Robust Accelerated ADMM

Fangda Gu, Jingqi Li, Ziye Ma*

May 2020

Abstract

Accelerated ADMM is a fast ADMM algorithm based on Nestrov accelerated gradient descent[1]. However, if part of the problem is noisy, then the algorithm can diverge easily, as shown in experiments in later sections. In order to address this problem, we propose an algorithm that has a single scalar parameter that can be tuned to trade off robustness to noise versus convergence gaurantees.

1 Introduction

ADMM is called the *Alternating Direction Method of Multipliers* because it alternates between solving for the primal problem (in its constraint-penalized augmented Lagrangian form) and the dual problem. The accelerated ADMM algorithm takes advantage of the fact that ADMM uses gradient ascent to solve for the dual problem by using Nestrov-accelerated gradient steps (also called the momentum method in some literature). However, accelerated gradient methods aren't as robust to noise (normally in the gradient) as vanilla gradient methods are. Therefore, it is conceivable that the accelerated ADMM also suffers from this problem, as corroborated by our own experiments. Towards this end, we propose a method that is similar to that of [2], enabling us to make the accelerated ADMM algorithm more robust to noise.

1.1 Notations

We denote l_2 norm as $\|\cdot\|$ and l_1 norm as $|\cdot|$. For any convex function f, f^* denotes its Fenchel conjugate, defined as $f^*(y) = \sup_x \langle x, y\rangle - f(x)$. We denote $q^P(\cdot)$ as the dual to the primal problem P (usually a mathematical programming, hence the use of capital P) and $q^P_{\text{aug}}(\cdot)$ as the augmented dual to the same primal problem. L^P and L^P_{aug} are lagrangians defined in a similar fashion. In this paper, Λ is used to denote eigenvalues as opposed to λ in order to avoid confusion with lagrange multipliers. Furthermore, $\Lambda_{\min}(\cdot)$ and $\Lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a symmetric matrix respectively. As per standard notation, $\mathcal{D}(f)$ is the domain of function f.

2 Problem Statement

First consider problem with this form

$$\begin{aligned} &\min H(u) + G(v) \\ &\text{s.t. } Au + Bv = b \end{aligned} \tag{1}$$

*Equal contribution, sorted in alphabetical order

where $u \in \mathbb{R}^{n_u}$, $v \in \mathbb{R}^{n_v}$, and both H, G are closed convex functions. $b \in \mathbb{R}^{n_b}$ and A, B are affine transformations of appropriate size.
ADMM is an algorithm tailored specifically to problems of form (1), and we present both ADMM and accelerated ADMM below:

Algorithm 1: ADMM

Input : $v_0 \in \mathbb{R}^{n_v}$,$\lambda_0 \in \mathbb{R}^{n_b}$, $\rho > 0$
for $k = 0, 1, 2, ...$ **do**
  $u_{k+1} = \mathrm{argmin}_u H(u) + \langle \lambda_k, -Au \rangle + \frac{\rho}{2}\|b - Au - Bv_k\|^2$
  $v_{k+1} = \mathrm{argmin}_u G(v) + \langle \lambda_k, -Bv \rangle + \frac{\rho}{2}\|b - Au_{k+1} - Bv\|^2$
  $\lambda_{k+1} = \lambda_k + \rho(b - Au_{k+1} - Bv_{k+1})$
end

Algorithm 2: Acclerated ADMM for strongly convex objective

Input : $v_{-1} = \hat{v}_0 \in \mathbb{R}^{n_v}$,$\lambda_{-1} = \hat{\lambda}_0 \in \mathbb{R}^{n_b}, \rho > 0, \alpha_1 = 1$
for $k = 0, 1, 2, ...$ **do**
  $u_k = \mathrm{argmin}_u H(u) + \langle \hat{\lambda}_k, -Au \rangle + \frac{\rho}{2}\|b - Au - B\hat{v}_k\|^2$
  $v_k = \mathrm{argmin}_u G(v) + \langle \hat{\lambda}_k, -Bv \rangle + \frac{\rho}{2}\|b - Au_k - Bv\|^2$
  $\lambda_k = \hat{\lambda}_k + \rho(b - Au_k - Bv_k)$
  $\alpha_{k+1} = \frac{1+\sqrt{1+4\alpha_k^2}}{2}$
  $\hat{v}_{k+1} = v_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(v_k - v_{k-1})$
  $\hat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(\lambda_k - \lambda_{k-1})$
end

2.1 Strong Concavity of Augmented Dual of Problem (1)

In this section we discuss when the augmented dual of problem (1) is strongly concave. This property is desirable since if it holds, many results in [2] can be applied directly.

