Derivation and Analysis of the Primal-Dual Method of Multipliers Based on Monotone Operator Theory

Sherson, Thomas William; Heusdens, Richard; Kleijn, W. Bastiaan

DOI
10.1109/TSIPN.2018.2876754

Publication date
2019

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Signal and Information Processing over Networks

Derivation and Analysis of the Primal-Dual Method of Multipliers Based on Monotone Operator Theory

Thomas Sherson, Richard Heusdens, and W. Bastiaan Kleijn

***Abstract*—In this paper we present a novel derivation of an existing algorithm for distributed optimization termed the primal-dual method of multipliers. In contrast to its initial derivation, monotone operator theory is used to connect PDMM with other first-order methods such as Douglas-Rachford splitting and the alternating direction method of multipliers thus providing insight into its operation. In particular, we show how PDMM combines a lifted dual form in conjunction with Peaceman-Rachford splitting to facilitate distributed optimization in undirected networks. We additionally demonstrate sufficient conditions for primal convergence for strongly convex differentiable functions and strengthen this result for strongly convex functions with Lipschitz continuous gradients by introducing a primal geometric convergence bound.**

***Index Terms*—Primal-Dual method of multipliers (PDMM), distributed optimization, monotone operator.**

I. Introduction

The world around us is evolving through the use of large scale networking. From the way we communicate via social media [1], to the revolution of utilities and services via the paradigm of the "Internet of Things" [2], networking is reshaping the way we operate as a society. Echoing this trend, the last three decades has seen a significant rise in the deployment of large scale *sensor networks* for a wide range of applications [3]–[5]. Such applications include environmental monitoring [6], [7], power grid management [8]–[10], as well being used as part of home health care systems [11], [12].

Where centralized network topologies were once the port of call for handling data processing of sensor networks, increasingly on-node computational capabilities of such systems are being exploited to parallelize or even fully distribute data processing and computation. In contrast to their centralized counterparts such *distributed networks* have a number of distinct advantages including robustness to node failure, scalability with network size and localized transmission requirements.

Unfortunately, these distributed networks are also often characterized by limited connectivity. This limited accessibility between nodes implicitly restricts data availability making classical signal processing operations impractical or infeasible to perform. Therefore, the desire to decentralize computation requires the design of novel signal processing approaches specifically tailored to the task of in-network computation.

Within the literature, a number of methods for performing distributed signal processing have been proposed including distributed consensus [13]–[15], belief propagation/message passing approaches [16]–[18], graph signal processing over networks [19]–[21] and more. An additional method of particular interest to this work, is to approach the task of signal processing via its inherent connection with convex optimization. In particular, over the last two decades, it has been shown that many classical signal processing problems can be recast in an equivalent convex form [22]. By defining methods to perform distributed optimization we can therefore facilitate distributed signal processing in turn.

Recently, a new algorithm for distributed optimization called the primal dual method of multipliers (PDMM) was proposed [23]. In [23], it was shown that PDMM exhibited guaranteed average convergence, which in some examples were faster than competing methods such as the alternating direction method of multipliers (ADMM) [24]. However, there are a number of open questions surrounding the approach. In particular, prior to this work, it was unclear how PDMM was connected with similar methods within the literature.

To clarify the link between PDMM and existing works, we present a novel viewpoint of the algorithm through the lens of monotone operator theory. By demonstrating how PDMM can be derived from this perspective, we link its operation with classic operator splitting algorithms. The major strength of this observation is the fact that we can leverage results from monotone operator theory to better understand the operation of PDMM. In particular we use this insight to demonstrate new and stronger convergence results for different classes of problems than those that currently exist within the literature.

A. Related Work

The work in this paper builds upon the extensive history within the field of convex optimization in the areas of parallel and decentralized processing. In the 1970's, Rockafellar's work in network optimization [25] and the relation between convex optimization and monotone operator theory [26]–[28] helped establish a foundation for the field. Importantly, Rockafellar showed how linearly constrained separable convex programs can be solved in parallel via Lagrangian duality.

In the field of parallel and distributed computation, further development was undertaken by Bertsekas and Tsitsiklis [29]–[31] throughout the 1980's, where again separability was used as a mechanism to design a range of new algorithms. Similarly,

Thomas Sherson is with the Department of Microelectronics, Circuits and Systems group, Delft University of Technology, The Netherlands. Email: t.sherson@tudelft.nl

Richard Heusdens is with the Department of Microelectronics, Circuits and Systems group, Delft University of Technology, The Netherlands. Email: r.heusdens@tudelft.nl

W. Bastiaan Kleijn is with the Department of Microelectronics, Circuits and Systems group, Delft University of Technology, The Netherlands. and with the School of Engineering and Computer Science, Victoria University of Wellington, New Zealand. Email: w.b.kleijn@tudelft.nl

Eckstein [32], [33] adopted an approach more reflective of Rockafellar, utilizing monotone operator theory and operator splitting to develop new distributed algorithms.

In recent years, there has been a renewed surge of interest in networked signal processing [34]–[36] due to the continued expansion of networked systems. This period has also seen the development of novel distributed optimization approaches for both convex and potentially non-convex problems. In the convex case, the works of [37], [38], echoing advances in three term operator splitting such as Vu-Condat splitting [39], [40], provide general frameworks for distributed convex optimization. Including classical approaches, such as ADMM, as special cases, these algorithms leverage primal-dual schemes and functional separability to create distributed implementations.

The work in [41], [42] focuses on the more general problem of potentially non-convex optimization. In particular, by at each iteration approximating both objective and constraints with specific strongly convex and smooth surrogates, the proposed methods have provable guarantees on convergence to local minima. Furthermore, in contrast to other methods, the proposed approach need not explicitly require functional separability, only the separability of the surrogates used. This allows for the optimization of problems typically outside of the scope of distributed algorithms.

B. Main Contribution

The main contributions of this paper are two-fold. Firstly we provide a novel derivation for PDMM from the perspective of monotone operator theory. In particular, we show how PDMM can be derived by combining a particular dual lifted problem with Peaceman-Rachford (PR) splitting. In contrast to its original derivation, this approach links PDMM with other classical first order methods from the literature including forward-backward splitting, Douglas-Rachford (DR) splitting and ADMM (see [43] for a recent overview).

The monotone operator perspective is also used to demonstrate a range of new convergence results for PDMM. We show how PDMM is guaranteed to converge to a primal optimal solution for strongly convex, differentiable objective functions. This result is strengthened for strongly convex functions with Lipschitz continuous gradients where a geometric convergence bound is demonstrated by linking the worst-case convergence of PDMM with that of a generalized alternating method of projections algorithm. Notably, while such results exist for PR splitting applied to dual domain optimization problems [44], they require an additional full row rank[1] assumption to ensure strong monotonicity which cannot be guaranteed in the case of PDMM. Furthermore, while a geometric convergence proof exists for distributed ADMM [45], currently there is no such result for PDMM. In this way the proposed work also strengthens the performance guarantees for PDMM, an important point for practical distributed optimization.

[1]*Row rank* refers to the dimension of the span of the row space of a matrix. *Row rank deficient* matrices have more rows than their row rank. The notions of *column rank* and *column rank deficiency* are defined equivalently.

C. Organization of the Paper

The remainder of this paper is organized as follows. Sec. II introduces appropriate nomenclature to support the manuscript. Sec. III introduces a monotone operator derivation of PDMM based on a specific dual lifting approach. Sec. IV demonstrates the guaranteed primal convergence of PDMM for strongly convex and differentiable functions. This is strengthened in Sec. V where we demonstrate primal geometric convergence for strongly convex functions with Lipschitz continuous gradients. Finally, Sec. VI includes simulation results to reinforce and verify the underlying claims of the document and the final conclusions are drawn in Sec. VII

II. Nomenclature

In this work we denote by $\mathbb{R}$ the set of real numbers, by $\mathbb{R}^N$ the set of real column vectors of length N and by $\mathbb{R}^{M\times N}$ the set of M by N real matrices. Let $\mathcal{X},\mathcal{Y}\subseteq\mathbb{R}^N$. A set valued operator $\mathbf{T}:\mathcal{X}\rightarrow\mathcal{Y}$ is defined by its graph, $\text{gra}\left(\mathbf{T}\right)=\{(\mathbf{x},\mathbf{y})\in\mathcal{X}\times\mathcal{Y}\mid\mathbf{y}\in\mathbf{T}\left(\mathbf{x}\right)\}$. Similarly, the notion of an inverse of an operator $\mathbf{T}^{-1}$ is defined via its graph so that $\text{gra}\left(\mathbf{T}^{-1}\right)=\{(\mathbf{y},\mathbf{x})\in\mathcal{Y}\times\mathcal{X}\mid\mathbf{y}\in\mathbf{T}\left(\mathbf{x}\right)\}$. $\mathbf{J}_{\mathbf{T},\rho}=\left(\mathbf{I}+\rho\mathbf{T}\right)^{-1}$ denotes the resolvent of an operator while $\mathbf{R}_{\mathbf{T},\rho}=2\mathbf{J}_{\mathbf{T},\rho}-\mathbf{I}$ denotes the reflected resolvent (Cayley operator). The fixed-point set of $\mathbf{T}$ is denoted by $\text{fix}\left(\mathbf{T}\right)=\{\mathbf{x}\in\mathcal{X}\mid\mathbf{T}\left(\mathbf{x}\right)=\mathbf{x}\}$. If $\mathbf{T}$ is a linear operator then $\text{ran}(\mathbf{T})$ and $\ker(\mathbf{T})$ denote its range and kernel respectively.

III. A Derivation of the Primal-Dual Method of Multipliers Based on Monotone Operator Theory

In this section we reintroduce a recently proposed algorithm for distributed optimization termed the Primal-Dual method of multipliers (PDMM) [23]. Unlike earlier efforts within the literature [23], [24], here we demonstrate how PDMM can be derived from the perspective of monotone operator theory. In particular we show how PDMM can be derived by applying PR splitting to a certain lifted dual problem. Additionally, we highlight a previously unknown connection between PDMM and a distributed ADMM variant.

A. Problem Statement: Node Based Distributed Optimization

Consider an undirected network consisting of N nodes with which we want to perform convex optimization in a distributed manner. The associated graphical model of such a network is given by $\mathbf{G}(V,E)$ where $V=\{1,...,N\}$ denotes the set of nodes and E denotes the set of undirected edges so that $(i,j)\in E$ if nodes i and j share a physical connection. Note that these are simple graphs as they do not contain self loops or repeated edges. We will assume that G forms a single connected component and will denote by $\mathcal{N}(i)=\{j\in V\mid(i,j)\in E\}$ the set of neighbors of node i, i.e. those nodes j so that i and j can communicate directly. An example of such a network is given in Figure 1.

As previously mentioned, we are interested in using this network to perform distributed convex optimization. In this way, assume that each node i is equipped with a function $f_i\in\Gamma_0\left(\mathbb{R}^{M_i}\right)$ parameterized by a local variable $\mathbf{x}_i\in\mathbb{R}^{M_i}$. Here

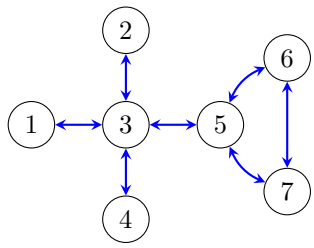


Fig. 1. The communication graph G of a seven node network. Numbered circles denote nodes and while the arrows denote the undirected edges. The neighborhood of node five is given by the set $\mathcal{N}(5) = \{3, 6, 7\}$.

Γ_0 denotes the family of closed, convex and proper (CCP) functions. Under this model, consider solving the following optimization problem in a distributed manner:

$$\begin{aligned} \min_{\mathbf{x}_i \,\forall\, i \in V} \quad & \sum_{i \in V} f_i(\mathbf{x}_i) \\ \text{s.t} \quad & \mathbf{A}_{i|j}\mathbf{x}_i + \mathbf{A}_{j|i}\mathbf{x}_j = \mathbf{b}_{i,j} \quad \forall\, (i,j) \in E. \end{aligned} \tag{1}$$

The matrices $\mathbf{A}_{i|j} \in \mathbb{R}^{M_{i,j} \times M_i}$ while the vectors $\mathbf{b}_{i,j} \in \mathbb{R}^{M_{i,j}}$. The identifier $i|j$ denotes a directed edge while i, j denotes an undirected edge. Furthermore, let $M_V = \sum_{i \in V} M_i$ and $M_E = \sum_{(i,j) \in E} M_{i,j}$. We will also assume that (1) is feasible. In such *distributed convex optimization problems* the terms $\mathbf{A}_{i|j}$ and $\mathbf{b}_{i,j}$ impose affine constraints between neighboring nodes.

The prototype problem in (1) includes, as a subset, the family of distributed consensus problems that minimize the sum of the local cost functions under network wide consensus constraints. The algorithm presented in this paper can therefore be used for this purpose.

