Principal component pursuit

In this project, we will apply several classic operator splitting schemes to solve the Principal Component Pursuit (PCP, a.k.a. Robust Principal Component Analysis) problem, proposed by Candès, Li, Ma and Wright, 2009.

1   Forward model

Suppose we have the following observation

$$f = x + y + \epsilon,$$

where $x \in \mathbb{R}^{m \times n}$ is a sparse matrix, $y \in \mathbb{R}^{m \times n}$ is a low-rank matrix and $\epsilon \in \mathbb{R}^{m \times n}$ is white Gaussian noise. The goal is to recover the sparse and low-rank components from the observation f. There are various formulations of this decomposition, in the following we mainly focus on several convex approaches.

2   Noiseless case

When there is no noise, *i.e.* $\epsilon = 0$, the decomposition problem can be achieved via the following constrained convex optimisation problem

$$\min_{x,y \in \mathbb{R}^{m \times n}} \mu \|x\|_1 + \|y\|_* \tag{2.1}$$
$$\text{s.t.} \quad x + y = f.$$

For the above problem, there are two methods can be applied: Douglas–Rachford splitting and Backward–Backward splitting with parameter continuation. Throughout the section, consider

$$\mu = \frac{1}{\sqrt{\max\{m, n\}}}.$$

2.1   Douglas–Rachford splitting

Let's first define several notations

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{2m \times n}, \quad R(z) = \mu \|x\|_1 + \|y\|_*, \quad A = \begin{bmatrix} \text{Id} & \text{Id} \end{bmatrix}. \tag{2.2}$$

Then (2.1) can be written as

$$\min_{z \in \mathbb{R}^{2m \times n}} R(z) \tag{2.3}$$
$$\text{s.t.} \quad Az = f.$$

Define the set $\Omega \stackrel{\text{def}}{=} \{z \in \mathbb{R}^{2m \times n} : Az = f\}$, then we further get

$$\min_{z \in \mathbb{R}^{2m \times n}} R(z) + \iota_\Omega(z), \tag{2.4}$$

which can be handled by Douglas–Rachford splitting easily. Let $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$

- The proximal mapping of R

$$\text{prox}_{\gamma R}(z) = \begin{pmatrix} \text{prox}_{\gamma \mu \|\cdot\|_1}(z_1) \\ \text{prox}_{\gamma \mu \|\cdot\|_*}(z_2) \end{pmatrix}.$$

- The projection mapping onto Ω

$$\text{P}_\Omega(z) = z + A^T (AA^T)^{-1}(f - Az).$$

Therefore, the iteration of Douglas–Rachford the reads: let $\gamma > 0$, $z_{1,0} = 0$, $z_{2,0} = f$ and $v = \text{P}_\Omega(z_0)$

$$u_{1,k+1} = \mathcal{T}_{\gamma\mu}(2v_{1,k} - z_{1,k})$$
$$u_{2,k+1} = U\mathcal{T}_\gamma(\Sigma)V^T$$
$$z_{1,k+1} = z_{1,k} + u_{1,k+1} - v_{1,k}$$
$$z_{2,k+1} = z_{2,k} + u_{2,k+1} - v_{2,k}$$
$$v_{k+1} = \text{P}_\Omega(z_{k+1}),$$

where $\mathcal{T}_\gamma(\cdot)$ is the soft-thresholding operation and $U\Sigma V^T$ is the SVD of $2v_{2,k} - z_{2,k}$.

Numerical goal The goal of this realisation is:
- Implement the Douglas–Rachford splitting method;
- Implement the multi-step inertial acceleration for Douglas–Rachford
  - 1-step inertial scheme: choose inertial parameter as 0.1;
  - 2-step inertial scheme: choose inertial parameters as $(0.2, -0.1)$.

For both implementations, observe the convergence of $\|z_k - z_{k-1}\|$.

3 Noisy case

When $\epsilon \neq 0$, then instead of the constrained optimisation problem, we need to consider the following regularised least square

$$\min_{x,y \in \mathbb{R}^{m \times n}} \nu \left(\frac{1}{\sqrt{\max\{m,n\}}} \|x\|_1 + \|y\|_* \right) + \frac{1}{2} \|x + y - f\|^2, \tag{3.1}$$

where $\nu > 0$ is tradeoff parameter. Follow the definitions in (2.2), we obtain the following simpler formulation

$$\min_{z \in \mathbb{R}^{2m \times n}} \nu R(z) + \frac{1}{2} \|Az - f\|^2, \tag{3.2}$$

which can be easily handled by Forward–Backward splitting and FISTA.

Now again, by applying the Moreau envelope trick, (3.1) is equivalent to

$$\min_{x \in \mathbb{R}^{m \times n}} \frac{\nu}{\sqrt{\max\{m,n\}}} \|x\|_1 + {}^1\left(\nu \| \cdot \|_1\right)(f - x), \tag{3.3}$$

where ${}^1\left(\nu \| \cdot \|_1\right)(f - x)$ is the Moreau envelope parameterised by 1, hence its gradient is 1-Lipschitz.

Numerical goal Solve either (3.2) or (3.3) with Forward–Backward splitting, FISTA and restarting FISTA to compare their performances in terms of $\|z_k - z_{k-1}\|$ (for (3.2)) or $\|x_k - x_{k-1}\|$ (for (3.3)).