The Application of Multi-block ADMM on Isotonic Regression Problems

Junxiang Wang JWANG40@GMU.EDU
Liang Zhao LZHAO9GMU.EDU
George Mason University, Fairfax, VA 22030

Abstract

The multi-block ADMM has received much attention from optimization researchers due to its excellent scalability. In this paper, the multi-block ADMM is applied to solve two large-scale problems related to isotonic regression. Numerical experiments show that the multi-block ADMM is convergent when the chosen parameter is small enough and the multi-block ADMM scales well compared with baselines.

1. Introduction

In recent years, the Alternating Direction Method of Multipliers (ADMM) has received considerable attention from the community of machine learning researchers. This is because it is a natural fit for wide large-scale data applications, including deep learning [23], phase retrieval [25], vaccine adverse event detection [21, 22] and compressive sensing [4]. The direct extension of the classic ADMM is multi-block ADMM (i.e. ADMM with no less than three variables), which is written mathematically as follows:

$$\min\nolimits_{x_1,\cdots,x_n} \sum\nolimits_{i=1}^n f_i(x_i), \;\; s.t. \sum\nolimits_{i=1}^n A_i x_i = 0$$

where $f_i : \mathbb{R}^{m_i} \to \mathbb{R}(i = 1, \cdots, n)$ are convex functions, $x_i \in \mathbb{R}^{m_i}(i = 1, \cdots, n)$ are vectors of length m_i. $A_i \in \mathbb{R}^{p\times m_i}(i = 1, \cdots, n)$ are matrices. The augmented Lagrangian function is formulated as $L_\rho(x_1, \cdots, x_n, y) = \sum_{i=1}^n f_i(x_i) + (\rho/2)\| \sum_{i=1}^n A_i x_i + y/\rho\|_2^2$ where $y \in \mathbb{R}^p$ is a dual variable and $\rho > 0$ is a penalty parameter. The multi-block ADMM is solved by the following steps:

$$\begin{aligned} &x_i^{k+1} \leftarrow \arg\min\nolimits_{x_i} L_\rho(\cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots)(i = 1, \cdots, n) \\ &y^{k+1} \leftarrow y^k + \rho \sum\nolimits_{i=1}^n A_i x_i^{k+1} \end{aligned}$$

A variety of related works on the convergence of the multi-block ADMM are detailed in Section B in the Appendix.

While the multi-block ADMM can solve problems with linear equality constraints, it cannot directly be applied to the problems with multiple inequality constraints such as the isotonic regression problem [9]. In this paper, we propose new strategies based on the multi-block ADMM to address existing computational challenges in the isotonic regression problems. The well-known isotonic regression aims to return a sequence of responses given a predictor and pre-defined order constraints, which has been addressed by many previous works. See [1, 2, 8, 9, 14, 16] for more information. However, The main drawback is that their computational cost is very expensive when solving large-scale problems.

To deal with the challenge of scalability, we leverge the advantage of parallel computing of the multi-block ADMM by integrating objective variables into several vectors. The following questions are addressed for the proposed multi-block ADMM frameworks on two isotonic regression problems: *1. Does the multi-block ADMM converge? 2. Does the multi-block ADMM scale well?*

2. Isotonic Regression Problems

Algorithm 1 The Multi-block ADMM Algorithm to Solve Problem 1

```
Initialize p, q, u, y_1, y_2, ρ > 0, k = 0.
repeat
    Update u^{k+1} in Equation (1).
    Update p^{k+1} in Equation (2).
    Update q^{k+1} in Equation (3).
    Update r_1^{k+1} ← p^{k+1} − q^{k+1} + u^{k+1}.
    Update r_{2,i}^{k+1} ← p_{i+1}^{k+1} − q_i^{k+1} (i = 1, ··· , n − 2), r_2^{k+1} ← [r_{2,1}^{k+1}; ··· ; r_{2,n−2}^{k+1}].
    Update s_1^{k+1} ← ρ(p^{k+1} − q^{k+1} − p^k + q^k).
    Update s_2^{k+1} ← ρ(q^k − q^{k+1}).
    Update r^{k+1} ← sqrt(||r_1^{k+1}||_2^2 + ||r_2^{k+1}||_2^2). # Calculate the primal residual.
    Update s^{k+1} ← sqrt(||s_1^{k+1}||_2^2 + ||s_2^{k+1}||_2^2). # Calculate the dual residual.
    Update y_1^{k+1} ← y_1^k + ρ r_1^{k+1}.
    Update y_2^{k+1} ← y_2^k + ρ r_2^{k+1}.
    k ← k + 1.
until convergence.
Output p, q and u.
```