B. Exploiting Separability Via Lagrangian Duality

Given the prototype problem in (1), the design of our distributed solver aims to address the coupling between the set of primal variables $\mathbf{x}_i$ due to the linear constraints. Echoing classic approaches in the literature, we can overcome this point via Lagrangian duality. In particular, the Lagrange dual problem of (1) is given by

$$\min_{\boldsymbol{\nu}} \sum_{i \in V} \left(f_i^* \left(\sum_{j \in \mathcal{N}(i)} \mathbf{A}_{i|j}^T \boldsymbol{\nu}_{i,j} \right) - \sum_{j \in \mathcal{N}(i)} \frac{\mathbf{b}_{i,j}}{2}^T \boldsymbol{\nu}_{i,j} \right), \tag{2}$$

where each $\boldsymbol{\nu}_{i,j} \in \mathbb{R}^{M_{i,j}}$ denotes the dual vector variable associated with the constraint at edge (i,j) and f_i^* is the Fenchel conjugate of f_i. By inspection, the resulting problem is still separable over the set of nodes but unfortunately each $\boldsymbol{\nu}_{i,j}$ in (2) is utilized in two conjugate functions, f_i^* and f_j^*, resulting in a coupling between neighboring nodes.

To decouple the objective terms, we can lift the dimension of the dual problem by introducing copies of each $\boldsymbol{\nu}_{i,j}$ at nodes i and j. The pairs of additional directed edge variables are denoted by $\boldsymbol{\lambda}_{i|j}, \boldsymbol{\lambda}_{j|i}$ $\forall (i,j) \in E$ and are associated with nodes i and j respectively. To ensure equivalence of the problems, these variables are constrained so that at optimality $\boldsymbol{\lambda}_{i|j} = \boldsymbol{\lambda}_{j|i}$. The resulting problem is referred to as the *extended dual* of Eq (1) and is given by

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & \sum_{i \in V} \left(f_i^* \left(\sum_{j \in \mathcal{N}(i)} \mathbf{A}_{i|j}^T \boldsymbol{\lambda}_{i|j} \right) - \sum_{j \in \mathcal{N}(i)} \frac{\mathbf{b}_{i,j}}{2}^T \boldsymbol{\lambda}_{i|j} \right) \\ \text{s.t.} \quad & \boldsymbol{\lambda}_{i|j} = \boldsymbol{\lambda}_{j|i} \quad \forall i \in V, j \in \mathcal{N}(i). \end{aligned} \tag{3}$$

The proposed lifting is appealing from the perspective of alternating minimization techniques as it partitions the resulting problem into two sections: a fully node separable objective function and a set of edge based constraints.

C. Simplification of Notation

To assist in the derivation of our algorithm, we firstly introduce a compact vector notation for Eq. (3). Specifically we will show that (3) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & f^*(\mathbf{C}^T \boldsymbol{\lambda}) - \mathbf{d}^T \boldsymbol{\lambda} \\ \text{s.t} \quad & (\mathbf{I} - \mathbf{P})\, \boldsymbol{\lambda} = \mathbf{0}. \end{aligned} \tag{4}$$

1) Dual Vector Notation: Firstly we introduce the dual variable $\boldsymbol{\lambda}$ as the stacked vector of the set of $\boldsymbol{\lambda}_{i|j}$ where the ordering of this stacking is given by $1|2 < 1|3 < \cdots < 1|N < 2|1 < 2|3 < \cdots < N|N-1$. In particular, $\boldsymbol{\lambda}$ is given by

$$\boldsymbol{\lambda} = \left[\boldsymbol{\lambda}_{1|2}^T, \cdots, \boldsymbol{\lambda}_{1|N}^T, \boldsymbol{\lambda}_{2|1}^T, \cdots, \boldsymbol{\lambda}_{N|N-1}^T \right]^T \in \mathbb{R}^{M_E}.$$

2) Compact Objective Notation: Given the definition of the dual vector $\boldsymbol{\lambda}$, we now move to simplifying the objective function. Firstly, we define the sum of local functions

$$f : \mathbb{R}^{M_V} \mapsto \mathbb{R}, \; \mathbf{x} \mapsto \sum_{i \in V} f_i(\mathbf{x}_i)$$

where $\mathbb{R}^{M_V} = \mathbb{R}^{M_1} \times \mathbb{R}^{M_2} \times ... \times \mathbb{R}^{M_N}$.

We can then define a matrix $\mathbf{C} \in \mathbb{R}^{M_E \times M_V}$ and vector $\mathbf{d} \in \mathbb{R}^{M_E}$ to rewrite our objective using $\boldsymbol{\lambda}$ and f. In particular,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{C}_N \end{bmatrix}, \; \mathbf{d} = \left[\mathbf{d}_1^T, \cdots, \mathbf{d}_N^T \right]^T,$$

where the components $\mathbf{C}_i$ and $\mathbf{d}_i$ are given by

$$\begin{aligned} \mathbf{C}_i &= \left[\mathbf{A}_{i|1}^T, \cdots, \mathbf{A}_{i|i-1}^T, \mathbf{A}_{i|i+1}^T, \cdots^T, \mathbf{A}_{i|N}^T \right]^T \forall i \in V, \\ \mathbf{d}_i &= \frac{1}{2} \left[\mathbf{b}_{i,1}^T, \cdots, \mathbf{b}_{i,i-1}^T, \mathbf{b}_{i,i+1}^T, \cdots, \mathbf{b}_{i,N}^T \right]^T \forall i \in V. \end{aligned}$$

The terms $\mathbf{A}_{i|j}$ and $\mathbf{b}_{i,j}$ are included in $\mathbf{C}_i$ and $\mathbf{d}_i$ respectively if only if $(i,j) \in E$.

The objective of Eq. (3) can therefore be rewritten as

$$f^*(\mathbf{C}^T \boldsymbol{\lambda}) - \mathbf{d}^T \boldsymbol{\lambda}.$$

3) Compact Constraints Notation: Similar to the objective, we can define an additional matrix to rewrite the constraint functions using our vector notation. For this task we introduce the symmetric permutation matrix $\mathbf{P} \in \mathbb{R}^{M_E \times M_E}$ that permutes each pair of variables $\boldsymbol{\lambda}_{i|j}$ and $\boldsymbol{\lambda}_{j|i}$. This allows the constraints in (3) to be rewritten as $(\mathbf{I} - \mathbf{P})\, \boldsymbol{\lambda} = \mathbf{0}$. The vector $\boldsymbol{\lambda}$ is therefore only feasible if it is contained in $\ker(\mathbf{I} - \mathbf{P})$.

D. From the Extended Dual Problem to a Nonexpansive PDMM Operator

Given the node and edge separable nature of the extended dual, we now move to forming a distributed optimization solver which takes advantage of this structure. In particular we aim to construct an operator of the form

$$\mathbf{S} = \mathbf{S}_E \circ \mathbf{S}_N,$$

where $\mathbf{S}_N$ and $\mathbf{S}_E$ are parallelizable over the nodes and edges respectively and $\circ$ is used to denote their composition so that $\forall\,(\mathbf{x},\mathbf{z}) \in \mathrm{gra}\,(\mathbf{S}_1 \circ \mathbf{S}_2)\,,\ \exists \mathbf{y} \mid (\mathbf{x},\mathbf{y}) \in \mathrm{gra}\,(\mathbf{S}_1)\,, (\mathbf{y},\mathbf{z}) \in \mathrm{gra}\,(\mathbf{S}_2)$. Furthermore, we would like such operators to be nonexpansive so that classic iterative solvers can be employed. The nonexpansiveness of an operator is defined as follows.

Definition III.1. *Nonexpansive Operators: An operator* $\mathbf{T} : \mathcal{X} \to \mathcal{Y}$ *is nonexpansive if*

$$\|\mathbf{u}-\mathbf{v}\| \leq \|\mathbf{x}-\mathbf{y}\| \quad (\mathbf{x},\mathbf{u})\,, (\mathbf{y},\mathbf{v}) \in \mathrm{gra}\,(\mathbf{T})\,,$$

We can construct such an $\mathbf{S}$ by making use of the relationship between monotone operators and the subdifferentials of convex functions. In particular, an operator is monotone if it satisfies the following definition.

Definition III.2. *Monotone Operators: An operator* $\mathbf{T} : \mathcal{X} \to \mathcal{Y}$ *is monotone iff*

$$\langle \mathbf{u}-\mathbf{v}, \mathbf{x}-\mathbf{y} \rangle \geq 0 \quad \forall\,(\mathbf{x},\mathbf{u})\,, (\mathbf{y},\mathbf{v}) \in \mathrm{gra}\,(\mathbf{T})\,,$$

Furthermore, $\mathbf{T}$ *is maximal monotone iff*

$$\nexists \text{ a monotone } \tilde{\mathbf{T}} : \mathcal{X} \to \mathcal{Y} \mid \mathrm{gra}(\mathbf{T}) \subset \mathrm{gra}(\tilde{\mathbf{T}}).$$

With these definitions in mind, consider the equivalent unconstrained form of (4) given by

$$\min_{\boldsymbol{\lambda}} \quad f^*(\mathbf{C}^T\boldsymbol{\lambda}) - \mathbf{d}^T\boldsymbol{\lambda} + \iota_{\ker(\mathbf{I}-\mathbf{P})}\,(\boldsymbol{\lambda})\,, \tag{5}$$

where $\iota_{\ker(\mathbf{I}-\mathbf{P})}$ is an indicator function defined as

$$\iota_{\ker(\mathbf{I}-\mathbf{P})}(\mathbf{y}) = \begin{cases} 0 & (\mathbf{I}-\mathbf{P})\mathbf{y} = \mathbf{0} \\ +\infty & \text{otherwise.} \end{cases}$$

As $\ker(\mathbf{I}-\mathbf{P})$ is a closed subspace, it follows from [46, Example 1.25] that $\iota_{\ker(\mathbf{I}-\mathbf{P})} \in \Gamma_0$. Furthermore, as $f \in \Gamma_0$, using [46, Theorem 13.32, Prop. 13.11], it follows that $f^*\left(\mathbf{C}^T\right) \in \Gamma_0$ as well. Due to our feasibility assumption of (1), the relative interiors of the domains of $f^*\left(\mathbf{C}^T\right)$ and $\iota_{\ker(\mathbf{I}-\mathbf{P})}$ share a common point. From [46, Theorem 16.3], it follows that $\boldsymbol{\lambda}^*$ is a minimizer of (5) if and only if

$$\mathbf{0} \in \mathbf{C}\partial f^*\left(\mathbf{C}^T\boldsymbol{\lambda}^*\right) - \mathbf{d} + \partial\iota_{\ker(\mathbf{I}-\mathbf{P})}\,(\boldsymbol{\lambda}^*)\,. \tag{6}$$

Note that the operators $\mathbf{T}_1 = \mathbf{C}\partial f^*\left(\mathbf{C}^T\right) - \mathbf{d}$ and $\mathbf{T}_2 = \partial\iota_{\ker(\mathbf{I}-\mathbf{P})}$ are by design separable over the set of nodes and edges respectively. Furthermore, $\mathbf{C}\partial f^*\left(\mathbf{C}^T\right)$ and $\partial\iota_{\ker(\mathbf{I}-\mathbf{P})}$ are the subdifferentials of CCP functions and thus are maximal monotone. A zero-point of (6) can therefore be found via a range of operator splitting methods (see [32] for an overview).

In this particular instance, we will use PR splitting to construct a nonexpansive PDMM operator by rephrasing the zero-point condition in (6) as a more familiar fixed-point condition. This equivalent condition, as demonstrated in [47] (Section 7.3), is given by

$$\mathbf{R}_{\mathbf{T}_2,\rho} \circ \mathbf{R}_{\mathbf{T}_1,\rho}\,(\mathbf{z}) = \mathbf{z},\ \boldsymbol{\lambda} = \mathbf{J}_{\mathbf{T}_1,\rho}\,(\mathbf{z})\,,$$

where $\mathbf{R}_{\mathbf{T}_i,\rho}$ and $\mathbf{J}_{\mathbf{T}_i,\rho}$ are the reflected resolvent and resolvent operators of $\mathbf{T}_i$ respectively. Here, the introduced $\mathbf{z}$ variables will be referred to as an *auxiliary* variables.

We define the PDMM operator as

$$\mathbf{T}_{P,\rho} = \mathbf{R}_{\mathbf{T}_2,\rho} \circ \mathbf{R}_{\mathbf{T}_1,\rho},$$

which will be used repeatedly throughout this work. Importantly given the nature of the operators considered, $\mathbf{T}_{P,\rho}$ is nonexpansive. Specifically, as both $\mathbf{T}_1$ and $\mathbf{T}_2$ are maximal monotone operators, $\mathbf{J}_{\mathbf{T}_1,\rho}$ and $\mathbf{J}_{\mathbf{T}_2,\rho}$ are both firmly nonexpansive. By [46, Proposition 4.2], it follows that $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{R}_{\mathbf{T}_2,\rho}$ are nonexpansive. The nonexpansiveness of $\mathbf{T}_{P,\rho}$ allows us to utilize fixed-point iterative methods to solve (3) and ultimately (1) in a distributed manner.

