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

Practical works on Chapter 3

In this practical work, our aim is to give an introduction to *low-rank nonnegative matrix completion problems*. This (still active) mathematical research field has many concrete applications, in particular for recommendation systems (used by Youtube and Netflix for example) as explained in the next paragraph.

Suppose that some users in an online survey provide nonnegative ratings of some movies. This yields a nonnegative matrix B (denoted by $B \geqslant 0$), with m users as rows and n movies as columns, whose (i,j)th entry B_{ij} is the rating given by the ith user to the jth movie. Since most users rate only a small portion of the movies, we typically only know a small subset $\{B_{ij} \mid (i,j) \in \Omega\}$ of the entries. Based on the known ratings of a user, we want to predict the user's ratings of the movies that the user did not rate, that is, we want to fill the missing entries of the matrix. It is commonly believed that only a few factors contribute to an individual's tastes or preferences for movies. Thus, the rating matrix B is likely to be of low rank. Therefore, finding the matrix B corresponds to solving the low-rank nonnegative matrix completion problem given by

$$\begin{aligned} \underset{X \in \mathbb{R}^{m \times n}}{\text{minimize}} \quad & \operatorname{rank}(X), \\ \text{subject to} \quad & X \geqslant 0, \\ & X_{ij} = B_{ij}, \quad \forall (i,j) \in \Omega. \end{aligned} \tag{P$_1$}$$

Due to the combinatorial nature of the rank operator, it has been proved in the literature (see references below) that rank minimization problems are NP-hard. Therefore, in the sequel, we will proceed to approximations and simplifications of Problem (P$_1$).

Before coming to this point, instead of working directly in the (quite difficult) general case of matrices, we will first work in Section 1 in the simpler case of vectors. We will go back to the general case of matrices only in the next Section 2.

Note that the present practical work has been inspired from several references (and references therein!) such as:

- M. Fazel. *Matrix rank minimization with applications.* 2002.
- B. Recht, M. Fazel, P. Parrilo. *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.* 2007.
- S. Ma, D. Goldfarb, L. Chen. *Fixed point and Bregman iterative methods for matrix rank minimization.* 2011.
- M. Javad Abdi. *Cardinality optimization problems.* 2013.
- F. Xu and G. He. *New algorithms for nonnegative matrix completion.* 2015.

1 Compressed sensing and cardinality minimization problems

Recall that, in this lecture, the proximal operator $\mathrm{prox}_f : \mathbb{R}^d \to \mathbb{R}^d$ of a function $f \in \Gamma_0(\mathbb{R}^d)$ has been defined as

$$\forall x \in \mathbb{R}^d, \quad \mathrm{prox}_f(x) := \mathrm{Argmin}\left(f + \frac{1}{2}\|\cdot - x\|_2^2\right),$$

with respect to the standard (auto-dual) Euclidean norm denoted by $\|\cdot\|_2$, deriving from the usual scalar product $\langle\cdot,\cdot\rangle$. Recall that other norms exist on the vector space $\mathbb{R}^d$, such as the 1-norm $\|\cdot\|_1$ which is the dual norm to the ∞-norm $\|\cdot\|_\infty$ (and vice-versa).

Compressed sensing is a signal processing technique for finding solutions to underdetermined linear systems with a minimal number of nonzero components (we speak of *sparse solutions*). To describe such a problem, we thus introduce the *cardinality function* $\|\cdot\|_0 : \mathbb{R}^d \to \mathbb{R}$ defined by

$$\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d, \quad \|x\|_0 := \mathrm{card}\Big(\{i \in \{1, \dots, d\} \mid x_i \neq 0\}\Big).$$

Therefore, a standard compressed sensing problem (also called *cardinality minimization problem*) is given by

$$\begin{aligned} \underset{x \in \mathbb{R}^d}{\text{minimize}} \quad & \|x\|_0, \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{Q$_1$}$$

where $A \in \mathbb{R}^{p \times d}$ is a given matrix (usually with $p << d$) and $b \in \mathbb{R}^p$ is a given vector.

1. **Problem (Q$_1$) is not a convex minimization problem.** Provide a counterexample showing that the cardinality function $\|\cdot\|_0$ is not convex. In particular, we deduce that $\|\cdot\|_0$ is not a norm (despite what the notation might lead us to believe). Therefore, since Problem (Q$_1$) is not a convex minimization problem, in the sequel, we will proceed to approximations and simplifications.

