Column Partition Based Distributed Algorithms for Coupled Convex Sparse Optimization: Dual and Exact Regularization Approaches

Jinglai Shen, Jianghai Hu, and Eswar Kumar Hathibelagal Kammara

***Abstract*—This paper develops column partition based distributed schemes for a class of convex sparse optimization problems, e.g., basis pursuit (BP), LASSO, basis pursuit denosing (BPDN), and their extensions, e.g., fused LASSO. We are particularly interested in the cases where the number of (scalar) decision variables is much larger than the number of (scalar) measurements, and each agent has limited memory or computing capacity such that it only knows a small number of columns of a measurement matrix. The problems in consideration are densely coupled and cannot be formulated as separable convex programs. To overcome this difficulty, we consider their dual problems which are separable or locally coupled. Once a dual solution is attained, it is shown that a primal solution can be found from the dual of corresponding regularized BP-like problems under suitable exact regularization conditions. A wide range of existing distributed schemes can be exploited to solve the obtained dual problems. This yields two-stage column partition based distributed schemes for LASSO-like and BPDN-like problems; the overall convergence of these schemes is established. Numerical results illustrate the performance of the proposed two-stage distributed schemes.**

***Index Terms*—Sparse optimization, distributed computation, duality theory, exact regularization, sensitivity analysis.**

I. Introduction

SPARSE modeling and approximation finds broad applications in numerous fields of contemporary interest. Various efficient schemes have been proposed for convex or nonconvex sparse signal recovery [7], [23], [24]. To motivate the work of this paper, consider the well-studied LASSO problem: $\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \lambda\,\|x\|_1$, where $A\in\mathbb{R}^{m\times N}$ is the measurement matrix, $b\in\mathbb{R}^m$ is the measurement vector, and $\lambda>0$ is the penalty parameter. In sparse recovery, N is much larger than m. Besides, the matrix A usually satisfies certain uniform recovery conditions, e.g., the RIP [7]. As such, A is often dense, i.e., (almost) all of its elements are nonzero. We aim to develop column partition based distributed algorithms to solve the LASSO and related problems. Specifically, let $\{\mathcal{I}_1,\dots,\mathcal{I}_p\}$ be a disjoint union of $\{1,\dots,N\}$ such that $\{A_{\bullet\mathcal{I}_i}\}_{i=1}^p$ forms a column partition of A. Each agent i knows $A_{\bullet\mathcal{I}_i}$ and b but does not know $A_{\bullet\mathcal{I}_j}$ with $j\neq i$. By running the proposed distributed scheme, it is expected that each agent i attains the subvector of an optimal solution x^* to the LASSO corresponding to the index set $\mathcal{I}_i$, i.e., $x^*_{\mathcal{I}_i}$, at the end.

The distributed optimization task described above is inspired by two scenarios: big data subject to high memory or computational cost, and network systems with limited access to data. In the context of big data, a practitioner may deal with a *ultra-large* data set, e.g., N is extremely large, so that it would be impossible to store $x\in\mathbb{R}^N$ in a single computing device, let alone the matrix A. Similarly, consider a multi-agent network system where each agent is operated by a low cost computing device which has limited memory and computing capacities. Even when N is moderately large, it would be impractical for the entire A to be stored or computed on such a device. Hence, when m is relatively small, the proposed distributed schemes can be exploited in these two scenarios. See [4], [17] and the references therein for more applications of such schemes.

Distributed or decentralized algorithms for the LASSO and related problems, e.g., fused LASSO, basis pursuit (BP), and basis pursuit denoising (BPDN), have been extensively studied, including ADMM schemes, (sub-)gradient methods, and operator splitting schemes, e.g., [10], [15], [17], [20], [26], [30]. Particularly, the paper [17] develops row and column partition based distributed ADMM (D-ADMM) schemes for the BP that are convergent over a bipartite graph. The row partitioned LASSO and column partitioned BPDN are formulated as separable convex optimization and solved via D-ADMM [18]. Consensus based distributed schemes are developed for the row partitioned LASSO-like problems [16], e.g., the consensus ADMM (C-ADMM). An inexact C-ADMM (IC-ADMM) is established for distributed computation of the row partitioned LASSO and column partitioned logistic regression [4]. A proximal dual consensus ADMM (PDC-ADMM) scheme is used for solving column partition based LASSO under separable polyhedral constraints [2]. A decentralized gradient decent scheme is proposed for the regularized BP using column partition [30]. Besides, distributed proximal gradient schemes, e.g., PG-EXTRA, are exploited to solve the row partitioned LASSO [14], [26]. Other

Manuscript received November 8, 2020; revised March 12, 2021 and May 27, 2021; accepted May 30, 2021. Date of publication June 8, 2021; date of current version June 21, 2021. This material is based upon work supported by the National Science Foundation under Grants CMMI-1902006 and No. 2014816. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ji Liu. *(Corresponding author: Jinglai Shen.)*

Jinglai Shen and Eswar Kumar Hathibelagal Kammara are with the Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: shenj@umbc.edu; eswar1@umbc.edu).

Jianghai Hu is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: jianghai@purdue.edu).

Digital Object Identifier 10.1109/TSIPN.2021.3087110

relevant distributed schemes include [3], [9], [11], [20], just to name a few.

Most distributed schemes in the literature deal with row partition based LASSO and BPDN. Further, column partition based distributed LASSO schemes often require the knowledge of different column blocks of A, e.g., [20], and thus cannot be implemented in a fully parallel manner. Exceptions include distributed BP [17], [30] and distributed BPDN [18] using the dual approach, and dual consensus ADMM (DC-ADMM) [4] and PDC-ADMM [2] for the LASSO with possible polyhedral constraints. However, exact regularization of the BPDN used in [18] generally fails as shown in Section III, and the DC-ADMM does not guarantee the convergence of the primal variables [4, Theorem 2]. In addition, it is hard for the PDC-ADMM to handle the fused LASSO with more coupling and the BPDN-like problems with non-polyhedral constraints.

This paper develops column partition based distributed algorithms for a wide range of LASSO-like and BPDN-like problems with possible polyhedral constraints by exploiting convex optimization techniques, e.g., dual problems, exact regularization, and distributed optimization. Motivated by the dual approach for distributed BP and BPDN [17], [18], [30], we consider the Lagrangian dual problems of the LASSO/BPDN-like problems, which are separable or locally coupled and can be solved via column partition based distributed schemes. By using the solution properties of the LASSO/BPDN-like problems, we show that a primal solution is a solution to a BP-like problem depending on a dual solution. Under exact regularization conditions, a primal solution can be obtained from the dual of a regularized BP-like problem which can be solved by another column partition based distributed scheme. This yields two-stage, column partition based distributed schemes, and many existing distributed schemes can be used at each stage. The proposed schemes are applicable to a large class of generalized BP, LASSO and BPDN under the assumption that the network is static, bidirectional and connected.

Specifically, the following theoretical and numerical tasks are addressed in the paper:

1) *Exact Regularization*. We show that when the ℓ_1-norm is used, the BP-like problem subject to a polyhedral constraint is exactly regularized, whereas the LASSO and BPDN are not in general (cf. Section III). These results lay a ground for using regularized BP-like problems to recover a desired primal solution in the second stage and justify why regularized LASSO-like and BPDN-like problems are not considered.

2) *Dual Formulations*. We derive dual problems of the above mentioned primal problems, e.g., the regularized BP-like problem, LASSO-like problem, and BPDN-like problem. These dual formulations are used in both stages of the LASSO-like and BPDN-like problems: in the first stage, we use it to obtain a dual solution to the LASSO-like (resp. BPDN-like) problem; in the second stage, we use the dual of a regularized BP-like problem to recover a primal solution to the LASSO-like (resp. BPDN-like) problem. Further, we study the relation between a primal solution and a dual solution via duality theory (cf. Lemmas IV.2 and IV.3). Along with Proposition II.2, this relation yields a regularized BP-like problem in the second stage of the LASSO-like (resp. BPDN-like) problem. Besides, we develop various reduced dual problems which facilitate developing distributed schemes.

3) *Distributed Scheme Development*. We show that the obtained dual problems can be formulated as separable or locally coupled convex consensus optimization problems. For example, consider the fused LASSO and fused BPDN. Their dual problems and those of the corresponding regularized BP's are given by locally coupled consensus optimization such that a wide range of existing methods, e.g., operator splitting methods [5] and consensus ADMM [4], [16], can be used to develop columned partitioned based distributed schemes over undirected and connected networks. Numerical tests are conducted to evaluate performance of these schemes.

4) *Overall Convergence*. Many distributed algorithms can be used in each stage and are convergent under suitable conditions. However, the first-stage iterative scheme generates an approximate solution to a true dual solution, and this raises the question of whether using an approximate dual solution leads to significant discrepancy when solving the regularized BP-like problem in the second stage. Using sensitivity analysis tools for the regularized BP-like problem, we establish continuous dependence of its solution on certain parameters and prove the overall convergence of the two-stage distributed algorithms.

The ℓ_1-norm will be considered for many convex sparse optimization problems in the paper. Nevertheless, the dual formulations and duality results can be obtained for an arbitrary norm. This allows us to handle the group LASSO and its extensions. We will treat this general framework in Section IV.

The paper is organized as follows. Section II presents problem formulations and solution properties. Exact regularization is addressed in Section III. Section IV formulates dual problems and establishes properties in connection with the primal problems. Column partition based distributed schemes are developed in Section V, whose overall convergence is shown in Section VI. Numerical results are given in Section VII with conclusion in Section VIII. Due to the paper length limit, many proofs and technical developments are omitted and can be found in the online version [22] of the paper.

Notation. Let $A \in \mathbb{R}^{m\times N}$, and $R(A)$ denote the range of A. For any index set $\mathcal{S} \subseteq \{1, \ldots, N\}$, let $A_{\bullet \mathcal{S}}$ be the matrix formed by the columns of A indexed by elements of $\mathcal{S}$. Similarly, $A_{\alpha \bullet}$ is defined for an index set $\alpha \subseteq \{1, \ldots, m\}$. Let $\{\mathcal{I}_i\}_{i=1}^p$ form a disjoint union of $\{1, \ldots, N\}$, and $\{x_{\mathcal{I}_i}\}_{i=1}^p$ form a partition of $x \in \mathbb{R}^N$. For a closed convex set $\mathcal{C}$ in $\mathbb{R}^n$, $\Pi_\mathcal{C}$ denotes the Euclidean projection operator onto $\mathcal{C}$. For $u, v \in \mathbb{R}^n$, $u \perp v$ stands for the orthogonality of u and v, i.e., $u^T v = 0$. Further, $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the ℓ_1-norm (or 1-norm), 2-norm, and ∞-norm, respectively. Let $\text{prox}_f(\cdot)$ denote the proximal operator for a proper, lower semicontinuous convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

II. Problem Formulations and Solution Properties

We introduce a class of convex sparse minimization problems and their generalizations in this section.

• **Basis Pursuit (BP) and Extensions**. The BP is given by

$$\text{BP}: \quad \min_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad Ax = b, \tag{1}$$

where we assume $b \in R(A)$. A generalization of the BP (1) is $\min_{x\in\mathbb{R}^N} \|Ex\|_1$ subject to $Ax = b$, where $E \in \mathbb{R}^{r\times N}$.

• **LASSO and Extensions**. The standard LASSO is given by

$$\text{LASSO}: \quad \min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \lambda\,\|x\|_1, \tag{2}$$

where $\lambda > 0$ is the penalty parameter. A generalized LASSO is given by $\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \|Ex\|_1$, where $E \in \mathbb{R}^{r\times N}$ is a given matrix. It includes several extensions and variations:

(i) **Fused LASSO**: $\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \lambda_1\|x\|_1 + \lambda_2\|D_1x\|_1$, where $D_1 \in \mathbb{R}^{(N-1)\times N}$ denotes the first order difference matrix. Letting $E := \begin{bmatrix}\lambda_1 I_N \\ \lambda_2 D_1\end{bmatrix}$, the fused LASSO can be converted to the generalized LASSO.

(ii) Generalized total variation denoising with $E = \lambda D_1$, and ℓ_1-trend filtering with $E = \lambda D_2$ [12], where D_2 is the second order difference matrix.

Another extension is **group LASSO** widely used in statistics for model selection [31]: given $\lambda_i > 0$ for $i = 1,\ldots,p$,

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \sum_{i=1}^{p}\lambda_i\|x_{\mathcal{I}_i}\|_2. \tag{3}$$

• **Basis Pursuit Denoising (BPDN) and Extensions**. The standard BPDN is given by

$$\text{BPDN}: \quad \min_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax-b\|_2 \le \sigma, \tag{4}$$

where $\sigma > 0$. The BPDN has several generalizations and extensions. For example, it can be extended to $\min_{x\in\mathbb{R}^N} \|Ex\|_1$ subject to $\|Ax-b\|_2 \le \sigma$, where $E \in \mathbb{R}^{r\times N}$.

We summarize some fundamental solution properties of the aforementioned problems in a general setting. Let $E \in \mathbb{R}^{r\times N}$, $\|\cdot\|_\star$ be a norm on the Euclidean space, and $\mathcal{C}$ be a polyhedral set. Consider the following problems:

$$(P_1): \min_{x\in\mathbb{R}^N} \|Ex\|_\star \ \text{ s.t. } \ Ax = b,\ x\in\mathcal{C}.$$

$$(P_2): \min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \|Ex\|_\star \text{ s.t. } x\in\mathcal{C}.$$

$$(P_3): \min_{x\in\mathbb{R}^N} \|Ex\|_\star \text{ s.t. } \|Ax-b\|_2\le\sigma,\ x\in\mathcal{C}.$$

We call (P_1), (P_2) and (P_3) the BP-like, LASSO-like, and BPND-like problems, respectively. All the models introduced above can be formulated within this framework. For example, letting $\|x\|_G := \sum_{i=1}^p \lambda_i\|x_{\mathcal{I}_i}\|_2$ in the group LASSO, then we have $\|\cdot\|_\star = \|\cdot\|_G$.

Proposition II.1: [22, Proposition 2.1] Assume that the problems (P_1)–(P_3) are feasible. The following hold:

(i) Each of the problems (P_1)–(P_3) attains a minimizer;

(ii) Let $\mathcal{H}_2$ be the solution set of (P_2). Then $Ax = Ax'$ and $\|Ex\|_\star = \|Ex'\|_\star$ for all $x, x' \in \mathcal{H}_2$;

(iii) In (P_3), suppose $\|b\|_2 > \sigma$, $0\in\mathcal{C}$, and the optimal value is positive. Then each minimizer x_* of (P_3) satisfies $\|Ax_* - b\|_2 = \sigma$ and Ax is constant on the solution set.

A sufficient condition for the optimal value of (P_3) to be positive, along with the conditions that $\|b\|_2 > \sigma$ and $0\in\mathcal{C}$, is that E has full column rank. The following result will be used; see [32, Theorem 2.1] or [19, Proposition 3.2] for a proof.

Proposition II.2: [22, Proposition 2.2] The following hold:

(i) Let x_* be a minimizer of (P_2). Then z_* is a minimizer of (P_2) if and only if z_* is a minimizer of the BP-like problem (P_1') : $\min_{z\in\mathbb{R}^N} \|Ez\|_\star$ subject to $Az = Ax_*$ and $z\in\mathcal{C}$. Furthermore, the optimal value of (P_1') equals $\|Ex_*\|_\star$.

(ii) Let x_* be a minimizer of (P_3) which satisfies: $\|b\|_2 > \sigma$, $0\in\mathcal{C}$, and the optimal value of (P_3) is positive. Then z_* is a minimizer of (P_3) if and only if z_* is a minimizer of the BP-like problem (P_1) with $b := Ax_*$, and the optimal value of this (P_1) equals $\|Ex_*\|_\star$.

III. Exact Regularization

We briefly review the exact regularization of general convex programs [8]. Consider the convex minimization problem (P) and its regularized problem (P_ε) for some $\varepsilon \ge 0$:

$$(P): \ \min_{x\in\mathcal{P}} f(x); \qquad (P_\varepsilon): \ \min_{x\in\mathcal{P}} f(x) + \varepsilon h(x),$$

where $f, h : \mathbb{R}^N \to \mathbb{R}$ are convex functions, and $\mathcal{P}$ is a closed convex set. It is assumed that (P) has a solution, and h is coercive such that (P_ε) has a solution for each $\varepsilon > 0$. We call the problem (P) *exactly regularized* if there exists $\overline{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \overline{\varepsilon}]$, any solution to (P_ε) is a solution to (P). To establish the exact regularization, consider the following convex program (P_h) : $\min_{x\in\mathcal{P},\, f(x)\le f_*} h(x)$, where f_* is the optimal value of (P). Clearly, the constraint set of (P_h) is equivalent to $\{x \,|\, x\in\mathcal{P}, f(x) = f_*\}$, which is the solution set of (P). It is shown in [8, Theorem 2.1] or [8, Corollary 2.2] that (P) is exactly regularized by h if and only if (P_h) has a Lagrange multiplier $\mu_* \ge 0$, i.e., $\min_{x\in\mathcal{P},\, f(x)\le f_*} h(x) = \min_{x\in\mathcal{P}} h(x) + \mu_*(f(x) - f_*)$ for some constant $\mu_* \ge 0$.

A. Motivation and Illustration Via the Standard LASSO

To motivate exact regularization and related results, consider the standard LASSO (2) first. Although the primal problem (2) of the LASSO is densely coupled, its dual problem

$$(D): \quad \min_y \frac{\|y\|_2^2}{2} + b^T y, \quad \text{subject to} \quad \|A^T y\|_\infty \le \lambda$$

can be formulated as a separable distributed consensus optimization problem, for which column partition based distributed schemes can be developed. Let y_* be the unique dual solution. A critical question is how to recover a primal solution from y_*. A possible way is to consider the regularized LASSO:

$$\text{r-LASSO}: \quad \min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \lambda\,\|x\|_1 + \frac{\alpha}{2}\|x\|_2^2,$$

where $\frac{\alpha}{2}\|x\|_2^2$ is the regularization term with the regularization parameter $\alpha > 0$. The dual of r-LASSO enjoys favorable properties for column partition based distributed computation, and there is a one-to-one correspondence between its primal solution and its dual solution. However, a primal solution to r-LASSO is generally *not* a solution to the original LASSO (2) for any small

$\alpha > 0$ (cf. Example III.1). In other words, *exact regularization* of the LASSO fails in general.