E. On the Link with the Primal Dual Method of Multipliers

We now demonstrate how PDMM, as defined in [23], can be linked with classical monotone operator splitting theory. For this purpose we will consider the fixed-point iteration of $\mathbf{T}_{P,\rho}$ given by

$$\mathbf{z}^{(k+1)} = \mathbf{T}_{P,\rho}\left(\mathbf{z}^{(k)}\right) = \mathbf{R}_{\mathbf{T}_2,\rho} \circ \mathbf{R}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^{(k)}\right). \tag{7}$$

To aid in the aforementioned relationship, the evaluation of the reflected resolvent operators $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{R}_{\mathbf{T}_2,\rho}$ are outlined in the following Lemmas.

Lemma III.1. $\mathbf{y}^{(k+1)} = \mathbf{R}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^{(k)}\right)$ *can be computed as*

$$\begin{aligned}
\mathbf{x}^{(k+1)} =& \arg\min_{\mathbf{x}} \left(f(\mathbf{x}) - \left\langle \mathbf{C}^T\mathbf{z}^{(k)}, \mathbf{x}\right\rangle + \frac{\rho}{2}||\mathbf{C}\mathbf{x}-\mathbf{d}||^2\right) \\
\boldsymbol{\lambda}^{(k+1)} =& \mathbf{z}^{(k)} - \rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right) \\
\mathbf{y}^{(k+1)} =& 2\boldsymbol{\lambda}^{(k+1)} - \mathbf{z}^{(k)}
\end{aligned}$$

A proof of this result can be found in Appendix A. Note that the block diagonal structure of $\mathbf{C}$ and the separability of f allow this reflected resolvent to be computed in parallel across the nodes.

Lemma III.2. $\mathbf{z}^{(k+1)} = \mathbf{R}_{\mathbf{T}_2,\rho}\left(\mathbf{y}^{(k+1)}\right)$ *can be computed as* $\mathbf{z}^{(k+1)} = \mathbf{P}\mathbf{y}^{(k+1)}$.

The proof for this result is included in Appendix B. The resulting permutation operation is equivalent to an exchange of auxiliary variables between neighboring nodes and is therefore distributable over the underlying network.

Utilizing Lemmas III.1 and III.2 it follows that

$$\mathbf{T}_{P,\rho} = \mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho}, \tag{8}$$

and thus that (7) is equivalent to

$$\mathbf{z}^{(k+1)} = \mathbf{P}\left(\mathbf{z}^{(k)} - 2\rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right)\right). \tag{9}$$

By noting that $\mathbf{z}^{(k+1)} = \mathbf{P}\left(\boldsymbol{\lambda}^{(k+1)} - \rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right)\right)$, the dependence on $\mathbf{y}^{(k+1)}$ and $\mathbf{z}^{(k+1)}$ can be removed, reducing the scheme to that given in Algorithm 1.

Algorithm 1 Simplified PDMM

Initialise: $\boldsymbol{\lambda}^{(0)} \in \mathbb{R}^{M_E},\ \mathbf{x}^{(0)} \in \mathbb{R}^{M_V}$
for k=0,..., **do**
  $\mathbf{x}^{(k+1)} = \underset{\mathbf{x}}{\operatorname{argmin}}\big(f(\mathbf{x}) - \big\langle \mathbf{C}^T\mathbf{P}\boldsymbol{\lambda}^{(k)}, \mathbf{x}\big\rangle + \frac{\rho}{2}||\mathbf{C}\mathbf{x} + \mathbf{P}\mathbf{C}\mathbf{x}^{(k)} - 2\mathbf{d}||^2\big)$
  $\boldsymbol{\lambda}^{(k+1)} = \mathbf{P}\boldsymbol{\lambda}^{(k)} - \rho\left(\mathbf{C}\mathbf{x}^{(k+1)} + \mathbf{P}\mathbf{C}\mathbf{x}^{(k)} - 2\mathbf{d}\right)$
end for

This algorithm is identical to a particular instance of PDMM proposed in [23]. Thus, PDMM is equivalent to the fixed-point iteration of the PR splitting of the extended dual problem, linking the approach with a plethora of existing algorithms within the literature [34], [38], [48], [49].

The connection with PR splitting motivates why PDMM may converge faster than ADMM for some problems, as demonstrated in [23]. In particular, [44, Remark 4] notes that PR splitting provides the fastest bound on convergence even though it may not converge for general problems. Specifically, the strong convexity and Lipschitz continuity of the averaging problem considered in [23] supports this link.

The distributed nature of PDMM can be more easily visualized in Algorithm 2 where we have utilized the definitions of $\mathbf{C}$ and $\mathbf{d}$. Here the notation $\mathbf{Node}_j \leftarrow \mathbf{Node}_i(\bullet)$ indicates the transmission of data from node i to node j.

Algorithm 2 Distributed PDMM

Initialise: $\mathbf{z}^{(0)} \in \mathbb{R}^{M_E}$
for k=0,..., **do**
  for all $i \in V$ **do** ▷ Primal Update
    $\mathbf{x}_i^{(k+1)} = \arg\min_{\mathbf{x}_i}\Big(f_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}(i)}\Big(-\big\langle \mathbf{A}_{i|j}^T\mathbf{z}_{i|j}^{(k)}, \mathbf{x}_i\big\rangle + \frac{\rho}{2}||\mathbf{A}_{i|j}\mathbf{x}_i - \frac{\mathbf{b}_{i,j}}{2}||^2\Big)\Big)$
    for all $j \in \mathcal{N}(i)$ **do** ▷ Dual Update
      $\mathbf{y}_{i|j}^{(k+1)} = \mathbf{z}_{i|j}^{(k)} - 2\rho\left(\mathbf{A}_{i|j}\mathbf{x}_i^{(k+1)} - \frac{\mathbf{b}_{i,j}}{2}\right)$
    end for
  end for
  for all $i \in V, j \in \mathcal{N}(i)$ **do** ▷ Transmit Variables
    $\mathbf{Node}_j \leftarrow \mathbf{Node}_i(\mathbf{y}_{i|j}^{(k+1)})$
  end for
  for all $i \in V, j \in \mathcal{N}(i)$ **do** ▷ Auxiliary Update
    $\mathbf{z}_{i|j}^{(k+1)} = \mathbf{y}_{j|i}^{(k+1)}$
  end for
end for

Each iteration of the algorithm only requires one-way transmission of the auxiliary $\mathbf{z}$ variables between neighboring nodes. Thus, no direct collaboration is required between nodes during the computation of each iteration leading to an appealing mode of operation for use in practical networks.

F. On the Link with the Distributed Alternating Direction Method of Multipliers

Using the proposed monotone interpretation of PDMM we can also link its behavior with ADMM. While in [23] it was suggested that these two methods were fundamentally different due to their contrasting derivations, in the following we demonstrate how they are more closely related than first thought. Interestingly, this link is masked via the change of variables typically used in the updating scheme for ADMM and PDMM (see [34, Sec. 3] and [23, Sec. 4] respectively for such representations). For this purpose we re-derive an ADMM variant from the perspective of monotone operator theory.

To begin, consider the prototype ADMM problem given by

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{c}. \end{aligned} \tag{10}$$

We can recast (1), in the form of (10) by introducing the additional variables $\mathbf{y}_{i|j}, \mathbf{y}_{j|i} \in \mathbb{R}^{M_{i,j}}\ \forall (i,j) \in E$ so that

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i\in V} f_i(\mathbf{x}_i) \\ \text{s.t} \quad & \left.\begin{aligned} & \mathbf{A}_{i|j}\mathbf{x}_i - \tfrac{\mathbf{b}_{i,j}}{2} = \mathbf{y}_{i|j} \\ & \mathbf{A}_{j|i}\mathbf{x}_j - \tfrac{\mathbf{b}_{i,j}}{2} = \mathbf{y}_{j|i} \\ & \mathbf{y}_{i|j} + \mathbf{y}_{j|i} = \mathbf{0} \end{aligned}\right\} \forall (i,j) \in E. \end{aligned} \tag{11}$$

Defining the stacked vector $\mathbf{y} \in \mathbb{R}^{M_E}$ and adopting the matrices $\mathbf{C}$, $\mathbf{P}$ and $\mathbf{d}$ as per Sec. III-C, (11) can be more simply written as

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) + \iota_{\ker(\mathbf{I}+\mathbf{P})}(\mathbf{y}) \\ \text{s.t} \quad & \mathbf{C}\mathbf{x} - \mathbf{d} = \mathbf{y}. \end{aligned} \tag{12}$$

Here, the indicator function is used to capture the final set of equality constraints in (11). It follows that (12) is exactly in the form of (10) so that ADMM can be applied.

The ADMM algorithm is equivalent to applying Douglas Rachford (DR) splitting [50] to the dual of (12), given by

$$\min_{\boldsymbol{\lambda}_i\ \forall\, i\in V} \quad f^*\left(\mathbf{C}^T\boldsymbol{\lambda}\right) - \mathbf{d}^T\boldsymbol{\lambda} + \iota^*_{\ker(\mathbf{I}+\mathbf{P})}(\boldsymbol{\lambda}), \tag{13}$$

where $\boldsymbol{\lambda}$, as in the case of PDMM, denotes the stacked vector of dual variables associated with the directed edges.

Comparing (13) and (6), we can note that the apparent difference in the dual problems is due to the use of $\iota_{\ker(\mathbf{I}-\mathbf{P})}$, in the case of PDMM, or $\iota^*_{\ker(\mathbf{I}+\mathbf{P})}$ in the case of ADMM. In actual fact these two functions are equal which stems from the definition of the Fenchel conjugate of an indicator function,

$$\begin{aligned} \iota^*_{\ker(\mathbf{I}+\mathbf{P})}(\boldsymbol{\lambda}) &= \sup_{\mathbf{y}}\left(\langle \mathbf{y}, \boldsymbol{\lambda}\rangle - \iota_{\ker(\mathbf{I}+\mathbf{P})}(\mathbf{y})\right) \\ &= \begin{cases} 0 & \boldsymbol{\lambda} \in \operatorname{ran}(\mathbf{I}+\mathbf{P}) \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

As $\operatorname{ran}(\mathbf{I}+\mathbf{P}) = \ker(\mathbf{I}-\mathbf{P})$, it follows that $\iota^*_{\ker(\mathbf{I}+\mathbf{P})} = \iota_{\ker(\mathbf{I}-\mathbf{P})}$. The problems in (5) and (13) are therefore identical.

As DR splitting is equivalent to a half averaged form of PR splitting [46], the operator form of ADMM is therefore given by $\mathbf{T}_{A,\rho} = \frac{1}{2}(\mathbf{I} + \mathbf{T}_{P,\rho})$. In this manner, despite their differences in earlier derivations, ADMM and PDMM are fundamentally linked. Within the literature, PDMM could therefore also be referred to as a particular instance of *generalised* [51] or *relaxed* ADMM [44].

IV. General Convergence Results for PDMM

Having linked PDMM with PR splitting, we now move to demonstrate convergence results for the algorithm. In particular we demonstrate a proof of convergence for PDMM for strongly convex and differentiable functions. This proof is required due to the fact that the strong monotonicity of either $\mathbf{T}_1$ or $\mathbf{T}_2$, usually required to guarantee convergence of PR splitting, cannot be guaranteed for PDMM due to the row rank deficiency of the matrix $\mathbf{C}$. We also highlight the use of operator averaging to guarantee convergence for all $f \in \Gamma_0$ and demonstrate its necessity with an analytic example where PDMM fails to converge.

A. Convergence of the Primal Error ($\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2$) of PDMM

The first result we demonstrate is that of the *primal convergence* of PDMM. In particular, we show that the sequence of primal iterates $\left(\mathbf{x}^{(k)}\right)_{k\in\mathbb{N}}$ converges to an optimal state, i.e.,

$$\exists \mathbf{x}^* \in \mathbf{X}^* \mid \|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2 \to 0. \tag{14}$$

where $\mathbf{X}^*$ denotes the set of primal optimizers of (1) and $\bullet \to \bullet$ denotes convergence. The term $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2$ will be referred to as the *primal error* from here on.

Many of the arguments used in this section make use of the notions of the kernel and range space of non-square matrices. These properties are defined below.

Definition IV.1. *Range Space and Kernel Space: Given a matrix* $\mathbf{A}$*, the range space of* $\mathbf{A}$ *is denoted by* $\operatorname{ran}(\mathbf{A})$ *where*

$$\forall \mathbf{y} \in \operatorname{ran}(\mathbf{A}), \exists \mathbf{u} \mid \mathbf{A}\mathbf{u} = \mathbf{y}.$$

Similarly, the kernel space of $\mathbf{A}$ *is denoted by* $\ker(\mathbf{A})$ *where*

$$\forall \mathbf{y} \in \ker(\mathbf{A}), \mathbf{A}\mathbf{y} = \mathbf{0}.$$

For any matrix, the subspaces $\operatorname{ran}(\mathbf{A})$ *and* $\ker\left(\mathbf{A}^T\right)$ *are orthogonal and, furthermore, their direct sum* $\operatorname{ran}(\mathbf{A}) + \ker\left(\mathbf{A}^T\right)$ *spans the entire space.*

To demonstrate that (14) holds, we can make use of the relationship between the primal $\mathbf{x}$ and auxiliary $\mathbf{z}$ variables of PDMM. In particular, we will demonstrate that both the primal and auxiliary variables converge by ultimately showing that

$$\exists \mathbf{z}^* \in \operatorname{fix}(\mathbf{T}_{P,\rho}) \mid \|\mathbf{z}^{(k)} - \mathbf{z}^*\|^2 \to 0,$$

which we will refer to as *auxiliary convergence*.