2. **The convex envelop of the cardinality function.** In the literature, it is commonly said, in a rough sense, that $\|\cdot\|_1$ is the convex envelop of $\|\cdot\|_0$. This is not exactly true. In this item, we will prove that $\|\cdot\|_1$ is the affine hull of $\|\cdot\|_0$ over the unit ball $\overline{\mathrm{B}}_\infty := \{x \in \mathbb{R}^d \mid \|x\|_\infty \leqslant 1\}$. Precisely, we will prove that

$$\mathrm{aff}\Big(\|\cdot\|_0 + \iota_{\overline{\mathrm{B}}_\infty}\Big) = \|\cdot\|_1 + \iota_{\overline{\mathrm{B}}_\infty}.$$

   To this aim, we introduce the function $g := \|\cdot\|_0 + \iota_{\overline{\mathrm{B}}_\infty}$ and our aim is to compute its conjugate g^* and its biconjugate $(g^*)^*$.

   (a) Prove that

$$\forall y = (y_1, \dots, y_d) \in \mathbb{R}^d, \quad \forall x = (x_1, \dots, x_d) \in \overline{\mathrm{B}}_\infty, \quad \langle y, x\rangle - \|x\|_0 \leqslant \sum_{x_i \neq 0}(|y_i| - 1) \leqslant \sum_{i=1}^{d}(|y_i| - 1)^+.$$

   (b) Deduce that

$$\forall y = (y_1, \dots, y_d) \in \mathbb{R}^d, \quad g^*(y) = \sum_{i=1}^{d}(|y_i| - 1)^+.$$

   (c) Prove that

$$\begin{aligned} &\forall x = (x_1, \dots, x_d) \in \overline{\mathrm{B}}_\infty, \quad \forall y = (y_1, \dots, y_d) \in \mathbb{R}^d, \\ &\qquad \langle x, y\rangle - g^*(y) \leqslant \sum_{|y_i| < 1} |x_i| + \sum_{|y_i| \geqslant 1} \Big(|y_i|(|x_i| - 1) + 1\Big) \leqslant \|x\|_1. \end{aligned}$$

   (d) Deduce that

$$\forall x \in \overline{\mathrm{B}}_\infty, \quad (g^*)^*(x) = \|x\|_1.$$

(e) Prove that
$$\forall x \in \mathbb{R}^d \backslash \overline{\mathrm{B}}_\infty, \quad (g^*)^*(x) = +\infty.$$

(f) Deduce that $(g^*)^* = \|\cdot\|_1 + \iota_{\overline{\mathrm{B}}_\infty}$.

(g) Deduce from the biconjugate theorem that $\|\cdot\|_1$ is the affine hull of $\|\cdot\|_0$ over $\overline{\mathrm{B}}_\infty$.

(h) Deduce that $\frac{1}{M}\|\cdot\|_1$ is the affine hull of $\|\cdot\|_0$ over the ball $\overline{\mathrm{B}}_\infty^M := M\overline{\mathrm{B}}_\infty$ for any $M > 0$.

3. **Several approximations and simplifications of Problem (Q$_1$).** Considering only the solutions to Problem (Q$_1$) that are bounded by some large $M > 0$ (with respect to the norm $\|\cdot\|_\infty$) and thanks to the previous item, we conclude that a convex approximation of Problem (Q$_1$) is given by

$$\begin{aligned} &\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \|x\|_1, \\ &\text{subject to} \quad Ax = b, \end{aligned} \tag{Q$_2$}$$

where we have removed the useless positive multiplicative factor $\frac{1}{M}$. Problem (Q$_2$) is usually called *basis pursuit problem* in the literature.

In practice, the data A and b are noisy (that is, subject to small measurement perturbations). Therefore, instead of solving Problem (Q$_2$), we usually consider the relaxed problem given by

$$\begin{aligned} &\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \|x\|_1, \\ &\text{subject to} \quad \|Ax - b\|_2 \leqslant \theta, \end{aligned} \tag{Q$_3$}$$

where $\theta > 0$ is a small positive parameter.

Finally, we introduce the relaxed version given by

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2, \tag{Q$_4$}$$

where $\mu > 0$ is a positive parameter.

By choosing appropriately the parameters θ and/or μ, or possibly by adding some hypotheses, provide a discussion on the fact that any solution to Problem (Q$_4$) is a solution to Problem (Q$_3$) (and possibly vice-versa).

4. **Setup of the ADMM algorithm.** To apply the ADMM algorithm, we finally rewrite Problem (Q$_4$) as the equivalent problem

$$\begin{aligned} &\underset{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d}{\text{minimize}} \quad \mu\|y\|_1 + \frac{1}{2}\|Ax - b\|_2^2, \\ &\text{subject to} \quad x - y = 0. \end{aligned} \tag{Q$_5$}$$

(a) For some $\lambda > 0$, give the expression of the augmented Lagrangian $\mathrm{L}^\lambda : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ associated with Problem (Q$_5$) and recall the associated ADMM algorithm (decomposed in three steps).