Despite this negative result, it follows from duality theory (cf. Lemma IV.2) that any solution x_* to the LASSO (2) satisfies $Ax_* = b + y_*$, where y_* is the unique dual solution to (D) indicated above. Moreover, in view of Statement (i) of Proposition II.2, each solution to the following BP: $\min_{z\in\mathbb{R}^N} \ \|z\|_1$ subject to $Az = b + y_*$ is a solution to the LASSO. Since the above BP is exactly regularized [17], one can solve the following regularized BP

$$\text{r-BP}: \quad \min_{x\in\mathbb{R}^N} \ \|x\|_1 + \frac{\alpha}{2}\|x\|_2^2 \ \text{ subject to } \ Ax = b + y_*$$

for a small $\alpha > 0$ from its dual problem given below via a column partition based distributed scheme:

$$(D_{\text{r-BP}}): \quad \min_y \ \left((b + y_*)^T y + \frac{1}{2\alpha}\sum_{i=1}^{p} \|S\left(-(A^T y)_{\mathcal{I}_i}\right)\|_2^2 \right),$$

where $S(\cdot)$ is the soft thresholding operator (cf. Section IV-B). Letting $\widehat{y}_*$ be a dual solution to $(D_{\text{r-BP}})$, the primal solution to r-BP is recovered from $\widehat{y}_*$ as $x^*_{\mathcal{I}_i} = -\frac{1}{\alpha}S((A^T\widehat{y}_*)_{\mathcal{I}_i}), \ \forall\, i = 1, \ldots, p$, which is a solution to the LASSO. This yields a two-stage column partitioned based distributed scheme for the LASSO: its dual solution y_* is solved in the first stage, and a primal solution is obtained from the dual of a regularized BP in the second stage using y_* (cf. Algorithm 1 in Section V-A).

B. Exact Regularization of Convex Piecewise Affine Function Based Optimization

We consider exact regularization of convex piecewise affine (PA) functions based convex optimization with applications to ℓ_1-minimization. A real-valued continuous function $f : \mathbb{R}^N \to \mathbb{R}$ is PA if there exists a finite family of affine functions $\{f_i\}_{i=1}^{\ell}$ such that $f(x) \in \{f_i(x)\}_{i=1}^{\ell}$ for each $x \in \mathbb{R}^N$. Convex PA functions represent an important class of nonsmooth convex functions, e.g., the ℓ_1-norm $\|\cdot\|_1$, $f(x) := \|Ex\|_1$ for a matrix E, a polyhedral gauge, and the ℓ_∞-norm; see [19]. The following result shows exact regularization for convex PA objective functions on a polyhedral set [22], [29].

Proposition III.1: [22, Proposition 3.1] Let $\mathcal{P}$ be a polyhedral set, and $f : \mathbb{R}^N \to \mathbb{R}$ be a convex PA function such that the problem $(P) : \min_{x\in\mathcal{P}} f(x)$ has a nonempty solution set, and let $h : \mathbb{R}^N \to \mathbb{R}$ be a convex coercive regularization function. Then there exists $\overline{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \overline{\varepsilon}]$, any optimal solution to (P_ε) is an optimal solution to (P).

In view of the above proposition, we have the exact regularization for the BP-like problem when the ℓ_1-norm is used.

Corollary III.1: Let $\mathcal{C}$ be a polyhedral set. Then the following problem attains the exact regularization of (P_1) for all sufficiently small $\alpha > 0$:

$$(P_{1,\alpha}): \ \min_{x\in\mathbb{R}^N} \ \|Ex\|_1 + \frac{\alpha}{2}\|x\|_2^2 \ \text{ s.t. } Ax = b, \ \ x \in \mathcal{C}.$$

We then consider the LASSO-like and BPDN-like problems when the ℓ_1-norm is used. For simplicity, we focus on the standard problems (i.e., $\mathcal{C} = \mathbb{R}^N$) although the results developed here can be extended. By Proposition II.1, the solution sets of the standard LASSO and BPDN are polyhedral. Hence, the constraint sets of (P_h) of the LASSO and BPDN are polyhedral. However, unlike the BP-like problem, exact regularization fails in general. The first example shows that the standard LASSO (2) is *not* exactly regularized by $h(x) = \|x\|_2^2$.

Example III.1: Let $A = [I_2 \ I_2 \ \cdots \ I_2] \in \mathbb{R}^{2\times N}$ with $N = 2r$ for some $r \in \mathbb{N}$, and $b \in \mathbb{R}^2$ with $b = (b_1, b_2)^T > 0$, i.e., $b_1 > 0, b_2 > 0$. We partition $x \in \mathbb{R}^N$ into $x = (x^1, \ldots, x^r)$, where each $x^i \in \mathbb{R}^2$. When $0 < \lambda < 1$, it follows from the KKT condition: $0 \in A^T(Ax_* - b) + \lambda\partial\|x_*\|_1$ and a straightforward computation that a particular optimal solution x_* is $x^i_* = \frac{1-\lambda}{r}b > 0$ for all $i = 1, \ldots, r$. Hence, the solution set $\mathcal{H} = \{x = (x^1, \ldots, x^r) \mid \sum_{i=1}^r x^i = (1-\lambda)b, \ \|x\|_1 \le (1-\lambda)\|b\|_1\}$. Consider the regularized LASSO for $\alpha > 0$: $\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\,\|x\|_1 + \frac{\alpha}{2}\|x\|_2^2$. For any $\alpha > 0$, its unique optimal solution $x_{*,\alpha}$ is given by $x^i_{*,\alpha} = \frac{1-\lambda}{r+\alpha}b$ for each $i = 1, \ldots, r$. Hence, $x_{*,\alpha} \notin \mathcal{H}$ for any $\alpha > 0$.

The second example shows that in general, the standard BPDN (4) is *not* exactly regularized by $h(x) = \|x\|_2^2$.

Example III.2: Let $A = [D \ D \ \cdots \ D] \in \mathbb{R}^{2\times N}$ with $N = 2r$ for some $r \in \mathbb{N}$, where $D = \text{diag}(1, \beta) \in \mathbb{R}^{2\times 2}$ for a positive constant β. We partition $x \in \mathbb{R}^N$ into $x = (x^1, \ldots, x^r)$ with each $x^i \in \mathbb{R}^2$. Let $b = (b_1, b_2)^T \in \mathbb{R}^2$ and $\sigma = 1$. We assume that $b \ge \mathbf{1} = (1, 1)^T$, which is a necessary and sufficient condition for $\|v - b\|_2 \le 1 \Rightarrow v \ge 0$.

We first consider the convex minimization problem: $\min_{u\in\mathbb{R}^2} \|u\|_1$ subject to $\|Du - b\|_2 \le 1$, which has a unique minimizer u_* as D is invertible for any $\beta > 0$. Further, we must have $\|Du_* - b\|_2 = 1$ and $u_* > 0$. In light of this, the necessary and sufficient optimality conditions for u_* are: there exists $\lambda \in \mathbb{R}_+$ such that $\partial\|u_*\|_1 + \lambda D^T(Du_* - b) = 0$, and $\|Du_* - b\|_2^2 = 1$. Since $u_* > 0$, we have $\lambda > 0$ and the first equation becomes $\mathbf{1} + \lambda D^T(Du_* - b) = 0$, which further gives rise to $Du_* = b - \frac{1}{\lambda}D^{-1}\mathbf{1}$. Substituting it into the equation $\|Du_* - b\|_2 = 1$, we obtain $\lambda = \frac{\sqrt{1+\beta^2}}{\beta}$. This yields $u_* = (b_1 - \frac{1}{\lambda}, \frac{1}{\beta}(b_2 - \frac{1}{\beta\lambda}))^T > 0$. Hence, the solution set of the BPDN: $\min_{\|Ax-b\|_2\le 1} \|x\|_1$ (with $\|b\|_2 > 1$) is

$$\begin{aligned}\mathcal{H} &= \{x_* = (x^1_*, \ldots, x^r_*) \mid \|x_*\|_1 = \|u_*\|_1, \ Ax_* = Du_*\} \\ &= \{(x^1_*, \ldots, x^r_*) \mid \sum_{i=1}^r \|x^i_*\|_1 = \|u_*\|_1, \ \ \sum_{i=1}^r x^i_* = u_*\} \\ &= \{x_* = (x^1_*, \ldots, x^r_*) \mid x^i_* = \lambda_i u_*, \sum_{i=1}^r \lambda_i = 1, \lambda_i \ge 0, \forall\, i\}.\end{aligned}$$

Therefore, it is easy to show that the regularized BPDN with $h(x) = \|x\|_2^2$ has the unique minimizer $x_* = (x^i_*)$ with $x^i_* = \frac{u_*}{r}$ for each $i = 1, \ldots, r$. Since $u_* > 0$, we have $x_* > 0$. Since A has full row rank and $x_* > 0$, it follows from [22, Proposition 3.2] that $(BPDN_h) : \min_{\|Ax-b\|_2\le 1, \, \|x\|_1\le f_*} \|x\|_2^2$, where f_* is the positive optimal value of the BPDN, has a Lagrange multiplier if and only if there exist constants $\mu \ge 0$, $\gamma > 0$ such that $x_* + \mu\mathbf{1} + \gamma g = 0$ for the unique minimizer x_*, where $g = A^T(Ax_* - b) = A^T(Du_* - b) = -\frac{1}{\lambda}\mathbf{1}$, and $\lambda = \sqrt{1 + \frac{1}{\beta^2}}$. Since $x_* = \frac{1}{r}(u_*, \ldots, u_*)$, constants $\mu \ge 0$ and $\gamma >$

0 exist if and only if $(u_*)_1 = (u_*)_2$ or equivalently $\beta(b_1 - \frac{1}{\lambda}) = b_2 - \frac{1}{\beta\lambda}$. The latter is further equivalent to $b_2 = \beta b_1 + \frac{1-\beta^2}{\sqrt{1+\beta^2}}$. Hence, for any $\beta > 0$, $(BPDN_h)$ has a Lagrange multiplier if and only if b satisfies $b_2 = \beta b_1 + \frac{1-\beta^2}{\sqrt{1+\beta^2}}$ and $b \geq \mathbf{1}$. The set of such b's has zero measure in $\mathbb{R}^2$. For instance, when $\beta = 1$, $(BPDN_h)$ has a Lagrange multiplier if and only if $b = \theta \cdot \mathbf{1}$ for all $\theta \geq 1$. Thus the BPDN is *not* exactly regularized by $h(x) = \|x\|_2^2$ in general.

C. Exact Regularization of Grouped BP From Group LASSO

Motivated by the group LASSO (3), we investigate exact regularization of the following BP-like problem: $\min \sum_{i=1}^p \|x_{\mathcal{I}_i}\|_2$ subject to $Ax = b$, which is called *grouped basis pursuit* or simply grouped BP. We set $\lambda_i = 1$ for all i in the original group LASSO (3), without loss of generality. It is shown in [22, Example 3.3] that its exact regularization fails in general. Nonetheless, we present a sufficient condition below for exact regularization to hold; see [22] for its proof.

Lemma III.1: Consider a nonzero $b \in \mathbb{R}^m$ and a column partition $\{A_{\bullet\mathcal{I}_i}\}_{i=1}^p$ of a matrix $A \in \mathbb{R}^{m\times N}$. Suppose $A_{\bullet\mathcal{I}_1}$ is invertible, $A_{\bullet\mathcal{I}_1}^{-1}A_{\bullet\mathcal{I}_i}$ is an orthogonal matrix for each $i = 1, \ldots, s$, and $\|(A_{\bullet\mathcal{I}_i})^T(A_{\bullet\mathcal{I}_1})^{-T}A_{\bullet\mathcal{I}_1}^{-1}b\|_2 < \|A_{\bullet\mathcal{I}_1}^{-1}b\|_2$ for each $i = s+1, \ldots, p$. Then the exact regularization holds for the group BP.

IV. Dual Problems: Formulations and Properties

We develop dual problems of the regularized BP-like, LASSO-like and BPDN-like problems in this section. These dual problems and their properties lay a foundation for the development of column partition based distributed algorithms.

Consider the convex minimization problems (P_1)–(P_3) given in Section II, where $E \in \mathbb{R}^{r\times N}$ and $\|\cdot\|_\star$ is a general norm on $\mathbb{R}^r$. Let $\|\cdot\|_\diamond$ be the dual norm of $\|\cdot\|_\star$, i.e., $\|z\|_\diamond := \sup\{z^T v \,|\, \|v\|_\star \leq 1\}, \ \forall\, z \in \mathbb{R}^r$. As an example, the dual norm of the ℓ_1-norm is the ℓ_∞-norm. For $\|x\|_G := \sum_{i=1}^p \|x_{\mathcal{I}_i}\|_2$ arising from the group LASSO, its dual norm is $\|z\|_{G,\diamond} = \max_{i=1,\ldots,p}\|z_{\mathcal{I}_i}\|_2$. Clearly, $\|x\|_\star = \sup\{x^T v \,|\, \|v\|_\diamond \leq 1\}, \forall\, x \in \mathbb{R}^r$, and the subdifferential of $\|\cdot\|_\star$ at $x = 0$ is $B_\diamond(0,1)$, where $B_\diamond(0,1) := \{v \,|\, \|v\|_\diamond \leq 1\}$.

A. Dual Problems: General Formulations

Strong duality will be exploited for the abovementioned problems and their dual problems. For this purpose, the following minimax result is needed; see Appendix for a proof.

Lemma IV.1: Consider the convex program (P): $\inf_{z\in\mathcal{P}, Az=b, Cz\leq d} J(z)$, where $J(z) := \|Ez\|_\star + f(z)$, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $\mathcal{P} \subseteq \mathbb{R}^n$ is a polyhedral set, A, C, E are matrices, and b, d are vectors. Suppose that (P) is feasible and has a finite infimum. Then $\inf_{z\in\mathcal{P}}(\sup_{y,\,\mu\geq 0,\,\|v\|_\diamond\leq 1}[(Ez)^T v + f(z) + y^T(Az - b) + \mu^T(Cz - d)]) = \sup_{y,\,\mu\geq 0,\,\|v\|_\diamond\leq 1}(\inf_{z\in\mathcal{P}}[(Ez)^T v + f(z) + y^T(Az - b) + \mu^T(Cz - d)])$.

In what follows, we consider a general polyhedral set of the following form unless otherwise stated

$$\mathcal{C} := \{x \in \mathbb{R}^N \,|\, Cx \leq d\}, \quad C \in \mathbb{R}^{\ell\times N}, \quad d \in \mathbb{R}^\ell. \tag{5}$$

As before, $\{\mathcal{I}_i\}_{i=1}^p$ is a disjoint union of $\{1, \ldots, N\}$.

• **Dual Problem of the Regularized BP-like Problem** Consider the regularized BP-like problem for a constant $\alpha > 0$:

$$\min_{Ax=b,\, x\in\mathcal{C}} \|Ex\|_\star + \frac{\alpha}{2}\|x\|_2^2, \tag{6}$$

where $b \in R(A) \cap A\mathcal{C}$ with $A\mathcal{C} := \{Ax \,|\, x \in \mathcal{C}\}$. Let $\mu \in \mathbb{R}_+^\ell$ be the Lagrange multiplier for the polyhedral constraint $Cx \leq d$. It follows from Lemma IV.1 with $z = x$ and $\mathcal{P} = \mathbb{R}^N$ that the dual problem is

$$\min_{y,\mu\geq 0,\|v\|_\diamond\leq 1} b^T y + d^T\mu + \frac{1}{2\alpha}\sum_{i=1}^p \left\|(A^T y + E^T v + C^T\mu)_{\mathcal{I}_i}\right\|_2^2. \tag{7}$$

Let $(y_*, \mu_*, v_*) \in \mathbb{R}^m \times \mathbb{R}_+^\ell \times B_\diamond(0,1)$ be an optimal solution to the dual problem; its existence is shown in the proof of Lemma IV.1. Consider the Lagrangian $L(x, y, \mu, v) := (Ex)^T v + \frac{\alpha}{2}\|x\|_2^2 + y^T(Ax - b) + \mu^T(Cx - d)$. Then by the strong duality given in Lemma IV.1, we see from $\nabla_x L(x_*, y_*, \mu_*, v_*) = 0$ that the unique primal solution is $x_{\mathcal{I}_i}^* = -\frac{1}{\alpha}(A^T y_* + E^T v_* + C^T\mu_*)_{\mathcal{I}_i}$ for $i = 1, \ldots, p$.

• **Dual Problem of the LASSO-like Problem** Consider the LASSO-like problem for $A \in \mathbb{R}^{m\times N}, b \in \mathbb{R}^m, E \in \mathbb{R}^{r\times N}$:

$$\min_{x\in\mathcal{C}} \frac{1}{2}\|Ax - b\|_2^2 + \|Ex\|_\star. \tag{8}$$

By Lemma IV.1 with $z = (x, u)$ and $\mathcal{P} = \mathbb{R}^N \times \mathbb{R}^m$, we obtain the dual problem:

$$\min_{y,\mu\geq 0,\|v\|_\diamond\leq 1} \left\{ \frac{\|y\|_2^2}{2} + b^T y + d^T\mu : (A^T y + E^T v + C^T\mu)_{\mathcal{I}_i} = 0, \forall i \right\}. \tag{9}$$

By Lemma IV.1, the dual problem attains an optimal solution $(y_*, \mu_*, v_*) \in \mathbb{R}^m \times \mathbb{R}_+^\ell \times B_\diamond(0,1)$. Since the objective function of (9) is strictly convex in y and convex in (μ, v), y_* is unique but (μ_*, v_*) may not.

The following lemma establishes a connection between a primal solution and a dual solution, which is critical to distributed algorithm development; its proof is given in Appendix.

Lemma IV.2: Let (y_*, μ_*, v_*) be an optimal solution to the dual problem (9). Then for any optimal solution x_* to the primal problem (8), $Ax_* - b = y_*$. Further, if $\mathcal{C}$ is a polyhedral cone (i.e., $d = 0$), then $\|Ex_*\|_\star = -(b + y_*)^T y_*$.