B. Primal Independence of a Non-Decreasing Subspace

To prove auxiliary convergence, other approaches in the literature often leverage additional operational properties such as strict nonexpansiveness. Unfortunately, in the case of PDMM, $\mathbf{T}_{P,\rho}$ is at best nonexpansive due to the presence of a non-decreasing component. Fortunately, this particular component does not influence the computation of the primary variables and ultimately can be ignored.

To demonstrate that PDMM is at best nonexpansive, consider the equation for two successive updates given by

$$\begin{aligned}\mathbf{z}^{(k+2)} =& \mathbf{T}_{P,\rho} \circ \mathbf{T}_{P,\rho}\left(\mathbf{z}^{(k)}\right)\\ =& \mathbf{T}_{P,\rho}\left(\mathbf{P}\left(\mathbf{z}^{(k)} - 2\rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right)\right)\right)\\ =& \mathbf{z}^{(k)} - 2\rho\left(\mathbf{P}\mathbf{C}\mathbf{x}^{(k+2)} + \mathbf{C}\mathbf{x}^{(k+1)} - 2\mathbf{d}\right),\end{aligned} \tag{15}$$

where the second and third lines use the PDMM update in (9). From our feasibility assumption of (1), $\exists \mathbf{x}^* \mid \mathbf{P}\mathbf{C}\mathbf{x}^* + \mathbf{C}\mathbf{x}^* = 2\mathbf{d}$ so that $\mathbf{d} \in \operatorname{ran}(\mathbf{P}\mathbf{C}) + \operatorname{ran}(\mathbf{C})$. Therefore, every two PDMM updates only affect the auxiliary variables in the subspace $\operatorname{ran}(\mathbf{P}\mathbf{C}) + \operatorname{ran}(\mathbf{C})$. By considering the projection of each iterate onto the orthogonal subspace of $\operatorname{ran}(\mathbf{P}\mathbf{C}) + \operatorname{ran}(\mathbf{C})$, which is given by $\ker\left(\mathbf{C}^T\right) \cap \ker\left(\mathbf{C}^T\mathbf{P}\right)$, it follows that, for all even k,

$$\begin{aligned}\prod_{\ker(\mathbf{C}^T)\cap\ker(\mathbf{C}^T\mathbf{P})}\left(\mathbf{z}^{(k+2)}\right) =& \prod_{\ker(\mathbf{C}^T)\cap\ker(\mathbf{C}^T\mathbf{P})}\left(\mathbf{z}^{(k)}\right)\\ =& \prod_{\ker(\mathbf{C}^T)\cap\ker(\mathbf{C}^T\mathbf{P})}\left(\mathbf{z}^{(0)}\right),\end{aligned}$$

where $\prod_{\mathcal{A}}$ denotes the orthogonal projection onto $\mathcal{A}$.

Every even-numbered auxiliary iterate $\mathbf{z}^{(k)}$ contains a non-decreasing component determined by our initial choice of $\mathbf{z}^{(0)}$. Fortunately, from Lemma III.1 it is clear that each $\mathbf{x}^{(k)}$ is independent of $\prod_{\ker(\mathbf{C}^T)}\left(\mathbf{z}^{(k)} + \rho\mathbf{d}\right)$. As $\ker\left(\mathbf{C}^T\right) \cap \ker\left(\mathbf{C}^T\mathbf{P}\right) \subseteq \ker\left(\mathbf{C}^T\right)$, any signal in the non-decreasing subspace of $\mathbf{T}_{P,\rho} \circ \mathbf{T}_{P,\rho}$ will not play a role in the primal updates. For proving primal convergence, we will therefore consider the *projected auxiliary error*

$$\|\prod_{\operatorname{ran}(\mathbf{C})+\operatorname{ran}(\mathbf{P}\mathbf{C})}\left(\mathbf{z}^{(k)} - \mathbf{z}^*\right)\|^2. \tag{16}$$

Such a projection can be easily computed for even iterates due to the structure noted in (15) by defining the vector

$$\mathbf{z}^\diamond = \mathbf{z}^* + \prod_{\ker(\mathbf{C}^T)\cap\ker(\mathbf{C}^T\mathbf{P})}\left(\mathbf{z}^{(0)}\right). \tag{17}$$

From the nonexpansiveness of PDMM, the projected auxiliary error satisfies

$$\|\mathbf{z}^{(k+2)} - \mathbf{z}^\diamond\| \leq \|\mathbf{z}^{(k)} - \mathbf{z}^\diamond\|.$$

The sequence $\left(\mathbf{z}^{(2k)}\right)_{k\in\mathbb{N}}$ is therefore Fejér monotone with respect to $\mathbf{z}^\diamond$ and thus the sequence $\left(\|\mathbf{z}^{(2k)} - \mathbf{z}^\diamond\|\right)_{k\in\mathbb{N}}$ converges [46, Proposition 5.4]. To prove projected auxiliary convergence, all that remains is to show that

$$\lim_{k\to\infty}\left(\mathbf{z}^{(2k)} - \mathbf{z}^\diamond\right) = 0. \tag{18}$$

C. Optimality of Auxiliary Limit Points

We will now demonstrate that (18) holds in the specific case of strongly convex and differentiable functions, in turn allowing us to prove primal convergence. While the differentiability of a function is straightforward, the notion of strong convexity is defined below.

Definition IV.2. *Strong Convexity: A function f is μ-strongly convex with $\mu > 0$ iff $\forall \theta \in [0,1], \mathbf{x}, \mathbf{y} \in dom\,(f)$,*

$$f\left(\theta \mathbf{x} + (1-\theta)\mathbf{y}\right) \leq \theta f\left(\mathbf{x}\right) + (1-\theta) f\left(\mathbf{y}\right) - \mu\theta(1-\theta)\|\mathbf{x}-\mathbf{y}\|^2$$

Additionally, if f is μ-strongly convex, ∂f is μ-strongly monotone.

Definition IV.3. *Strongly Monotone: An operator $\mathbf{T} : \mathcal{X} \to \mathcal{Y}$ is μ-strongly monotone with $\mu > 0$, if*

$$\langle \mathbf{u}-\mathbf{v}, \mathbf{x}-\mathbf{y}\rangle \geq \mu\|\mathbf{x}-\mathbf{y}\|^2 \quad \forall\left(\mathbf{x},\mathbf{u}\right),\left(\mathbf{y},\mathbf{v}\right) \in \operatorname{gra}\left(\mathbf{T}\right).$$

To verify that (18) holds under the aforementioned assumptions, we make use of the following Lemma relating to the limit points of the primal and dual variables.

Lemma IV.1. *If f is differentiable and μ-strongly convex then*

$$\lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}^*,$$
$$\lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})}\left(\boldsymbol{\lambda}^{(k)}\right) = \prod_{\operatorname{ran}(\mathbf{C})}\left(\boldsymbol{\lambda}^*\right).$$

The proof for this Lemma can be found in Appendix C.

Using Lemma IV.1, and rearranging the dual update equation in Lemma III.1, it follows that

$$\begin{aligned}\lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})}\left(\mathbf{z}^{(k)}\right) &= \lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})}\Big(\boldsymbol{\lambda}^{(k+1)} \\ &\quad + \rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right)\Big) \\ &= \prod_{\operatorname{ran}(\mathbf{C})}\left(\boldsymbol{\lambda}^* + \rho\left(\mathbf{C}\mathbf{x}^* - \mathbf{d}\right)\right) \\ &= \prod_{\operatorname{ran}(\mathbf{C})}\left(\mathbf{z}^*\right).\end{aligned} \tag{19}$$

From (19), if also follows that

$$\begin{aligned}\mathbf{0} &= \lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})}\left(\mathbf{z}^{(k+1)} - \mathbf{z}^*\right) \\ &= \lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})} \mathbf{P}\left(\mathbf{z}^{(k)} - \mathbf{z}^* - 2\rho\mathbf{C}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^*\right)\right) \\ &= \lim_{k\to\infty} \mathbf{P} \prod_{\operatorname{ran}(\mathbf{C})} \mathbf{P}\left(\mathbf{z}^{(k)} - \mathbf{z}^*\right) \\ &= \lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{P}\mathbf{C})}\left(\mathbf{z}^{(k)} - \mathbf{z}^*\right),\end{aligned} \tag{20}$$

where the second line uses Eq. (9), the third line uses that $\lim_{k\to\infty} \mathbf{x}^{(k+1)} = \mathbf{x}^*$ and that $\mathbf{P}$ is full rank, while the last line exploits that $\mathbf{P} = \mathbf{P}^{-1}$ such that $\mathbf{P} \prod_{\operatorname{ran}(\mathbf{C})} \mathbf{P} = \prod_{\operatorname{ran}(\mathbf{P}\mathbf{C})}$. Combining (19) and (20), finally demonstrates that, under the restrictions of strong convexity and differentiability of f, that

$$\lim_{k\to\infty} \prod_{\operatorname{ran}(\mathbf{C})+\operatorname{ran}(\mathbf{P}\mathbf{C})}\left(\mathbf{z}^{(2k)} - \mathbf{z}^*\right) = \lim_{k\to\infty}\left(\mathbf{z}^{(2k)} - \mathbf{z}^\diamond\right) = \mathbf{0}.$$

Primal convergence follows from Lemma III.1, by noting,

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \left(\nabla f + \rho\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\left(\mathbf{z}^{(k)} + \rho\mathbf{d}\right) \\ \mathbf{x}^* &= \left(\nabla f + \rho\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\left(\mathbf{z}^* + \rho\mathbf{d}\right).\end{aligned} \tag{21}$$

The equality in this case follows from the fact that ∇f is μ-strongly monotone such that $\left(\nabla f + \rho\mathbf{C}^T\mathbf{C}\right)^{-1}$ is Lipschitz continuous and thus single-valued. Substituting (21) into the primal error, it follows that

$$\begin{aligned}\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|^2 =& \|\left(\nabla f + \rho\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\left(\mathbf{z}^{(k)} + \rho\mathbf{d}\right) \\ &- \left(\nabla f + \rho\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\left(\mathbf{z}^* + \rho\mathbf{d}\right)\|^2 \\ \leq& \frac{1}{\mu^2}\|\mathbf{C}^T\left(\mathbf{z}^{(k)} - \mathbf{z}^*\right)\|^2 \\ \leq& \frac{\sigma^2_{\max}\left(\mathbf{C}\right)}{\mu^2}\|\mathbf{z}^{(k)} - \mathbf{z}^\diamond\|^2,\end{aligned} \tag{22}$$

where, $\sigma_{\max}$ denotes the largest singular value of a matrix.

The primal error $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|^2$ is therefore upper bounded by the projected auxiliary error and thus converges.

D. Averaged PDMM Convergence

As with other operator splitting methods, PDMM can be combined with an averaging stage to guarantee convergence $\forall f \in \Gamma_0$, even those which do not satisfy the strong convexity or differentiability assumptions introduced in Sec. IV-C. The general form of the averaged PDMM operator is given by

$$\mathbf{T}_{P,\rho,\alpha} = (1-\alpha)\mathbf{I} + \alpha\mathbf{T}_{P,\rho},$$

where the scalar $\alpha \in (0,1)$. In the particular case that $\alpha = \frac{1}{2}$, averaged PDMM is equivalent to ADMM, as was previously noted in Sec. III-F. In this case, by [46, Proposition 4.4], the operator $\mathbf{T}_{P,\rho,\alpha}$ is firmly nonexpansive.

The fixed-point iteration of $\mathbf{T}_{P,\rho,\alpha}$ is therefore given by

$$\mathbf{z}^{(k+1)} = (1-\alpha)\mathbf{z}^{(k)} + \alpha\mathbf{T}_{P,\rho}\mathbf{z}^{(k)}.$$

This is referred to as the α-Krasnosel'skiĭ-Mann iteration [46] of the operator $\mathbf{T}_{P,\rho}$ which is a well documented method of guaranteeing convergence for nonexpansive operators. Notably, recursively applying [46, Eq. 5.16], it follows that the fixed-point residual $\left(\mathbf{T}_{P,\rho} - \mathbf{I}\right)\left(\mathbf{z}^{(k)}\right)$ converges at an asymptotic rate of $\mathcal{O}\left(\frac{1}{k}\right)$ and thus $\mathbf{z}^{(k)}$ converges to a point in $\operatorname{fix}\left(\mathbf{T}_{P,\rho}\right)$ for finite dimensional problems.

E. Lack of Convergence of PDMM for $f \in \Gamma_0$

Without the use of averaging, the convergence results demonstrated so far require f to be both strongly convex and differentiable. While such a result is well known in the case of PR splitting, it is not noted in the existing analysis of PDMM within the literature [23].

