Lecture: Splitting methods for convex optimization
Master ACSYON, second year, 2022/2023
University of Limoges, France
Teacher: Loïc Bourdin

Practical works on Chapter 2

Tikhonov regularization problem and LASSO problem

Exercise 1. In this exercise, we consider the *Tikhonov regularization problem* given by

$$\underset{x\in\mathbb{R}^d}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|^2+\frac{\gamma}{2}\|x\|^2,$$

where $A\in\mathbb{R}^{n\times d}$ is a given matrix (usually with $d>>n$), $b\in\mathbb{R}^n$ is a given vector and $\gamma\geqslant 0$ is a given nonnegative real number.

1. Discuss the objective of the Tikhonov regularization problem, and give an interpretation of the parameter γ.
2. Provide four codes, respectively based on the gradient descent method, on the proximal point algorithm, on the proximal gradient algorithm and on the Douglas-Rachford algorithm, allowing to solve the above Tikhonov regularization problem.
3. Bonus: compare the convergence rates of the four codes.
4. Provide a code allowing to solve the *generalized Tikhonov regularization problem* given by
$$\underset{x\in\mathbb{R}^d}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|^2+\frac{1}{2}\|Bx\|^2,$$
where $B\in\mathbb{R}^{m\times d}$ is another given matrix.
5. Discuss the objective of the generalized Tikhonov regularization problem, and give an interpretation of the matrix B (with a relevant example).

Exercise 2. In this exercise, we consider the *LASSO problem* (Least Absolute Shrinkage and Selection Operator) given by

$$\underset{x\in\mathbb{R}^d}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|^2+\gamma\|x\|_1,$$

where $A\in\mathbb{R}^{n\times d}$ is a given matrix (usually with $d>>n$), $b\in\mathbb{R}^n$ is a given vector and $\gamma\geqslant 0$ is a given nonnegative real number.

1. Why is the last term $\gamma\|x\|_1$ usually called the *sparsity enforcement*? Make a drawing in the two-dimensional case for illustration.
2. Discuss the objective of the LASSO problem, and give an interpretation of the parameter γ.
3. Provide two codes, respectively based on the proximal gradient algorithm and on the Douglas-Rachford algorithm, allowing to solve the above LASSO problem. Note that the first code (based on the forward-backward algorithm) corresponds to the classical *ISTA* (Iterative Shrinkage Thresholding Algorithm).

4. Bonus: compare the convergence rates of the two codes.

5. Provide a code allowing to solve the *generalized LASSO problem* given by

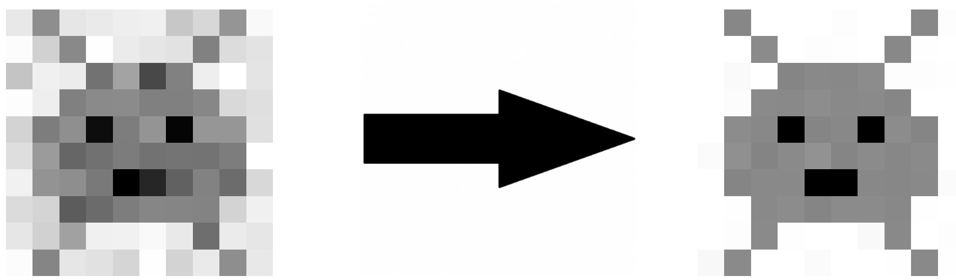
$$\underset{x\in\mathbb{R}^d}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|^2+\|Bx\|_1,$$

   where $B\in\mathbb{R}^{m\times d}$ is another given matrix.

6. Discuss the objective of the generalized LASSO problem, and give an interpretation of the matrix B (with a relevant example).

Exercise 3: Application to image deblurring

The aim of this exercise is to provide a Matlab code, based on the Davis-Yin algorithm, allowing to deblur a (very small!) grayscale image of 10*10 pixels.



The original image is on the right, while the blurred image is on the left.