• **Dual Problem of the BPDN-like Problem** Consider the BPDN-like problem with $\sigma > 0$:

$$\min_{x\in\mathcal{C},\,\|Ax-b\|_2\leq\sigma} \|Ex\|_\star = \inf_{x\in\mathcal{C},\,u=Ax-b,\,\|u\|_2\leq\sigma} \|Ex\|_\star, \tag{10}$$

where we assume that the problem is feasible and has a positive optimal value, $\|b\|_2 > \sigma$, and the polyhedral set $\mathcal{C}$ satisfies $0 \in \mathcal{C}$. Note that $0 \in \mathcal{C}$ holds if and only if $d \geq 0$.

To establish the strong duality, we also assume that there is an $\widetilde{x}$ in the relative interior of $\mathcal{C}$ (denoted by $\mathrm{ri}(\mathcal{C})$)

such that $\|A\widetilde{x} - b\|_2 < \sigma$ or equivalently, by [21, Theorem 6.6], there exists $\widetilde{u} \in A(\mathrm{ri}(\mathcal{C})) - b$ such that $\|\widetilde{u}\|_2 < \sigma$. A sufficient condition for this assumption to hold is that $b \in A(\mathrm{ri}(\mathcal{C}))$. Under this assumption, it follows from [21, Theorem 28.2] that there exist $y_* \in \mathbb{R}^m$, $\mu_* \geq 0$, and $\lambda_* \geq 0$ such that $\inf_{x\in\mathcal{C},\, u=Ax-b, \|u\|_2\leq\sigma} \|Ex\|_\star = \inf_{x\in\mathcal{C}, u} \|Ex\|_\star + y_*^T(Ax-b-u) + \lambda_*(\|u\|_2^2 - \sigma^2) + \mu_*^T(Cx-d)$. By the similar argument for Lemma IV.1, we obtain the dual problem:

$$\min_{y,\mu,v} \ \{b^T y + \sigma\|y\|_2 + d^T\mu : (A^T y + E^T v + C^T\mu)_{\mathcal{I}_i} = 0,$$
$$\mu \geq 0, \ \|v\|_\diamond \leq 1, \ \forall\, i = 1,\ldots,p\}. \tag{11}$$

Moreover, the dual problem attains an optimal solution $(y_*, \mu_*, v_*) \in \mathbb{R}^m \times \mathbb{R}^\ell_+ \times B_\diamond(0,1)$ along with $\lambda_* \geq 0$. The following lemma establishes certain solution properties of the dual problem and a connection between primal and dual solutions, which is crucial to distributed algorithm development; see Appendix for its proof. Particularly, it shows that the y-part of a dual solution is unique when $\mathcal{C}$ is a polyhedral cone.

Lemma IV.3: Consider the BPDN (10), where $\|b\|_2 > \sigma$, $0 \in \mathcal{C}$, and its optimal value is positive. Assume that the strong duality holds. The following hold:

(i) Let (y_*, μ_*, v_*) be a dual solution to (11). Then $y_* \neq 0$, and for any solution x_* to (10), $Ax_* - b = \frac{\sigma y_*}{\|y_*\|_2}$. Further, if $\mathcal{C}$ is a polyhedral cone (i.e., $d = 0$), then $\|Ex_*\|_\star = -b^T y_* - \sigma\|y_*\|_2$.

(ii) Suppose $d = 0$. Let (y_*, μ_*, v_*) and (y'_*, μ'_*, v'_*) be two arbitrary solutions of (11). Then $y_* = y'_*$.

Remark IV.1: The above dual formulations for a general polyhedral set $\mathcal{C}$ are useful for distributed computation when $\ell \ll N$, even if $C \in \mathbb{R}^{\ell\times N}$ is a dense matrix; see Section V. When both N and ℓ are large, e.g., $\mathcal{C} = \mathbb{R}^N_+$, decoupling properties of $\mathcal{C}$ are preferred. In particular, consider the following polyhedral set of certain decoupling structure:

$$\mathcal{C} := \{x = (x_{\mathcal{I}_i})_{i=1}^p \in \mathbb{R}^N \,|\, C_{\mathcal{L}_i\mathcal{I}_i} x_{\mathcal{I}_i} \leq d_{\mathcal{L}_i}, \forall\, i\}, \tag{12}$$

where $(\mathcal{L}_i)_{i=1}^p$ is a disjoint union of $\{1,\ldots,\ell\}$, $C_{\mathcal{L}_i\mathcal{I}_i} \in \mathbb{R}^{\ell_i\times|\mathcal{I}_i|}$ and $d_{\mathcal{L}_i} \in \mathbb{R}^{\ell_i}$ with $\ell_i = |\mathcal{L}_i|$ for each $i = 1,\ldots,p$. Also, let $\mu = (\mu_{\mathcal{L}_i})_{i=1}^p$ with $\mu_{\mathcal{L}_i} \in \mathbb{R}^{\ell_i}_+$ be the Lagrange multiplier for $\mathcal{C}$. The dual problems in (7), (9), and (11) can be extended to the above $\mathcal{C}$ by replacing $\mu^T d$ with $\sum_{i=1}^p \mu_{\mathcal{L}_i}^T d_{\mathcal{L}_i}$ and $(A^T y + E^T v + C^T\mu)_{\mathcal{I}_i}$ with $(A^T y + E^T v)_{\mathcal{I}_i} + (C_{\mathcal{L}_i\mathcal{I}_i})^T\mu_{\mathcal{L}_i}$, respectively. Lemmas IV.2 and IV.3 also hold for a primal solution x_* and a dual solution y_*.

Remark IV.2: Consider the box constraint set $\mathcal{C} := [l_1, u_1] \times \cdots \times [l_N, u_N]$ with $0 \in \mathcal{C}$, where $-\infty \leq l_i < u_i \leq +\infty$ with $l_i \leq 0 \leq u_i$ for each $i = 1,\ldots,N$. We may write $\mathcal{C} = \{x \in \mathbb{R}^N \,|\, \mathbf{l} \leq x \leq \mathbf{u}\}$, where $\mathbf{l} := (l_1,\ldots,l_N)^T$ and $\mathbf{u} := (u_1,\ldots,u_N)^T$. For any i, define the function $\theta_i : \mathbb{R} \to \mathbb{R}$

$$\theta_i(t) := t^2 - (t - \Pi_{[l_i,u_i]}(t))^2, \ \ \forall\, t \in \mathbb{R}. \tag{13}$$

Hence, θ_i is C^1 and convex [6, Theorem 1.5.5, Exercise 2.9.13], and θ_i is increasing on $\mathbb{R}_+$ and decreasing on $\mathbb{R}_-$, and its minimal value on $\mathbb{R}$ is zero. When $\mathcal{C} = \mathbb{R}^N$, $\theta_i(s) = s^2, \forall\, i$; when $\mathcal{C} = \mathbb{R}^N_+$, $\theta_i(s) = (s_+)^2, \forall\, i$. The dual problems for such $\mathcal{C}$ can be reduced by removing the dual variable μ. For example, the dual of the regularized BP-like problem (6) becomes $\min_{y,\|v\|_\diamond\leq 1} b^T y + \frac{\alpha}{2}\sum_{i=1}^N \theta_i(-\frac{1}{\alpha}(A^T y + E^T v)_i)$. More examples and details can be found in [22, Section 4.1].

These dual problems can be reduced for specific norms or polyhedral constraints shown in the following subsections.

B. Applications to the ℓ_1-Norm Based Problems

Let $\|\cdot\|_\star$ be the ℓ_1-norm; its dual norm is the ℓ_∞-norm. As before, $\mathcal{C}$ is a general polyhedral set given by (5). In this case, the dual variable v can be removed via the soft thresholding operator $S : \mathbb{R} \to \mathbb{R}$, i.e., $S(t) = t - \mathrm{sgn}(t)$ if $|t| \geq 1$; $S(t) = 0$ if $|t| \leq 1$. For a vector $v = (v_1,\ldots,v_k)^T \in \mathbb{R}^k$, let $S(v) := (S(v_1),\ldots,S(v_k))^T \in \mathbb{R}^k$ (cf. [22, Section 4.2]).

• **Reduced Dual Problem of the Regularized BP-like Problem** Consider two cases as follows:

Case (a): $E = I_N$. The dual problem (7) reduces to

$$\min_{y,\,\mu\geq 0} \left(b^T y + \mu^T d + \frac{1}{2\alpha}\sum_{i=1}^p \|S\left(-(A^T y + C^T\mu)_{\mathcal{I}_i}\right)\|_2^2 \right). \tag{14}$$

Letting (y_*, μ_*) be a dual solution, the unique primal solution

$$x^*_{\mathcal{I}_i} = -\frac{1}{\alpha} S\left((A^T y_* + C^T\mu_*)_{\mathcal{I}_i}\right), \quad \forall\, i = 1,\ldots,p. \tag{15}$$

When $\mathcal{C}$ is a box constraint with $0 \in \mathcal{C}$, the dual problem is

$$\min_{y\in\mathbb{R}^m} \left[b^T y + \frac{\alpha}{2}\sum_{i=1}^N \theta_i \circ \left(-\frac{1}{\alpha} S\left((A^T y)_i\right)\right)\right]. \tag{16}$$

Letting y_* be a dual solution, the unique primal solution is $x^*_i = \max\{l_i, \, \min(-\frac{1}{\alpha}S((A^T y_*)_i), \, u_i)\}$ for each i.

Case (b): $E = \begin{bmatrix} I_N \\ F \end{bmatrix}$ for some matrix $F \in \mathbb{R}^{k\times N}$. Such an E appears in the ℓ_1 penalty of the fused LASSO or BPDN (cf. Remark V.7). Let $v = (v,' \widetilde{v})$. The dual problem is

$$\min_{y,\mu\geq 0, \|\widetilde{v}\|_\infty\leq 1} b^T y + \mu^T d$$
$$+ \frac{1}{2\alpha}\sum_{i=1}^p \left\|S\left(-(A^T y + F^T\widetilde{v} + C^T\mu)_{\mathcal{I}_i}\right)\right\|_2^2. \tag{17}$$

Letting $(y_*, \mu_*, \widetilde{v}_*)$ be a dual solution, the primal solution is

$$x^*_{\mathcal{I}_i} = -\frac{1}{\alpha} S\left((A^T y_* + F^T\widetilde{v}_* + C^T\mu_*)_{\mathcal{I}_i}\right), \forall\, i. \tag{18}$$

When $\mathcal{C}$ is a box constraint with $0 \in \mathcal{C}$, the dual problem is $\min_{y,\|\widetilde{v}\|_\infty\leq 1} b^T y + \frac{\alpha}{2}\sum_{i=1}^N \theta_i \circ (-\frac{1}{\alpha}S((A^T y + F^T\widetilde{v})_i))$, and the unique primal solution is: for each $i = 1,\ldots,N$, $x^*_i = \max\{l_i, \min(-\frac{1}{\alpha}S((A^T y_* + F^T\widetilde{v}_*)_i), u_i)\}$.

• **Reduced Dual Problem of the LASSO-like Problem** Consider the following cases:

Case (a): $E = \lambda I_N$ with $\lambda > 0$. Its dual problem becomes

$$\min_{y,\mu\geq 0} \left\{ \frac{\|y\|_2^2}{2} + b^T y + d^T\mu : \|(A^T y + C^T\mu)_{\mathcal{I}_i}\|_\infty \leq \lambda, \forall\, i \right\}. \tag{19}$$

Case (b): $E = \begin{bmatrix} \lambda I_N \\ F \end{bmatrix}$ for some $F \in \mathbb{R}^{k\times N}$ and $\lambda > 0$. Such an E appears in the fused LASSO. The dual problem (9) reduces to

$$\min_{y,\mu,\widetilde{v}} \Big\{ \frac{\|y\|_2^2}{2} + b^T y + d^T \mu \,:\, \mu \geq 0,\ \|\widetilde{v}\|_\infty \leq 1,$$
$$\|(A^T y + F^T \widetilde{v} + C^T \mu)_{\mathcal{I}_i}\|_\infty \leq \lambda,\ i = 1, \ldots, p \Big\}. \quad (20)$$

• **Reduced Dual Problem of the BPDN-like Problem** Consider the following cases under the similar assumptions given below (10) in Section IV-A:

Case (a): $E = I_N$. The equivalent dual problem (11) is

$$\min_{y,\,\mu\geq 0} \big\{ b^T y + \sigma \|y\|_2 + d^T \mu : \|(A^T y + C^T \mu)_{\mathcal{I}_i}\|_\infty \leq 1,\ \forall i \big\}. \quad (21)$$

Case (b): $E = \begin{bmatrix} I_N \\ F \end{bmatrix}$ for some $F \in \mathbb{R}^{k\times N}$. The equivalent dual problem (11) reduces to

$$\min_{y,\mu,\widetilde{v}} \big\{ b^T y + \sigma\|y\|_2 + d^T \mu \,:\, \mu \geq 0,\ \|\widetilde{v}\|_\infty \leq 1,$$
$$\|(A^T y + F^T \widetilde{v} + C^T \mu)_{\mathcal{I}_i}\|_\infty \leq 1, i = 1, \ldots, p \big\}. \quad (22)$$

C. Applications to Problems Associated With the Norm $\|\cdot\|_G$ From Group LASSO

Consider the norm $\|x\|_G := \sum_{i=1}^p \|x_{\mathcal{I}_i}\|_2$ from the group LASSO; its dual norm $\|x\|_{G,\diamond} = \max_{i=1,\ldots,p} \|x_{\mathcal{I}_i}\|_2$.

• **Reduced Dual Problem of the Regularized BP-like Problem under** $\|\cdot\|_G$ Consider $E = I_N$ as follows.

Case (a): $\mathcal{C}$ is given by (5). Given a vector w, we see that $\min_{\|v\|_{G,\diamond}\leq 1} \sum_{i=1}^p \big\|(v-w)_{\mathcal{I}_i}\big\|_2^2 = \sum_{i=1}^p \min_{\|v_{\mathcal{I}_i}\|_2\leq 1} \big\|v_{\mathcal{I}_i} - w_{\mathcal{I}_i}\big\|_2^2$. Let $S_{\|\cdot\|_2}(z) := (1 - \frac{1}{\|z\|_2})_+ z, \forall\, z \in \mathbb{R}^n$ denote the soft thresholding operator with respect to the ℓ_2-norm, and let $B_2(0,1) := \{z \,|\, \|z\|_2 \leq 1\}$. It is known that given w, $z_* := \Pi_{B_2(0,1)}(w) = w - S_{\|\cdot\|_2}(w)$ and $\|z_* - w\|_2^2 = \|S_{\|\cdot\|_2}(w)\|_2^2 = [(\|w\|_2 - 1)_+]^2$. Applying these results to (7), we obtain the reduced dual problem

$$\min_{y,\,\mu\geq 0} \left(b^T y + \mu^T d + \frac{1}{2\alpha} \sum_{i=1}^p \Big[\big(\big\|(A^T y + C^T \mu)_{\mathcal{I}_i}\big\|_2 - 1\big)_+ \Big]^2 \right). \quad (23)$$

Letting (y_*, μ_*) be a dual solution, the primal solution is

$$x^*_{\mathcal{I}_i} = -\frac{1}{\alpha} S_{\|\cdot\|_2}\left((A^T y_* + C^T \mu_*)_{\mathcal{I}_i} \right), \quad \forall\, i = 1, \ldots, p. \quad (24)$$

The above results can be easily extended to the decoupled polyhedral constraint set given by (12).

Case (b): $\mathcal{C}$ is a box constraint with $0 \in \mathcal{C}$. In this case, the dual variable μ can be removed. By the results at the end of Section IV-A, the reduced dual problem is

$$\min_{y,\,(v_{\mathcal{I}_i})_{i=1}^p} \sum_{i=1}^p \left[\frac{b^T y}{p} + \frac{\alpha}{2} \sum_{j\in\mathcal{I}_i} \theta_j \left(-\frac{1}{\alpha} \left((A_{\bullet\mathcal{I}_i})^T y + v_{\mathcal{I}_i} \right)_j \right) \right], \quad (25)$$

subject to $\|v_{\mathcal{I}_i}\|_2 \leq 1$ for $i = 1, \ldots, p$, where θ_j's are defined in (13). Given a dual solution (y_*, v_*), the primal solution $x^*_{\mathcal{I}_i} = \max(\mathbf{l}_{\mathcal{I}_i}, \min(-\frac{(A_{\bullet\mathcal{I}_i})^T y_* + (v_*)_{\mathcal{I}_i}}{\alpha}, \mathbf{u}_{\mathcal{I}_i}))$ for $i = 1, \ldots, p$. When $\mathcal{C}$ is a cone, the dual can be further reduced by removing v. For example, when $\mathcal{C} = \mathbb{R}^N$, the dual problem becomes $\min_{y\in\mathbb{R}^m} (b^T y + \frac{1}{2\alpha}\sum_{i=1}^p [(\|(A_{\bullet\mathcal{I}_i})^T y\|_2 - 1)_+]^2)$, and the primal solution $x^*_{\mathcal{I}_i} = -\frac{1}{\alpha} S_{\|\cdot\|_2}((A_{\bullet\mathcal{I}_i})^T y_*)$, $i = 1, \ldots, p$ for a dual solution y_*. Similar results for $\mathcal{C} = \mathbb{R}^N_+$ are given in [22, Section 4.3].

• **Reduced Dual Problem of the LASSO-like Problem under** $\|\cdot\|_G$ Let $E = \lambda I_N$ for $\lambda > 0$. Then (9) becomes

$$\min_{y,\mu\geq 0} \left\{ \frac{\|y\|_2^2}{2} + b^T y + d^T \mu : \|(A^T y + C^T \mu)_{\mathcal{I}_i}\|_2 \leq \lambda, \forall i \right\}. \quad (26)$$

If $\mathcal{C} = \mathbb{R}^N$, it becomes $\min_y (b^T y + \frac{\|y\|_2^2}{2})$ subject to $\|(A_{\bullet\mathcal{I}_i})^T y\|_2 \leq \lambda, i = 1, \ldots, p$. If $\mathcal{C} = \mathbb{R}^N_+$, it is $\min_y (b^T y + \frac{\|y\|_2^2}{2})$ subject to $(A^T y)_{\mathcal{I}_i} \in B_2(0,\lambda) + \mathbb{R}^{|\mathcal{I}_i|}_+, i = 1, \ldots, p$.