(b) Prove that the (unique) solution x_{k+1} to the first step of the above ADMM algorithm is given by
$$x_{k+1} = (\lambda \mathrm{Id} + A^\top A)^{-1}(A^\top b + \lambda y_k - z_k).$$

(c) Prove that the (unique) solution y_{k+1} to the second step of the above ADMM algorithm is given by
$$y_{k+1} = \mathrm{prox}_{\frac{\mu}{\lambda}\|\cdot\|_1}\left(x_{k+1} + \frac{1}{\lambda}z_k\right).$$

5. **Matlab code.** Write a Matlab code allowing to solve Problem (Q$_5$) (and thus Problem (Q$_4$)) thanks to the above ADMM algorithm and proceed to some numerical tests (keeping in mind the initial Problem (Q$_1$)).

2 *Going back to the low-rank nonnegative matrix completion problem*

An usual scalar product on the matrix space $\mathbb{R}^{m\times n}$ is given by

$$\forall X,Y\in\mathbb{R}^{m\times n},\quad \langle X,Y\rangle=\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}Y_{ij},$$

and leads to the so-called *Frobenius norm* defined by $\|X\|_F:=\sqrt{\langle X,X\rangle}$ for all $X\in\mathbb{R}^{m\times n}$. Similarly to the vector space $\mathbb{R}^d$ (see Section 1), these basic tools allow to introduce, for any function $h\in\Gamma_0(\mathbb{R}^{m\times n})$, the corresponding proximal operator $\mathrm{prox}_h:\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ defined by

$$\forall X\in\mathbb{R}^{m\times n},\quad \mathrm{prox}_h(X):=\mathrm{Argmin}\left(h+\frac{1}{2}\|\cdot-X\|_F^2\right).$$

Recall that a lot of different norms are possible on the matrix space $\mathbb{R}^{m\times n}$. For example, there are the *matrix norms induced by vector norms* (such as the usual ones denoted by $|||\cdot|||_1$, $|||\cdot|||_2$ and $|||\cdot|||_\infty$). Nevertheless, there are also norms which are not induced by vector norms, such as the above (auto-dual) Frobenius norm but not only. Indeed, in what follows, we will introduce the so-called *nuclear norm* and *operator norm* based on the process of Singular Value Decomposition (SVD).

1. **SVD theorem.** Recall the proof of SVD theorem asserting that any matrix $X\in\mathbb{R}^{m\times n}$ can be written as

$$X=U^X\Sigma^X(V^X)^\top,$$

where $U^X\in\mathbb{R}^{m\times m}$ and $V^X\in\mathbb{R}^{n\times n}$ are unitary matrices, and where $\Sigma^X=\mathrm{diag}(\sigma^X)\in\mathbb{R}^{m\times n}$ where $\sigma^X\in\mathbb{R}^{\max\{m,n\}}$ is a vector with nonincreasing nonnegative components (that is, $\sigma_1^X\geqslant\sigma_2^X\geqslant\ldots\geqslant\sigma_{\max\{m,n\}}^X\geqslant 0$) which is uniquely determined by X (while U^X and V^X are not). Then, provide a simple example of SVD.

2. **Link between the Frobenius norm and the Euclidean norm via SVD.** Prove that

$$\forall X\in\mathbb{R}^{m\times n},\quad \|X\|_F=\sqrt{\mathrm{Trace}(X^\top X)}=\|\sigma^X\|_2.$$

3. **Link between the rank operator and the cardinality function via SVD.** Prove that

$$\forall X\in\mathbb{R}^{m\times n},\quad \mathrm{rank}(X)=\mathrm{rank}(X^\top X)=\|\sigma^X\|_0.$$

4. **A list of admitted results (no answer required for this item).** In the references provided at the beginning of this document (and in the references therein), it has been proved that the functions defined via SVD by

$$\forall X\in\mathbb{R}^{m\times n},\quad \|X\|_N:=\|\sigma^X\|_1\qquad\text{and}\qquad \|X\|_O:=\|\sigma^X\|_\infty,$$

both constitute norms on the matrix space $\mathbb{R}^{m\times n}$, which are respectively called *nuclear norm* and *operator norm* in the literature. It can be proved that they are dual to each other. Furthermore, similarly to Section 1, it can be proved that $\|\cdot\|_N$ is the affine hull of $\mathrm{rank}(\cdot)$ over the unit ball $\overline{\mathrm{B}}_O:=\{X\in\mathbb{R}^{m\times n}\mid\|X\|_O\leqslant 1\}$.

Finally, it can be proved that

$$\forall\gamma>0,\quad\forall X\in\mathbb{R}^{m\times n},\quad \mathrm{prox}_{\gamma\|\cdot\|_N}(X)=U^X\mathrm{diag}\Big(\mathrm{prox}_{\gamma\|\cdot\|_1}(\sigma^X)\Big)(V^X)^\top,$$

where $U^X\mathrm{diag}(\sigma^X)(V^X)^\top$ stands for a SVD of X.