1. Download the blurred image at

   https://www.unilim.fr/pages_perso/loic.bourdin/Documents/blurred.jpg

2. Let us denote by $b\in[0,255]^{100}$ the vector encoding the blurred image on Matlab (see Practical work 1 for the corresponding Matlab commands). In order to deblur the image, our method in this exercise consists in solving the convex optimization problem given by

$$\underset{x\in[0,255]^{100}}{\text{minimize}}\ \frac{1}{2}\|x-b\|^2+\gamma\|Fx\|_1,$$

   where $\gamma\geqslant 0$ is a nonnegative parameter and $F\in\mathbb{R}^{200\times 100}$ is the *total variation matrix* constructed such that

$$\forall x\in[0,255]^{100},\quad \|Fx\|_1=\sum_{i=1}^{100}\Big(|x_i-x_{\mathrm{ho}(i)}|+|x_i-x_{\mathrm{ve}(i)}|\Big),$$

   where:

   - if the index $i\in\{1,\dots,100\}$ corresponds to a pixel that does not belong to the last row of pixels, neither to the last column of pixels, then the index $\mathrm{ho}(i)\in\{1,\dots,100\}$ corresponds to the pixel located just to the right of the one of i, and $\mathrm{ve}(i)\in\{1,\dots,100\}$ to the pixel located just below the one of i;
   - if the index $i\in\{1,\dots,100\}$ corresponds to a pixel that belongs to the last row of pixels, then $\mathrm{ve}(i)=i$;
   - if the index $i\in\{1,\dots,100\}$ corresponds to a pixel that belongs to the last column of pixels, then $\mathrm{ho}(i)=i$.

   Provide a drawing illustrating the three above situations.

3. Explain why the above convex optimization problem is suitable in order to deblur the image. In particular, explain the role of the matrix F, as well as of the parameter γ.

4. Write on a paper the global structure of the matrix F, and write a Matlab code allowing to construct it.

5. By increasing the number of variables in order to deal with the term $\|Fx\|_1$, rewrite the above convex optimization problem as

$$\underset{(x,y)\in\mathbb{R}^{100}\times\mathbb{R}^{200}}{\text{minimize}}\ f(x,y)+g(x,y)+h(x,y),$$

   where:

   - the function $f \in \Gamma_0(\mathbb{R}^{100}\times\mathbb{R}^{200})$ can be expressed as

$$\forall (x,y)\in\mathbb{R}^{100}\times\mathbb{R}^{200},\quad f(x,y)=f_1(x)+f_2(y),$$

     where $f_1 \in \Gamma_0(\mathbb{R}^{100})$ is an indicator function and $f_2 \in \Gamma_0(\mathbb{R}^{200})$ is a real-valued nondifferentiable function;
   - the function $g \in \Gamma_0(\mathbb{R}^{100}\times\mathbb{R}^{200})$ is an indicator function;
   - and the function $h \in \Gamma_0(\mathbb{R}^{100}\times\mathbb{R}^{200})$ is a real-valued differentiable function.

6. Provide an expression of prox_f, as well as of prox_g and of ∇h.

7. Recall the Davis-Yin algorithm (and its assumptions), depending on a positive parameter $\lambda > 0$, and then provide a Matlab code implementing it in the present context.

8. After convergence, show the image obtained. You can use the following Matlab commands, right after the imshow one, in order to get a (relatively) large image:

```
set(gcf,'MenuBar','none')
set(gca,'DataAspectRatioMode','auto')
set(gca,'Position',[0 0 1 1])
```

   If the image obtained is not satisfactory, then try different values for the parameters λ and γ.

In this exercise, we have only dealt with a (very small!) image of 10*10 pixels in order to avoid long computations and techniques of parallelization. Of course, numerous techniques have been developed in the image processing literature in order to deal with larger images (even in colour) and even with videos.

original   noisy/blurred   restored

Numerous references exist in the literature, among which:

- P.L. Combettes and J.-C. Pesquet. *A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery.* IEEE Journal of Selected Topics in Signal Procesing (2007).

- D. O'Connor and L. Vandenberghe. *Primal–dual decomposition by operator splitting and applications to image deblurring.* SIAM J. Imaging Sciences (2014).

- Simon Setzer. *Splitting methods in image processing.* PhD thesis, University of Mannheim (2009).

- etc., etc.