• **Reduced Dual Problem of the BPDN-like Problem under** $\|\cdot\|_G$ Let $E = I_N$. Suppose the similar assumptions indicated in Section IV-A hold. The dual problem (11) reduces to

$$\min_{y,\mu\geq 0} \left\{ b^T y + \sigma\|y\|_2 + d^T \mu : \|(A^T y + C^T \mu)_{\mathcal{I}_i}\|_2 \leq 1, \forall i \right\}. \quad (27)$$

When $\mathcal{C} = \mathbb{R}^N$, the dual problem becomes $\min_y (b^T y + \sigma \|y\|_2)$ subject to $\|(A_{\bullet\mathcal{I}_i})^T y\|_2 \leq 1, i = 1, \ldots, p$.

V. Development of Column Partition Based Distributed Algorithms

In this section, we develop column partition based distributed schemes for the LASSO-like problem (8) and the BPDN like problem (10), which include a board class of convex sparse optimization problems as special cases. As a by-product, column partition based distributed schemes are also developed for the regularized BP-like problem (6).

Consider a network of p agents modeled by a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, p\}$ is the set of agents, and $\mathcal{E}$ denotes the set of edges, each of which connects two agents in $\mathcal{V}$. For each $i \in \mathcal{V}$, $\mathcal{N}_i$ denotes the set of neighbors of agent i. The following assumptions are made throughout this section:

A.1 The graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is undirected and connected;

A.2 The matrix $A \in \mathbb{R}^{m\times N}$ attains a column partition $\{A_{\bullet\mathcal{I}_i}\}_{i=1}^p$, where $\{\mathcal{I}_1, \ldots, \mathcal{I}_p\}$ is a disjoint union of $\{1, \ldots, N\}$. Each agent i knows $A_{\bullet\mathcal{I}_i}$, b, and possibly other information but does not know $A_{\bullet\mathcal{I}_j}$'s with $j \neq i$.

We also consider a general polyhedral set $\mathcal{C}$ given by (5) satisfying $\ell \ll N$ or having decoupling structure given by (12), e.g., the box constraints.

A. Structure of Column Partition Based Distributed Schemes

We first present a general structure of the proposed column partition based distributed schemes for the LASSO/BPDN-like problems. Recall that these problems are densely coupled but not exactly regularized in general (cf. Section III-B). However, it follows from Proposition II.2 that if Ax_* is known, where x_* is a minimizer of the LASSO/BPDN-like problem, then an exact primal solution can be solved via the dual of a regularized BP-like problem using column partition of A, under

Algorithm 1: Two-stage Distributed Algorithm for LASSO-like or BPDN-like Problem: General Structure.

1: Initialization

2: Stage 1 Compute a dual solution y_* to the LASSO-like problem (8) or BPDN-like problem (10) using a column partition based distributed scheme;

3: Stage 2 Solve the dual of the following regularized BP-like problem for a sufficiently small $\alpha > 0$ using y_* and a column partition based distributed scheme:

$$\text{r-BP}_{\text{LASSO}}: \min_{Ax=b+y_*,\ x\in\mathcal{C}} \|Ex\|_\star + \frac{\alpha}{2}\|x\|_2^2, \quad (28)$$

or

$$\text{r-BP}_{\text{BPDN}}: \min_{Ax=b+\frac{\sigma y_*}{\|y_*\|_2},\ x\in\mathcal{C}} \|Ex\|_\star + \frac{\alpha}{2}\|x\|_2^2, \quad (29)$$

4: Output: obtain the subvector $x^*_{\mathcal{I}_i}$ from a dual solution to (28) or (29) for each $i = 1, \ldots, p$

exact regularization. To find Ax_*, it follows from Lemmas IV.2 and IV.3 that $Ax_* = b + y_*$ (resp. $Ax_* = b + \frac{\sigma y_*}{\|y_*\|_2}$), where y_* is a dual solution to the LASSO (resp. BPDN)-like problem. Since the dual of the LASSO/BPDN-like problem can be solved distributively using column partition of A, this yields column partition based two-stage distributed schemes summarized in Algorithm 1. See Section III-A for more illustration.

The dual problems used in each stage of Algorithm 1 have been derived in Section IV. We will show that these dual problems can be formulated as separable or locally coupled convex optimization problems to which a wide range of existing distributed schemes can be applied. For the purpose of illustration, we consider operator splitting method based schemes including Douglas-Rachford (D-R) algorithm and its variations [5], [10], consensus ADMM (C-ADMM) schemes [16], and inexact C-ADMM (IC-ADMM) schemes [4]. Specific distributed schemes in each stage are given in the next subsections; see Table I for a summary of dual problems and related schemes. It should be noted that it is *not* our goal to improve the performance of the existing schemes or seek the most efficient existing scheme but rather to demonstrate their applicability to the obtained dual problems. Many other distributed schemes can be exploited under weaker assumptions.

Remark V.1: Before ending this subsection, we discuss a variation of the BP formulation in the second stage for an important special case by exploiting solution properties of (8) and (10). Consider $E = \lambda I_N$ with $\lambda > 0$ and $\mathcal{C}$ is a polyhedral cone (i.e., $d = 0$), and let y_* be the unique dual solution to (8) or (10). For (8), by Lemma IV.2 and $E = \lambda I_N$, we have $Ax_* = b + y_*$ and $\lambda\|x_*\|_1 = -y_*^T(b + y_*)$ for any minimizer x_* of (8), noting that $\|x_*\|_1$ is constant on the solution set by Proposition II.1. Suppose $x_* \neq 0$ or equivalently $b + y_* \neq 0$. Then $\|x_*\|_1 = -\frac{1}{\lambda}y_*^T(y_* + b)$, and $\frac{Ax_*}{\|x_*\|_1} = -\frac{\lambda(y_*+b)}{y_*^T(y_*+b)}$. Consider the scaled regularized BP (or scaled r-BP) for $\alpha > 0$:

$$\min_{z\in\mathbb{R}^N} \|z\|_1 + \frac{\alpha}{2}\|z\|_2^2 \ \text{ s.t. } Az = -\frac{\lambda(y_* + b)}{y_*^T(y_* + b)},\ z \in \mathcal{C}. \quad (30)$$

Once the unique minimizer z_* of the above r-BP is obtained (satisfying $\|z_*\|_1 = 1$), the least 2-norm minimizer x_* of the LASSO-like problem is given by $x_* = -\frac{1}{\lambda}y_*^T(y_* + b)z_*$. Similarly, for (10), by Lemma IV.3 and the assumption that the optimal value of (10) is positive, we have $-b^Ty_* - \sigma\|y_*\|_2 > 0$. Hence, x_* can be solved from the following scaled r-BP:

$$\min_{z\in\mathbb{R}^N} \|z\|_1 + \frac{\alpha}{2}\|z\|_2^2 \ \text{ s.t. } Az = -\frac{b + \frac{\sigma y_*}{\|y_*\|_2}}{b^Ty_* + \sigma\|y_*\|_2},\ x \in \mathcal{C}. \quad (31)$$

Once the unique minimizer z_* is obtained (satisfying $\|z_*\|_1 = 1$), the least 2-norm minimizer x_* of the BPDN-like problem is given by $x_* = -(b^Ty_* + \sigma\|y_*\|_2)z_*$.

The advantages of using the scaled r-BP (30) or (31) are two folds. First, since $\|x_*\|_1$ may be small or near zero in some applications, a direct application of the r-BP$_{\text{LASSO}}$ or r-BP$_{\text{BPDN}}$ using y_* in Algorithm 1 may be sensitive to round-off errors. Using the scaled r-BP (30) or (31) can avoid such a problem. Second, the suitable value of α achieving exact regularization is often unknown. A simple rule for choosing such an α is [13]: $\alpha \leq \frac{1}{10\|\widehat{x}\|_\infty}$, where $\widehat{x} \neq 0$ is a sparse vector to be recovered. An estimate of the upper bound of α is $\frac{1}{10\|\widehat{x}\|_1}$ in view of $\|\widehat{x}\|_1 \geq \|\widehat{x}\|_\infty$. When the scaled r-BP (30) or (31) is used, we can simply choose $\alpha \leq \frac{1}{10}$ as $\|z_*\|_1 = 1$.

B. Column Partition Based Distributed Schemes for the Standard LASSO-Like Problem

Consider the standard LASSO-like problem, i.e., the LASSO-like problem (8) with $E = \lambda I_N$ for a constant $\lambda > 0$, $\|\cdot\|_\star = \|\cdot\|_1$, and a general polyhedral set $\mathcal{C}$ given by (5).

Stage One. We solve the dual problem (19), i.e.,

$$\min_{y,\mu\geq 0} \left\{ \frac{\|y\|_2^2}{2} + b^Ty + d^T\mu : \|(A^Ty + C^T\mu)_{\mathcal{I}_i}\|_\infty \leq \lambda, \forall i \right\}.$$

Consider $\ell \ll N$ first. Let $\mathbf{y} := (\mathbf{y}_1, \ldots, \mathbf{y}_p) \in \mathbb{R}^{mp}$ and $\boldsymbol{\mu} := (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_p) \in \mathbb{R}^{\ell p}$. Define the consensus subspace $\mathcal{A}_{\mathbf{y}}$ and the consensus cone $\mathcal{A}_{\boldsymbol{\mu}}$:

$$\mathcal{A}_{\mathbf{y}} := \{\mathbf{y} \,|\, \mathbf{y}_i = \mathbf{y}_j,\ \forall\,(i,j) \in \mathcal{E}\}, \quad (32)$$

$$\mathcal{A}_{\boldsymbol{\mu}} := \{\boldsymbol{\mu} \geq 0 \,|\, \boldsymbol{\mu}_i = \boldsymbol{\mu}_j,\ \forall\,(i,j) \in \mathcal{E}\}. \quad (33)$$

Hence, the dual problem (19) is equivalent to the consensus convex optimization problem:

$$\min_{(\mathbf{y},\boldsymbol{\mu})\in\mathcal{A}_{\mathbf{y}}\times\mathcal{A}_{\boldsymbol{\mu}}} \sum_{i=1}^p J_i(\mathbf{y}_i, \boldsymbol{\mu}_i), \text{ s.t. } (\mathbf{y}_i, \boldsymbol{\mu}_i) \in \mathcal{W}_i,\ \forall i, \quad (34)$$

where for each $i = 1, \ldots, p$, the function

$$J_i(\mathbf{y}_i, \boldsymbol{\mu}_i) := \frac{1}{p}\left(\frac{\|\mathbf{y}_i\|_2^2}{2} + b^T\mathbf{y}_i + d^T\boldsymbol{\mu}_i\right), \quad (35)$$

and the set $\mathcal{W}_i := \{(\mathbf{y}_i, \boldsymbol{\mu}_i) \,|\, \|(A_{\bullet\mathcal{I}_i})^T\mathbf{y}_i + (C_{\bullet\mathcal{I}_i})^T\boldsymbol{\mu}_i\|_\infty \leq \lambda\}$. To present distributed schemes for (34), the following notation is used: let $\mathbf{w} := (\mathbf{y}, \boldsymbol{\mu}) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p}$ and $\mathbf{w}_i := (\mathbf{y}_i, \boldsymbol{\mu}_i) \in \mathbb{R}^m \times \mathbb{R}^\ell$ for each i. Let $\mathbf{w}^k := (\mathbf{y}^k, \boldsymbol{\mu}^k) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p}$, $\mathbf{z}^k = (\mathbf{z}^k_{\mathbf{y}}, \mathbf{z}^k_{\boldsymbol{\mu}}) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p}$, $\mathbf{w}^k_i = (\mathbf{y}^k_i, \boldsymbol{\mu}^k_i) \in \mathbb{R}^m \times \mathbb{R}^\ell$, and $\mathbf{z}^k_i = ((\mathbf{z}^k_{\mathbf{y}})_i, (\mathbf{z}^k_{\boldsymbol{\mu}})_i) \in \mathbb{R}^m \times \mathbb{R}^\ell$ for each $i = 1, \ldots, p$.

TABLE I
DUAL PROBLEMS AND CORRESPONDING DISTRIBUTED SCHEMES. STANDARD PROBLEMS: $E = \lambda I_N$; FUSED PROBLEMS: $E = [\lambda I_n \ F^T]^T$

Stage	S1 of Standard Problems	S1 of Fused Problems	S2 of Standard Problems	S2 of Fused Problems
Dual	Eqn.(19), (21), (26), (27)	Eqn.(20), (22)	Eqn.(14), (23), (25)	Eqn.(17)
Scheme	Algorithms 2-4	Algorithm 6	Algorithm 5, Eqn.(40)-(41)	Algorithm 7

Algorithm 2: Distributed Averaging Based Operator Splitting Scheme [5] for Solving (34).

1: Initialization with suitable constants $\eta > 0$ and $\varrho > 0$
2: **repeat**
3: Compute $\widetilde{\mathbf{w}}^{k+1} = (\overline{\mathbf{z}_{\mathbf{y}}^k}, (\overline{\mathbf{z}_{\boldsymbol{\mu}}^k})_+)$ via a distributed averaging scheme
4: $\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \varrho[\Pi_{\mathcal{W}_i}(2\widetilde{\mathbf{w}}_i^{k+1} - \mathbf{z}_i^k - \eta\nabla J_i((\widetilde{\mathbf{w}}^{k+1})_i)) - \widetilde{\mathbf{w}}_i^{k+1}]$ for each $i = 1, \dots, p$
5: $k \leftarrow k+1$
6: **until** Stopping criterion is met
7: Output: obtain a dual solution $y_* = (\overline{\mathbf{z}_{\mathbf{y}}^k})_i$ for each i.

Algorithm 3: Distributed C-ADMM Scheme [4] for (34).

1: Initialization with a suitable constant $\eta > 0$
2: **repeat**
3: $\mathbf{p}_i^{k+1} = \mathbf{p}_i^k + \eta \sum_{j\in\mathcal{N}_i}(\mathbf{w}_i^k - \mathbf{w}_j^k)$ for $i = 1, \dots, p$
4: $\mathbf{w}_i^{k+1} = \arg\min_{\mathbf{w}_i = (\mathbf{y}_i, \boldsymbol{\mu}_i)} J_i(\mathbf{w}_i) + \mathbf{w}_i^T \mathbf{p}_i^{k+1} + \eta \sum_{j\in\mathcal{N}_i} \left\|\mathbf{w}_i - \frac{\mathbf{w}_i^k + \mathbf{w}_j^k}{2}\right\|_2^2$ subject to $\mathbf{w}_i = (\mathbf{y}_i, \boldsymbol{\mu}_i) \in \mathcal{W}_i,\ \boldsymbol{\mu}_i \geq 0$ for each $i = 1, \dots, p$
5: $k \leftarrow k+1$
6: **until** Stopping criterion is met
7: Output: obtain a dual solution $y_* = \mathbf{y}_i^k$ for each i.

1.1) *Distributed Averaging based Operator Splitting Scheme*. Given $\mathbf{y} = (\mathbf{y}_i)_{i=1}^p$, let $\overline{\mathbf{y}} := \mathbf{1} \otimes [\frac{1}{p}\sum_{i=1}^p \mathbf{y}_i]$ denote the averaging of $\mathbf{y}$. Similarly, $\overline{\boldsymbol{\mu}}$ denotes the averaging of $\boldsymbol{\mu} = (\boldsymbol{\mu}_i)_{i=1}^p$. Such averaging can be computed via a distributed averaging scheme, e.g., [28]. A distributed averaging based operator splitting scheme [5] is shown in Algorithm 2.

1.2) *Consensus-ADMM (C-ADMM) Scheme*. A distributed consensus-ADMM scheme [4] is given in Algorithm 3.

1.3) *Local Averaging based Douglas-Rachford (D-R) Scheme*. Recall $\mathbf{w}_i = (\mathbf{y}_i, \boldsymbol{\mu}_i)$ for each i. For each $i = 1, \dots, p$, define $\mathbf{x}_i := (\mathbf{w}_i, (\mathbf{w}_{ij})_{j\in\mathcal{N}_i})$ with $\mathbf{w}_{ij}$ denoting local copies of $\mathbf{w}_j$'s for agent i [10], the set $\mathcal{A}_i^{\mathrm{L}} := \{\mathbf{x}_i \,|\, \boldsymbol{\mu}_i \geq 0,\ \mathbf{w}_i = \mathbf{w}_{ij},\ \forall j \in \mathcal{N}_i\}$, and $\widehat{J}_i(\mathbf{x}_i) := J_i(\mathbf{w}_i) + \boldsymbol{\delta}_{\mathcal{A}_i^{\mathrm{L}}}(\mathbf{x}_i) + \boldsymbol{\delta}_{\mathcal{W}_i}(\mathbf{w}_i)$, where $\boldsymbol{\delta}$ denotes the indicator function. Similarly, define $\mathbf{u}_i := (\mathbf{w}'_i, (\mathbf{w}'_{ij})_{j\in\mathcal{N}_i})$ for each i. Given $\mathbf{u} = (\mathbf{u}_i)_{i=1}^p$, we also define $\overline{\mathbf{u}}_i^{\mathrm{LA}} = (\overline{\mathbf{w}'}_i^{\mathrm{LA}}, (\overline{\mathbf{w}'}_{ij}^{\mathrm{LA}})_{j\in\mathcal{N}_i})$, where $\overline{\mathbf{w}'}_i^{\mathrm{LA}} = \overline{\mathbf{w}'}_{ji}^{\mathrm{LA}} = \frac{1}{|\mathcal{N}_i|+1}(\mathbf{w}'_i + \sum_{s\in\mathcal{N}_i}\mathbf{w}'_{si})$ for all $j \in \mathcal{N}_i$ denotes the local averaging [10]. This leads to Algorithm 4.

Remark V.2: The values of the parameters for convergence of the above schemes are as follows: (i) $0 < \eta < \frac{2}{L}$ and $0 < \varrho < 2 - \frac{L\eta}{2}$ for Algorithm 2, where $L > 0$ is the Lipschtiz constant of $\sum_{I=1}^p \nabla J_i$ [5]; (ii) $\eta > 0$ for the C-ADMM scheme (Algorithm 3) [16]; and (iii) $\eta \in (0,1)$ and $\rho > 0$ for the local averaging based D-R scheme (Algorithm 4). Similar values can be found for the subsequent schemes.

