A Proximal Approach for Solving Matrix Optimization Problems Involving a Bregman Divergence

Alessandro Benfenati, Emilie Chouzenoux, Jean-Christophe Pesquet

A Proximal Approach for Solving Matrix Optimization Problems Involving a Bregman Divergence

Alessandro Benfenati*, Emilie Chouzenoux*†, and Jean–Christophe Pesquet†,
* LIGM, University Paris–Est Marne–la–Vallée
† Center for Visual Computing, CentraleSupelec, University Paris-Saclay

***Abstract*—In recent years, there has been a growing interest in problems such as shape classification, gene expression inference, inverse covariance estimation. Problems of this kind have a common underlining mathematical model, which involves the minimization in a matrix space of a Bregman divergence function coupled with a linear term and a regularization term. We present an application of the Douglas-Rachford algorithm which allows to easily solve the optimization problem.**

In recent years, some applications such as shape classification models [1], gene expression [2], or inverse covariance estimation [3] have led to matrix variational formulations of the form:

$$\underset{C\in\mathcal{S}_+}{\text{minimize}}\ \ D_f(C,S)+g(C) \tag{1}$$

where $\mathcal{S}_+$ is the cone of symmetric semidefinite positive matrices of size $n\times n$, S is a given matrix in $\mathcal{S}_+$, f and g are proper lower-semicontinuous (lsc) convex functions defined on the space of $n\times n$ matrices, and D_f is the Bregman divergence associated with f. Recall that

$$D_f(C,S)=f(C)-f(S)-\mathrm{tr}\,(T(C-S)) \tag{2}$$

where $T\in\partial f(S)\neq\varnothing$. Note also that solving (1) amounts to computing the proximity operator of $g+\iota_{\mathcal{S}_+}$ at S,[1] with respect to the divergence D_f, which has also been found to be useful in a number of recent works [4], [5].

Very often, due to the nature of the problems, the regularization functional g has to promote the sparsity of C. A generic class of regularization is obtained by assuming that $g=g_0+g_1$ where

$$g_0(C)=\begin{cases}\psi(d) & \text{if } C\in\mathcal{S}_+\\ +\infty & \text{otherwise,}\end{cases} \tag{3}$$

where $\psi\colon\mathbb{R}^n\to]-\infty,+\infty]$ is a proper lsc function and d is the vector of eigenvalues of C, whereas g_1 is a function which cannot be expressed under this form. Typical examples are the nuclear norm $\|\cdot\|_*$ (or any Schatten norm) for g_0 and the ℓ_1 norm $\|\cdot\|_1$ (of the matrix elements) for g_1 [6].

In this paper, we will assume that function f can be expressed similarly to g_0 as $f(C)=\varphi(d)$ if $C\in\mathcal{S}_+$, $f(C)+\infty$ otherwise, where $\varphi\colon\mathbb{R}^n\to]-\infty,+\infty]$ is a proper lsc convex function. In particular, this assumption is satisfied when

$$f(C)=\begin{cases}-\log\det(C) & \text{if } C\succ 0\\ +\infty & \text{otherwise.}\end{cases} \tag{4}$$

Various algorithm have been proposed to solve Problem (1) when f is the above function and some specific choices of the function g are made: the popular GLASSO algorithm [3], a Gradient Projection method [1], and a splitting technique on the regularization term [6]. Here we propose to employ the Douglas–Rachford algorithm [7], which enables us to solve (1) in a fast manner, as soon as an efficient procedure for the eigenvalue decomposition is provided. The Douglas–Rachford approach alternates proximity steps on $D_f(\cdot,S)+g_0+\iota_{\mathcal{S}_+}$ and on g_1. For many functions g_1 of practical interest, the proximity operator of g_1 (e.g, $g_1=\|\cdot\|_1$) has a closed form solution [7] . Let us define $F(C)=f(C)+g_0(C)$. Let $\gamma\in]0,+\infty[$. It can be noted that computing the proximity operator of $\gamma\big(D_f(\cdot,S)+g_0+\iota_{\mathcal{S}_+}\big)$ w.r.t. the Frobenius metric $\|\cdot\|_{\mathrm{F}}$, at some symmetric matrix $\overline{C}$, is equivalent to find

$$\widehat{C}=\underset{C\in\mathbb{R}^{n\times n}}{\text{argmin}}\ \left(F(C)-\mathrm{tr}\,(TC)+\frac{1}{2\gamma}\|C-\overline{C}\|_{\mathrm{F}}^2\right).$$

Classical properties of the proximity operator [7] state that

$$\widehat{C}=\mathrm{prox}_{\gamma F-\gamma\mathrm{tr}(T\cdot)}(\overline{C})=\mathrm{prox}_{\gamma F}(\overline{C}+\gamma T).$$

Moreover, if $\overline{C}+\gamma T=U\mathrm{Diag}(\sigma)U^\top$ where U is an orthogonal matrix and $\sigma\in\mathbb{R}^n$, then $\widehat{C}=UDU^\top$ with $D=\mathrm{Diag}(\mathrm{prox}_{\gamma(\varphi+\psi)}(\sigma))$. For example, if f is the log-det function (4) and $g_0=\mu\|\cdot\|_*$ where $\mu\in[0,+\infty[$, according to [8], the diagonal matrix of eigenvalues of $\widehat{C}$ is given by

$$D=\frac{1}{2}\left(\Sigma-\gamma\mu I_n+\sqrt{(\Sigma-\gamma\mu I_n)^2+4\gamma I_n}\right)$$

where $\Sigma=\mathrm{Diag}(\sigma)$. The operations to compute $\mathrm{prox}_{\gamma(\varphi+\psi)}$ are thus component–wise.

The proposed Douglas–Rachford approach is easy to implement: if an efficient procedure for the eigenvalue decomposition is available, according to our numerical experiments, it is also very fast.

Acknowledgements: This work was supported by the Agence Nationale de la Recherche under grant ANR-14-CE27-0001 GRAPHSIP.

[1] ι_E designates the indicator function of a set E.

References

[1] J. Duchi, S. Gould, and D. .Koller, "Projected subgradient methods for learning sparse gaussians," in *Proceedings of the Twenty-fourth Conference on Uncertainty in AI (UAI)*, 2008.
[2] S. Ma, L. Xue, and H. .Zou, "Alternating direction methods for latent variable gaussian graphical model selection." *Neural Computation*, vol. 25, no. 8, pp. 2172–2198, 2013.
[3] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, jul 2008.
[4] H. H. Bauschke, P. L. Combettes, and D. Noll, "Joint minimization with alternating bregman proximity operators," *Pacific Journal of Optimization*, vol. 2, no. 3, pp. 401–424, 2006.
[5] A. Benfenati and V. Ruggiero, "Inexact Bregman iteration with an application to Poisson data reconstruction," *Inverse Problems*, vol. 29, no. 6, pp. 1–32, 2013.
[6] V. Chandrasekaran, P. Parrilo, and A. S.Willsky, "Latent variable graphical model selection via convex optimization," *Ann. Statist.*, vol. 40, no. 4, pp. 1935–1967, 08 2012.
[7] P. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, B. B. C. E. L. W. (Eds.), Ed. Springer, 2011, pp. 185–212.
[8] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, p. 1495, 2007.