Assignment 3

Convex Optimization SS2022

Martin Zach (martin.zach@icg.tugraz.at)

May 16, 2022

Submission: Upload your report as a single pdf–file ($\leq$ 10MB) to the TeachCenter and include your implementation (main.py) in your submission.
Deadline: June, 6^{th} 2022, 18:00

The ROF model for image denoising (10P)

In this task, we would like to reconstruct images degraded by additive Gaussian noise, see Figure 1. We use the well-studied Rudin, Osher and Fatemi (ROF) model [1], also known as TV-ℓ^2 model, which yields good results despite its simplicity. Let $\mathbf{u}_0 \in \mathbb{R}^{3n}$ be the observed noisy image of size $n = n_1 \times n_2$ with 3 color channels, the discrete version of the ROF model using a finite difference discretization of images is defined as

$$\min_{\mathbf{u}\in\mathbb{R}^n} \left\{ \mathrm{E}_P(\mathbf{u}) := \lambda \|D\mathbf{u}\|_{2,1} + \frac{1}{2}\|\mathbf{u} - \mathbf{u}_0\|_2^2 \right\}, \tag{P}$$

where $\mathbf{u} \in \mathbb{R}^{3n}$ is the denoised image and λ is a parameter balancing the tradeoff between regularization and data fidelity. The regularization term is given by the isotropic total variation for color images that utilizes the finite difference operator $D \in \mathbb{R}^{6n\times 3n}$ defined as

$$D = \begin{pmatrix} D_x & 0 & 0 \\ D_y & 0 & 0 \\ 0 & D_x & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_x \\ 0 & 0 & D_y \end{pmatrix}$$

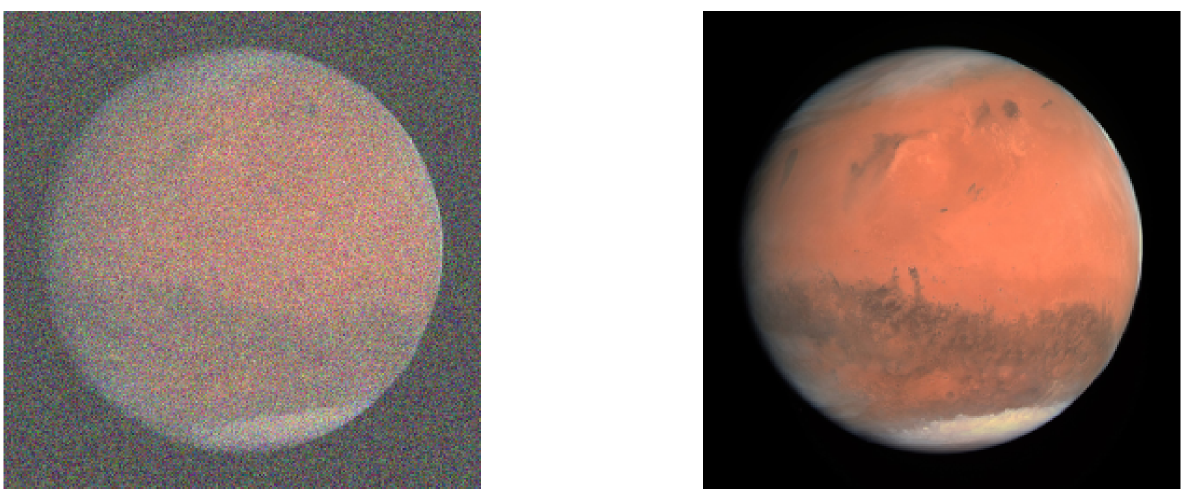

Figure 1: Left noisy observed image $\mathbf{u}_0$ and right the original uncorrupted image of Mars.

using the operators $D_x, D_y \in \mathbb{R}^{n\times n}$

$$D_x = \begin{pmatrix} -1 & 1 & 0 & \cdots \\ 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \qquad D_y = \begin{pmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots \\ 0 & -1 & \cdots & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

which extract image gradients in x/y direction assuming a lexicographic ordering of the image pixels and Neumann boundary conditions. Let $\mathbf{z} \in \mathbb{R}^{6n}$, we define the $\ell^{2,1}$-norm as follows:

$$\|\mathbf{z}\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=0}^{5} (z_{i+jn})^2}.$$