2.1. The Multi-block ADMM for Smoothed Isotonic Regression

The classic isotonic regression is a problem to return a non-decreasing response given a predictor. However, the fitted response resembles a step function while a response is expected to be smooth and continuous in many applications [18]. To achieve this, Sysoev and Burdakov proposed a smoothed isotonic regression problem to eliminate sharp 'jumps' of the response function given a predictor $x_i(i = 1, \cdots, n)$ [18]:

Problem 1 (Smoothed Isotonic Regression)

$$\begin{aligned}&\min_{\beta_1,\cdots,\beta_n} \sum\nolimits_{i=1}^{n} w_i(x_i - \beta_i)^2 + \lambda \sum\nolimits_{i=1}^{n-1} (\beta_i - \beta_{i+1})^2\\ &s.t.\ \beta_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_n\end{aligned}$$

where $w_i > 0(i = 1, \cdots, n)$ are assigned weights, $\beta_i(i = 1, \cdots, n)$ are fitted predictors, and $\lambda \geq 0$ is a penalty parameter. The multi-block ADMM is applied to realize parallel computing: we introduce two vectors p and q of length $n-1$ such that $p_i = \beta_i(i = 1, \cdots, n-1)$ and $q_i = \beta_{i+1}(i = 1, \cdots, n-1)$, respectively. The problem can be reformulated as

$$\begin{aligned}&\min_{p,q,u} w_1(x_1 - p_1)^2 + \sum\nolimits_{i=2}^{n-1} ((w_i/2)(x_i - p_i)^2 + (w_i/2)(x_i - q_{i-1})^2) + w_n(x_n - q_{n-1})^2 + \lambda\|u\|_2^2\\ &s.t.\ p - q + u = 0, u \geq 0, p_{i+1} = q_i,\ i = 1, 2, \cdots, n-2\end{aligned}$$

where $p = [p_1, \cdots, p_{n-1}]$ and $q = [q_1, \cdots, q_{n-1}]$. The augmented Lagrangian is $L_\rho(u, p, q, y_1, y_2) = w_1(x_1 - p_1)^2 + \sum_{i=2}^{n-1}((w_i/2)(x_i - p_i)^2 + (w_i/2)(x_i - q_{i-1})^2) + w_n(x_n - q_{n-1})^2 + \lambda\|u\|_2^2 + (\rho/2)\|p - q + u + y_1/\rho\|_2^2 + (\rho/2)\sum_{i=1}^{n-2}(p_{i+1} - q_i + y_{2,i}/\rho)^2$ where $y_1 = [y_{1,1}, \cdots, y_{1,n-1}]$, $y_2 = [y_{2,1}, \cdots, y_{2,n-2}]$ and $\rho > 0$. The multi-block ADMM to solve Problem 1 is shown in Algorithm 1. Each subproblem has a closed-form solution and can be implemented in parallel, which is shown as follows:

1. Update u.

The variable u is updated as follows:

$$\begin{aligned}u^{k+1} &\leftarrow \arg\min_u (\rho/2)\|p^k - q^k + u + y_1^k/\rho\|_2^2 + \lambda\|u\|_2^2,\ s.t.\ u \geq 0.\\ &\leftarrow \max((\rho(q^k - p^k) - y_1^k)/(\rho + 2\lambda), 0).\end{aligned} \tag{1}$$

2. Update p.
The variable p is updated as follows:

$$\begin{aligned}
p_1^{k+1} &\leftarrow \arg\min_{p_1} w_1(x_1-p_1)^2+(\rho/2)\|p_1-q_1^k+u_1^{k+1}+y_1^k/\rho\|_2^2\\
&\leftarrow (2w_1x_1+\rho q_1^k-\rho u_1^{k+1}-y_{1,1}^k)/(2w_1+\rho)\\
p_i^{k+1} &\leftarrow \arg\min_{p_i}(w_i/2)(x_i-p_i)^2+(\rho/2)\|p_i-q_i^k+u_i^{k+1}+y_i^k/\rho\|_2^2\\
&+(\rho/2)\|p_i-q_{i-1}^k+y_{2,i-1}^k/\rho\|_2^2(i=2,\cdots,n-1).\\
&\leftarrow (w_ix_i+\rho q_i^k-\rho u_i^{k+1}-y_{1,i}^k+\rho q_{i-1}^k-y_{2,i-1}^k)/(w_i+2\rho)(i=2,\cdots,n-1). \qquad (2)
\end{aligned}$$

