Assignment: Total Variation Regularization

TV Regularization

The total variation (TV) functional $\varphi : \mathbb{R}^n \to \mathbb{R}$ is defined

$$\forall\, s = (s_1, \ldots, s_n) \in \mathbb{R}^n : \quad \varphi(s) = \sum_{i=2}^{n} |s_i - s_{i-1}|. \tag{1}$$

If we represent a one-dimensional signal as a vector $s \in \mathbb{R}^n$, where n is the length of the signal, then $\varphi(s)$ is the ℓ_1 norm of a discrete form of the derivative of s. It measures the size of "jumps" in a signal. In particular, $\varphi(s)$ is small if s is a piecewise constant signal with small jumps.

Exercise 1 *Show that we can write* (1) *as* $\varphi(s) = \|\mathbf{H}s\|_1$*, where* $\mathbf{H}$ *is a matrix of size* $(n-1) \times n$ *and* $\|\cdot\|_1$ *is the* ℓ_1 *norm on* $\mathbb{R}^{n-1}$*. Compute the spectral norm* $\|\mathbf{H}\|_2$*.*

The TV denoising operator $\mathrm{D}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ for some $\lambda > 0$ is defined as

$$\mathrm{D}_\lambda(s) = \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|s - \boldsymbol{z}\|_2^2 + \lambda \varphi(\boldsymbol{z}) \right\}. \tag{2}$$

This can be seen as the "denoising" of s in the following sense: we wish to find the closest signal to s under the constraint that its total variation is small. This is called *TV denoising* in the imaging community[1].

Exercise 2 *Write code that simulates* D_λ*. The input to the code are* s *and* $\lambda > 0$ *and the output is* $\mathrm{D}_\lambda(s)$*.*

We can go beyond denoising — once we can denoise using some form of "prior" on signals, we can apply the same prior to general inverse problems[2]. The prior in the present case is that the signals should have a small total variation. We will focus on the compressed sensing problem[3], where we have an (unknown) ground truth signal $\boldsymbol{\xi}$ for which we have compressed measurements of the form

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{\xi} + \sigma\boldsymbol{\eta}, \tag{3}$$

where

- $\mathbf{A}$ is an $(m \times n)$ Gaussian random matrix,
- $\boldsymbol{\eta} \in \mathbb{R}^m$ is white Gaussian noise, i.e., $\eta_1, \ldots, \eta_m$ are iid Gaussian with zero mean and unit variance, and
- σ is the noise level.

By Gaussian random matrix, we mean that each component $\mathbf{A}_{ij}$ is drawn from a Gaussian distribution with zero mean and unit variance.

Exercise 3 *Write code to simulate the measurement process in* (3)*. The input to the code is* m *and* n*, the ground truth signal* $\boldsymbol{\xi}$*, and the noise level* σ*.*

Compressed sensing aims to recover the unknown signal $\boldsymbol{\xi}$ from the measurement $\boldsymbol{x}$. In particular, this can be done using the reconstruction operator $\mathrm{R}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\mathrm{R}_\lambda(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \mathbf{A}\boldsymbol{z}\|_2^2 + \lambda \varphi(\boldsymbol{z}) \right\}. \tag{4}$$

This is called *TV reconstruction* in the literature — we want a reconstruction $\boldsymbol{z}$ with small total variation that conforms to the measurement model in (3).

[1] Rudin et al. (1992). Nonlinear total variation based noise removal algorithms. Physica D.

[2] Osher et al. (2005). An iterative regularization method for total variation-based image restoration. Multiscale Modeling and Simulation.

[3] Donoho (2006). Compressed sensing. IEEE Transactions on Information Theory.

Exercise 4 *Using the code developed in Exercise 2, write code to compute* R_λ. *The input to the code is* $\boldsymbol{x}, \mathbf{A}$ *and* $\lambda > 0$ *and the output is* $\mathrm{R}_\lambda(\boldsymbol{x})$.

Algorithms

In the rest of the note, we describe the algorithms for computing (2) and (4).

TV denoising: We can solve (4) in an iterative manner using (2). We will use an algorithm called the *proximal gradient descent*[4]. We start with an arbitrary $\boldsymbol{z}_0 \in \mathbb{R}^n$ and generate a sequence of estimates $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$ as follows:

$$\forall\, k \geqslant 0: \quad \boldsymbol{z}_{k+1} = \mathrm{D}_\lambda\big(\boldsymbol{z}_k - \gamma \nabla f(\boldsymbol{z}_k)\big), \tag{5}$$

where $\gamma > 0$ is a step size, D_λ is the denoising operator in (2), and

$$f(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{x} - \mathbf{A}\boldsymbol{z}\|_2^2.$$

We can use $\gamma = 1/\|\mathbf{A}\|_2^2$, where $\|\mathbf{A}\|_2$ is the spectral norm of $\mathbf{A}$. For this particular choice of γ, it can be shown that the sequence $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$ generated by (5) converges to a solution of (4).

TV regularization: There are different iterative methods for solving (2). We will use Chambolle's algorithm[5] that uses duality along with projected gradient descent.