Assumption 1. *$H(u)$ and $G(v)$ are l_H and l_G smooth respectively.*

We first proceed to show that if assumption 1 holds, then the dual of problem (1) is strongly concave. First observe that:

$$\begin{aligned} q^{(1)}(\lambda) &= \inf_{u,v} H(u) + G(v) + \langle \lambda, b - Au - Bv \rangle \\ &= \inf_u \{H(u) - \langle A^T\lambda, u \rangle\} + \inf_v \{G(v) - \langle B^T\lambda, v \rangle\} + \langle \lambda, b \rangle \\ &= -H^*(A^T\lambda) - G^*(B^T\lambda) + \langle \lambda, b \rangle \end{aligned} \tag{2}$$

According to Theorem 1 of [3], $H^*(\cdot)$ is $(\frac{1}{l_H})$-strongly convex and $G^*(\cdot)$ is $(\frac{1}{l_G})$-strongly convex. For H^*, this means $(s_y - s_x)^T(y - x) \geq \frac{1}{l_H}\|y - x\|^2 \; \forall x, y$ and any $s_x \in \partial H^*(x), s_y \in \partial H^*(y)$. WLOG, make $x = A^T\lambda_x$ and $y = A^T\lambda_y$ $\forall \lambda_x, \lambda_y$ s.t. $x, y \in \mathcal{D}(H^*)$. Then,

$$\begin{aligned}(s_y - s_x)^T(y-x) &= (\partial H^*(A^T\lambda_y) - \partial H^*(A^T\lambda_x))^T(A^T\lambda_y - A^T\lambda_x) \\ &= (A\partial H^*(A^T\lambda_y) - A\partial H^*(A^T\lambda_x))^T(\lambda_x - \lambda_y) \\ &= (\frac{\partial H^*(A^T\lambda_y)}{\partial\lambda_y} - \frac{\partial H^*(A^T\lambda_x)}{\partial\lambda_x})^T(\lambda_x - \lambda_y) \\ &\geq \frac{1}{l_H}\|y-x\|^2 = \frac{1}{l_H}(\lambda_y-\lambda_x)^T AA^T(\lambda_y - \lambda_x) \\ &\geq \frac{\Lambda_{\min}(AA^T)}{l_H}\|\lambda_y - \lambda_x\|^2\end{aligned} \tag{3}$$

Therefore, it is easy to see that $H^*(A^T\lambda)$ is $(\frac{\Lambda_{\min}(AA^T)}{l_H})$-strongly convex in λ. Following exactly the same procedures, we also arrive at the fact that $G^*(B^T\lambda)$ is $(\frac{\Lambda_{\min}(BB^T)}{l_G})$-strongly convex in λ. Denote $\mu_1 = \frac{\Lambda_{\min}(AA^T)}{l_H} + \frac{\Lambda_{\min}(BB^T)}{l_G}$, then if follows that $H^*(A^T\lambda) + G^*(B^T\lambda)$ is (μ_1)-strongly convex in λ. Now, define

$$\begin{aligned}F'(\lambda) &= H^*(A^T\lambda) + G^*(B^T\lambda) \\ F(\lambda) &= H^*(A^T\lambda) + G^*(B^T\lambda) - \langle\lambda, b\rangle\end{aligned} \tag{4}$$

Given $F'(\lambda)$ is (μ_1)-strongly convex, it's obvious that by using the $(s_y - s_x)^T(y-x) \geq \gamma\|y-x\|^2$ definition of (γ)-strongly convex function we can verify that $F(\lambda)$ is also (μ_1)-strongly convex in λ. This is equivalent to saying that $q^{(1)}(\lambda) = -F(\lambda)$ is (μ_1)-strongly concave.
Now we consider the augmented dual. Note that augmented dual is actually just a Moreau envelope of the original dual. For a concave function f, the Moreau envelope is defined as such:

$$M_f(\lambda) = \sup_x\{f(x) - \frac{1}{2\rho}\|x-\lambda\|^2\} \tag{5}$$

Specifically, for $q^{(1)}(\lambda)$,

$$\begin{aligned}M_{q^{(1)}}(\lambda) &= \sup_x \inf_{u,v}\{H(u) + G(v) + \langle x, b - Au - Bv\rangle\} - \sup_x \frac{1}{2\rho}\|x-\lambda\|^2 \\ &= \inf_{u,v}\sup_x\{H(u) + G(v) + \langle x, b - Au - Bv\rangle\} - \sup_x \frac{1}{2\rho}\|x-\lambda\|^2 \\ &= \inf_{u,v}\{H(u) + G(v) + \sup_x\{\langle x, b-Au-Bv\rangle - \frac{1}{2\rho}\|x-\lambda\|^2\}\} \\ &= \inf_{u,v}\{H(u)+G(v)+\langle\lambda, b-Au-Bv\rangle + \frac{\rho}{2}\|b-Au-Bv\|^2\} = q^{(1)}_{\text{aug}}(\lambda)\end{aligned} \tag{6}$$

Note that the minimax theorem is used here to swap inf and sup since $H(u)+G(v)+\langle x, b-Au-Bv\rangle$ is convex in u, v and affine in x.
From Fact 3.11 of [4] we know that if $\rho = 1$, $q^{(1)}_{\text{aug}}(\lambda)$ is $(\frac{1}{K})$-strongly concave with $\frac{1}{\mu_1} - 1 \leq K \leq \frac{1}{\mu_1} + 1$. Therefore the augmented dual of problem (1) is strongly convex when $\rho = 1$ and assumption 1 holds.

3 Robust Accelerated ADMM

For our work, we adapt an algorithm for robust accelerated first-order optimization from [2]. In essence, this algorithm hopes to achieve characteristics of both the accelerated GD and vanilla GD, albeit not at the same time, but can be thought of a compromise between the two.
We adapted this algorithm for the dual ascent part of our ADMM:

Algorithm 3: Robust Accelerated ADMM for strongly convex objective

Input : $v_{-1} = \hat{v}_0 \in \mathbb{R}^{n_v}, \lambda_{-1} = \hat{\lambda}_0 \in \mathbb{R}^{n_b}, \rho = 1$

for $k = 0, 1, 2, \ldots$ **do**

  $u_k = \text{argmin}_u H(u) + \langle \hat{\lambda}_k, -Au \rangle + \frac{\rho}{2} \|b - Au - B\hat{v}_k\|^2$

  $v_k = \text{argmin}_u G(v) + \langle \hat{\lambda}_k, -Bv \rangle + \frac{\rho}{2} \|b - Au_k - Bv\|^2$

  $\lambda_k = \lambda_{k-1} + \beta(\lambda_{k-1} - \lambda_{k-2}) + \alpha(b - Au_k - Bv_k)$

  $\hat{\lambda}_{k+1} = \lambda_k + \gamma(\lambda_k - \lambda_{k-1})$

  $\hat{v}_{k+1} =$

end

3.1 Analysis of Robust A-ADMM

Given results derived in section 2.1, we know that when $\rho = 1$ and both H, G are smooth, the augmented lagrangian of primal is strongly concave. Moreover, from the properties of augmented dual we know that they are also $(\frac{1}{\rho})$-smooth. If $K \geq 1$, then we have the augmented dual is (m)-strongly concave and (L)-smooth with $0 < m \leq L$

If either A or B is zero, then all the convergence guarantees from [2] can be trivially applied. This can be verified by realizing:

$$v_k, u_k = \text{argmin}_{u,v} H(u) + G(v) + \langle \hat{\lambda}_k, -Au - Bv \rangle + \frac{\rho}{2} \|b - Au - Bv\|^2 \ \forall k \quad \text{if}\{A = 0 || B = 0\} \tag{7}$$

4 Conclusions and Future Work

Acknowledgements

A special thanks goes to Professor Wainwright and TA Armin Askari.

References

[1] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.

[2] S. Cyrus, B. Hu, B. Van Scoy, and L. Lessard, "A robust accelerated optimization algorithm for strongly convex functions," in *2018 Annual American Control Conference (ACC)*, pp. 1376–1381, IEEE, 2018.

[3] X. Zhou, "On the fenchel duality between strong convexity and lipschitz continuous gradient," *arXiv preprint arXiv:1803.06573*, 2018.

[4] C. Planiden and X. Wang, "Proximal mappings and moreau envelopes of single-variable convex piecewise cubic functions and multivariable gauge functions," in *Nonsmooth Optimization and Its Applications*, pp. 89–130, Springer, 2019.