In the following, we reinforce the importance of this result by demonstrating a problem instance were PDMM does not converge despite $f \in \Gamma_0$. For this purpose we consider solving the following problem over two nodes.

$$\begin{aligned}\min_{x_1,x_2} \quad & |x_1 - 1| + |x_2 + 1| \\ \text{s.t.} \quad & x_1 - x_2 = 0.\end{aligned} \tag{23}$$

The objective in (23) is neither differentiable nor strongly convex. From Lemmas III.1 and III.2, the primal and auxiliary updates for PDMM are given respectively by

$$\begin{aligned}x_1^{(t+1)} =& \operatorname{argmin}_x\left(|x-1| - z_{1|2}^{(t)}x + \frac{\rho}{2}\|x\|^2\right), \\ x_2^{(t+1)} =& \operatorname{argmin}_x\left(|x+1| + z_{2|1}^{(t)}x + \frac{\rho}{2}\|x\|^2\right), \\ z_{1|2}^{(t+1)} =& z_{2|1}^{(t)} + 2\rho x_2^{(t+1)},\ z_{2|1}^{(t+1)} = z_{1|2}^{(t)} - 2\rho x_1^{(t+1)},\end{aligned} \tag{24}$$

By setting $z_{1|2}^{(0)} = z_{2|1}^{(0)} = 0$ and $\rho = 1$ it follows from (24) that after the first iteration $x_1^{(1)} = -x_2^{(1)} = 1$ and $z_{1|2}^{(1)} = z_{2|1}^{(1)} = 2$. Note that $x_1 \neq x_2$ such that $\mathbf{x}$ is not primal feasible.

For the second iteration $x_1^{(2)} = -x_2^{(2)} = -1$ and $z_{1|2}^{(2)} = z_{2|1}^{(2)} = 0$. Again, $x_1 \neq x_2$ and furthermore the auxiliary variables are back to their original configuration. The auxiliary variables of PDMM are therefore stuck in a limit cycle and can never converge for this problem. The primal variables also exhibit a limit cycle in this case. As such, $f \in \Gamma_0$ is not a sufficient condition for the convergence of PDMM without the use of operator averaging.

V. Geometric Convergence and Distributed Parameter Selection

While PR splitting is well known to converge geometrically under the assumption of strong monotonicity and Lipschitz continuity, such conditions cannot be guaranteed in the case of PDMM due to the row rank deficiency of $\mathbf{C}$. However, by assuming that f is strongly convex and has a Lipschitz continuous gradient, we can demonstrate a geometrically contracting upper bound for the primal error of PDMM despite this fact.

A. A Primal Geometric Convergence Bound for Strongly Convex and Smooth Functions

In the following we demonstrate that for strongly convex functions with Lipschitz continuous gradients, the primal variables of PDMM converge at a geometric rate. More formally we show that $\exists\, \epsilon \geq 0, \gamma \in [0,1)$ so that

$$\forall k \in \mathbb{N},\ \|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2 \leq \gamma^k \epsilon.$$

As in the case of Section IV-A, this is achieved by firstly forming a geometric bound for the projected auxiliary error

$$\|\underset{\mathrm{ran}(\mathbf{C})+\mathrm{ran}(\mathbf{PC})}{\Pi}\left(\mathbf{z}^{(k)} - \mathbf{z}^*\right)\|^2 = \|\mathbf{z}^{(k)} - \mathbf{z}^\diamond\|^2,$$

before linking back to the primal variables.

The process of bounding the projected auxiliary error is broken down into two stages. Firstly, in Sections V-B and V-C we demonstrate how, for strongly convex functions with Lipschitz continuous gradients, PDMM is contractive over a subspace. In Sections V-D and V-E we then show how a geometric convergence bound can be found by linking PDMM with a generalized form of the alternating method of projections allowing us to derive the aforementioned γ and ϵ.

B. Contractive Nature of PDMM Over a Subspace

Proving that the projected auxiliary error of PDMM converges geometrically relies on strong monotonicity and the additional notion of Lipschitz continuity. This is defined as follows.

Definition V.1. *Lipschitz Continuous: An operator $\mathbf{T} : \mathcal{X} \to \mathcal{Y}$ is L-Lipschitz if*

$$\|\mathbf{u} - \mathbf{v}\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\, (\mathbf{x}, \mathbf{u}), (\mathbf{y}, \mathbf{v}) \in \mathrm{gra}\,(\mathbf{T}).$$

If $L = 1$, $\mathbf{T}$ is nonexpansive while if $L < 1$ it is contractive.

Given this notion, we demonstrate the contractive nature of the PDMM operator over $\mathrm{ran}\,(\mathbf{C})$ by showing that $\mathbf{C}\nabla f^*(\mathbf{C}^T\bullet)$ is strongly monotone and Lipschitz continuous over this subspace. This is summarized in Lemma V.1.

Lemma V.1. *If f is μ-strongly convex and ∇f is β-Lipschitz continuous then $\mathbf{C}\nabla f^*(\mathbf{C}^T\bullet)$ is*

(i) $\frac{\sigma_{\max}^2(\mathbf{C})}{\mu}$-Lipschitz continuous

(ii) $\frac{\sigma_{\min\neq 0}^2(\mathbf{C})}{\beta}$-strongly monotone $\forall \mathbf{z} \in \mathrm{ran}\,(\mathbf{C})$,

where $\sigma_{\min\neq 0}$ denotes the smallest non-zero singular value.

The proof of this lemma can be found in Appendix D. Lemma V.1 reflects a similar approach in [44] for general PR splitting problems. Note that the result demonstrated therein does not hold in this context due to the row-rank deficiency of $\mathbf{C}$. Specifically, [44, Assumption 2] is violated.

As $\mathbf{C}\nabla f^*(\mathbf{C}^T\bullet)$ is both strongly monotone and Lipschitz continuous over $\mathrm{ran}\,(\mathbf{C})$, from [44], $\mathbf{R}_{\mathbf{T}_1,\rho}$ is contractive $\forall \mathbf{z} \in \mathrm{ran}\,(\mathbf{C})$ with an upper bound on this contraction given by

$$\delta = \max\left(\frac{\rho\frac{\sigma_{\max}^2(\mathbf{C})}{\mu} - 1}{\rho\frac{\sigma_{\max}^2(\mathbf{C})}{\mu} + 1}, \frac{1 - \rho\frac{\sigma_{\min\neq 0}^2(\mathbf{C})}{\beta}}{1 + \rho\frac{\sigma_{\min\neq 0}^2(\mathbf{C})}{\beta}}\right) \in [0, 1).$$

By the same arguments, the operator $\mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}$ is δ contractive over $\mathrm{ran}\,(\mathbf{PC})$. Using the definition of the PDMM operator (8), the two-step PDMM updates given in (15), can equivalently be written as

$$\mathbf{z}^{(k+2)} = (\mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}) \circ \mathbf{R}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^{(k)}\right).$$

Every two PDMM iterations is therefore the composition of the operators $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}$ with each being δ-contractive over $\mathrm{ran}\,(\mathbf{C})$ and $\mathrm{ran}\,(\mathbf{PC})$ respectively.

C. Inequalities due to the Contraction of PDMM

The contractive nature of $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}$ leads to two important inequalities. In this case we will assume that k is even and that $\mathbf{z}^\diamond$ is defined as per (17).

Beginning with the operator $\mathbf{R}_{\mathbf{T}_1,\rho}$, consider the updates $\mathbf{y}^\diamond = \mathbf{R}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^\diamond\right)$ and $\mathbf{y}^{(k+1)} = \mathbf{R}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^{(k)}\right)$. Using Lemma III.1, it follows that

$$\begin{aligned}\mathbf{y}^{(k+1)} - \mathbf{y}^\diamond =& 2\boldsymbol{\lambda}^{(k+1)} - \mathbf{z}^{(k)} - (2\boldsymbol{\lambda}^\diamond - \mathbf{z}^\diamond)\\ =& \mathbf{z}^{(k)} - \mathbf{z}^\diamond - 2\rho\mathbf{C}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^*\right),\end{aligned}$$

so that the projection onto $\ker(\mathbf{C}^T)$ satisfies

$$\underset{\ker(\mathbf{C}^T)}{\Pi}\left(\mathbf{y}^{(k+1)} - \mathbf{y}^\diamond\right) = \underset{\ker(\mathbf{C}^T)}{\Pi}\left(\mathbf{z}^{(k)} - \mathbf{z}^\diamond\right).$$

Combining with the δ-contractive nature of $\mathbf{R}_{\mathbf{T}_1,\rho}$ over $\mathrm{ran}\,(\mathbf{C})$, it follows that,

$$\begin{aligned}\|\mathbf{y}^{(k+1)} - \mathbf{y}^\diamond\|^2 \leq \delta^2\|&\underset{\mathrm{ran}(\mathbf{C})}{\Pi}\left(\mathbf{z}^{(k)} - \mathbf{z}^\diamond\right)\|^2\\ &+\|\underset{\ker(\mathbf{C}^T)}{\Pi}\left(\mathbf{z}^{(k)} - \mathbf{z}^\diamond\right)\|^2.\end{aligned}$$

For the operator $\mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}$, as $\mathbf{z}^\diamond = \mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}\left(\mathbf{y}^\diamond\right)$ by the results of Section IV-B and $\mathbf{z}^{(k+2)} = \mathbf{P} \circ \mathbf{R}_{\mathbf{T}_1,\rho} \circ \mathbf{P}\left(\mathbf{y}^{(k+1)}\right)$, it can be similarly shown that

$$\underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\left(\mathbf{z}^{(k+2)} - \mathbf{z}^\diamond\right) = \underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\left(\mathbf{y}^{(k+1)} - \mathbf{y}^\diamond\right),$$

and furthermore that

$$\|\mathbf{z}^{(k+2)}-\mathbf{z}^{\diamond}\|^2 \leq \delta^2 \|\underset{\mathrm{ran}(\mathbf{PC})}{\Pi}\left(\mathbf{y}^{(k+1)}-\mathbf{y}^{\diamond}\right)\|^2 + \|\underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\left(\mathbf{y}^{(k+1)}-\mathbf{y}^{\diamond}\right)\|^2.$$

While the contractive nature of $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{P}\circ\mathbf{R}_{\mathbf{T}_1,\rho}\circ\mathbf{P}$ suggests the geometric convergence of PDMM, it is unclear what this convergence rate may be. In the following, this will be addressed by deriving a geometric error bound for two-step PDMM by connecting it with the method of alternating projections.

D. A Geometric Rate Bound for PDMM Interpreted as an Optimization Problem

Using the results of Section V-C we now demonstrate that $\exists\gamma$ so that the projected auxiliary error satisfies

$$\|\mathbf{z}^{(k+2)}-\mathbf{z}^{\diamond}\|^2 \leq \gamma^2\|\mathbf{z}^{(k)}-\mathbf{z}^{\diamond}\|^2, \tag{25}$$

where γ^2 can be computed via a non-convex optimization problem. Specifically, it is the maximum objective value of

$$\begin{aligned}
\max_{\mathbf{y},\mathbf{z},\hat{\mathbf{z}}} \quad & \|\hat{\mathbf{z}}-\mathbf{z}^{\diamond}\|^2 && \text{(26a)}\\
\text{s.t.} \quad & \mathbf{y}=\mathbf{R}_{\mathbf{T}_1,\rho}(\mathbf{z}) && \text{(26b)}\\
& \hat{\mathbf{z}}=\mathbf{P}\circ\mathbf{R}_{\mathbf{T}_1,\rho}\circ\mathbf{P}(\mathbf{y}) && \text{(26c)}\\
& \|\mathbf{z}-\mathbf{z}^{\diamond}\|^2 \leq 1. && \text{(26d)}
\end{aligned}$$

Here, (26a) captures the worst case improvement in the distance between the two-step iterates ($\hat{\mathbf{z}}$) and the projected fixed point ($\mathbf{z}^{\diamond}$). Due to (26d), the maximum of this objective exactly determines the worst case convergence rate. The vector $\mathbf{z}$ corresponds to the initial auxiliary variable, $\mathbf{y}$ and $\hat{\mathbf{z}}$ are generated via the one and two step PDMM updates imposed by (26b) and (26c), and (26d) defines the feasible set of $\mathbf{z}$. In a similar manner to (17), $\mathbf{z}^{\diamond}=\mathbf{z}^{*}+\underset{\ker(\mathbf{C}^T)\cap\ker(\mathbf{C}^T\mathbf{P})}{\Pi}(\mathbf{z})$ so that $\mathbf{z}-\mathbf{z}^{\diamond}\in\mathrm{ran}(\mathbf{PC})+\mathrm{ran}(\mathbf{C})$.