Exercise 5 *Show that*

$$\forall\, \boldsymbol{\theta} \in \mathbb{R}^n: \quad \|\boldsymbol{\theta}\|_1 = \max_{\boldsymbol{y} \in \mathbb{B}^n} \boldsymbol{y}^\top \boldsymbol{\theta}, \tag{6}$$

where $\mathbb{B}^n$ *is the unit* ℓ_∞ *ball in* $\mathbb{R}^n$*:*

$$\mathbb{B}^n = \big\{\boldsymbol{y} \in \mathbb{R}^n : \, |y_i| \leqslant 1, \, i = 1, \ldots, n\big\}.$$

On substituting $\boldsymbol{\theta} = \mathbf{H}\boldsymbol{z}$ in (6), we can write the optimization problem in (2) as

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \max_{\boldsymbol{y} \in \mathbb{B}^{n-1}} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \lambda\, \boldsymbol{y}^\top \mathbf{H}\boldsymbol{z} \right\}.$$

Using Sion's minimax theorem[6], we can interchange the min and max to get

$$\max_{\boldsymbol{y} \in \mathbb{B}^{n-1}} \min_{\boldsymbol{z} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \lambda\, \boldsymbol{y}^\top \mathbf{H}\boldsymbol{z} \right\}. \tag{7}$$

Exercise 6 *Let*

$$\psi(\boldsymbol{y}, \boldsymbol{x}) = \min_{\boldsymbol{z} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \lambda\, \boldsymbol{y}^\top \mathbf{H}\boldsymbol{z} \right\}$$

Find an exact formula for ψ *and check that it is a concave function. In particular, compute* $\nabla_{\boldsymbol{y}} \psi(\boldsymbol{y}, \boldsymbol{x})$ *and find* $L > 0$ *such that*

$$\forall\, \boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^{n-1}: \quad \|\nabla_{\boldsymbol{y}} \psi(\boldsymbol{y}_1, \boldsymbol{x}) - \nabla_{\boldsymbol{y}} \psi(\boldsymbol{y}_2, \boldsymbol{x})\|_2 \leqslant L \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_2.$$

We can write (7) as

$$\max_{\boldsymbol{y} \in \mathbb{B}^{n-1}} \psi(\boldsymbol{y}, \boldsymbol{x}). \tag{8}$$

Exercise 7 *If* $\boldsymbol{y}^*$ *is the optimal solution of* (8)*, then show that the optimal solution of* (2) *is*

$$\boldsymbol{z}^* = \boldsymbol{x} - \lambda\, \mathbf{H}^\top \boldsymbol{y}^*. \tag{9}$$

[4] Beck (2017). First-order methods in optimization. SIAM.
[5] Chambolle (2004). An algorithm for total variation minimization and applications. JMIV.
[6] M. Sion (1958). On general minimax theorems. Pacific J. Math.

We have thus reduced the original problem (2) to (8). The advantage with (8) is that the objective function is differentiable in $\boldsymbol{y}$, and a simple closed-form expression exists for projecting a point onto the constraint set $\mathbb{B}^{n-1}$.

Exercise 8 *Consider the projection operator* $\Pi_{\mathbb{B}^n} : \mathbb{R}^n \to \mathbb{R}^n$ *given by*

$$\forall\, \boldsymbol{x} \in \mathbb{R}^n : \quad \Pi_{\mathbb{B}^n}(\boldsymbol{x}) = \underset{\boldsymbol{z}\in\mathbb{B}^n}{\operatorname{argmin}}\ \|\boldsymbol{z} - \boldsymbol{x}\|^2.$$

Find a closed-form expression for $\Pi_{\mathbb{B}^n}(\boldsymbol{x})$.

We will solve (8) using projected gradient ascent. We start with an arbitrary $\boldsymbol{y}_0 \in \mathbb{R}^n$ and generate a sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ as as follows:

$$\forall\, k \geqslant 0 : \quad \boldsymbol{y}_{k+1} = \Pi_{\mathbb{B}^{n-1}}\Big(\boldsymbol{y}_k + L^{-1}\nabla\psi(\boldsymbol{y}_k, \boldsymbol{x})\Big). \tag{10}$$

As with proximal gradient descent, it can be shown that the sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ generated by (10) converges to a solution of (8). Subsequently, we can use this to get the solution of (2) via (9).

This completes the description of the iterative solvers for TV denoising (2) and TV regularization (4).

Assignment

1. Solve exercise (1).
2. ***Signal Generation***

   Generate and plot a one-dimensional signal of length $n = 256$ for the following categories:

   (a) **Piecewise constant**: The signal has 128 samples with a magnitude of 1, 64 samples with a magnitude of 2, and 64 samples with a magnitude of 3.

   (b) **Sawtooth**: The signal should be a sawtooth signal with four cycles and an amplitude of 1.

   (c) **Sinusoid**: The signal is a sinusoid with four cycles and an amplitude of 1.
3. ***Signal Denoising***

   (a) Consider the measurement model where the measurement matrix A is the identity matrix ($A = I$) in measurement model (3) and plot the corresponding noisy signals for noise level $\sigma \in \{0, 0.5, 1\}$.

   (b) Solve exercise (2) for each of the measured signals. Experiment with different values λ to find a good reconstruction. Report the value of λ and plot the corresponding reconstructed signal.
4. ***Signal Reconstruction***

   (a) Apply (3) as the measurement model on the generated signals. Consider different noise levels $\sigma \in \{0, 0.5, 1\}$ and plot the resulting noisy, compressed signals for the measurement ratios $m \in \{0.5 \times n, 0.75 \times n, n\}$.

   (b) Solve exercises (7), (8), and (9).

   (c) Solve exercise (4) for each of the measured signals. In each case, tune λ to find a good reconstruction. Report the optimal λ and plot the reconstructed signal.