3. Update q.
The variable q is updated as follows:

$$\begin{aligned}
q_i^{k+1} &\leftarrow \arg\min_{q_i}(w_{i+1}/2)(x_{i+1}-q_i)^2+(\rho/2)\|p_i^{k+1}-q_i+u_i^{k+1}+y_{1,i}^k/\rho\|_2^2\\
&+(\rho/2)(p_{i+1}^{k+1}-q_i+y_{2,i}^k/\rho)(i=1,\cdots,n-2). \qquad (3)\\
&\leftarrow (w_{i+1}x_{i+1}+\rho p_i^{k+1}+\rho u_i^{k+1}+y_{1,i}^k+\rho p_{i+1}^{k+1}+y_{2,i}^k)/(w_{i+1}+2\rho)(i=1,\cdots,n-2)\\
q_{n-1}^{k+1} &\leftarrow \arg\min_{q_{n-1}} w_n(x_n-q_{n-1})^2+(\rho/2)\|p_{n-1}^{k+1}-q_{n-1}+u_{n-1}^{k+1}+y_{1,n-1}^k/\rho\|_2^2.\\
&\leftarrow (2w_nx_n+\rho p_{n-1}^{k+1}+\rho u_{n-1}^{k+1}+y_{1,n-1}^k)/(2w_n+\rho).
\end{aligned}$$

Due to space limit, the convergence of Algorithm 1 is discussed in Section A in the Appendix.

2.2. The Multi-block ADMM for Multi-dimensional Ordering

The previous smoothed isotonic regression only considers linear orders, while multi-dimensional orders are more general in isotonic regression applications [17]. A multi-dimensional order is defined in a $m-$dimensional space $Z_i=(z_{i,1},\cdots,z_{i,m})$. $Z_i\leq Z_j$ if and only if $z_{i,1}\leq z_{j,1},\cdots,z_{i,m}\leq z_{j,m}$. It can be represented equivalently as Directed Acyclic Graph (DAG) $G=(V,E)$ where $(Z_i,Z_j)\in E$ if $Z_i\leq Z_j$ [17]. Formally, the multi-dimensional ordering problem is formulated as follows:

Problem 2 (Multi-dimensional Ordering)

$$\begin{aligned}
&\min_{\alpha_1,\cdots,\alpha_n}\sum\nolimits_{i=1}^n w_i(Y_i-\alpha_i)^2\\
&s.t.\ \alpha_i\leq\alpha_j \textit{ iff } Z_i\leq Z_j(1\leq i,j\leq n).
\end{aligned}$$

where $w_i>0(i=1,\cdots,n)$ are assigned weights, $Y_i(i=1,\cdots,n)$ are predictors, and $\alpha_i(i=1,\cdots,n)$ are fitted predictors.

To handle large-scale multi-dimensional ordering problems, we leverage the advantage of parallel computing of the multi-block ADMM to solve it in an exact form. By introducing two vectors g and h, this problem is equivalent of

$$\begin{aligned}
&\min_{g,h} W^T((Y-g)\odot(Y-g))/2+W^T((Y-h)\odot(Y-h))/2\\
&s.t.\ E_1g-E_2h+v=0,\ v\geq 0,\ g=h.
\end{aligned}$$

where $W=(w_1,\cdots,w_n)$, $Y=(Y_1,\cdots,Y_n)$ and $\odot$ is the Hadamard product. $E_1\in R^{|E|\times n}$ and $E_2\in R^{|E|\times n}$ are representations of the edge set E: the k-th edge $(i,j)\in E$ means that $E_{1,k,i}=1$ and $E_{2,k,j}=1$ while $E_{1,k,p}=0(1\leq p\leq n,p\neq i)$ and $E_{2,k,q}=0(1\leq q\leq n,q\neq j)$. The augmented Lagrangian is $L_\rho(v,g,h,y_1,y_2)=W^T((Y-g)\odot(Y-g))/2+W^T((Y-h)\odot(Y-h))/2+(\rho/2)\|E_1g-E_2h+v+y_1/\rho\|_2^2+(\rho/2)\|g-h+y_2/\rho\|_2^2$. where $\rho>0$. The multi-block ADMM to solve Problem 2 is shown in Algorithm 2. Each subproblem has a closed-form solution