Using the properties of $\mathbf{R}_{\mathbf{T}_1,\rho}$ and $\mathbf{P}\circ\mathbf{R}_{\mathbf{T}_1,\rho}\circ\mathbf{P}$ from Sec. V-C, the optimum of (26) can be equivalently computed via

$$\begin{aligned}
\max_{\mathbf{y},\mathbf{z}} \quad & \|\left(\delta\underset{\mathrm{ran}(\mathbf{PC})}{\Pi}+\underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\right)(\mathbf{y}-\mathbf{y}^{\diamond})\|^2\\
\text{s.t.} \quad & \|\underset{\mathrm{ran}(\mathbf{C})}{\Pi}(\mathbf{y}-\mathbf{y}^{\diamond})\|^2 \leq \delta^2\|\underset{\mathrm{ran}(\mathbf{C})}{\Pi}(\mathbf{z}-\mathbf{z}^{\diamond})\|^2 && \text{(27a)}\\
& \underset{\ker(\mathbf{C}^T)}{\Pi}(\mathbf{y}-\mathbf{y}^{\diamond}) = \underset{\ker(\mathbf{C}^T)}{\Pi}(\mathbf{z}-\mathbf{z}^{\diamond}) && \text{(27b)}\\
& \|\mathbf{z}-\mathbf{z}^{\diamond}\|^2 \leq 1, && \text{(27c)}
\end{aligned}$$

where $\mathbf{y}^{\diamond}=\mathbf{R}_{\mathbf{T}_1,\rho}(\mathbf{z}^{\diamond})$ and in the objective we have exploited the orthogonality of $\mathrm{ran}(\mathbf{PC})$ and $\ker\left(\mathbf{C}^T\mathbf{P}\right)$. The constraints of (27) increase the feasible sets of $\mathbf{y}$ and $\hat{\mathbf{z}}$ while including the true updates due to $\mathbf{R}_{\mathbf{T}_1,\rho}$ as special cases.

The constraints (27a), (27b) and (27c) collectively define the feasible set of the vectors $\mathbf{y}-\mathbf{y}^{\diamond}$. We can further simplify (27) by considering the form of this feasible set. In particular, as (27c) denotes a sphere, the constraints (27a) and (27b) restrict the vectors $\mathbf{y}-\mathbf{y}^{\diamond}$ to lie in an ellipsoid given by

$$\mathbf{y}-\mathbf{y}^{\diamond}\in\left\{\left(\delta\underset{\mathrm{ran}(\mathbf{C})}{\Pi}+\underset{\ker(\mathbf{C}^T)}{\Pi}\right)\mathbf{u}\mid\|\mathbf{u}\|\leq 1\right\}.$$

By defining the additional variable $\mathbf{u}=\mathbf{z}-\mathbf{z}^{\diamond}$, the optimization problem in (26) is therefore equivalent to

$$\begin{aligned}
\max_{\mathbf{u}} \quad & \|\left(\delta\underset{\mathrm{ran}(\mathbf{PC})}{\Pi}+\underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\right)\left(\delta\underset{\mathrm{ran}(\mathbf{C})}{\Pi}+\underset{\ker(\mathbf{C}^T)}{\Pi}\right)\mathbf{u}\|^2\\
\text{s.t.} \quad & \|\mathbf{u}\|^2\leq 1, \mathbf{u}\in\mathrm{ran}(\mathbf{PC})+\mathrm{ran}(\mathbf{C}), && \text{(28)}
\end{aligned}$$

where the additional domain constraint stems from the definition of $\mathbf{z}^{\diamond}$. In the following we demonstrate how (28) exhibits an analytic expression for γ, ultimately allowing us to form our primal convergence rate bound.

E. Relationship with the Method Alternating of Projections

To compute the contraction factor γ in (25), we can exploit the relationship between (28) and the method of alternating projections. Optimal rate bounds for generalizations of the classic alternating projections algorithm has been an area of recent attention in the literature with two notable papers on the subject being [52] and [53]. Our analysis below follows in the spirit of these methods.

Consider the particular operator from Eq. (28),

$$\mathbb{G}=\left(\delta\underset{\mathrm{ran}(\mathbf{PC})}{\Pi}+\underset{\ker(\mathbf{C}^T\mathbf{P})}{\Pi}\right)\left(\delta\underset{\mathrm{ran}(\mathbf{C})}{\Pi}+\underset{\ker(\mathbf{C}^T)}{\Pi}\right).$$

Given the domain constraint also from (28), it follows that γ corresponds to the largest singular value of the matrix $\underset{\mathrm{ran}(\mathbf{C})+\mathrm{ran}(\mathbf{PC})}{\Pi}\mathbb{G}^T\mathbb{G}\underset{\mathrm{ran}(\mathbf{C})+\mathrm{ran}(\mathbf{PC})}{\Pi}$. We can therefore compute γ by taking advantage of the structure of $\mathbb{G}$. In particular, from [53], there exists an orthonormal matrix $\mathbf{D}$ such that

$$\underset{\mathrm{ran}(\mathbf{PC})}{\Pi}=\mathbf{D}\begin{bmatrix}\boldsymbol{\mathcal{C}}^2 & \boldsymbol{\mathcal{CS}} & \mathbf{0} & \mathbf{0}\\ \boldsymbol{\mathcal{CS}} & \boldsymbol{\mathcal{S}}^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{D}^H,\ \underset{\mathrm{ran}(\mathbf{C})}{\Pi}=\mathbf{D}\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{D}^H,$$

where $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{S}}$ denote diagonal matrices of the cosines and sines of the principal angles between $\mathrm{ran}(\mathbf{C})$ and $\mathrm{ran}(\mathbf{PC})$, respectively. It follow that for the considered operator

$$\mathbb{G}=\mathbf{D}\begin{bmatrix}\delta^2+\delta(1-\delta)\boldsymbol{\mathcal{S}}^2 & -(1-\delta)\boldsymbol{\mathcal{CS}} & \mathbf{0} & \mathbf{0}\\ -\delta(1-\delta)\boldsymbol{\mathcal{CS}} & (1-\delta)\boldsymbol{\mathcal{C}}^2+\delta & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \delta\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\end{bmatrix}\mathbf{D}^H.$$

Note that the bottom right identity matrix corresponds to those vectors that lie outside our feasible set.

Given the structure of $\mathbb{G}$ and the diagonal nature of $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{S}}$, it follows that γ is either given by δ or by $\sigma_{\max}$ of any of the two by two submatrices

$$\mathbb{G}_i=\begin{bmatrix}\delta^2+\delta(1-\delta)\mathcal{S}_i^2 & -(1-\delta)\mathcal{C}_i\mathcal{S}_i\\ -\delta(1-\delta)\mathcal{C}_i\mathcal{S}_i & \delta+(1-\delta)\mathcal{C}_i^2\end{bmatrix},$$

where $\mathcal{S}_i=\sin(\theta_i)$, $\mathcal{C}_i=\cos(\theta_i)$ and $\theta_i\in(0,\frac{\pi}{2}]$ is the ith principal angle. The singular values of such a submatrix can be computed via the following lemma.

Lemma V.2. *The singular values of $\mathbb{G}_i$ are given by*

$$\sigma(\mathbb{G}_i)=\sqrt{\delta^2+(1-\delta^2)\mathcal{C}_i\left(\frac{(1-\delta^2)\mathcal{C}_i}{2}\pm\sqrt{\frac{(1-\delta^2)^2\mathcal{C}_i^2}{4}+\delta^2}\right)}.$$

The proof for this lemma can be found in Appendix E. As the singular values are a nondecreasing function of $\mathcal{C}_i$ and thus a nonincreasing function of θ_i, it follows that

$$\gamma = \max\{\delta, \{\sigma_{\max}(\mathbb{G}_i)\ \forall i\}\} = \sigma_{\max}(\mathbb{G}_F). \tag{29}$$

Here $\mathbb{G}_F$ refers to the submatrix associated with the smallest non-zero principal angle θ_F, which is referred to as the Friedrichs angle. Therefore, given δ and $\mathcal{C}_F = \cos(\theta_F)$,

$$\gamma = \sqrt{\delta^2 + (1-\delta^2)\mathcal{C}_F\left(\frac{(1-\delta^2)\mathcal{C}_F}{2} + \sqrt{\frac{(1-\delta^2)^2\mathcal{C}_F^2}{4} + \delta^2}\right)}.$$

F. From an Auxiliary Error Bound to a Geometric Primal Convergence Bound

Using (29), our primal convergence bound can finally be constructed. For two-step PDMM we already know that

$$\|\mathbf{z}^{(k+2)} - \mathbf{z}^{\diamond}\|^2 \leq \gamma^2 \|\mathbf{z}^{(k)} - \mathbf{z}^{\diamond}\|^2.$$

By recursively applying this result, it follows that, for even k,

$$\begin{aligned}\|\mathbf{z}^{(k+1)} - \mathbf{T}_{P,\rho}(\mathbf{z}^{\diamond})\|^2 &\leq \gamma^k \|\mathbf{z}^1 - \mathbf{T}_{P,\rho}(\mathbf{z}^{\diamond})\|^2 \\ &\leq \gamma^k \|\mathbf{z}^0 - \mathbf{z}^{\diamond}\|^2,\end{aligned}$$

so that the projected auxiliary error of PDMM satisfies

$$\|\mathbf{z}^{(k+2)} - \mathbf{z}^{\diamond}\|^2 \leq \gamma^{k+2} \frac{\|\mathbf{z}^0 - \mathbf{z}^{\diamond}\|^2}{\gamma}. \tag{30}$$

By applying (22) to (30), the final primal bound is given by

$$\begin{aligned}\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|^2 &\leq \frac{\sigma_{\max}^2(\mathbf{C})}{\mu^2} \|\mathbf{z}^{(k)} - \mathbf{z}^{\diamond}\|^2 \\ &\leq \gamma^{k+2} \frac{\sigma_{\max}^2(\mathbf{C})}{\mu^2 \gamma} \|\mathbf{z}^{(0)} - \mathbf{z}^{\diamond}\|^2\end{aligned} \tag{31}$$

The primal error $\|\mathbf{x}^{(k+2)} - \mathbf{x}^*\|^2$ is therefore upper bounded by a geometrically contracting sequence and thus converges at a geometric rate. To the best of the authors knowledge, this is the fastest rate for PDMM proven within the literature.

VI. Numerical Experiments

In this section, we verify the analytical results of Sec. IV and V with numerical experiments. These results are broken down into two subsections: the convergence of PDMM for strongly convex and differentiable functions and the geometric convergence of PDMM for strongly convex functions with Lipschitz continuous gradients.

A. PDMM for Strongly Convex and Differentiable Functions

The first set of simulations validate the sufficiency of strong convexity and differentiability to guarantee primal convergence, as introduced in Sec. IV. For these simulations, as testing all such functions would be computationally infeasible, we instead considered the family of m-th power of m-norms for $m \in \{3, 4, 5, \cdots\}$ combined with an additive squared Euclidean norm term to enforce strong convexity. The prototype problem for these simulations is given by

$$\begin{aligned}\min_{\mathbf{x}} \quad & \sum_{i \in V} \left(\|\mathbf{x}_i - \mathbf{a}_i\|_m^m + \mu \|\mathbf{x}_i - \mathbf{a}_i\|^2\right) \\ \text{s.t.} \quad & \mathbf{x}_i - \mathbf{x}_j = \mathbf{0} \quad \forall (i,j) \in E,\end{aligned}$$

where $\mathbf{a}_i$ are local observation vectors, μ controls the strong convexity parameter and, for simplicity, edge based consensus constraints were chosen.

An $N = 10$ node undirected Erdős-Rényi network [54] was considered for these simulations. Such networks are randomly generated graphs where $\forall\ i,\ j \in V \setminus i$, there is equal probability that $(i,j) \in E$. This probability determines the density of the connectivity in the network and in this case was set to $\frac{\log(N)}{N}$. The resulting network had 12 undirected edges and was verified as forming a single connected component as per the assumptions in Sec. III. Additionally, a randomly generated initial $\mathbf{z}^{(0)}$ was also used for all problem instances. Finally the strong convexity parameter was set to $\mu = 10^{-3}$.

For $m = 3, \cdots 10$, 150 iterations of PDMM were performed and the resulting primal error computed. The squared Euclidean distance between the primal iterates and the primal optimal set was used as an error measure. Figure 2 demonstrates the convergence of this error with respect to iteration count. For each m the step sizes ρ were empirically selected to optimize convergence rate. Note that the finite precision stems

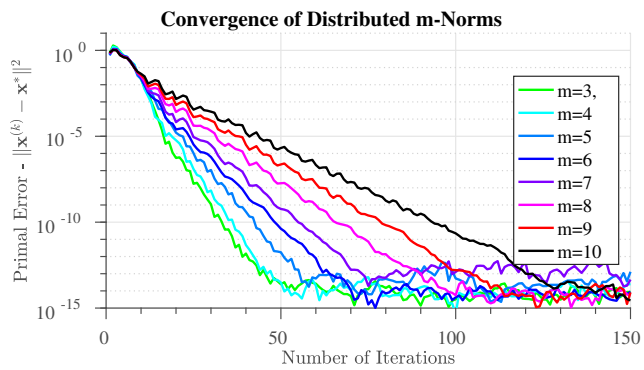


Fig. 2. The primal convergence of different m-norm^m consensus problem for a 10 node Erdős-Rényi network.

from the use of MATLABs $fminunc$ function.