Algorithm 4: Local Averaging Based Douglas-Rachford (D-R) Scheme [10] for Solving (34).

1: Initialization with suitable constants $\eta \in (0,1)$ and $\rho > 0$
2: **repeat**
3: $\mathbf{x}_i^{k+1} = \overline{\mathbf{u}^k}_i^{\mathrm{LA}}, \forall\, i = 1, \dots, p$
4: $\mathbf{u}_i^{k+1} = \mathbf{u}_i^k + 2\eta(\mathrm{prox}_{\rho\widehat{J}_i}(2\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) - \mathbf{x}_i^{k+1}), \forall\, i$
5: $k \leftarrow k+1$
6: **until** Stopping criterion is met
7: Output: obtain a dual solution $y_* = (\overline{\mathbf{u}^k})_i$ for each i.

Remark V.3: Consider a large ℓ and $\mathcal{C} \in \mathbb{R}^{\ell\times N}$ given by (12), i.e. $\mathcal{C} := \{x = (x_{\mathcal{I}_i})_{i=1}^p \in \mathbb{R}^N \,|\, C_{\mathcal{L}_i\mathcal{I}_i} x_{\mathcal{I}_i} \leq d_{\mathcal{L}_i}, \forall\, i\}$. Recall that $\mu := (\mu_{\mathcal{L}_i})_{i=1}^p \in \mathbb{R}^\ell$ with $\mu_{\mathcal{L}_i} \in \mathbb{R}^{\ell_i}$. By Remark IV.1 and Section IV-B, the dual problem (19) is equivalent to the consensus convex optimization problem:

$$\min_{(\mathbf{y},\mu)\in\mathcal{A}_{\mathbf{y}}\times\mathbb{R}_+^\ell} \sum_{i=1}^p F_i(\mathbf{y}_i, \mu_{\mathcal{L}_i}), \text{ s.t. } (\mathbf{y}_i, \mu_{\mathcal{L}_i}) \in \mathcal{U}_i,\ \forall\, i, \quad (36)$$

where, fore each $i = 1, \dots, p$, the function

$$F_i(\mathbf{y}_i, \mu_{\mathcal{L}_i}) := \frac{1}{p}\left(\frac{\|\mathbf{y}_i\|_2^2}{2} + b^T\mathbf{y}_i\right) + d_{\mathcal{L}_i}^T \mu_{\mathcal{L}_i}, \quad (37)$$

and $\mathcal{U}_i := \{(\mathbf{y}_i, \mu_{\mathcal{L}_i}) \,|\, \|(A_{\bullet\mathcal{I}_i})^T\mathbf{y}_i + (C_{\mathcal{L}_i\mathcal{I}_i})^T\mu_{\mathcal{L}_i}\|_\infty \leq \lambda\}$. Several distributed schemes can be developed for solving (36). For example, a distributed averaging based operator splitting scheme similar to Algorithm 2 can be used. Specifically, let $\mathbf{w} := (\mathbf{y}, \mu)$, $\mathbf{w}_i := (\mathbf{y}_i, \mu_{\mathcal{L}_i})$, and $\mathbf{z} = (\mathbf{z}_{\mathbf{y}}, \mathbf{z}_\mu)$. Replacing $\widetilde{\mathbf{w}}^{k+1} = (\overline{\mathbf{z}_{\mathbf{y}}^k}, (\overline{\mathbf{z}_{\boldsymbol{\mu}}^k})_+)$ in Line 3 by $\widetilde{\mathbf{w}}^{k+1} = (\overline{\mathbf{z}_{\mathbf{y}}^k}, (\mathbf{z}_\mu^k)_+)$ and $\mathcal{W}_i$ and J_i in Line 4 by $\mathcal{U}_i$ and F_i respectively, Algorithm 2 is applicable whose output is given by a dual solution $y_* = (\overline{\mathbf{z}_{\mathbf{y}}^k})_i, \forall\, i$. A C-ADMM scheme similar to Algorithm 3 can also be developed, noting that for any fixed $\mathbf{y}_i$, $h_i(\mathbf{y}) := \min_{\mu_{\mathcal{L}_i}\geq 0} \sum_{i=1}^p F_i(\mathbf{y}_i, \mu_{\mathcal{L}_i})$ subject to $(\mathbf{y}_i, \mu_{\mathcal{L}_i}) \in \mathcal{U}_i$ is a real-valued convex function, and the sub-gradient of h_i exists.

Stage Two. The 2nd stage is defined by the regularized BP-like problem (6) with the regularization parameter $\alpha > 0$ and b replaced by $b + y_*$. Further, $\lambda = 1$ such that $E = I_N$. When $\|\cdot\|_\star$ is the ℓ_1-norm, Corollary III.1 shows that exact regularization holds, i.e., the regularized problem attains a solution to the original BP-like problem for all small $\alpha > 0$.

Consider $\ell \ll N$ first. Using the consensus subspace $\mathcal{A}_{\mathbf{y}}$ in (32) and the consensus cone $\mathcal{A}_{\boldsymbol{\mu}}$ in (33), the reduced dual

Algorithm 5: Distributed Averaging Based D-R Scheme [10] for Solving (38).

1: Initialization with suitable constants $\eta \in (0,1)$ and $\rho > 0$
2: **repeat**
3: Compute $\mathbf{w}^{k+1} = (\overline{\mathbf{z}^k_{\mathbf{y}}}, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+)$ via a distributed averaging scheme
4: $\mathbf{z}_i^{k+1} =$ $\mathbf{z}_i^k + 2\eta(\operatorname{prox}_{\rho G_i}(2\mathbf{w}_i^{k+1} - \mathbf{z}_i^k) - \mathbf{w}_i^{k+1}), \forall i$
5: $k \leftarrow k+1$
6: **until** Stopping criterion is met
7: Output: obtain $(y_*, \mu_*) = ((\overline{\mathbf{z}^k_{\mathbf{y}}})_i, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+)_i)$ for each i.

problem (14) becomes the consensus convex optimization:

$$\min_{(\mathbf{y},\boldsymbol{\mu}) \in \mathcal{A}_{\mathbf{y}} \times \mathcal{A}_{\boldsymbol{\mu}}} \sum_{i=1}^p G_i(\mathbf{y}_i, \boldsymbol{\mu}_i), \tag{38}$$

where for each $i = 1, \ldots, p$, the function

$$G_i(\mathbf{y}_i, \boldsymbol{\mu}_i) := \frac{1}{p}\left(b^T \mathbf{y}_i + d^T \boldsymbol{\mu}_i\right) + \frac{1}{2\alpha} \left\| S\left(-(A_{\bullet \mathcal{I}_i})^T \mathbf{y}_i - (C_{\bullet \mathcal{I}_i})^T \boldsymbol{\mu}_i\right) \right\|_2^2. \tag{39}$$

The same notation introduced below (35) is used, e.g., $\mathbf{w}, \mathbf{w}_i, \mathbf{w}^k, \mathbf{w}_i^k, \mathbf{z}^k, \mathbf{z}_i^k$ for each $i = 1, \ldots, p$. Further, given $\mathbf{y} = (\mathbf{y}_i)_{i=1}^p$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_i)_{i=1}^p$, let $\overline{\mathbf{y}}$ and $\overline{\boldsymbol{\mu}}$ be the averaging of $\mathbf{y}$ and $\boldsymbol{\mu}$, respectively.

We present a distributed averaging based D-R scheme first.

An alternative algorithm for solving (38) is a distributed C-ADMM scheme [4] for a suitable constant $\eta > 0$ given below.

$$\mathbf{p}_i^{k+1} = \mathbf{p}_i^k + \eta \sum_{j \in \mathcal{N}_i} (\mathbf{y}_i^k - \mathbf{y}_j^k), \ \forall\, i = 1, \ldots, p \tag{40a}$$

$$\mathbf{w}_i^{k+1} = \operatorname*{arg\,min}_{\mathbf{w}_i = (\mathbf{y}_i, \boldsymbol{\mu}_i)} \ G_i(\mathbf{w}_i) + \mathbf{w}_i^T \mathbf{p}_i^{k+1} + \eta \sum_{j \in \mathcal{N}_i} \left\| \mathbf{w}_i - \frac{\mathbf{w}_i^k + \mathbf{w}_j^k}{2} \right\|_2^2 \text{ s.t. } \boldsymbol{\mu}_i \geq 0, \forall\, i \tag{40b}$$

The output is a dual solution $(y_*, \mu_*) = (\mathbf{w}^k)_i$ for each i.

In the above two algorithms, once a dual solution (y_*, μ_*) is found, it follows from (15) that the primal solution x^* is given by $x^*_{\mathcal{I}_i} = -\frac{1}{\alpha} S((A_{\bullet \mathcal{I}_i})^T y_* + (C_{\bullet \mathcal{I}_i})^T \mu_*)$ for $i = 1, \ldots, p$.

Remark V.4: Since the function $G_i(\cdot)$ given by (39) involves the soft thresholding operator S, it may be difficult to solve the subproblem in (40b). In practice, we formulate this subproblem as: $\mathbf{w}_i^{k+1} = (\mathbf{y}_i^*, \boldsymbol{\mu}_i^*)$, where $\mathbf{w}_i = (\mathbf{y}_i, \boldsymbol{\mu}_i)$ and $(\mathbf{y}_i^*, \boldsymbol{\mu}_i^*, v^*_{\mathcal{I}_i}) = \arg\min_{(\mathbf{w}_i, v_{\mathcal{I}_i})} \frac{1}{p}(b^T \mathbf{y}_i + d^T \boldsymbol{\mu}_i) + \frac{1}{2\alpha} \left\| (A_{\bullet \mathcal{I}_i})^T \mathbf{y}_i + (C_{\bullet \mathcal{I}_i})^T \boldsymbol{\mu}_i + v_{\mathcal{I}_i} \right\|_2^2 + \mathbf{w}_i^T \mathbf{p}_i^{k+1} + \eta \sum_{j \in \mathcal{N}_i} \left\| \mathbf{w}_i - \frac{\mathbf{w}_i^k + \mathbf{w}_j^k}{2} \right\|_2^2$ subject to $\boldsymbol{\mu}_i \geq 0$ and $\|v_{\mathcal{I}_i}\|_\infty \leq 1$, for each i. This new subproblem can be efficiently solved via a quadratic program. Besides, the subproblem in Line 4 of Algorithm 5 can be solved in a similar way.

Remark V.5: Another scheme for solving (38) is the distributed averaging based operator splitting scheme [5]:

$$\mathbf{w}^{k+1} = (\overline{\mathbf{z}^k_{\mathbf{y}}}, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+), \tag{41a}$$

$$\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \varrho \left[\mathbf{w}_i^{k+1} - \mathbf{z}_i^k - \eta \nabla G_i\left(\mathbf{w}_i^{k+1}\right)\right], \ \forall\, i, \tag{41b}$$

where (41a) is solved via distributed averaging, and ∇G_i is easy to compute. When $\mathcal{C}$ is a box constraint with $0 \in \mathcal{C}$, the dual problem can be reduced to (16) depending on y only, and a distributed scheme similar to (41) can be developed. A drawback of (41) is that the Lipschitz constant of $\sum_i \nabla G_i$ is given by $(\|A\|_F^2 + \|C\|_F^2)/\alpha$, which is large for a large N. This yields a small $\eta > 0$ and thus slow convergence. Nonetheless, the scheme (41) can be used for a small or moderate N.

We then consider a large ℓ with $\mathcal{C}$ given by (12). It follows from Remark IV.1 and Section IV-B that the equivalent dual problem is given by: recalling that $\mu := (\mu_{\mathcal{L}_i})_{i=1}^p \in \mathbb{R}^\ell$,

$$\min_{(\mathbf{y},\mu) \in \mathcal{A}_{\mathbf{y}} \times \mathbb{R}^\ell_+} \sum_{i=1}^p \widetilde{F}_i(\mathbf{y}_i, \mu_{\mathcal{L}_i}),$$

where $\widetilde{F}_i(\mathbf{y}_i, \mu_{\mathcal{L}_i}) := \frac{1}{2\alpha} \left\| S(-(A_{\bullet \mathcal{I}_i})^T \mathbf{y}_i - (C_{\mathcal{L}_i \mathcal{I}_i})^T \mu_{\mathcal{L}_i}) \right\|_2^2 + \frac{1}{p}(b^T \mathbf{y}_i) + d^T_{\mathcal{L}_i} \mu_{\mathcal{L}_i}$. Let $\mathbf{w} := (\mathbf{y}, \mu)$ and $\mathbf{w}_i := (\mathbf{y}_i, \mu_{\mathcal{L}_i})$, and $\mathbf{z} = (\mathbf{z}_{\mathbf{y}}, \mathbf{z}_\mu)$. Distributed schemes similar to Algorithms 5 or (40) can be developed by replacing G_i with $\widetilde{F}_i$.

C. Column Partition Based Distributed Schemes for the Standard BDPN-Like Problem

Consider the standard BPDN-like problem, i.e., the BPDN-like problem (10) with $E = I_N$, $\|\cdot\|_\star = \|\cdot\|_1$, and a polyhedral set $\mathcal{C}$ given by (5). Suppose the assumptions given below (10) in Section IV-A hold. Consider the dual problem (11). As shown in Lemma IV.3, a dual solution $y_* \neq 0$. Hence, the function $\|y\|_2$ is differentiable near y_*.

Stage One. Consider $\ell \ll N$ first. In light of (21), it is easy to verify that the distributed Algorithms 2-4 can be applied by replacing the functions J_i in (35) with

$$\widehat{J}_i(\mathbf{y}_i, \boldsymbol{\mu}_i) := \frac{1}{p}\left(\sigma \|\mathbf{y}_i\|_2 + b^T \mathbf{y}_i + d^T \boldsymbol{\mu}_i\right), \quad \forall\, i = 1, \ldots, p,$$

and setting $\lambda = 1$ in $\mathcal{W}_i$. Moreover, an inexact C-ADMM scheme [4] can be applied; its details are omitted. When ℓ is large and $\mathcal{C}$ is given by (12), letting $\mu := (\mu_{\mathcal{L}_i})_{i=1}^p \in \mathbb{R}^\ell$, the schemes discussed in Remark V.3 can be used by replacing F_i's in (37) by $\widehat{F}_i(\mathbf{y}_i, \mu_{\mathcal{L}_i}) := \frac{1}{p}(\sigma \|\mathbf{y}_i\|_2 + b^T \mathbf{y}_i) + d^T_{\mathcal{L}_i} \mu_{\mathcal{L}_i}$ and setting $\lambda = 1$ in $\mathcal{U}_i$. When $\mathcal{C} = \mathbb{R}^N$ or $\mathcal{C} = \mathbb{R}^N_+$, μ or $\boldsymbol{\mu}$ can be removed; see the discussions below (21).

Stage Two. The 2nd stage is defined by the regularized BP-like problem (6) with the regularization parameter $\alpha > 0$, b replaced by $b + \frac{\sigma y_*}{\|y_*\|_2}$, and $E = I_N$. Hence, all the results for the 2nd stage of the standard LASSO-like problem given in Section V-B apply.

D. Column Partition Based Distributed Schemes for the Fused LASSO-Like and Fused BDPN-Like Problems

Through this subsection, let $\| \cdot \|_\star$ be the ℓ_1-norm, $D_1 \in \mathbb{R}^{(N-1)\times N}$ be the first order difference matrix, and we assume in addition that the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ satisfies $(i, i+1) \in \mathcal{E}, \forall\, i = 1, \ldots, p-1$. Consider the fused LASSO-like problem first, i.e., the LASSO-like problem (8) with $E = \begin{bmatrix} \lambda I_N \\ \gamma D_1 \end{bmatrix}$ for positive constants λ and γ and a general polyhedral set $\mathcal{C}$ as before.

Stage One. Consider $\ell \ll N$ first. To solve the dual problem (20) with $F = \gamma D_1$ and $\widetilde{v} = (\widetilde{v}_1, \ldots, \widetilde{v}_{N-1}) \in \mathbb{R}^{N-1}$, define $n_s := \sum_{i=1}^{s} |\mathcal{I}_i|$ for $s = 1, \ldots, p$. Without loss of generality, let $\mathcal{I}_1 = \{1, \ldots, n_1\}$, and $\mathcal{I}_{i+1} = \{n_i + 1, \ldots, n_i + |\mathcal{I}_{i+1}|\}$ for each $i = 1, \ldots, p-1$. Define the index sets $\mathcal{S}_1 := \mathcal{I}_1$, $\mathcal{S}_i := \{n_{i-1}\} \cup \mathcal{I}_i$ for $i = 2, \ldots, p-1$, and $\mathcal{S}_p := \{n_{p-1}, \ldots, N-1\}$. Define $r_i := |\mathcal{S}_i|$ and $\mathbf{v}_i := \widetilde{v}_{\mathcal{S}_i}$ for each $i = 1, \ldots, p$. Thus for $i = 1, \ldots, p-1$, $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ overlap on one variable $\widetilde{v}_{n_i}$. Let $\mathbf{v} := (\mathbf{v}_i)_{i=1}^p \in \mathbb{R}^{N+p-2}$, and the local coupling constraint $\mathcal{A}_{LC} := \{\mathbf{v} \in \mathbb{R}^{N+p-2} \,|\, (\mathbf{v}_i)_{r_i} = (\mathbf{v}_{i+1})_1, \ \forall\, i = 1, \ldots, p-1\}$. For each $i = 1, \ldots, p$, define the function

$$H_i(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) \;:=\; \frac{1}{p}\left(\frac{\|\mathbf{y}_i\|_2^2}{2} + b^T \mathbf{y}_i + d^T \boldsymbol{\mu}_i \right), \qquad (42)$$

and the set $\mathcal{V}_i := \{(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) \,|\, \|\mathbf{v}_i\|_\infty \leq 1, \ \ \|(A_{\bullet\mathcal{I}_i})^T \mathbf{y}_i + (C_{\bullet\mathcal{I}_i})^T \boldsymbol{\mu}_i + \gamma[(D_1)_{\mathcal{S}_i\mathcal{I}_i}]^T \mathbf{v}_i\|_\infty \leq \lambda\,\}$. The dual problem (20) is formulated as the locally coupled convex program:

$$\min_{(\mathbf{y},\boldsymbol{\mu},\mathbf{v}) \in \mathcal{A}_{\mathbf{y}} \times \mathcal{A}_{\boldsymbol{\mu}} \times \mathcal{A}_{LC}} \sum_{i=1}^{p} H_i(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i), \qquad (43)$$

subject to $(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) \in \mathcal{V}_i, \ \forall\, i = 1, \ldots, p$.