5. **Convex approximation of Problem (P$_1$).** We first rewrite Problem (P$_1$) as

$$\begin{aligned}&\underset{X\in\mathbb{R}^{m\times n}}{\text{minimize}} && \mathrm{rank}(X),\\ &\text{subject to} && X\geqslant 0,\\ & && P_\Omega(X)=P_\Omega(B),\quad\forall(i,j)\in\Omega,\end{aligned}\tag{P$_2$}$$

where $P_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the linear projection operator defined by

$$\forall X \in \mathbb{R}^{m\times n}, \quad P_\Omega(X)_{ij} := \begin{cases} X_{ij} & \text{if} \quad (i,j) \in \Omega, \\ 0 & \text{if} \quad (i,j) \in \Omega^{\mathrm{c}}, \end{cases}$$

where Ω^{c} stands for the complementary of Ω.

Define similarly the linear projection operator $P_{\Omega^{\mathrm{c}}} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and note that $X = P_\Omega(X) + P_{\Omega^{\mathrm{c}}}(X)$ with $\langle P_\Omega(X), P_{\Omega^{\mathrm{c}}}(X)\rangle_F = 0$ for all $X \in \mathbb{R}^{m\times n}$.

Following exactly the same steps as in Section 1, we will consider in the sequel the convex approximation of Problem (P$_2$) given by

$$\begin{aligned} &\underset{X\in\mathbb{R}^{m\times n}}{\text{minimize}} \quad \mu\|X\|_N + \frac{1}{2}\|P_\Omega(X) - P_\Omega(B)\|_F^2, \\ &\text{subject to} \quad X \geqslant 0, \end{aligned} \tag{P$_3$}$$

where $\mu > 0$ is a positive parameter. Note that the only difference with Section 1 is the presence of the nonlinear constraint $X \geqslant 0$.

6. **Setup of the ADMM algorithm.** To apply the ADMM algorithm, we rewrite Problem (P$_3$) as the equivalent problem

$$\begin{aligned} &\underset{(X,Y)\in\mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}}{\text{minimize}} \quad \mu\|Y\|_N + \frac{1}{2}\|P_\Omega(X) - P_\Omega(B)\|_F^2, \\ &\text{subject to} \quad X \geqslant 0, \\ &\qquad\qquad\quad\; X - Y = 0. \end{aligned} \tag{P$_4$}$$

   (a) For some $\lambda > 0$, give the expression of the augmented Lagrangian $\mathcal{L}^\lambda : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}$ associated with Problem (P$_4$) and recall the associated ADMM algorithm (decomposed in three steps). *Recall that only the linear constraint $X - Y = 0$ should appear in the expression of $\mathcal{L}^\lambda$, but not the nonlinear constraint $X \geqslant 0$.*

   (b) Decomposing $X_{k+1} = P_\Omega(X_{k+1}) + P_{\Omega^{\mathrm{c}}}(X_{k+1})$, prove that the (unique) solution X_{k+1} to the first step of the above ADMM algorithm is given by

$$X_{k+1} = P_+\left(\frac{1}{1+\lambda} P_\Omega(B + \lambda Y_k - Z_k)\right) + P_+\left(P_{\Omega^{\mathrm{c}}}\left(Y_k - \frac{1}{\lambda}Z_k\right)\right),$$

   where $P_+ : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the (nonlinear) projection operator onto nonnegative matrices defined by

$$\forall X \in \mathbb{R}^{m\times n}, \quad P_+(X)_{ij} := \begin{cases} X_{ij} & \text{if} \quad X_{ij} \geqslant 0, \\ 0 & \text{if} \quad X_{ij} < 0. \end{cases}$$

   (c) Prove that the (unique) solution Y_{k+1} to the second step of the above ADMM algorithm is given by

$$Y_{k+1} = \operatorname{prox}_{\frac{\mu}{\lambda}\|\cdot\|_N}\left(X_{k+1} + \frac{1}{\lambda}Z_k\right) = U^k \operatorname{diag}\left(\operatorname{prox}_{\frac{\mu}{\lambda}\|\cdot\|_1}(\sigma^k)\right)(V^k)^\top,$$

   where $U^k \operatorname{diag}(\sigma^k)(V^k)^\top$ stands for a SVD of $(X_{k+1} + \frac{1}{\lambda}Z_k)$.

7. **Matlab code.** Using the Matlab command [U,S,V]=svd(X), write a Matlab code allowing to solve Problem (P$_4$) (and thus Problem (P$_3$)) thanks to the above ADMM algorithm and proceed to some numerical tests (keeping in mind the initial Problem (P$_1$) and the applications to recommendation systems for Youtube and Netflix).