Here, we would like to solve this method by means of dualization and the application of the proximal gradient method as well as FISTA. Your tasks are as follows:

- Show that the dual problem of (P) is given by

$$\max_{\mathbf{p}\in\mathbb{R}^{6n}} \left\{ \mathrm{E}_D(\mathbf{p}) := -\frac{1}{2}\|D^\top \mathbf{p}\|_2^2 + \left\langle D^\top \mathbf{p}, \mathbf{u}_0 \right\rangle - \delta_{B_{\|\cdot\|_{2,\infty}}[0,\lambda]}(\mathbf{p}) \right\}, \tag{D}$$

  where $\mathbf{p} \in \mathbb{R}^{6n}$ is the dual variable and

$$B_{\|\cdot\|_{2,\infty}}[0,\lambda] = \left\{ \mathbf{p} \in \mathbb{R}^{6n} : \sqrt{\sum_{j=0}^{5} (p_{i+jn})^2} \leq \lambda \text{ for } i = 1, \ldots, n \right\}.$$

  Explain each step in your derivation.

- Show that the primal solution $\mathbf{u}$ can be obtained from the dual $\mathbf{p}$ by

$$\mathbf{u} = \mathbf{u}_0 - D^\top \mathbf{p}. \tag{1}$$

- Prove that

$$\|D\|_2 \leq \sqrt{8}.$$

- Implement proximal gradient ascent as well as FISTA to solve the dual ROF problem (D).

Compare both algorithms for $\lambda \in \{0.25, 1.0, 4.0\}$:

- Include a logarithmic plot of the dual energy (D) and the primal energy (P) for both algorithms in your report. *Hint:* Use (1) to convert dual variables to corresponding primal variables.
- Discuss the convergence rates of both algorithms and recommend the faster one.
- What is the influence of λ on the convergence?
- Include the resulting image $\mathbf{u}$ and the dual variable $\mathbf{p}$ in your report and discuss their structure.
- Plot the primal-dual gap $\mathrm{G}(\mathbf{u}, \mathbf{p}) = \mathrm{E}_P(\mathbf{u}) - \mathrm{E}_D(\mathbf{p})$ over the iterations for both algorithms.

Fenchel conjugate (5P)

- Compute the Fenchel conjugate of $f : \mathbb{R} \to \mathbb{R}$, where f is given by
  1. $f(\mathbf{x}) = \frac{|\mathbf{x}|^p}{p}$ for $p \in (1, \infty)$,
  2. $f(\mathbf{x}) = \exp(\mathbf{x})$,
  3. $f(\mathbf{x}) = \cosh(\mathbf{x})$,
  4. $f(\mathbf{x}) = \sqrt{1 + \mathbf{x}^2}$.
- Let $f, g : \mathbb{E} \to (-\infty, \infty]$ and $\alpha \in (0, 1)$. Show that

$$(\alpha f + (1-\alpha) g)^* \leq \alpha f^* + (1-\alpha) g^*.$$

Infimal convolution (6P)

- Let $f : \mathbb{E} \to (-\infty, \infty]$ be proper, lower semicontinuous and convex. Set $g_0 = \delta_{\{0\}}$ and $g_{n+1} = f \square g_n$ for $n \in \mathbb{N}$. Prove that
$$g_n(\mathbf{x}) = n f(\tfrac{\mathbf{x}}{n})$$
for all $n \geq 1$.

- Let $f : \mathbb{E} \to (-\infty, \infty]$ be subadditive such that $f(\mathbf{0}) = 0$. Show that $f \square f = f$. Recall that a function $f : \mathbb{E} \to (-\infty, \infty]$ is subadditive if $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$.

Proximal map (4P)

Compute the proximal map for $f : \mathbb{R}^n \to \mathbb{R}$ given by

1. $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$, where $A \in \mathbb{R}^{n\times n}$ is positive definite, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,
2. $f(\mathbf{x}) = \|\mathbf{x}\|_2$,
3. $f(\mathbf{x}) = \|\mathbf{x}\|_1$,
4. $f(\mathbf{x}) = -\sum_{i=1}^n \log x_i$.

References

[1] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.