Let $\mathbf{z}^k = (\mathbf{z}^k_{\mathbf{y}}, \mathbf{z}^k_{\boldsymbol{\mu}}, \mathbf{z}^k_{\mathbf{v}}) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p} \times \mathbb{R}^{N+p-2}$, and η, ρ are suitable positive constants depending on the Lipschitz constant of $\sum_{i=1}^{p} \nabla H_i$; see [5, Thoerem 1] for details. For any $\mathbf{v} = (\mathbf{v}_i)_{i=1}^p \in \mathbb{R}^{N+p-2}$ defined above, $\widetilde{\mathbf{v}} = (\widetilde{\mathbf{v}}_i)_{i=1}^p := \Pi_{\mathcal{A}_{LC}}(\mathbf{v})$ is $(\widetilde{\mathbf{v}}_i)_{r_i} = (\widetilde{\mathbf{v}}_{i+1})_1 = \frac{1}{2}[(\mathbf{v}_i)_{r_i} + (\mathbf{v}_{i+1})_1]$ for $i = 1, \ldots, p-1$, and for each i, $(\widetilde{\mathbf{v}}_i)_j = (\mathbf{v}_i)_j$ for the other indices j. Clearly, this local averaging can be computed distributively. A distributed averaging based operator splitting scheme is given in Algorithm 6. A local averaging based operating splitting scheme can be developed in a similar way. These schemes can be extended to a large ℓ with $\mathcal{C}$ given by (12) and be extended to the generalized total variation denoising or ℓ_1-trend filtering with $E = \lambda D_1$ or $E = \lambda D_2$.

Stage Two. The 2nd stage is given by the regularized BP-like problem (6) with the parameter $\alpha > 0$, b replaced by $b + y_*$, and $E = \begin{bmatrix} I_N \\ \gamma D_1 \end{bmatrix}$ for a constant $\gamma > 0$ after scaling. Consider a general polyhedral set $\mathcal{C}$ with $\ell \ll N$ first.

We follow the same notation used in the first stage. Hence, the reduce dual problem (17) can be formulated as the following locally coupled convex program:

$$\min_{(\mathbf{y},\boldsymbol{\mu},\mathbf{v}) \in \mathcal{A}_{\mathbf{y}} \times \mathcal{A}_{\boldsymbol{\mu}} \times \mathcal{A}_{LC}} \sum_{i=1}^{p} P_i(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i), \qquad (44)$$

Algorithm 6: Distributed Averaging Based Operator Splitting Scheme [5] for Solving (43).

1: Initialization with suitable constants $\eta > 0$ and $\varrho > 0$
2: **repeat**
3: Compute $\widehat{\mathbf{w}}^{k+1} = (\overline{\mathbf{z}^k_{\mathbf{y}}}, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+, \widetilde{\mathbf{z}^k_{\mathbf{v}}})$, where $\overline{\mathbf{z}^k_{\mathbf{y}}}$ and $\overline{\mathbf{z}^k_{\boldsymbol{\mu}}}$ are solved via a distributed averaging scheme, and $\widetilde{\mathbf{z}^k_{\mathbf{v}}} = ((\widetilde{\mathbf{z}^k_{\mathbf{v}}})_i)_{i=1}^p$ with $(\widetilde{\mathbf{z}^k_{\mathbf{v}}})_i$ computed distributively
4: $\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \varrho[\Pi_{\mathcal{V}_i}(2\widehat{\mathbf{w}}_i^{k+1} - \mathbf{z}_i^k - \eta \nabla H_i(\widehat{\mathbf{w}}_i^{k+1})) - \widehat{\mathbf{w}}_i^{k+1}]$ for each $i = 1, \ldots, p$
5: $k \leftarrow k + 1$
6: **until** Stopping criterion is met
7: Output: obtain a dual solution $y_* = (\overline{\mathbf{z}^k_{\mathbf{y}}})_i$ for each i.

Algorithm 7: Distributed Averaging Based D-R Scheme [10] for (44).

1: Initialization with suitable constants $\eta > 0$ and $\rho > 0$
2: **repeat**
3: Compute $\widehat{\mathbf{w}}^{k+1} = (\overline{\mathbf{z}^k_{\mathbf{y}}}, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+, \widetilde{\mathbf{z}^k_{\mathbf{v}}})$, where $\overline{\mathbf{z}^k_{\mathbf{y}}}$ and $\overline{\mathbf{z}^k_{\boldsymbol{\mu}}}$ are solved via a distributed averaging scheme, and $\widetilde{\mathbf{z}^k_{\mathbf{v}}} = ((\widetilde{\mathbf{z}^k_{\mathbf{v}}})_i)_{i=1}^p$ with $(\widetilde{\mathbf{z}^k_{\mathbf{v}}})_i$ computed distributively
4: $\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + 2\eta(\operatorname{prox}_{\rho P_i}(2\widehat{\mathbf{w}}_i^{k+1} - \mathbf{z}_i^k) - \widehat{\mathbf{w}}_i^{k+1}), \forall\, i$
5: $k \leftarrow k + 1$
6: **until** Stopping criterion is met
7: Output: obtain a dual solution $(y_*, \mu_*, \widetilde{v}_*)$ from $(\overline{\mathbf{z}^k_{\mathbf{y}}}, (\overline{\mathbf{z}^k_{\boldsymbol{\mu}}})_+, \widetilde{\mathbf{z}^k_{\mathbf{v}}})$

---

where, for each $i = 1, \ldots, p$, the function

$$P_i(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) \;:=\; \frac{1}{p}\left(b^T \mathbf{y}_i + d^T \boldsymbol{\mu}_i\right) + \frac{1}{2\alpha}\left\| S\left(-(A_{\bullet\mathcal{I}_i})^T \mathbf{y}_i - (C_{\bullet\mathcal{I}_i})^T \boldsymbol{\mu}_i - \gamma[(D_1)_{\mathcal{S}_i\mathcal{I}_i}]^T \mathbf{v}_i\right) \right\|_2^2.$$

Let $\mathbf{z}^k = (\mathbf{z}^k_{\mathbf{y}}, \mathbf{z}^k_{\boldsymbol{\mu}}, \mathbf{z}^k_{\mathbf{v}}) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p} \times \mathbb{R}^{N+p-2}$, $\mathbf{w} := (\mathbf{y}, \boldsymbol{\mu}, \mathbf{v}) \in \mathbb{R}^{mp} \times \mathbb{R}^{\ell p} \times \mathbb{R}^{N+p-2}$ and $\mathbf{w}_i := (\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) \in \mathbb{R}^m \times \mathbb{R}^\ell \times \mathbb{R}^{r_i}$ for each i. This leads to Algorithm 7.

Once a dual solution $(y_*, \mu_*, \widetilde{v}_*)$ is obtained from Algorithm 7, the primal solution $(x^*_{\mathcal{I}_i})_{i=1}^p$ is computed using (18). Moreover, to solve the subproblem in Line 4 of Algorithm 7, we apply the similar technique given in Remark V.4 to formulate it as a quadratic program.

Remark V.6: Another scheme for solving (44) is the distributed averaging based operator splitting scheme [5] by replacing $\mathcal{V}_i$ and H_i in Line 4 of Algorithm 6 by $\mathcal{R}_i$ and P_i respectively, where the set $\mathcal{R}_i := \mathbb{R}^m \times \mathbb{R}^\ell \times \{\mathbf{v}_i \,|\, \|\mathbf{v}_i\|_\infty \leq 1\}$ for each i. This scheme is suitable for a small or moderate N. Similar distributed schemes can be developed for the decoupled constraint given by (12). Moreover, they can be extended to the generalized total variation denoising and ℓ_1-trend filtering where $E = \lambda D_1$ or $E = \lambda D_2$ with $\lambda > 0$.

Remark V.7: Consider the fused BPDN-like problem, i.e., the BPDN-like problem (10) with $E = \begin{bmatrix} I_N \\ \gamma D_1 \end{bmatrix}$ for a constant $\gamma > 0$.

Suppose the assumptions given below (10) in Section IV-A hold. In the first stage, to solve the dual problem (22) with $F = \gamma D_1$ and $\ell \ll N$, define the function $\widehat{H}_i(\mathbf{y}_i, \boldsymbol{\mu}_i, \mathbf{v}_i) := (\sigma\|\mathbf{y}_i\|_2 + b^T\mathbf{y}_i + d^T\boldsymbol{\mu}_i)/p$. Then Algorithm 6 can be applied by replacing H_i with $\widehat{H}_i$. Similar results can be made for a large ℓ with $\mathcal{C}$ given by (12). The second stage is almost identical to that of the fused LASSO-like problem, except that $b + y_*$ is replaced by $b + \frac{\sigma y_*}{\|y_*\|_2}$.

E. LASSO-Like, BPDN-Like, and Regularized BP-Like Problems With the Norm From the Group LASSO

Consider $\|\cdot\|_\star = \|\cdot\|_G$ arising from the group LASSO (3); its dual norm $\|x\|_{G,\diamond} = \max_{i=1,\ldots,p}\|x_{\mathcal{I}_i}\|_2$. Many preceding results for the ℓ_1-norm can be extended to this case.

For illustration, consider the standard LASSO-like problem with $E = \lambda I_N$ for $\lambda > 0$ and a small ℓ. In the first stage, by virtue of the dual problem (26), Algorithms 2-4 can be used by replacing the set $\mathcal{W}_i$ with the set $\widehat{\mathcal{W}}_i := \{(\mathbf{y}_i, \boldsymbol{\mu}_i) \,|\, \|(A_{\bullet\mathcal{I}_i})^T\mathbf{y}_i + (C_{\bullet\mathcal{I}_i})^T\boldsymbol{\mu}_i\|_2 \le \lambda\}$, which has nonempty interior. When ℓ is large and $\mathcal{C} \in \mathbb{R}^{\ell\times N}$ is given by (12), the schemes in discussed in Remark V.3 can be used after the same replacement. In the second stage, we assume that exact regularization holds (cf. Section III-C). When $E = I_N$ and $\mathcal{C}$ is a general polyhedral set, the reduced dual problem (23) is formulated as the convex consensus optimization problem: $\min_{(\mathbf{y},\boldsymbol{\mu})\in\mathcal{A}_\mathbf{y}\times\mathcal{A}_{\boldsymbol{\mu}}} \sum_{i=1}^p J_i(\mathbf{y}_i, \boldsymbol{\mu}_i)$, where $J_i(\mathbf{y}_i, \boldsymbol{\mu}_i) := (b^T\mathbf{y}_i + d^T\boldsymbol{\mu}_i)/p + \frac{1}{2\alpha}[(\big\|(A_{\bullet\mathcal{I}_i})^T\mathbf{y}_i + (C_{\bullet\mathcal{I}_i})^T\boldsymbol{\mu}_i\big\|_2 - 1)_+]^2$ for $i = 1, \ldots, p$, and $\mathcal{A}_\mathbf{y}, \mathcal{A}_{\boldsymbol{\mu}}$ are defined in (32)–(33). Thus a distributed scheme similar to (41) can be applied.

In the second stage, when $\mathcal{C}$ is a box constraint set, consider the reduced dual problem (25). By introducing p copies of y's given by $\mathbf{y}_i$ and imposing the consensus condition on $\mathbf{y}_i$'s, this problem can be converted to a convex program of the variable $(\mathbf{y}_i, v_{\mathcal{I}_i})_{i=1}^p$ with a separable objective function and separable constraint sets with nonempty interiors. By Slater's condition, the D-R scheme or operator splitting schemes similar to the scheme (41) can be developed. If, in addition, $\mathcal{C}$ is a cone, the dual problems can be further reduced to unconstrained problems of the variable y only, e.g., those for $\mathcal{C} = \mathbb{R}^N$ and $\mathcal{C} = \mathbb{R}^N_+$ given in Case (b) of Section IV-C. These problems can be formulated as consensus convex programs and solved by column partition based distributed schemes. Finally, the primal solution $x^*_{\mathcal{I}_i}$ can be computed distributively using a dual solution y_* and the operator $S_{\|\cdot\|_2}$ (cf. Section IV-C).

The above results can be easily extended to the standard BPDN-like and fused LASSO/BPDN-like problems.

VI. Overall Convergence of the Two-Stage Distributed Algorithms

In this section, we analyze the overall convergence of the two-stage distributed algorithms proposed in Section V, assuming that a distributed algorithm in each stage is convergent. See Remark V.2 for the convergence conditions.

Consider the following regularized BP-like problem:

$$\min_{x\in\mathcal{C},\, Ax=b} \|Ex\|_\star + \frac{\alpha}{2}\|x\|_2^2, \tag{45}$$

where $\alpha > 0$, $E \in \mathbb{R}^{r\times N}$, $0 \neq A \in \mathbb{R}^{m\times N}$, $\mathcal{C}$ is a polyhedral set given by (5) with $C \in \mathbb{R}^{\ell\times N}$ and $d \in \mathbb{R}^\ell$, and $b \in \mathbb{R}^m$ with $b \in A\mathcal{C} := \{Ax \,|\, x \in \mathcal{C}\}$. We shall show that its unique optimal solution is continuous in b. To achieve this goal, consider the necessary and sufficient optimality condition for the unique solution x_* to (45), i.e., there exist (possibly non-unique) multipliers $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^\ell_+$ such that $0 \in E^T\partial\|Ex_*\|_\star + \alpha x_* + A^T\lambda + C^T\mu$, $Ax_* = b$, and $0 \le \mu \perp Cx_* - d \le 0$. When we need to emphasize the dependence of x_* on b, we write it as $x_*(b)$. For a given $b \in A\mathcal{C}$, define the set $\mathcal{S}(x_*) := \{(w, \lambda, \mu) \,\big|\, w \in \partial\|Ex_*\|_\star,\ E^Tw + \alpha x_* + A^T\lambda + C^T\mu = 0,\ 0 \le \mu \perp Cx_* - d \le 0\}$. This set contains all the sub-gradients w and the multipliers λ, μ satisfying the optimality condition at x_*, and it is often unbounded due to possible unboundeness of λ and μ (noting that w's are bounded). To overcome this difficulty in continuity analysis, we present the following proposition.

Proposition VI.1: [22, Proposition 6.1] The following hold for the problem (45):

(i) Let $\mathcal{B}$ be a bounded set in $\mathbb{R}^m$. Then $\{x_*(b) \,|\, b \in A\mathcal{C} \cap \mathcal{B}\}$ is a bounded set;

(ii) Let (b^k) be a convergent sequence in $A\mathcal{C} \cap \mathcal{B}$. Then there exist a constant $\gamma > 0$ and an index subsequence (k_s) such that for each k_s, there exists $(w^{k_s}, \lambda^{k_s}, \mu^{k_s}) \in \mathcal{S}(x_*(b^{k_s}))$ satisfying $\|(\lambda^{k_s}, \mu^{k_s})\| \le \gamma$.

The proof of this proposition is technical and relies on the Lipschitz property of the linear complementarity problem (LCP) under certain singleton property [6, Proposition 4.2.2], and [25]. See [22, Proposition 6.1] for details.

Theorem VI.1: Let $\alpha > 0$, $E \in \mathbb{R}^{r\times N}$ $A \in \mathbb{R}^{m\times N}$, $\mathcal{C} := \{x \in \mathbb{R}^N \,|\, Cx \le d\}$ for some $C \in \mathbb{R}^{\ell\times N}$ and $d \in \mathbb{R}^\ell$, and $b \in \mathbb{R}^m$ with $b \in A\mathcal{C}$. Then the unique solution x_* of the minimization problem (45) is continuous in b on $A\mathcal{C}$.

Proof: Fix an arbitrary $b \in A\mathcal{C}$. Suppose $x_*(\cdot)$ is discontinuous at this b. Then there exist $\varepsilon_0 > 0$ and a sequence (b^k) in $A\mathcal{C}$ such that (b^k) converges to b but $\|x^k_* - x_*(b)\| \ge \varepsilon_0$ for all k, where $x^k_* := x_*(b^k)$. By Statement (i) of Proposition VI.1, (x^k_*) is bounded and attains a convergent subsequence which, without loss of generality, can be itself. Let the limit of (x^k_*) be $\widehat{x}$. Further, as shown in Statement (ii) of Proposition VI.1, there exists a bounded subsequence $((w^{k_s}, \lambda^{k_s}, \mu^{k_s}))$ such that $(w^{k_s}, \lambda^{k_s}, \mu^{k_s}) \in \mathcal{S}(x^{k_s}_*)$ for each k_s. Without loss of generality, we assume that $((w^{k_s}, \lambda^{k_s}, \mu^{k_s}))$ converges to $(\widehat{w}, \widehat{\lambda}, \widehat{\mu})$. Since $(Ex^{k_s}_*) \to E\widehat{x}$ and $(w^{k_s}) \to \widehat{w}$ with $w^{k_s} \in \partial\|Ex^{k_s}_*\|_\star$ for each k_s, it follows from [1, Proposition B.24(c)] that $\widehat{w} \in \partial\|E\widehat{x}\|_\star$. By taking the limit, we deduce that $(\widehat{x}, \widehat{w}, \widehat{\lambda}, \widehat{\mu})$ satisfies $E^T\widehat{w} + \alpha\widehat{x} + A^T\widehat{\lambda} + C^T\widehat{\mu} = 0$, $A\widehat{x} = b$, and $0 \le \widehat{\mu} \perp C\widehat{x} - d \le 0$, i.e., $(\widehat{w}, \widehat{\lambda}, \widehat{\mu}) \in \mathcal{S}(\widehat{x})$. This shows that $\widehat{x}$ is a solution to (45) for the given b. Since this solution is unique, we must have $\widehat{x} = x_*(b)$. Hence, $(x^{k_s}_*)$ converges to $x_*(b)$, a contradiction to $\|x^{k_s}_* - x_*(b)\| \ge \varepsilon_0$ for all k_s. This yields the continuity of x_* in b on $A\mathcal{C}$. ■

When the norm $\|\cdot\|_\star$ in the problem (45) is the ℓ_1-norm or a convex PA function in general, the continuity property in Theorem VI.1 can be enhanced to Lipschitz continuity, which is useful in deriving the overall convergence rate.