Algorithm 2 The Multi-block ADMM Algorithm to Solve Problem 2

Initialize g, h, v, y_1, y_2, $\rho > 0$, $k = 0$.
repeat
  Update v^{k+1} in Equation (4).
  Update g^{k+1} in Equation (5).
  Update h^{k+1} in Equation (6).
  Update $r_1^{k+1} \leftarrow E_1 g^{k+1} - E_2 h^{k+1} + v^{k+1}$.
  Update $r_2^{k+1} \leftarrow g^{k+1} - h^{k+1}$.
  Update $s_1^{k+1} \leftarrow \rho(E_1 g^{k+1} - E_2 h^{k+1} - E_1 g^k + E_2 h^k)$.
  Update $s_2^{k+1} \leftarrow \rho E_2 (h^k - h^{k+1})$.
  Update $s_3^{k+1} \leftarrow \rho(h^k - h^{k+1})$.
  Update $r^{k+1} \leftarrow \sqrt{\|r_1^{k+1}\|_2^2 + \|r_2^{k+1}\|_2^2}$. # Calculate the primal residual.
  Update $s^{k+1} \leftarrow \sqrt{\|s_1^{k+1}\|_2^2 + \|s_2^{k+1}\|_2^2 + \|s_3^{k+1}\|_2^2}$. # Calculate the dual residual.
  Update $y_1^{k+1} \leftarrow y_1^k + \rho r_1^{k+1}$.
  Update $y_2^{k+1} \leftarrow y_2^k + \rho r_2^{k+1}$.
  $k \leftarrow k + 1$.
until convergence.
Output g, h and v.

and can be implemented in parallel, which is shown as follows:

1. Update v.

The variable v is updated as follows:

$$\begin{aligned} v^{k+1} &\leftarrow \arg\min_v (\rho/2)\|E_1 g^k - E_2 h^k + v + y_1^k/\rho\|_2^2, \ s.t. \ v \geq 0. \\ &\leftarrow \max(-E_1 g^k + E_2 h^k - y_1^k/\rho, 0). \end{aligned} \tag{4}$$

2. Update g.

The variable g is updated as follows:

$$\begin{aligned} g^{k+1} &\leftarrow \arg\min_g W^T((Y-g)\odot(Y-g))/2 + (\rho/2)\|E_1 g - E_2 h^k + v^{k+1} + y_1^k/\rho\|_2^2 \\ &\quad + (\rho/2)\|g - h^k + y_2^k/\rho\|_2^2. \\ &\leftarrow (\text{diag}(W) + \rho E_1^T E_1 + \rho I)^{-1}(\text{diag}(W)Y + \rho E_1^T E_2 h^k - \rho E_1^T v^{k+1} - E_1^T y_1^k + \rho h^k - y_2^k) \end{aligned} \tag{5}$$

where I is an identity matrix and diag(W) is a diagonal matrix where the main diagonal is W.

3. Update h.

The variable h is updated as follows:

$$\begin{aligned} h^{k+1} &\leftarrow \arg\min_h W^T((Y-h)\odot(Y-h))/2 + (\rho/2)\|E_1 g^{k+1} - E_2 h + v^{k+1} + y_1^k/\rho\|_2^2 \\ &\quad + (\rho/2)\|g^{k+1} - h + y_2^k/\rho\|_2^2. \\ &\leftarrow (\text{diag}(W) + \rho E_2^T E_2 + \rho I)^{-1}(\text{diag}(Y) + \rho E_2^T E_1 g^{k+1} + \rho E_2^T v^{k+1} + E_2^T y_1^k + \rho g^{k+1} + y_2^k). \end{aligned} \tag{6}$$

Due to space limit, the convergence of Algorithm 2 is discussed in Section A in the Appendix.

3. Experiment

In this section, we validate the multi-block ADMM using simulated datasets and compare it with existing state-of-the-art methods. All experiments were conducted on a 64-bit machine with Intel(R) core(TM)processor (i7-6820HQ CPU@ 2.70GHZ) and 16.0GB memory.

