MADMM: A GENERIC NON-SMOOTH OPTIMIZATION ON MANIFOLDS

{Artiom.Kovnatsky, Klaus.Glashoff, Michael.Bronstein}@usi.ch
Faculty of Informatics, Università della Svizzera italiana (USI), Lugano, Switzerland

Università della Svizzera italiana
Faculty of Informatics

Problem & Contributions

Problem

$$\min_{X \in \mathcal{M}} f(X) + g(AX),$$

where f and g are *smooth* (e.g., $\mathrm{tr}(\cdot)$, $\|\cdot\|_2^2$) and *non-smooth* (e.g., $\|\cdot\|_1$, $\|\cdot\|_{2,1}$, $\|\cdot\|_*$) real-valued functions, respectively, A is a $k \times m$ matrix, and $\mathcal{M}$ is a Riemannian matrix-valued manifold.

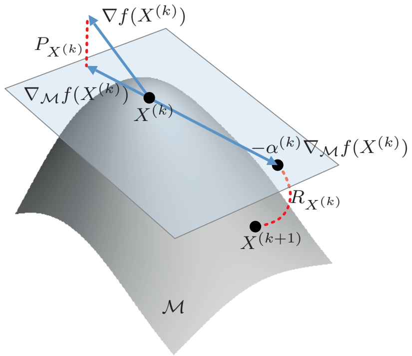
Contributions

1. **MADMM**: the first generic algorithm for non-smooth optimization on manifolds
2. Remarkably simple implementation
3. Applicable to any manifold (read constraint)
4. Straightforwardly modified for a task at hand
5. **MADMM** converges faster than previous methods in a broad range of applications

Mathematical background

Manifold Optimization

repeat
  Compute the extrinsic gradient $\nabla f(X^{(k)})$
  Projection: $\nabla_{\mathcal{M}} f(X^{(k)}) = P_{X^{(k)}}(\nabla f(X^{(k)}))$
  Compute the step size $\alpha^{(k)}$ along the descent direction
  Retraction:
  $X^{(k+1)} = R_{X^{(k)}}(-\alpha^{(k)} \nabla_{\mathcal{M}} f(X^{(k)}))$
until *convergence*;



$f : \mathbb{R}^{n \times m} \to \mathbb{R}$ is smooth
$\mathcal{M} \subseteq \mathbb{R}^{n \times m}$ is a Riemannian submanifold

ADMM

Let us be given an optimization: $\min_{x,z} f(x) + g(z)$, s.t. $Ax + Bz = c$, with variables $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, where $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{p \times m}$, and $c \in \mathbb{R}^p$.
The *augmented Lagrangian*: $L_\rho(x, z, y) = f(x) + g(z) + y^\top(Ax + By - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2$.
Applying *method of multipliers* (scaled form is obtained by introducing *scaled dual variable* $u = (1/\rho)y$):
(1) $(x^{k+1}, z^{k+1}) := \mathrm{argmin}_{x,z} L_\rho(x, z, y)$ (2) $y^{k+1} := y^k + \rho(Ax^{k+1} + Bz^{k+1} - c)$

Method Summary

$$\min_{X \in \mathcal{M}} f(X) + g(AX) \iff \min_{X \in \mathcal{M}, Z} f(X) + g(Z), \text{ s.t. } Z = AX \iff \min_{X \in \mathcal{M}, Z} f(X) + g(Z) + \tfrac{\rho}{2}\|AX - Z + U\|_F^2$$

Initialize $k \leftarrow 1$, $Z^{(1)} = AX^{(1)}$, $U^{(1)} = 0$.
repeat
  X-step: $X^{(k+1)} = \mathrm{argmin}_{X \in \mathcal{M}} f(X) + \frac{\rho}{2}\|AX - Z^{(k)} + U^{(k)}\|_F^2$
  Z-step: $Z^{(k+1)} = \mathrm{argmin}_{Z} g(Z) + \frac{\rho}{2}\|AX^{(k+1)} - Z + U^{(k)}\|_F^2$
  $U^{(k+1)} = U^{(k)} + AX^{(k+1)} - Z^{(k+1)}$
  $k \leftarrow k + 1$
until *convergence*;

Algorithm 1: Generic MADMM method for non-smooth optimization on manifold $\mathcal{M}$.

X-step: a smooth optimization, only a few iterations are done [1]

Z-step: a proximity operator of $\frac{1}{\rho} g(Z)$ at $AX + U$, often has a closed form solution (e.g., for $\|\cdot\|_1$, $\|\cdot\|_*$, $\|\cdot\|_{2,1}$)

Acknowledgement

This research was supported by the ERC Starting Grant No.307047 (COMET).

Results

Compressed modes ($f = \mathrm{tr}(\cdot)$, $g = \mu\|\cdot\|_1$, $\mathcal{M} = \{X \in \mathbb{R}^{n \times k} : X^\top X = I\}$, Stiefel manifold)

Ozoliņš et al. [2] proposed a construction of localized Laplacian quasi-eigenbases by solving

$$\min_{\Phi \in \mathbb{R}^{n \times k}} \mathrm{tr}(\Phi^\top \Delta \Phi) + \mu\|\Phi\|_1 \quad \text{s.t.} \quad \Phi^\top \Phi = I,$$

where Δ is a Laplacian represented as $n \times n$ sparse symmetric matrix, $\Phi = (\phi_1, \ldots, \phi_k)$ is the $n \times k$ matrix of the first quasi-eigenvectors arranged as columns, and $\mu > 0$ is a parameter, controlling sparseness.