Theorem VI.2: [22, Theorem 6.3] Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex PA function, $A\in\mathbb{R}^{m\times N}$, $\mathcal{C}$ be a polyhedral set given by (5), and $b\in\mathbb{R}^m$ with $b\in A\mathcal{C}$. Then for any $\alpha>0$, $\min_{x\in\mathcal{C}} f(x)+\frac{\alpha}{2}\|x\|_2^2$ subject to $Ax=b$ has a unique minimizer x_*. Further, x_* is Lipschitz continuous in b on $A\mathcal{C}$, i.e., there exists a constant $L>0$ such that $\|x_*(b')-x_*(b)\|\le L\|b'-b\|$ for any $b,b'\in A\mathcal{C}$.

The proof of this theorem is technical and uses the global Lipschitz property of the LCP with singleton property on a convex set; see [22, Theorem 6.3] for details.

For a polyhedral set $\mathcal{C}$, it follows from Lemmas IV.2 and IV.3 that $y_*+b\in A\mathcal{C}$ (respectively $\frac{\sigma y_*}{\|y_*\|_2}+b\in A\mathcal{C}$), where y_* is a solution to the dual problem (9) (respectively (11)). Practically, y_* is approximated by (y^k) generated in the first stage. For the LASSO-like problem (8), one uses y^k+b (with a large k) instead of y_*+b in the $\mathrm{BP}_{\mathrm{LASSO}}$ (28) in the second stage. This raises the question of whether $y^k+b\in A\mathcal{C}$ for all large k. The same question also arises for the BPDN-like problem (10). We give a mild sufficient condition for the feasibility under perturbations to a given b. Suppose $\mathcal{C}$ has a nonempty interior and A has full row rank. In view of $\mathrm{ri}(A\mathcal{C})=A\,\mathrm{ri}(\mathcal{C})=A\,\mathrm{int}(\mathcal{C})$ [21, Theorem 6.6], we see that $A\mathcal{C}$ has nonempty interior given by $A\,\mathrm{ri}(\mathcal{C})=A\,\mathrm{int}(\mathcal{C})$. Thus if $\widehat{b}:=y_*+b$ is such that $\widehat{b}=A\widehat{x}$ for some $\widehat{x}\in\mathrm{int}(\mathcal{C})$, then there exists a neighborhood $\mathcal{N}$ of $\widehat{b}$ such that $b\in A\mathcal{C}$ for any $b\in\mathcal{N}$. Additional sufficient conditions independent of b can also be established. For instance, when $\mathcal{C}=\mathbb{R}^N$, A need to have full row rank; when $\mathcal{C}=\mathbb{R}^N_+$, A need to have full row rank and $A(I+Q)=0$ for a nonnegative matrix Q.

Based on the above results, we establish the overall convergence of the two-stage algorithms.

Theorem VI.3: Consider the two-stage distributed algorithms for the LASSO-like problem (8) (resp. the BPDN-like problem (10)) with the norm $\|\cdot\|_\star$. Let (y^k) be a sequence generated in the first stage such that $(y^k)\to y_*$ as $k\to\infty$ and $b+y^k\in A\mathcal{C}$ (resp. $b+\frac{\sigma y^k}{\|y^k\|_2}\in A\mathcal{C}$) for all large k, where y_* is a solution to the dual problem (9) (resp. (11)), and (x^s) be a convergent sequence in the second stage for solving (28) (resp. (29)). Then

(i) $(x^s)\to x_*$ as $k,s\to\infty$, where x_* is the unique solution to the regularized $\mathrm{BP}_{\mathrm{LASSO}}$ (28) (resp. $\mathrm{BP}_{\mathrm{BPDN}}$ (29)).

(ii) Let $\|\cdot\|_\star$ be the ℓ_1-norm. Suppose (y^k) has the convergence rate $O(\frac{1}{k^q})$ and (x^s) has the convergence rate $O(\frac{1}{s^r})$. Then (x^s) converges to x_* in the rate of $O(\frac{1}{k^q})+O(\frac{1}{s^r})$.

Proof: We consider the LASSO-like problem only; the similar argument holds for the BPDN-like problem.

(i) For each k, let $\widehat{b}^k:=b+y^k$, where (y^k) is a sequence generated from the first stage that converges to y_*. When $\widehat{b}^k$ is used in the $\mathrm{BP}_{\mathrm{LASSO}}$ (28) in the second stage, i.e., the constraint $Ax=b+y_*$ is replaced by $Ax=\widehat{b}^k$, we have $\|x^s(\widehat{b}^k)-x_*\|\le\|x^s(\widehat{b}^k)-x_*(\widehat{b}^k)\|+\|x_*(\widehat{b}^k)-x_*\|$, where $x_*(\widehat{b}^k)$ is the unique solution to the $\mathrm{BP}_{\mathrm{LASSO}}$ (28) corresponding to the constraint $Ax=\widehat{b}^k$ (and $x\in\mathcal{C}$). Since $(x^s(\widehat{b}^k))$ converges to $x_*(\widehat{b}^k)$ as $s\to\infty$ (for a fixed k), $\|x^s(\widehat{b}^k)-x_*(\widehat{b}^k)\|$ converges to zero. Further, note that $x_*=x_*(\widehat{b}_*)$ with $\widehat{b}_*:=b+y_*$. It follows from Theorem VI.1 that $\|x_*(\widehat{b}^k)-x_*\|=\|x_*(\widehat{b}^k)-x_*(\widehat{b}_*)\|$ converges to zero as $k\to\infty$ in view of the convergence of (y^k) to y_*. This establishes the convergence of the two-stage algorithm.

(ii) When $\|\cdot\|_\star$ is the ℓ_1-norm, we deduce via Theorem VI.2 that x_* is Lipschitz continuous in b on $A\mathcal{C}$, i.e., there exists a constant $L>0$ such that $\|x_*(b)-x_*(b')\|\le L\|b-b'\|$ for any $b,b'\in A\mathcal{C}$. Hence, $\|x^s(\widehat{b}^k)-x_*\|\le\|x^s(\widehat{b}^k)-x_*(\widehat{b}^k)\|+\|x_*(\widehat{b}^k)-x_*(\widehat{b}_*)\| \le \|x^s(\widehat{b}^k)-x_*(\widehat{b}^k)\|+L\|\widehat{b}^k-\widehat{b}_*\|=\|x^s(\widehat{b}^k)-x_*(\widehat{b}^k)\|+L\|y^k-y_*\|=O(\frac{1}{s^r})+O(\frac{1}{k^q})$. ■

VII. Numerical Results

We present numerical results to demonstrate the performance of the proposed two-stage column partition based distributed algorithms for the standard LASSO/BPDN, fused LASSO/BPDN, group LASSO, and their extensions, e.g., those subject to polyhedral constraints. Distributed algorithms are implemented on MATLAB and run on a computer of the following processor: Intel(R) Core(TM) i7-8550 U CPU with 4 cores @ 1.80 GHz and RAM: 16.0 GB. We consider a network of $p=40$ agents with two topologies: the first is a cyclic graph, and the second is a random graph. The matrix $A\in\mathbb{R}^{100\times4000}$ is random normal (i.e., $m=100$ and $N=4000$), and $b\in\mathbb{R}^{100}$ is a random normal vector. For the standard/fused BPDN and its extensions, $\|b\|_2=11.63$ and the parameter $\sigma=0.2$, satisfying $\|b\|_2>\sigma$. We consider even column partitioning, i.e., each agent has 100 columns.

In each scheme, the stopping criterion is measured by the absolute error of two neighboring iterates, and its termination tolerance is given below. Further, to simplify notation, we use the following abbreviations: DA for distributed averaging, LA for local averaging, DR for Douglas-Rachford, and OS for operator splitting. For instance, a DA-OS scheme represents a distributed averaging based operator splitting scheme. In each table below, *Time* stands for the computation time *per agent*.

In each DA based scheme, we use the distributed averaging scheme with optimal constant edge weight [28, Section 4.1] for consensus computation. Numerical experiments show that this scheme is highly efficient. For instance, to compute the average of $\mathbf{y}=(\mathbf{y}_i)_{i=1}^p$ with $p=40$ and $\mathbf{y}_i\in\mathbb{R}^{100}$, it takes 0.0061 (resp. 0.001) seconds per agent to converge on the cyclic (resp. the random) graph with the relative error less than 10^{-7}. For the standard and fused LASSO/BPDN-like problems involving the ℓ_1-norm, the subproblem in each scheme, e.g., the projection step in an OS scheme, the proximal operator in DR scheme, or the subproblem in C-ADMM, is solved via a quadratic program; see Remark V.4. For the group LASSO, the projection step is formulated as a second order cone program (SOCP) and solved by SeDuMi.

To evaluate the accuracy of the proposed schemes, let J denote the objective function of each (primal) problem, and x^*_{dist} be a numerical primal solution obtained from a proposed 2-stage distributed scheme. Let $J^*_{\mathrm{dist}}:=J(x^*_{\mathrm{dist}})$, J^*_{true} be the true optimal

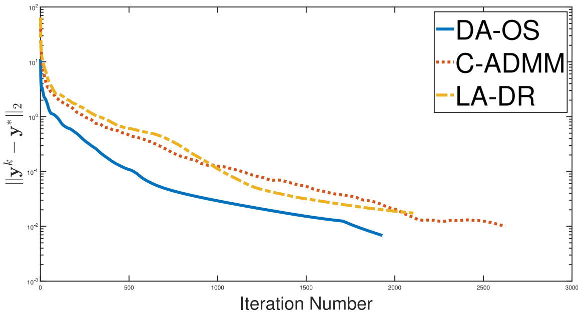

Fig. 1. Convergence behaviors in stage one of standard LASSO.

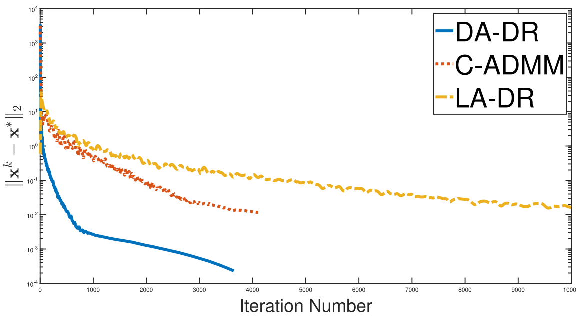

Fig. 2. Convergence behaviors in stage two of standard LASSO.

value obtained from a high-precision centralized scheme, and $J_{\text{RE, o}} := \frac{|J^*_{\text{dist}} - J^*_{\text{true}}|}{|J^*_{\text{true}}|}$ be the *overall* relative error of the optimal value. We also use $J_{\text{RE, s}}$ to denote the similar relative error for stage $s = 1, 2$.

A. Numerical Results for the LASSO-Like Problems

Consider the cyclic graph and $\mathcal{C} = \mathbb{R}^N$ unless otherwise stated.

• **Standard LASSO** The ℓ_1-penalty parameter $\lambda = 1.8$, and the regularization parameter in the second stage $\alpha = 0.18$. We apply three schemes for each stage: DA-OS (Algorithm 2), C-ADMM (Algorithm 3), LA-DR (Algorithm 4) for stage one, and DA-DR (Algorithm 5), C-ADMM (40), LA-DR (similar to Algorithm 4) for stage two. The termination tolerances for stages one and two are 10^{-4} and 10^{-5}, respectively. See the following table for the numerical results.

Standard LASSO: Stage One

Scheme	DA-OS	C-ADMM	LA-DR
Parameter	$\varrho = 1.2, \eta = 3$	$\eta = 1.5$	$\eta = 1.2, \rho = 0.6$
Time (sec)	58.8	64.9	51.2
$J_{\text{RE, 1}}$	8.1×10^{-6}	4.2×10^{-5}	4.5×10^{-5}

Standard LASSO: Stage Two

Scheme	DA-DR	C-ADMM	LA-DR
Parameter	$\rho = 0.6, \eta = 0.8$	$\eta = 5.5$	$\eta = 0.8, \rho = 0.1$
Time (sec)	41.2	26.7	66.2
$J_{\text{RE, 2}}$	6.2×10^{-6}	5.2×10^{-4}	5.6×10^{-3}
$J_{\text{RE, o}}$	3.5×10^{-7}	1.9×10^{-5}	6.2×10^{-4}

The convergence behaviors of these schemes in the two stages are displayed in Figs. 1 and 2. In the first stage, the errors of the dual variable is shown; in the second stage, we compute the corresponding primal variables from its numerical dual solutions and display its convergence behavior, where y^* and x^* are the true dual and primal solutions, respectively.

We also test the standard LASSO on the random graph via the DA-OS and C-ADMM for stage one, and the DA-DR for stage two, which is also used for the scaled regularized BP (cf. Remark V.1) with the regularization constant $\alpha = 0.1$ in stage two. The same termination tolerances are used. See the following table for a summary of the numerical results.

Standard LASSO: Stage One

Scheme	DA-OS	C-ADMM
Parameter	$\varrho = 1.2, \eta = 3$	$\eta = 1.5$
Time (sec)	54.7	73.3
$J_{\text{RE, 1}}$	8.3×10^{-6}	3.2×10^{-5}

Standard LASSO: Stage Two

Scheme	DA-DR	DA-DR for scaled r-BP
Parameter	$\rho = 0.6, \eta = 0.8$	$\rho = 0.6, \eta = 0.8$
Time (sec)	23.6	43.9
$J_{\text{RE, 2}}$	1.5×10^{-6}	2.9×10^{-4}
$J_{\text{RE, o}}$	4.7×10^{-7}	2.1×10^{-6}

• **LASSO with** $\mathcal{C} = \mathbb{R}^N_+$ This problem is known as the non-negative garrote in the literature [31]. We apply the DA-OS and C-ADMM for stage one, and the DA-DR and C-ADMM for stage two with $\alpha = 0.18$. The termination tolerances are 10^{-4} for both the schemes in stage one, 10^{-4} for the DA-DR in stage two, and 5×10^{-5} for the C-ADMM in stage two. See the following table for the numerical results.

Constrained LASSO: Stage One

Scheme	DA-OS	C-ADMM
Parameter	$\varrho = 1.2, \eta = 3$	$\eta = 1.5$
Time (sec)	12.4	24.9
$J_{\text{RE, 1}}$	1.2×10^{-5}	7.5×10^{-5}

Constrained LASSO: Stage Two

Scheme	DA-DR	C-ADMM
Parameter	$\rho = 0.6, \eta = 0.95$	$\eta = 5.5$
Time (sec)	71.5	32.8
$J_{\text{RE, 2}}$	3.5×10^{-3}	3.5×10^{-3}
$J_{\text{RE, o}}$	1.3×10^{-3}	7.8×10^{-3}

• **Fused LASSO** The matrix $E = \begin{bmatrix} \lambda I \\ \gamma D_1 \end{bmatrix}$ with $\lambda = 0.6$ and $\gamma = 0.4$, and the regularization constant $\alpha = 0.18$. We apply DA-OS (Algorithm 6) for stage one and the DA-DR (Algorithm 7) for stage two with termination tolerances 8×10^{-4} and 10^{-4}, respectively. We obtain $J_{\text{RE, o}} = 1.6 \times 10^{-4}$.

Fused LASSO: Stage One

Scheme	DA-OS
Parameter	$\varrho = 0.3, \eta = 3$
Time (sec)	213.8
$J_{\text{RE, 1}}$	8.5×10^{-5}

Fused LASSO: Stage Two

Scheme	DA-DR
Parameter	$\rho = 0.4, \eta = 0.6$
Time (sec)	271.9
$J_{\text{RE, 2}}$	6.5×10^{-4}

• **Group LASSO** The penalty constant $\lambda = 1.8$, and the regularization parameter $\alpha = 0.18$. For stage one, a DA-OS scheme similar to Algorithm 2 is used by replacing the set $\mathcal{W}_i$ with $\widehat{\mathcal{W}_i} := \{\mathbf{y}_i \,|\, \|(A_{\bullet \mathcal{I}_i})^T \mathbf{y}_i\|_2 \le \lambda\}$ as in Section V-E. Its projection step is formulated as a SOCP and solved by SeDuMi. For stage two, we exploit the reduced dual problem via the soft thresholding operator $S_{\|\cdot\|_2}$ and apply the DA-OS scheme similar to that in (41) by dropping $\boldsymbol{\mu}$ and replacing G_i with $J_i(\mathbf{y}_i) := (b^T \mathbf{y}_i)/p + \frac{1}{2\alpha}[(\|(A_{\bullet \mathcal{I}_i})^T \mathbf{y}_i - 1)_+]^2$ for each i. The termination tolerances for stages one and two are 10^{-5} and 8×10^{-7}, respectively. We obtain $J_{\text{RE, o}} = 3.6 \times 10^{-4}$.

Group LASSO: Stage One

Scheme	DA-OS
Parameter	$\varrho = 1.2, \eta = 2$
Time (sec)	52.8
$J_{\text{RE, 1}}$	3.6×10^{-4}

Group LASSO: Stage Two

Scheme	DA-OS
Parameter	$\rho = 0.005, \eta = 0.3$
Time (sec)	19.5
$J_{\text{RE, 2}}$	4.5×10^{-5}

B. Numerical Results for the BPDN-Like Problems

Consider the cyclic graph and $\mathcal{C} = \mathbb{R}^N$ unless otherwise stated.

• **Standard BPDN** The regularization parameter in the second stage $\alpha = 0.15$. We apply DA-OS and IC-ADMM [4] (with the parameters c and β_i's) for stage one, and DA-DR, C-ADMM, LA-OS for stage two. The termination tolerances for the first and second stages are 4×10^{-4} and 10^{-5} respectively.