3.1. Data Generation and Parameter Settings

Due to space limit, data and parameters are detailed in Section C in the Appendix.

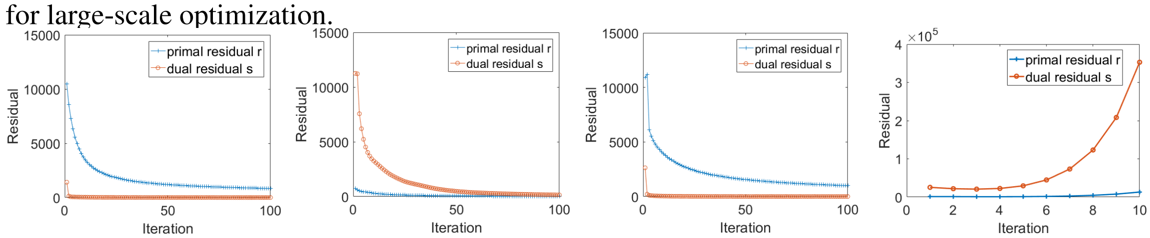
3.2. Baselines

Two methods Smoothed Pool-Adjacent-Violators (SPAV) [18] and Interior Point Method (IPM) [9] are used for comparison. The details can be found in Section D in the Appendix.

3.3. Experimental Results

In this section, the experimental results on two problems are explained in detail.

1. Does the multi-block ADMM converge? Figure 1 illustrates the convergence properties of the multi-block ADMM when $n = 1000$ on the smoothed isotonic regression problem and the multi-dimensional ordering problem. Two choices of $\rho = 0.1$ and $\rho = 10$ are shown on Figure 1. Overall, Figure 1 (a)-(c) shows that the multi-block ADMM converges while Figure 1(d) shows the divergence: r and s drop drastically at the beginning and then decrease smoothly through the end in the Figures 1 (a)-(c); however, Figure 1(d) displays a surge of r. Moreover, when $\rho = 0.1$, r is located above s while when $\rho = 10$ the situation is the opposite. Obviously, the multi-block ADMM can obtain a reasonable solution within tens of iterations as long as ρ is small and hence it is suitable for large-scale optimization.



(a).$\rho = 0.1$ on the smoothed isotonic regression problem.

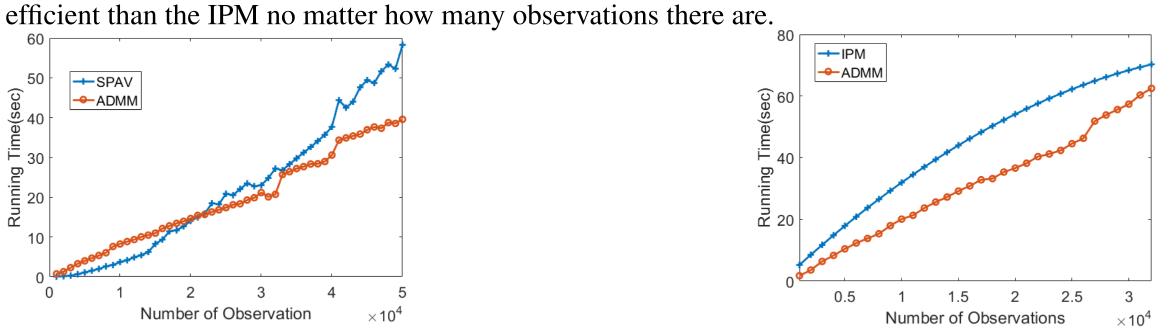
(b).$\rho = 10$ on the smoothed isotonic regression problem.

(c).$\rho = 0.1$ on the multi-dimensional ordering problem.

(c).$\rho = 10$ on the multi-dimensional ordering problem.

Figure 1: Convergence of the multi-block ADMM when $n = 1000$ on two problems.

2. Does the multi-block ADMM scale well? As Figure 2 shown, the running time of the multi-block ADMM increases linearly with the number of observations on two problems. In Figure 2(a), the SPAV is more efficient than the multi-block ADMM when the number of observations n is less than 20,000 but needs more time since then; as for Figure 2(b), the multi-block ADMM is more efficient than the IPM no matter how many observations there are.



(a).Scalability on the smoothed isotonic regression problem.

(b).Scalability on the multi-dimensional ordering problem.

Figure 2: Scalability of the multi-block ADMM on two problems.

