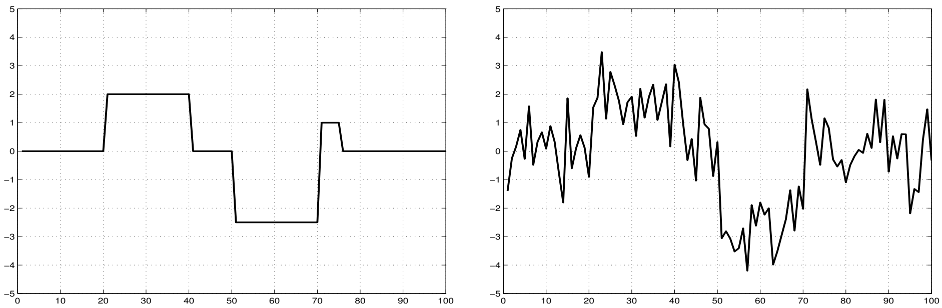
TP : Non-smooth optimization

1 Piecewise constant denoising

Let $\overline{x} = (\overline{x}^{(i)})_{1 \leq i \leq N} \in \mathbb{R}^N$ be a sampled piecewise constant noisy signal. We denote $y = \overline{x} + \varepsilon$ a noisy version of $\overline{x}$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. An illustration of $\overline{x}$ and y is displayed in Figure 1.



Original signal$\overline{x}$ Noisy signal y

FIGURE 1 – Illustration of a piecewise constant signal with $N = 100$ samples degraded with a white Gaussian noise of variance $\sigma^2 = 1$.

The objective of this first exrecice is to obtain a piecewise constante estimate $\widehat{x}$ the closest from the original $\overline{x}$ from data y. A solution consists in minimizing the following objective function :

$$\widehat{x}_\lambda = \arg \min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - y\|_2^2 + \lambda \|Lx\|_1$$

where $(Lx)^{(i)} = x^{(i+1)} - x^{(i)}$ for every $i \in \{1, \ldots, N-1\}$ and $\lambda > 0$ is the regularization parameter. $L \in \mathbb{R}^{N \times N}$ denotes the finite difference operator.

1) Discuss the impact of the parameter λ on the solution $\widehat{x}_\lambda$.

2) Prove that

$$\widehat{u}_\lambda \in \mathrm{Argmin}_{u \in \mathbb{R}^N} \frac{1}{2}\|y - L^* u\|_2^2 \quad \text{s.t.} \quad \|u\|_\infty \leq \lambda,$$

et que la relation entre les solutions primale et duale est

$$\widehat{x}_\lambda = y - L^* \widehat{u}_\lambda.$$

3) Solve the dual problem using forward-backward iterations (cf. course).

- Which is the Lipschitz differentiable function ? Specify its Lipchitz constant.

- Considering the relation between the proximal operator of a function $f \in \Gamma_0(\mathbb{R}^N)$ and its conjugate f^*, i.e.,

$$(\forall x \in \mathbb{R}^N) \quad \mathrm{prox}_{\gamma f^*} x = x - \gamma \mathrm{prox}_{\gamma^{-1} f}(\gamma^{-1} x),$$

  give the closed form expression of $P_{\|\cdot\|_\infty \leq \lambda}$.

- In MATLAB/Python, load the signal_exo1.mat

  $MATLAB$ >> load signal_ex1.mat ;

- Create a noisy version of this signal :

  $MATLAB$ >> y = x + randn(size(x)) ;

- Implement the forward-backward iterations and deduce $\widehat{x}$. Use the functions opL.m, opL_adj.m, opL.m allowing us to compute L, L^* and the proximity operator of the ℓ_1-norm, respectively.

- Plot the primal and the dual objective functions with respect to the iteration number. Change the algorithme parameters (regualrization parameter, step-size). Comment.

- Plot the duality gap with respect to the iteration number. Comment

- Plot the mean square error between the original siginal $\overline{x}$ and the estimated signal $\widehat{x}_\lambda$ with respect to λ. Comment.

2 Sparse logistic regression

« Prenons l'exemple d'une maladie à composante génétique dont on cherche à trouver les gènes impliqués. Il est devenu tellement facile et bon marché d'accumuler les données sur un sujet que l'on mesure tout ce qu'on peut. Au final, le signal qui nous intéresse, c'est-à-dire les variables qui ont quelque chose à voir avec la maladie, est noyé dans un océan de variables qui ne nous intéresse pas. On se retrouve avec des modèles comprenant des centaines de millions de variables dont on sait bien que seul une petite partie a un rapport avec le phénomène que l'on recherche. Du coup les techniques d'optimisation convexe (minimisation de la norme L_1) peuvent s'appliquer et on arrive à résoudre effectivement le système – par exemple à trouver les sites de génome impliqués dans la maladie – en supposant que la solution est parcimonieuse. »

– Emmanuel Candès, La Recherche, Fév. 2014 –

In this second exercice, we focus on the feature selection problem in classification. We consider a database with N patients. Among these N patients, a subset $S \subset \{1, \ldots, N\}$ belongs to the classe of « healthy patients » and the complementary set stands for the « sick patients » classe ($M \subset \{1, \ldots, N\}$). We denote $b = (b_i)_{1 \leq i \leq N}$ a vector allowing to describe a patient as healthy (*resp.* sick), i.e., for every $i \in S$, $b_i \equiv 1$ (*resp.*, for every $i \in M$, $b_i \equiv -1$). For each patient, we consider a set of information gather in $y_i = (y_{i,\ell})_{1 \leq \ell \leq K} \in \mathbb{R}^K$. The N samples provided for the training step are denoted $\mathcal{D} = \{(b_1, y_1), \ldots, (b_N, y_N)\}$.

We consider here, the estimation of the weight matrix $\widehat{x}_\lambda \in \mathbb{R}^K$ such that :

$$\widehat{x}_\lambda \in \underset{x \in \mathbb{R}^K}{\operatorname{Argmin}} \sum_{i=1}^{N} \log(1 + \exp(-b_i x^\top y_i) + \lambda \|x\|_1$$

where $\lambda > 0$.

1) Discuss the impact of the parameter λ on the solution $\widehat{x}_\lambda$.
2) With MATLAB/Python, download the vector b and the matrix y from the file data_exo2_training.mat
3) How many patients are listed in the training database. How many are healthy (resp. sick) ?
4) Implement forard-bacward to estimate $\widehat{x}_\lambda$. Plot the primal objective function with respect to the iteration number. Check the estimation validity $\widehat{x}_\lambda$ by computing

$$\widehat{b}_\lambda = \operatorname{sign}(\widehat{x}_\lambda^\top y)$$

for $\lambda = \{0.1, 1, 10\}$. Comment the obtained results.

5) Evaluate the estimator performance on the test database data_exo2_estimation.mat.