy observation. It is also known as the ROF model by the names of the authors (Rudin, Osher and Fatemi) who first proposed the model [3]. Since the total variation norm penalizes edges in the image, the solution to the above optimization problem is a piecewise constant image, with λ controlling the trade-off between the data-fidelity term $\frac{1}{2}\|u-g\|^2$ and the TV norm $\|\nabla u\|_{2,1}$.

2.3 Primal-Dual ROF

Casting (13) in the form of (1), we identify $K = \nabla$, $F = ||\cdot||_{2,1}$ and $G = \frac{\lambda}{2}||\cdot - g||_2^2$. Using the indicator function

$$\delta_A(v) = \begin{cases} 0 & \text{if } v \in A, \\ \infty & \text{else,} \end{cases} \tag{14}$$

we can detail the convex conjugate $(||\cdot||_{2,1})^* = \delta_P$, where the set $P = \{p \in Y : \|p\|_{2,\infty} \leq 1\}$ is the product of point-wise ℓ_2 balls: Here, $\|\cdot\|_{2,\infty}$ denotes the point-wise maximum of the 2-norms $\|p\|_{2,\infty} = \max_{i,j}\|p_{i,j}\|_2$. Thus the primal-dual formulation of (13) reads

$$\min_{u \in X}\max_{p \in Y} -\langle u, \operatorname{div} p\rangle_X + \frac{\lambda}{2}\|u-g\|_2^2 - \delta_P(p). \tag{15}$$

To apply (3), it remains to detail the proximal maps of G and F^*. The proximal operator of G is a point-wise quadratic minimization problem solved by

$$\operatorname{prox}_{\tau\frac{\lambda}{2}\|\cdot - g\|_2^2}(\tilde{u}) = \frac{\tilde{u} + \tau\lambda g}{1 + \tau\lambda}. \tag{16}$$

The proximal operator of F^* reduces to a point-wise projection onto $\|p_{i,j}\|_2 \leq 1$:[1]

$$p = \operatorname{prox}_{\sigma\delta_P}(\tilde{p}) \iff p_{i,j} = \frac{\tilde{p}_{i,j}}{\max(1, \|p_{i,j}\|_2)}. \tag{17}$$

To select the step-sizes, we utilize $\|\nabla\|^2 < 8$ (see [4]) and take the standard choice $\tau = \sigma = \frac{1}{\sqrt{8}}$.

[1] By (14), δ_P is invariant w.r.t. rescaling, i.e. $\delta_P = \sigma\delta_P$. Thus, the dual step-size σ vanishes in the computation of the proximal map.

3 Tasks

3.1 Implementation (13P)

Implement the proximal maps (16) and (17). We provide the skeleton file `pd.py` as well as reference reconstructions after 200 iterations using $\lambda = 10$.

3.2 Discussion (12P)

Show some results for a range of λ (the initial provided λ is a good starting point). How does it influence the output? Do you think the total variation is a good prior for natural images? What could be its drawbacks? Do you know any other priors that are commonly used for natural images?

3.3 Bonus Challenge

Find the optimal regularization parameter λ. As in Assignment 2, we award $\{5, 4, 3\}$ bonus points for the best three groups. If you want to participate in this challenge, please upload your results in a seperate folder by the name of "test_out". Hint: Nowadays, it is well known that the best image quality is achieved before (13) reaches a minimum. Thus, you may also play around with the number of iterations of the optimization algorithm. You can get some inspiration from [5] and use your own images (or any images from the internet) as a "training set". The noisy test images were generated with Gaussian noise with standard deviation 0.1 (see `utils.py` from Assignment 2).

References

[1] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, December 2010.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[4] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, Jan 2004.

[5] Alexander Effland, Erich Kobler, Karl Kunisch, and Thomas Pock. Variational networks: An optimal control approach to early stopping variational methods for image restoration. *Journal of Mathematical Imaging and Vision*, 62(3):396–416, March 2020.