4. Conclusion

The multi-block ADMM is an interesting topic in the optimization community in recent years. In this paper, we apply the multi-block ADMM to two problems related to isotonic regression: smoothed isotonic regression and multi-dimensional ordering. Most existing methods are not efficient enough to run on large-scale datasets. However, the main advantage of the multi-block ADMM is parallel computing and hence it is scalable to large datasets. We find that the multi-block ADMM converges when ρ is small and its running time increases linearly with the scalability of observations.

References

[1] RE Barlow and HD Brunk. The isotonic regression problem and its dual. *Journal of the American Statistical Association*, 67(337):140–147, 1972.

[2] Michael J Best and Nilotpal Chakravarti. Active set algorithms for isotonic regression; a unifying framework. *Mathematical Programming*, 47(1-3):425–439, 1990.

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[4] Rick Chartrand and Brendt Wohlberg. A nonconvex admm algorithm for group sparsity with sparse groups. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6009–6013. IEEE, 2013.

[5] Caihua Chen, Bingsheng He, Yinyu Ye, and Xiaoming Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1-2):57–79, 2016.

[6] Wei Deng, Ming-Jun Lai, Zhimin Peng, and Wotao Yin. Parallel multi-block admm with o (1/k) convergence. *Journal of Scientific Computing*, 71(2):712–736, 2017.

[7] Bingsheng He and Xiaoming Yuan. On the direct extension of admm for multi-block separable convex programming and beyond: from variational inequality perspective. *Manuscript, http://www. optimizationonline. org/DB_HTML/2014/03/4293. html*, 2014.

[8] Adam Tauman Kalai and Ravi Sastry. The isotron algorithm: High-dimensional isotonic regression. In *COLT*. Citeseer, 2009.

[9] Rasmus Kyng, Anup Rao, and Sushant Sachdeva. Fast, provable algorithms for isotonic regression in all l_p-norms. In *Advances in Neural Information Processing Systems*, pages 2719–2727, 2015.

[10] Tianyi Lin, Shiqian Ma, and Shuzhong Zhang. On the convergence rate of multi-block admm. *arXiv preprint arXiv:1408.4265*, 229, 2014.

[11] Tianyi Lin, Shiqian Ma, and Shuzhong Zhang. Global convergence of unmodified 3-block admm for a class of convex minimization problems. *arXiv preprint arXiv:1505.04252*, 2015.

[12] Tianyi Lin, Shiqian Ma, and Shuzhong Zhang. Iteration complexity analysis of multi-block admm for a family of convex minimization without strong convexity. *Journal of Scientific Computing*, 69(1):52–81, 2016.

[13] Sindri Magnússon, Pradeep Chathuranga Weeraddana, Michael G Rabbat, and Carlo Fischione. On the convergence of alternating direction lagrangian methods for nonconvex structured optimization problems. *IEEE Transactions on Control of Network Systems*, 3(3):296–309, 2016.

[14] Taesup Moon, Alex Smola, Yi Chang, and Zhaohui Zheng. Intervalrank: isotonic regression with listwise and pairwise constraints. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 151–160. ACM, 2010.

[15] Daniel P Robinson and Rachael Tappenden. A flexible admm algorithm for big data applications. *Journal of Scientific Computing*, 71(1):435–467, 2017.

[16] Quentin F Stout. Isotonic regression via partitioning. *Algorithmica*, 66(1):93–112, 2013.

[17] Quentin F Stout. Isotonic regression for multiple independent variables. *Algorithmica*, 71(2): 450–470, 2015.

[18] Oleg Sysoev and Oleg Burdakov. A smoothed monotonic regression via l2 regularization. *Knowledge and Information Systems*, pages 1–22, 2018.

[19] Min Tao and Xiaoming Yuan. Convergence analysis of the direct extension of admm for multiple-block separable convex minimization. *Advances in Computational Mathematics*, pages 1–41, 2016.

[20] Fenghui Wang, Zongben Xu, and Hong-Kun Xu. Convergence of bregman alternating direction method with multipliers for nonconvex composite problems. *arXiv preprint arXiv:1410.8625*, 2014.

[21] Junxiang Wang and Liang Zhao. Multi-instance domain adaptation for vaccine adverse event detection. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 97–106. International World Wide Web Conferences Steering Committee, 2018.