Figure 3 further demonstrates that the choice of ρ does not effect the guarantee of convergence which in this instance was modeled via the number of iterations required to reach an auxiliary precision of $1e^{-5}$. This measure was chosen as the auxiliary error is monotonically decreasing with iteration count. In contrast the primal error need not satisfy this point, as can be observed in Figure 2. Note that while there is a clear variation in the rate of convergence for different choices of ρ, the guarantee of convergence of the algorithms are unaffected.

B. Geometric Convergence of PDMM for Strongly Convex and Smooth Functions

The final simulations verify the geometric bound from Sec. V by comparing the convergence of multiple problem instances to (31). Specifically, 10^4 random quadratic optimisation problems were generated, each of the form

$$\begin{aligned}\min_{\mathbf{x}} \quad & \sum_{i \in V} \left(\frac{1}{2}\mathbf{x}_i^T \mathbf{Q}_i \mathbf{x}_i - \mathbf{q}_i^T \mathbf{x}_i\right) \\ \text{s.t.} \quad & \mathbf{x}_i - \mathbf{x}_j = \mathbf{0} \quad \forall (i,j) \in E.\end{aligned}$$

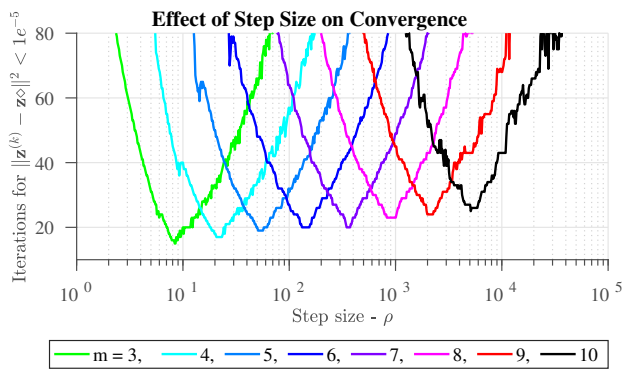


Fig. 3. A comparison to the required iterations for $\|\mathbf{z}^{(k)} - \mathbf{z}^{\diamond}\|^2 \leq 1e^{-5}$ for various step sizes (ρ). The step size is plotted on a log scale to better demonstrate the convergence characteristics of the different problems.

For each problem, the local variables were configured so that $\mathbf{x}_i \in \mathbb{R}^3 \ \forall i \in V$ and the resulting objective was paired with a random 10 node Erdős-Rényi network. The connection probability of each network was set to $\frac{\log(N)}{N}$ and the networks were verified as forming single connected components.

For each problem instance, the matrices $\mathbf{Q}_i \succeq \mathbf{0}$ were generated in such a way that a constant convergence rate bound was achieved. In this case the contraction factor of this rate bound was specified as $\gamma = 0.9$. Furthermore, the initial vector $\mathbf{z}^{(0)}$ was generated randomly and for each the associated $\mathbf{z}^{\diamond}$ was computed as per Eq. (17). This randomization procedure was implemented so that $\frac{\sigma^2_{\max}(\mathbf{C})}{\mu^2\gamma}\|\mathbf{z}^{(0)} - \mathbf{z}^{\diamond}\|^2 = 1$ for all instances.

For each problem instance, a total of 120 iterations of PDMM, were performed and the auxiliary errors, $\|\mathbf{z}^{(k)} - \mathbf{z}^{\diamond}\|^2$ for k even and $\|\mathbf{z}^{(k)} - \mathbf{T}_{P,\rho}\left(\mathbf{z}^{\diamond}\right)\|^2$ for k odd, were computed. The distribution of the resulting data is demonstrated in Figure 4 which highlights the spread of the convergence curves across all problem instances.

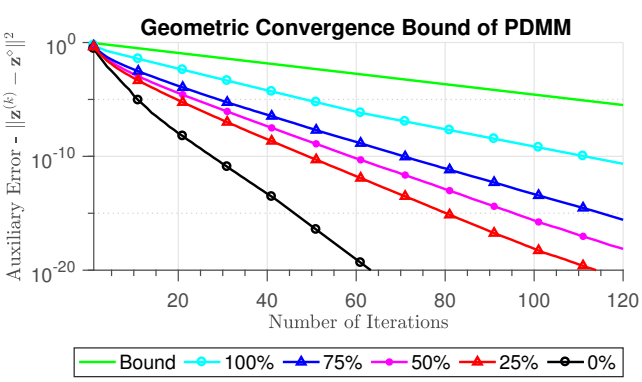


Fig. 4. Convergence of simulated PDMM problem instances. From top to bottom, the solid green line denotes the convergence rate bound while the remaining 5 lines denote the 100%, 75%, 50%, 25% and 0% quantiles respectively.

As expected, (30) provides a strict upper bound for all problem instances, with the smoothness of the curves stemming from the linear nature of the PDMM update equations. Furthermore, the rate of the worst case sequence (100% quantile) does not exceed that of the bound. Interestingly, while (30) holds for the worst case functions, most problem instances exhibit far faster convergence. This suggests that, for more restrictive problem classes, stronger bounds may exist.

VII. Conclusions

In this paper we have presented a novel derivation of the node-based distributed algorithm termed the primal-dual method of multipliers (PDMM). Unlike existing efforts within the literature, monotone operator theory was used for this purpose, providing both a succinct derivation for PDMM while highlighting the relationship between it and other existing first order methods such as PR splitting and ADMM. Using this derivation, primal convergence was demonstrated for strongly convex, differentiable functions and, in the case of strongly convex functions with Lipschitz continuous gradients, a geometric primal convergence bound was presented. This is despite the loss of a full row-rank assumption required by existing approaches and is a first for PDMM. In conclusion, the demonstrated results unify PDMM with existing solvers in the literature while providing new insight into its operation and convergence characteristics.

Appendix

A. Proof of Lemma III.1

As $\mathbf{R}_{\mathbf{T}_1,\rho} = 2\mathbf{J}_{\mathbf{T}_1,\rho} - \mathbf{I}$, we begin by defining a method for computing the update $\boldsymbol{\lambda}^{(k+1)} = \mathbf{J}_{\mathbf{T}_1,\rho}\left(\mathbf{z}^{(k)}\right)$. Firstly, by the definition of the resolvent,

$$\boldsymbol{\lambda}^{(k+1)} = \left(\mathbf{I} + \rho\mathbf{T}_1\right)^{-1}\left(\mathbf{z}^{(k)}\right)$$
$$\boldsymbol{\lambda}^{(k+1)} \in \mathbf{z}^{(k)} - \rho\mathbf{T}_1\left(\boldsymbol{\lambda}^{(k+1)}\right).$$

From the definition of the operator $\mathbf{T}_1$, it follows that

$$\boldsymbol{\lambda}^{(k+1)} \in \mathbf{z}^{(k)} - \rho\left(\mathbf{C}\partial f^*\left(\mathbf{C}^T\boldsymbol{\lambda}^{(k+1)}\right) - \mathbf{d}\right).$$

Let $\mathbf{x} \in \partial f^*\left(\mathbf{C}^T\boldsymbol{\lambda}\right)$. For $f \in \Gamma_0$, it follows from Proposition 16.10 [46], that $\mathbf{x} \in \partial f^*\left(\mathbf{C}^T\boldsymbol{\lambda}\right) \iff \partial f\left(\mathbf{x}\right) \ni \mathbf{C}^T\boldsymbol{\lambda}$ so that

$$\begin{aligned}\boldsymbol{\lambda}^{(k+1)} &= \mathbf{z}^{(k)} - \rho\left(\mathbf{C}\mathbf{x}^{(k+1)} - \mathbf{d}\right)\\ \mathbf{C}^T\boldsymbol{\lambda}^{(k+1)} &\in \partial f\left(\mathbf{x}^{(k+1)}\right).\end{aligned} \tag{32}$$

Thus, $\mathbf{x}^{(k+1)}$ can be computed as

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}}\left(f(\mathbf{x}) - \left\langle \mathbf{C}^T\mathbf{z}^{(k)}, \mathbf{x}\right\rangle + \frac{\rho}{2}||\mathbf{C}\mathbf{x} - \mathbf{d}||^2\right)$$

Combining (32) with the fact that $\mathbf{y}^{(k+1)} = \left(2\mathbf{J}_{\mathbf{T}_1,\rho} - \mathbf{I}\right)\left(\mathbf{z}^{(k)}\right)$ completes the proof. □

B. Proof of Lemma III.2

As $\mathbf{R}_{\mathbf{T}_2,\rho} = 2\mathbf{J}_{\mathbf{T}_2,\rho} - \mathbf{I}$, we again begin by defining a method for computing the update $\mathbf{J}_{\mathbf{T}_2,\rho}\left(\mathbf{y}^{(k+1)}\right)$,

From [48, Eq. 1.3], the resolvent of $\iota_{\ker(\mathbf{I}-\mathbf{P})}$, is given by

$$\mathbf{J}_{\mathbf{T}_2,\rho}\left(\mathbf{y}^{(k+1)}\right) = \underset{\ker(\mathbf{I}-\mathbf{P})}{\Pi}\mathbf{y}^{(k+1)}.$$

It follows that the reflected resolvent can be computed as

$$\mathbf{z}^{(k+1)} = \left(2\underset{\ker(\mathbf{I}-\mathbf{P})}{\Pi} - \mathbf{I}\right)\mathbf{y}^{(k+1)} = \mathbf{P}\mathbf{y}^{(k+1)},$$

completing the proof. □

C. Proof of Lemma IV.1

Reconsider the auxiliary PDMM updates given in Eq. (9). Substituting (9) into (16), it follows that

$$\begin{aligned}\|\mathbf{z}^{(k+1)} - \mathbf{z}^*\|^2 =& \|\mathbf{P}\left(\mathbf{z}^{(k)} - \mathbf{z}^* - 2\rho\mathbf{C}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^*\right)\right)\|^2\\ =& \|\mathbf{z}^{(k)} - \mathbf{z}^* - 2\rho\mathbf{C}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^*\right)\|^2\\ =& \|\mathbf{z}^{(k)} - \mathbf{z}^*\|^2 - 4\rho\left\langle\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^*, \mathbf{C}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^*\right)\right\rangle\\ \leq& \|\mathbf{z}^{(k)} - \mathbf{z}^*\|^2,\end{aligned} \tag{33}$$

where the penultimate line uses the dual update in Lemma III.1 and the final line uses the nonexpansiveness of $\mathbf{T}_{P,\rho}$.

As $\mathbf{C}^T\boldsymbol{\lambda} = \nabla f(\mathbf{x})$ (32), by Definition IV.3 it follows that,

$$\begin{aligned}&\langle\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2, \mathbf{C}(\mathbf{x}_1 - \mathbf{x}_2)\rangle \geq \mu\|\mathbf{x}_1 - \mathbf{x}_2\|^2\\ &\forall\mathbf{x}_1 \neq \mathbf{x}_2, \underset{\mathrm{ran}(\mathbf{C})}{\Pi}(\boldsymbol{\lambda}_1) \neq \underset{\mathrm{ran}(\mathbf{C})}{\Pi}(\boldsymbol{\lambda}_2).\end{aligned} \tag{34}$$

Recursively applying (33) and by using (34), it follows that

$$\lim_{k\to\infty} 4\rho\sum_{i=1}^{k}\mu\|\mathbf{x}^{(i)} - \mathbf{x}^*\|^2 \leq \|\mathbf{z}^{(0)} - \mathbf{z}^*\|^2 - \lim_{k\to\infty}\|\mathbf{z}^{(k)} - \mathbf{z}^*\|^2$$

so that $\left(\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2\right)_{k\in\mathbb{N}}$ is finitely summable. If $\left(\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2\right)_{k\in\mathbb{N}}$ is non-zero infinitely often then $\lim_{k\to\infty}\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2 = 0$ and thus $\lim_{k\to\infty}\mathbf{x}^{(k)} = \mathbf{x}^*$.

To demonstrate this point note that if $\exists k \mid \mathbf{x}^{(k+2)} = \mathbf{x}^{(k+1)} = \mathbf{x}^*$ then by the two-step PDMM update given in (15), $\mathbf{z}^{(k+2)} = \mathbf{z}^{(k)}$. Thus, $\forall M \geq 1$ the same primal updates will be computed so that $\mathbf{x}^{(k+M)} = \mathbf{x}^{(k+M-1)} = \mathbf{x}^*$.

Any $\mathbf{z}^{(k)}$ which produces two successive primal optimal updates therefore guarantees primal convergence. Thus, given our assumptions on f, any sequence which does not guarantee primal convergence in finite iterations has to be non-zero infinitely often so that $\lim_{k\to\infty}\mathbf{x}^{(k)} = \mathbf{x}^*$. As ∇f is single-valued, it also follows that $\lim_{k\to\infty}\underset{\mathrm{ran}(\mathbf{C})}{\Pi}\left(\boldsymbol{\lambda}^{(k)}\right) = \underset{\mathrm{ran}(\mathbf{C})}{\Pi}(\boldsymbol{\lambda}^*)$. □

D. Proof of Lemma V.1

Under the assumption that $f \in \Gamma_0$ is μ-strongly convex and ∇f is β-Lipschitz, from Theorem 18.15 [46], f^* is both $\frac{1}{\beta}$-strongly convex and $\frac{1}{\mu}$-smooth. It follows that ∇f^* is both $\frac{1}{\beta}$ strongly monotone and $\frac{1}{\mu}$ Lipschitz continuous.