Standard BPDN: Stage One

Scheme	DA-OS	IC-ADMM
Parameter	$\varrho = 1.2, \eta = 3$	$c = 1.5, \beta_i = 1.7, \forall\, i$
Time (sec)	26.6	58.5
$J_{\text{RE, 1}}$	5.9×10^{-4}	1.7×10^{-5}

Standard BPDN: Stage Two

Scheme	DA-DR	C-ADMM	LA-OS
Parameter	$\rho = 0.5, \eta = 0.95$	$\eta = 5.5$	$\eta = 0.95, \rho = 0.1$
Time (sec)	13.2	24.4	33.9
$J_{\text{RE, 2}}$	1.68×10^{-5}	1.3×10^{-3}	5.1×10^{-4}
$J_{\text{RE, o}}$	1.7×10^{-5}	1.3×10^{-3}	5.1×10^{-4}

We also test the standard BPDN on the random graph with the regularization parameter $\alpha = 0.18$. We apply DA-OS for stage one, and DA-DR for stage two. The table below shows the numerical results with $J_{\text{RE,o}} = 5.6 \times 10^{-4}$.

Standard BPDN: Stage One

Scheme	DA-OS
Parameter	$\varrho = 1.2, \eta = 3$
Time (sec)	19.9
$J_{\text{RE, 1}}$	7.04×10^{-4}

Standard BPDN: Stage Two

Scheme	DA-DR
Parameter	$\rho = 0.5, \eta = 0.95$
Time (sec)	6.8
$J_{\text{RE, 2}}$	1.28×10^{-3}

• **BPDN with** $\mathcal{C} = \mathbb{R}^N_+$ We apply the DA-OS for stage one, and the DA-DR for stage two with $\alpha = 0.18$. The termination tolerances for stage one and stage two are 10^{-5} and 10^{-4}, respectively. We obtain $J_{\text{RE, o}} = 4.9 \times 10^{-4}$.

Constr. BPDN: Stage One

Scheme	DA-OS
Parameter	$\varrho = 1.2, \eta = 3$
Time (sec)	37.1
$J_{\text{RE, 1}}$	1.5×10^{-4}

Constr. BPDN: Stage Two

Scheme	DA-DR
Parameter	$\rho = 0.95, \eta = 0.6$
Time (sec)	49.0
$J_{\text{RE, 2}}$	7.4×10^{-4}

• **Fused BPDN** The matrix $E = \begin{bmatrix} I_N \\ \gamma D_1 \end{bmatrix}$ with $\gamma = 2/3$, and the regularization constant $\alpha = 0.18$. We apply DA-OS and DA-DR for stage one and stage two with the termination tolerances 10^{-4} and 2×10^{-4}, respectively. We obtain $J_{\text{RE, o}} = 6.9 \times 10^{-4}$.

Fused BPDN: Stage One

Scheme	DA-OS
Parameter	$\varrho = 1.2, \eta = 3$
Time (sec)	248.4
$J_{\text{RE, 1}}$	1.5×10^{-6}

Fused BPDN: Stage Two

Scheme	DA-DR
Parameter	$\rho = 0.4, \eta = 0.8$
Time (sec)	205.3
$J_{\text{RE, 2}}$	7.7×10^{-4}

C. Discussions and Comparison

We compare the proposed two-stage schemes with two existing distributed schemes: the DC-ADMM [4], [16] for the standard LASSO with $\mathcal{C} = \mathbb{R}^N$, and the PDC-ADMM [2] for the standard LASSO with $\mathcal{C} = \mathbb{R}^N_+$. These two schemes cannot handle the non-polyhedral constraints in the BPDN-like problems and the additional coupling in the fused LASSO/BPDN. Hence, we focus on the standard LASSO and its constrained case. The results of the DC-ADMM and PDC-ADMM over the cyclic graph on the same computer with the tolerance 10^{-4} are given below.

Standard LASSO

Scheme	DC-ADMM
Parameter	$c = 0.05$
Time (sec)	307.1
$J_{\text{RE, o}}$	4.1×10^{-4}
Iteration No.	12974

Constrained LASSO

Scheme	PDC-ADMM
Parameter	$c = \tau_i = 0.01$
Time (sec)	323.5
$J_{\text{RE, o}}$	3.6×10^{-2}
Iteration No.	4632

The overall computation time of the proposed two-stage schemes for the standard (resp. constrained) LASSO is 78-121 seconds (resp. 57-94 seconds) with smaller $J_{\text{RE, o}}$. Hence, the proposed schemes outperform the DC-ADMM and PDC-ADMM in both computation time and numerical accuracy.

Since the communication cost is proportional to the number of iterations, we compare the number of iterations of the proposed two-stage schemes vs. DC-ADMM (resp. PDC-ADMM) for the standard LASSO (resp. constrained LASSO) on the cyclic graph. The following table summarizes the number of iterations for the proposed two-stage schemes.

Stage	Standard LASSO: Iteration No.			Constr. LASSO: Iteration No.	
	DA-OS/DR	C-ADMM	LA-DR	DA-OS/DR	C-ADMM
One	OS: 1930	2624	2105	OS: 1175	2864
Two	DR: 3483	4104	9999	DR: 2683	2846

For the standard LASSO, the total iteration numbers of the two-stage DA based scheme, C-ADMM, and LA-DR scheme are 5413, 6728, and 12 094, respectively. For the constrained LASSO, the total iteration number is 3958 for the DA-based scheme and is 5710 for the C-ADMM. Note that the DA-based schemes need additional iterations for distributed averaging computation, leading to extensive communications. Nevertheless, the proposed two-stage C-ADMM takes fewer or a similar number of iterations and less computation time while achieving better numerical accuracy in comparison with the DC-ADMM and PDC-ADMM. Finally, the memory costs of the proposed two-stage schemes are same or similar to those of the DC-ADMM or PDC-ADMM.

VIII. Conclusions

Column partition based distributed schemes are developed for a class of densely coupled convex sparse optimization problems, and overall convergence is proved. Future research is to extend these schemes to a broader class of problems.

A. Proof of Lemma IV.1

Proof: Let $J_* > -\infty$ be the finite infimum of (P). Since $\mathcal{P}$ is polyhedral, it follows from [1, Proposition 5.2.1] that the strong duality holds, i.e., $J_* = \inf_{z\in\mathcal{P}}[\sup_{y,\mu\geq 0} J(z) + y^T(Az - b) + \mu^T(Cz-d)] = \sup_{y,\mu\geq 0}[\inf_{z\in\mathcal{P}} J(z) + y^T(Az-b) + \mu^T(Cz-d)]$, and the dual problem of (P) attains an optimal solution (y_*, μ_*) with $\mu_* \geq 0$ such that $J_* = \inf_{z\in\mathcal{P}} J(z) + y_*^T(Az-b) + \mu_*^T(Cz-d)$. Therefore,

$$\begin{aligned} J_* &= \inf_{z\in\mathcal{P}} \|Ez\|_\star + f(z) + y_*^T(Az-b) + \mu_*^T(Cz-d) \\ &= \inf_{z\in\mathcal{P}} \sup_{\|v\|_\diamond\leq 1} \big[(Ez)^T v + f(z) + y_*^T(Az-b) + \mu_*^T(Cz-d)\big] \\ &= \sup_{\|v\|_\diamond\leq 1} \inf_{z\in\mathcal{P}} \big[(Ez)^T v + f(z) + y_*^T(Az-b) + \mu_*^T(Cz-d)\big] \\ &\leq \sup_{y,\,\mu\geq 0,\,\|v\|_\diamond\leq 1} \inf_{z\in\mathcal{P}} \big[(Ez)^T v + f(z) + y^T(Az-b) \\ &\qquad + \mu^T(Cz-d)\big] \\ &\leq \inf_{z\in\mathcal{P}} \sup_{y,\,\mu\geq 0,\,\|v\|_\diamond\leq 1} \big[(Ez)^T v + f(z) + y^T(Az-b) \\ &\qquad + \mu^T(Cz-d)\big] = J_*, \end{aligned}$$

where the third equation follows from Sion's minimax theorem [27, Corollary 3.3] and the compactness of $B_\diamond(0,1)$, and the second inequality is due to the weak duality. ■

B. Proof of Lemma IV.2

Proof: Consider the equivalent primal problem for (8): $\min_{x\in\mathcal{C}, Ax-b=u} \frac{1}{2}\|u\|_2^2 + \|Ex\|_\star$, and let (x_*, u_*) be its optimal solution. Consider the Lagrangian $L(x,u,y,\mu,v) := \frac{\|u\|_2^2}{2} + (Ex)^T v + y^T(Ax-b-u) + \mu^T(Cx-d)$. In view of the strong duality shown in Lemma IV.1, $(x_*, u_*, y_*, \mu_*, v_*)$ is a saddle point of L. Hence, $L(x_*, u_*, y, \mu, v) \leq L(x_*, u_*, y_*, \mu_*, v_*)$ for all $y\in\mathbb{R}^m, \mu\in\mathbb{R}^\ell_+, v\in B_\diamond(0,1)$, and $L(x_*, u_*, y_*, \mu_*, v_*) \leq L(x, u, y_*, \mu_*, v_*)$ for all $x\in\mathbb{R}^N, u\in\mathbb{R}^m$. The former inequality implies that $\nabla_y L(x_*, u_*, y_*, \mu_*, v_*) = 0$ such that $Ax_* - b - u_* = 0$; the latter inequality shows that $\nabla_u L(x_*, u_*, y_*, \mu_*, v_*) = 0$, which yields $u_* - y_* = 0$. These results lead to $Ax_* - b = y_*$. Lastly, when $d = 0$, it follows from the strong duality that $\frac{1}{2}\|Ax_* - b\|_2^2 + \|Ex_*\|_\star = -b^T y_* - \frac{1}{2}\|y_*\|_2^2$. Using $Ax_* - b = y_*$, we have $\|Ex_*\|_\star = -b^T y_* - \|y_*\|_2^2 = -(b+y_*)^T y_*$. ■

C. Proof of Lemma IV.3

Proof: (i) Consider the equivalent primal problem for (10): $\min_{x\in\mathcal{C},\, Ax=b=u,\, \|u\|_2\leq\sigma} \|Ex\|_\star$, and let (x_*, u_*) be its optimal solution. For a dual solution (y_*, μ_*, v_*), we deduce that $y_* \neq 0$ since otherwise, we have $-(b^T y_* + \sigma\|y_*\|_2 + d^T\mu_*) \leq 0$, which contradicts its positive optimal value by the strong duality. Let the Lagrangian $L(x,u,y,\mu,v,\lambda) := (Ex)^T v + y^T(Ax-b-u) + \lambda(\|u\|_2^2 - \sigma^2) + \mu^T(Cx-d)$. By the strong duality, $(x_*, u_*, y_*, \mu_*, v_*, \lambda_*)$ is a saddle point of L such that $L(x_*, u_*, y, \mu, v, \lambda) \leq L(x_*, u_*, y_*, \mu_*, v_*, \lambda_*)$, for all $y\in\mathbb{R}^m, \mu\in\mathbb{R}^\ell_+, v\in B_\diamond(0,1), \lambda\in\mathbb{R}_+$, and $L(x_*, u_*, y_*, \mu_*, v_*, \lambda_*) \leq L(x, u, y_*, \mu_*, v_*, \lambda_*)$ for all $x\in\mathbb{R}^N, u\in\mathbb{R}^m$. The former inequality implies that $\nabla_y L(x_*, u_*, y_*, \mu_*, v_*, \lambda_*) = 0$, yielding $Ax_* - b - u_* = 0$, and the latter shows that $\nabla_u L(x_*, u_*, y_*, \mu_*, v_*, \lambda_*) = 0$, which gives rise to $2\lambda_* u_* = y_*$. Since $y_* \neq 0$, we have $\lambda_* > 0$ which implies $\|u_*\|_2 - \sigma = 0$ by the complementarity relation. It thus follows from $2\lambda_* u_* = y_*$ and $\|u_*\|_2 = \sigma$ that $\lambda_* = \frac{\|y_*\|_2}{2\sigma}$. This leads to $u_* = \frac{y_*}{2\lambda_*} = \frac{\sigma y_*}{\|y_*\|_2}$. Therefore, $Ax_* - b = u_* = \frac{\sigma y_*}{\|y_*\|_2}$. Finally, when $d = 0$, we deduce via the strong duality that $\|Ex_*\|_\star = -b^T y_* - \sigma\|y_*\|_2$.

(ii) Let $d = 0$. Let (y_*, μ_*, v_*) and (y'_*, μ'_*, v'_*) be two solutions of (11), where $y_* \neq 0$ and $y'_* \neq 0$. Then $b^T y_* + \sigma\|y_*\|_2 = b^T y'_* + \sigma\|y'_*\|_2 = -\|Ex_*\|_\star < 0$. Therefore, $\|y_*\|_2(b^T\frac{y_*}{\|y_*\|_2} + \sigma) = \|y'_*\|_2(b^T\frac{y'_*}{\|y'_*\|_2} + \sigma)$, and $b^T\frac{y_*}{\|y_*\|_2} + \sigma < 0$. Proposition II.1 shows that for any solution x_* of the primal problem (10), $Ax_* - b$ is constant. By the argument for Part (i), $Ax_* - b = u_*$ and $Ax'_* - b = u'_*$ such that $u_* = u'_*$, and $u_* = \frac{\sigma y_*}{\|y_*\|_2}$ and $u'_* = \frac{\sigma y'_*}{\|y'_*\|_2}$. Thus $\frac{y_*}{\|y_*\|_2} = \frac{y'_*}{\|y'_*\|_2}$ so that $b^T\frac{y_*}{\|y_*\|_2} + \sigma = b^T\frac{y'_*}{\|y'_*\|_2} + \sigma < 0$. By $\|y_*\|_2(b^T\frac{y_*}{\|y_*\|_2} + \sigma) = \|y'_*\|_2(b^T\frac{y'_*}{\|y'_*\|_2} + \sigma)$, we have $\|y_*\|_2 = \|y'_*\|_2$. By $\frac{y_*}{\|y_*\|_2} = \frac{y'_*}{\|y'_*\|_2}$ again, we have $y_* = y'_*$. ■

References

[1] D. P. Bertsekas, *Nonlinear Program.*, 2nd ed. Athena Scientific, Belmont, MA, 1999.
[2] T.-H. Chang, "A proximal dual consensus ADMM method for multi-agent constrained optimization," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3719–3734, Jul. 2016.
[3] T.-H. Chang, M. Hong, W.-C. Liao, and X. Wang, "Asynchronous distributed ADMM for large-scale optimization-part i: Algorithm and convergence analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3118–3130, Jun. 2016.
[4] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 482–497, Jan. 2015.
[5] D. Davis and W. Yin, "A three-operator splitting scheme and its optimization applications," *Set-valued Variational Anal.*, vol. 25, no. 4, pp. 829–858, 2017.
[6] F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*. Berlin, Germany: Springer-Verlag, 2003.
[7] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Cambridge, MA, USA: Birkhäuser, Basel, 2013.
[8] M. P. Friedlander and P. Tseng, "Exact regularization of convex programs," *SIAM J. Optim.*, vol. 18, no. 4, pp. 1326–1350, 2007.
[9] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo, "A block successive upper-bound minimization method of multipliers for linearly constrained convex optimization," *Math. Operations Res.*, vol. 45, no. 3, pp. 833–861, 2020.
[10] J. Hu, Y. Xiao, and J. Liu, "Distributed algorithms for solving locally coupled optimization problems on agent networks," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 2420–2425.
[11] D. Jakovetić, J. Xavier, and J. M. Moura, "Fast distributed gradient methods," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1131–1146, May 2014.

[12] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," *SIAM Rev.*, vol. 51, no. 2, pp. 339–360, 2009.
[13] M.-J. Lai and W. Yin, "Augmented ℓ_1 and nuclear-norm models with a globally linearly convergent algorithm," *SIAM J. Imag. Sci.*, vol. 6, no. 2, pp. 1059–1091, 2013.
[14] Z. Li, W. Shi, and M. Yan, "A decentralized proximal-gradient method with network independent step-sizes and separated convergence rates," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4494–4506, Sep. 2019.
[15] J. Liu and S. J. Wright, "Asynchronous stochastic coordinate descent: Parallelism and convergence properties," *SIAM J. Optim.*, vol. 25, no. 1, pp. 351–376, 2015.
[16] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5262–5276, Oct. 2010.
[17] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Puschel, "Distributed basis pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1942–1956, Apr. 2012.
[18] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Püschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, May 2013.
[19] S. Mousavi and J. Shen, "Solution uniqueness of convex piecewise affine functions based optimization with applications to constrained ℓ_1 minimization," *ESAIM: Control, Optimisation Calculus Variations*, vol. 25, Art. no. 56, 2019.
[20] Z. Peng, M. Yan, and W. Yin, "Parallel and distributed sparse optimization," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, 2013, pp. 659–664.
[21] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.
[22] J. Shen, J. Hu, and E. K. H. K., "Column partition based distributed algorithms for coupled convex sparse optimization: Dual and exact regularization approaches," 2020, *arXiv:2002.08561*.
[23] J. Shen and S. Mousavi, "Least sparsity of p-norm based optimization problems with $p > 1$," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2721–2751, 2018.
[24] J. Shen and S. Mousavi, "Exact support and vector recovery of constrained sparse vectors via constrained matching pursuit," 2019, *arXiv:1903.07236*.
[25] J. Shen and J.-S. Pang, "Linear complementarity systems with singleton properties: Non-Zenoness," in *Proc. IEEE Amer. Control Conf.*, 2007, pp. 2769–2774.
[26] W. Shi, Q. Ling, G. Wu, and W. Yin, "A proximal gradient algorithm for decentralized composite optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6013–6023, Nov. 2015.
[27] M. Sion, "On general minimax theorems," *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, 1958.
[28] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.
[29] W. Yin, "Analysis and generalizations of the linearized bregman method," *SIAM J. Imag. Sci.*, vol. 3, no. 4, pp. 856–877, 2010.
[30] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM J. Optim.*, vol. 26, no. 3, pp. 1835–1854, 2016.
[31] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc.: Ser. B (Stat. Methodol.)*, vol. 68, no. 1, pp. 49–67, 2006.
[32] H. Zhang, W. Yin, and L. Cheng, "Necessary and sufficient conditions of solution uniqueness in 1-norm minimization," *J. Optim. Theory Appl.*, vol. 164, no. 1, pp. 109–122, 2015.