[22] Junxiang Wang, Liang Zhao, and Yanfang Ye. Semi-supervised multi-instance interpretable models for flu shot adverse event detection. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 851–860. IEEE, 2018.

[23] Junxiang Wang, Fuxun Yu, Xiang Chen, and Liang Zhao. Admm for efficient deep learning with global convergence. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 111–119, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330936. URL http://doi.acm.org/10.1145/3292500.3330936.

[24] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *arXiv preprint arXiv:1511.06324*, 2015.

[25] Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, 2012.

Appendix

Appendix A. The Convergence of Algorithms 1 and 2

In addition to the advantage of parallel computing, the convergence of the Algorithm 1 can be guaranteed by the following theorem:

Theorem 1 *There exists $d > 0$ such that if $\rho < d$, the Algorithm 1 converges globally to the optimal point with sublinear convergence rate.*

Proof Because the objective function is strongly convex in p and q, all assumptions in [10] are satisfied and hence the Theorem 3.2 and Theorem 4.2 in [10] ensure the global sublinear convergence if ρ is smaller than a threshold d. ■

As for the convergence properties of Algorithm 2, the below theorem guarantees the convergence if ρ is small.

Theorem 2 *There exists $d > 0$ such that if $\rho < d$, then the Algorithm 2 converges globally to the optimal point with sublinear convergence rate.*

Proof The proof is the same as Theorem 1. ■

Appendix B. Related Work on the Multi-block ADMM

The multi-block ADMM was firstly studied by Chen et al. [5] when they concluded that the multi-block ADMM does not necessarily converge by giving a counterexample. He and Yuan explained why the multi-block ADMM may diverge from the perspective of variational inequality framework [7]. Since then, many researchers studied sufficient conditions to ensure the global convergence of the multi-block ADMM. For example, Robinson and Tappenden and Lin et al. imposed strong convexity assumption on the objective function $f_i(x_i)(i = 1, \cdots, n)$ [10, 15]. Tao and Yuan did not require all objective functions $f_i(x_i)(i = 1, \cdots, n)$ to be strongly convex, but imposed full-rank assumption on $A_i(i = 1, \cdots, n)$ [19]. Lin et al. weakened the strong convexity assumption on the objective function $f_i(x_i)(i = 1, \cdots, n)$ by imposing the additional cocercity assumption on $f_i(x_i)(i = 1, \cdots, n)$ [12]. Deng et al. proved that the multi-block ADMM preserved convergence if multiple variables $x_i(i = 1, \cdots, n)$ were updated in the Jacobian fashion rather than Gauss-Seidel fashion [6]. Lin et al. proved that the multi-block ADMM converged for any $\rho > 0$ in the regularized least squares decomposition problem [11]. Moreover, some work extended the multi-block ADMM into the nonconvex settings. See [13, 20, 24] for more information.

Appendix C. Data Generation and Parameter Settings

The experimental data and parameters for two applications are explained in this section. Commonly, the maximal number of iteration was set to $10,000$. We set the tolerance $\varepsilon = 0.01\sqrt{n}$ according to the recommendation by Boyd et al. [3]. For the smoothed isotonic regression problem, we generated simulated observations $x_i(i = 1, \cdots, n)$ from a uniform distribution in $(0, 1000)$; we set $w_i = 1(i = 1, \cdots, n)$ and $\lambda = 1$. For the multi-dimensional ordering problem, we generated simulated observations $Y_i(i = 1, \cdots, n)$ from a uniform distribution in $(0, 1000)$. $w_i(i = 1, \cdots, n)$ were set to 1. E_1 and E_2 were generated from a 2-dimensional random grid graph to simulate the ordering of the 2-dimensional space.

Appendix D. Baselines

In order to test the scalability of the multi-block ADMM on two applications, two baselines were utilized for comparison:

1. Smoothed Pool-Adjacent-Violators(SPAV) [18]. The SPAV algorithm is a extension of the PAV algorithm designed for the smoothed isotonic regression problem. It partitions all $\beta_i(i = 1, \cdots, n)$ into many blocks according to the feasibility of inequality constraints. The iteration ends when all constraints are feasible.

2. Interior Point Method(IPM) [9]. The IPM method is an efficient algorithm to solve multi-dimensional ordering with convergence guarantee. Its time complexity is $O(|E|^{1.5} \log^2 |V| \log^2(|V|/\delta))$ where δ is a tolerance.