In the case of *(i)*, due to the Lipschitz continuity of ∇f^*

$$\begin{aligned}&||\mathbf{C}\left(\nabla f^*(\mathbf{C}^T\mathbf{z}_1) - \nabla f^*(\mathbf{C}^T\mathbf{z}_2)\right)||\\ &\leq \sigma_{\max}(\mathbf{C})\,||\nabla f^*(\mathbf{C}^T\mathbf{z}_1) - \nabla f^*(\mathbf{C}^T\mathbf{z}_2)||\\ &\leq \frac{\sigma_{\max}(\mathbf{C})}{\mu}||\mathbf{C}^T(\mathbf{z}_1 - \mathbf{z}_2)||\\ &\leq \frac{\sigma^2_{\max}(\mathbf{C})}{\mu}||\mathbf{z}_1 - \mathbf{z}_2||,\end{aligned}$$

Therefore, $\mathbf{C}\nabla f^*(\mathbf{C}^T\bullet)$ is $\frac{\sigma_{\max}(\mathbf{C})^2}{\mu}$-Lipschitz continuous. In the case of *(ii)*, due to the strong monotonicity of ∇f^*

$$\left\langle\mathbf{C}\big(\nabla f^*(\mathbf{C}^T\mathbf{z}_1) - \nabla f^*(\mathbf{C}^T\mathbf{z}_2)\big), \mathbf{z}_1 - \mathbf{z}_2\right\rangle \geq \frac{||\mathbf{C}^T(\mathbf{z}_1 - \mathbf{z}_2)||^2}{\beta}.$$

For all $\mathbf{z}_1, \mathbf{z}_2 \in \mathrm{ran}(\mathbf{C})$ it follows that

$$\frac{||\mathbf{C}^T(\mathbf{z}_1 - \mathbf{z}_2)||^2}{\beta} \geq \frac{\sigma^2_{\min\neq 0}(\mathbf{C})\,||\mathbf{z}_1 - \mathbf{z}_2||^2}{\beta},$$

completing the proof. □

E. Proof of Lemma V.2

Consider the two by two matrix

$$\mathbb{G}_i = \begin{bmatrix}\delta^2 + \delta(1-\delta)\mathcal{S}_i^2 & -(1-\delta)\mathcal{C}_i\mathcal{S}_i\\ -\delta(1-\delta)\mathcal{C}_i\mathcal{S}_i & \delta + (1-\delta)\mathcal{C}_i^2\end{bmatrix}$$

The squared singular values of this matrix are given by the eigenvalues of the matrix

$$\mathbb{G}_i^T\mathbb{G}_i = \begin{bmatrix}\delta^4 + \delta^2(1-\delta^2)\mathcal{S}_i^2 & -\delta(1-\delta^2)\mathcal{C}_i\mathcal{S}_i\\ -\delta(1-\delta^2)\mathcal{C}_i\mathcal{S}_i & \delta^2 + (1-\delta^2)\mathcal{C}_i^2\end{bmatrix} \tag{35}$$

The eigenvalues of (35) can be computed via its trace and determinant. With some manipulation, these are given by

$$\mathrm{tr}\left(\mathbb{G}_i^T\mathbb{G}_i\right) = 2\delta^2 + (1-\delta^2)^2\mathcal{C}_i^2,\ \det\left(\mathbb{G}_i^T\mathbb{G}_i\right) = \delta^4$$

It follows that the squared singular values of $\mathbb{G}_i$ are given by

$$\begin{aligned}\sigma^2(\mathbb{G}_i) &= \frac{\mathrm{tr}\left(\mathbb{G}_i^T\mathbb{G}_i\right)}{2} \pm \sqrt{\frac{\mathrm{tr}\left(\mathbb{G}_i^T\mathbb{G}_i\right)^2}{4} - \det\left(\mathbb{G}_i^T\mathbb{G}_i\right)}\\ &= \delta^2 + (1-\delta^2)\mathcal{C}_i\left(\frac{(1-\delta^2)\mathcal{C}_i}{2} \pm \sqrt{\frac{(1-\delta^2)^2\mathcal{C}_i^2}{4} + \delta^2}\right)\end{aligned}$$

completing the proof. □

References

[1] R. Hanna, A. Rohm, and V. L. Crittenden, "We're all connected: The power of the social media ecosystem," *Business horizons*, vol. 54, no. 3, pp. 265–273, 2011.

[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE Int. Conf. on*, vol. 4. IEEE, 2001, pp. 2033–2036.

[4] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[5] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, *Wireless sensor networks: Signal processing and communications*. John Wiley & Sons, 2007.

[6] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM Int. workshop on Wireless sensor networks and applications*. Acm, 2002, pp. 88–97.

[7] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat monitoring: Application driver for wireless communications technology," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 2 supplement, pp. 20–41, 2001.

[8] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Trans. on industrial electronics*, vol. 57, no. 10, pp. 3557–3564, 2010.

[9] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. Timbus, "Overview of control and grid synchronization for distributed power generation systems," *IEEE Trans. Industrial Electronics*, vol. 53, no. 5, pp. 1398–1409, 2006.

[10] M. Erol-Kantarci and H. T. Mouftah, "Wireless sensor networks for cost-efficient residential energy management in the smart grid," *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 314–325, 2011.

[11] C. R. Baker, K. Armijo, S. Belka, M. Benhabib, V. Bhargava, N. Burkhart, A. Der Minassians, G. Dervisoglu, L. Gutnik, M. B. Haick *et al.*, "Wireless sensor networks for home health care," in *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st Int. Conf. on*, vol. 2. IEEE, 2007, pp. 832–837.

[12] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2688–2710, 2010.

[13] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, 2009.

[14] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. on Networking (TON)*, vol. 14, no. SI, pp. 2508–2530, 2006.

[15] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Information theory proceedings (isit), 2010 IEEE Int. symposium on*. IEEE, 2010, pp. 1753–1757.

[16] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth Conf. on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.

[17] A. Schwing, T. Hazan, M. Pollefeys, and R. Urtasun, "Distributed message passing for large scale graphical models," in *Computer vision and pattern recognition (CVPR), 2011 IEEE Conf. on*. IEEE, 2011, pp. 1833–1840.

[18] Y. Weiss and W. Freeman, "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 736–744, 2001.

[19] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[20] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1931–1935, 2015.

[21] E. Isufi, A. Simonetto, A. Loukas, and G. Leus, "Stochastic graph filtering on time-varying graphs," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th Int. Workshop on*. IEEE, 2015, pp. 89–92.

[22] Z. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on selected areas in communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

[23] G. Zhang and R. Heusdens, "Distributed optimization using the primal-dual method of multipliers," *IEEE Trans. on Signal and Information Processing over Networks*, 2016, accepted for publication.

[24] ——, "On simplifying the primal-dual method of multipliers," in *2016 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4826–4830.

[25] R. Rockafellar, "Network flows and monotropic optimization," 1984.

[26] ——, *Convex analysis*. Princeton, NJ: Princeton University Press, 1970.

[27] ——, "Monotone operators and the proximal point algorithm," *SIAM journal on control and Opt.*, vol. 14, no. 5, pp. 877–898, 1976.

[28] ——, "On the maximal monotonicity of subdifferential mappings," *Pacific Journal of Mathematics*, vol. 33, no. 1, pp. 209–216, 1970.

[29] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice hall Englewood Cliffs, NJ, 1989, vol. 23.

[30] J. Tsitsiklis, "Problems in decentralized decision making and computation." Massachusetts Inst of Tech Cambridge lab for information and decision systems, Tech. Rep., 1984.

[31] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. on automatic control*, vol. 31, no. 9, pp. 803–812, 1986.

[32] J. Eckstein, "Splitting methods for monotone operators with applications to parallel optimization," Ph.D. dissertation, Massachusetts Institute of Technology, 1989.

[33] ——, "Parallel alternating direction multiplier decomposition of convex programs," *Journal of Opt. Theory and Applications*, vol. 80, no. 1, pp. 39–62, 1994.

[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[35] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Trans. on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[36] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[37] P. Bianchi, W. Hachem, and I. Franck, "A stochastic coordinate descent primal-dual algorithm and applications," in *Machine Learning for Signal Processing (MLSP), 2014 IEEE Int. Workshop on*. IEEE, 2014, pp. 1–6.

[38] P. Latafat and P. Patrinos, "Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators," *Computational Opt. and Applications*, pp. 1–37, 2016.

[39] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Opt. Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[40] B. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[41] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization-part i: Theory." *IEEE Trans. Signal Processing*, vol. 65, no. 8, pp. 1929–1944, 2017.

[42] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song, "Parallel and distributed methods for constrained nonconvex optimization-part ii: Applications in communications and machine learning," *IEEE Trans. on Signal Processing*, vol. 65, no. 8, pp. 1945–1960, 2017.

[43] J. Eckstein and W. Yao, "Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," *RUTCOR Research Reports*, vol. 32, p. 3, 2012.

[44] P. Giselsson and S. Boyd, "Linear convergence and metric selection for Douglas-Rachford splitting and ADMM," *IEEE Trans. on Automatic Control*, vol. 62, no. 2, pp. 532–544, 2017.

[45] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization." *IEEE Trans. Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[46] H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. New York, NY.: Springer NY, 2017, vol. 408.

[47] E. Ryu and S. Boyd, "Primer on monotone operator methods," *Appl. Comput. Math*, vol. 15, no. 1, pp. 3–43, 2016.

[48] N. Parikh, S. P. Boyd *et al.*, "Proximal algorithms." *Foundations and Trends in Opt.*, vol. 1, no. 3, pp. 127–239, 2014.

[49] P. Bianchi, W. Hachem, and F. Iutzeler, "A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization," *ArXiv:1407:0898*, 2014.

[50] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

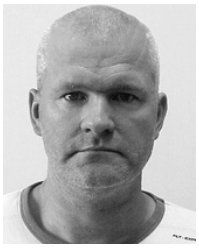
[51] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *Journal of Scientific Computing*, vol. 66, no. 3, pp. 889–916, 2016.

[52] M. Fält and P. Giselsson, "Optimal convergence rates for generalized alternating projections," *arXiv preprint arXiv:1703.10547*, 2017.

[53] H. Bauschke, J. Cruz, T. Nghia, H. Pha, and X. Wang, "Optimal rates of linear convergence of relaxed alternating projections and generalized douglas-rachford methods for two subspaces," *Numerical Algorithms*, vol. 73, no. 1, pp. 33–76, 2016.

[54] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

Thomas W. Sherson received his Bachelor of Engineering with First Class Honors, majoring in Electrical and Computer Systems Engineering, from Victoria University of Wellington in New Zealand, in 2015. He was also awarded the Victoria University Medal of Academic Excellence in the same year. Following his graduation, he joined the Department of Microelectronics at Delft University of Technology to continue his studies towards a Doctor of Philosophy (PhD) in the field of Electrical Engineering. His research interests include the likes of signal processing in wireless sensor networks, distributed/decentralized optimization, monotone operator theory and audio signal processing. Additionally, he is an avid outdoorsman with a passion for nature and a love for music.



Richard Heusdens received the M.Sc. and Ph.D. degrees from Delft University of Technology, Delft, The Netherlands, in 1992 and 1997, respectively. Since 2002, he has been an Associate Professor in the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. In the spring of 1992, he joined the digital signal processing group at the Philips Research Laboratories, Eindhoven, The Netherlands. He has worked on various topics in the field of signal processing, such as image/video compression and VLSI architectures for image processing algorithms. In 1997, he joined the Circuits and Systems Group of Delft University of Technology, where he was a Postdoctoral Researcher. In 2000, he moved to the Information and Communication Theory (ICT) Group, where he became an Assistant Professor responsible for the audio/speech signal processing activities within the ICT group. He held visiting positions at KTH (Royal Institute of Technology, Sweden) in 2002 and 2008 and is a part-time professor at Aalborg University. He is involved in research projects that cover subjects such as audio and acoustic signal processing, speech enhancement, and distributed signal processing for sensor networks.

W. Bastiaan Klein is a Professor at Victoria University of Wellington, New Zealand and Professor (part-time) at Delft University of Technology (DUT). He was Professor and Head of the Sound and Image Processing Laboratory at KTH in Stockholm, 1996-2010. Kleijn was a founder of Global IP Solutions, a company that provided the enabling audio technology to Skype. It was acquired by Google in 2010. He has served on a number of editorial Boards including those of the IEEE Trans. Audio Speech Language Processing, Signal Processing, IEEE Signal Processing Letters, and IEEE Signal Processing Magazine. He was the Technical Chair of ICASSP 1999 and EUSIPCO 2010, and two IEEE workshops. Kleijn received a PhD degree in Electrical Eng. from TU Delft, a PhD in Soil Science and an MSc in Physics from the University of California, Riverside, and an MSEE from Stanford University. He is Fellow of the IEEE since 1999.