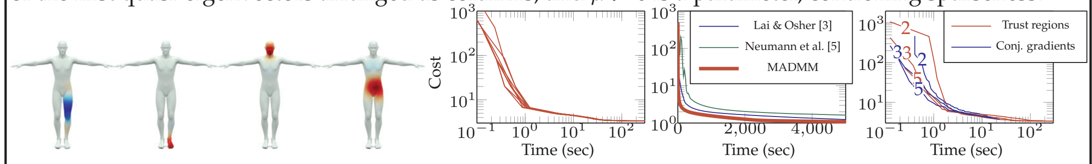


Figure 2. Compressed modes problem. From left to right: first compressed modes computed on a human mesh containing $n = 8$K points using MADMM with parameter $\mu = 10^{-3}$ and three manifold optimization iterations in the X-step; convergence of MADMM on a problem of size $n = 500$, $k = 10$ with different random initialization; comparison of convergence of different splitting methods and MADMM on a problem of size $n = 8$K; convergence of MADMM using different solvers and number of iterations at X-step on the same problem.

Robust Euclidean embedding (REE) ($f \equiv 0$, $g = \|\cdot\|_1$, $\mathcal{M}$ = fixed-rank positive semi-definite matrices)

Cayton and Dasgupta [6] treated an L_1 formulation of the multidimensional scaling (MDS) problem: given an $n \times n$ matrix D of squared distances, the goal is to find a k-dimensional configuration of points $X \in \mathbb{R}^{n \times k}$ preserving D as close as possible. The L_1 norm allows to cope with outliers and noise in data.

RRE problem (notation as in Figure 2):

$$\min_{B \in \mathbb{R}^{n \times n}} \|D - \mathrm{dist}(B)\|_1 \quad \text{s.t.} \quad B \succeq 0, \ \mathrm{rank}(B) \leq k,$$

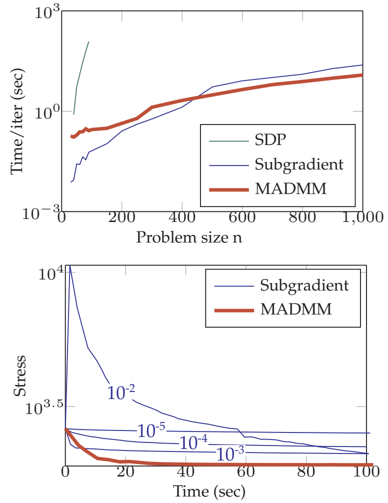


Figure 3. REE problem. Top: scalability of different algorithms; shown is single iteration complexity as functions of the problem size n using random distance data. SDP did not scale beyond $n = 100$. Bottom: example of convergence of MADMM and subgradient algorithm of [6] on the US cities problem of size $n = 500$. The subgradient algorithm is very sensitive to the choice of the initial step size c (choosing too large c breaks the convergence, while too small c slows down the convergence).

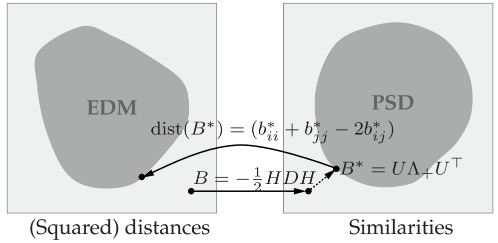


Figure 2. Illustration of the classical MDS approach and the equivalence between Euclidean distance matrices (EDM; D) and positive semi-definite (PSD; B) similarity matrices; $H = I - \frac{1}{n}11^\top$ and $B^* = XX^\top$, where X is the desired embedding.

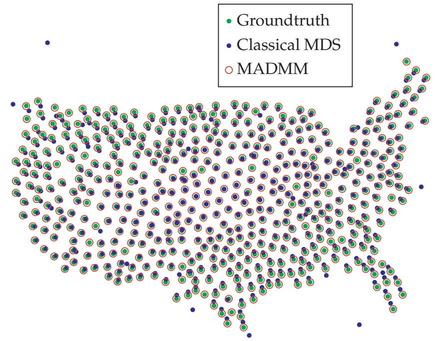


Figure 4. Embedding of the noisy distances between 500 US cities in the plane using classical MDS (blue) and REE solved using MADMM (red). The distance matrix was contaminated by sparse noise by doubling the distance between some cities.

References

[1] N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a Matlab toolbox for optimization on manifolds JMLR, 2014
[2] V. Ozoliņš, R. Lai, R. Caflisch, S. Osher. Compressed modes for variational problems in mathematics and physics PNAS, 2013
[3] R. Lai, S. Osher. A splitting method for orthogonality constrained problems J. Scientific Computing, 2014
[4] G. Rosman, Y. Wang, X.-C. Tai, R. Kimmel, A. M. Bruckstein. Fast Regularization of Matrix-Valued Images Proc. EAOMICV, 2011
[5] T. Neumann, K. Varanasi, C. Theobalt, M. Magnor, M. Wacker. Compressed Manifold Modes for Mesh Processing CGF, 2014
[6] L. Cayton, S. Dasgupta. Robust Euclidean embedding Proc. ICML, 2006