Convergence of Stochastic PDMM

Jordan, Sebastian O. ; Sherson, Thomas W.; Heusdens, Richard

DOI
10.1109/ICASSP49357.2023.10095808

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095808

CONVERGENCE OF STOCHASTIC PDMM

Sebastian O. Jordan[1], *Thomas W. Sherson*[1] *and Richard Heusdens*[1,2]

[1] Delft University of Technology, The Netherlands
[2] Netherlands Defence Academy, The Netherlands

ABSTRACT

In recent years, the large increase in connected devices and the data that are collected by these devices have caused a heightened interest in distributed processing. Many practical distributed networks are of heterogeneous nature, because different devices in the network can have different specifications. Because of this, it is highly desirable that algorithms operating within these networks can operate asynchronously, since in that case there is no need for clock synchronisation between the nodes, and the algorithm is not slowed down by the slowest device in the network. In this paper, we focus on the primal-dual method of multipliers (PDMM), which is a promising distributed optimisation algorithm that is suitable for distributed optimisation in heterogeneous networks. Most theoretical work that can be found in existing literature focuses on synchronous versions of PDMM. In this work, we prove the convergence of stochastic PDMM, which is a general framework that can model variations such as asynchronous PDMM and PDMM with transmission losses.

Index Terms— distributed optimisation, convex optimisation, PDMM, asynchronous algorithms, transmission loss

1. INTRODUCTION

The world around us is becoming increasingly connected. More and more electronic devices are being used and these devices are producing more and more data. Many devices have the ability to (wirelessly) connect to other devices and thus form large distributed networks. Distributed processing could leverage the full potential of such large scale distributed networks. An important aspect of these types of networks is that they are often of heterogeneous nature, with connected devices having different computation, communication, power and clock specifications. Furthermore, in many real world implementations the wireless links between nodes will not be ideal, so that the distributed processing needs to be robust against transmission losses.

A lot of research has been done in the context of distributed average consensus, where the average of noisy measurements is calculated over the entire network. As presented in [1], gossip based algorithms can be used to solve these types of problems. These algorithms are relatively simple but can be performed asynchronously. A more general type of distributed algorithms is the class of convex optimisation based algorithms. One of these algorithms is the primal-dual method of multipliers (PDMM), first introduced in [2]. In [3] it was shown that PDMM is closely related to the more commonly used ADMM algorithm, in the sense that ADMM is a 1/2-averaged version of PDMM. Therefore, PDMM can achieve faster convergence rates, provided it converges. Convergence has been proved for synchronous implementations in [3] and simulations show it also converges when implemented asynchronously. Asynchronous algorithms have the advantage over synchronous ones that there is no need for clock synchronisation between the nodes, and that the algorithm is not slowed down by the slowest device in the network.

In this paper we give convergence conditions for stochastic PDMM using monotone operator theory, fixed point theory of nonexpansive operators [4] and probability theory [5]. This framework is general in the sense that it includes asynchronous PDMM and PDMM with transmission losses. To the best of our knowledge, there only exist convergence results for averaged operators in literature, see [4, Def. 4.33] for Definition. In [6], convergence of asynchronous ADMM (1/2-averaged operator) is shown, a result generalised to arbitrary θ-averaged ADMM algorithms in [7]. An alternative approach to the convergence proof of this class of mathematical problems is given in [8]. In [9] robustness against transmission loss is explicitly mentioned in the context of asynchronous θ-averaged ADMM.

2. BACKGROUND

In this section, we formulate the mathematical problem statement of distributed optimisation and introduce the PDMM updating equations that will be used in the proof in our work. For more details regarding PDMM we refer to [3].

2.1. Problem Statement

Consider a graphical model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., n\}$ denotes the set of vertices, or nodes, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of m undirected edges (unordered paired vertices) representing communication links in the network. Additionally, let $\mathcal{E}_{\text{dir}} \subseteq \mathcal{V} \times \mathcal{V}$ be the set of *ordered pairs* of nodes, denoting the set of *directed edges* in $G(\mathcal{V}, \mathcal{E})$. Clearly, for every edge in $\mathcal{E}$ we have two ordered pairs in $\mathcal{E}_{\text{dir}}$, so that $|\mathcal{E}_{\text{dir}}| = 2m$. Furthermore, we use the following notational conventions: a variable $\mathbf{x}_i$ is related to node i; a variable $\mathbf{x}_{i|j}$ is related to edge (i, j) but held by node i. Moreover, the neighbourhood of node i is defined as $\mathcal{N}_i = \{j \in \mathcal{V} \,|\, (i, j) \in \mathcal{E}\}$.

Each node i is equipped with an arbitrary convex, closed and proper (CCP) cost function $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}$, where each function is dependent on a local optimisation variable $\mathbf{x}_i \in \mathbb{R}^{n_i}$. These local variables are stacked as $\mathbf{x} = [\mathbf{x}_1^T, ..., \mathbf{x}_n^T]^T$.

We would like to solve the following optimisation problem:

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

with constraints along edges defined by $\mathbf{A}_{i|j} \in \mathbb{R}^{m_{i,j} \times n_i}$ and $\mathbf{b}_{i,j} \in \mathbb{R}^{m_{i,j}}$, where $m_{i,j}$ is the dimension of the constraint along edge (i, j). Let $n_{\mathcal{V}} = \sum_{i \in \mathcal{V}} n_i$ and $m_{\mathcal{E}} = \sum_{(i,j) \in \mathcal{E}} m_{i,j}$.

For many practical use cases, consensus is required. That is, after convergence we have $\mathbf{x}_i = \mathbf{x}_j$ for all $i, j \in \mathcal{V}$. This means that for all $(i,j) \in \mathcal{E}$ the constraints in (1) are defined using $\mathbf{A}_{i|j} = \mathbf{I}$, $\mathbf{A}_{j|i} = -\mathbf{I}$ and $\mathbf{b}_{i,j} = \mathbf{0}$. Because consensus problems are very common, throughout this paper we will assume the network constraints correspond to these consensus constraints. The results can be easily generalised to problems where this is not the case.

2.2. PDMM

PDMM is an iterative algorithm that can solve the optimisation problem stated in (1). An insightful derivation of the algorithm using monotone operator theory is given in [3]. As derived in [3], PDMM can be formulated as a monotone inclusion problem that can be iteratively solved using Peaceman-Rachford splitting (see [4, Sec. 26.4]). This leads to the following PDMM update equations:

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

$$\mathbf{y}^{(k+1)} = \mathbf{z}^{(k)} + 2\rho \mathbf{C}\mathbf{x}^{(k+1)}, \quad (3)$$

$$\mathbf{z}^{(k+1)} = \mathbf{P}\mathbf{y}^{(k+1)}, \quad (4)$$

where k denotes the iteration index, $\mathbf{x} \in \mathbb{R}^{n\mathcal{V}}$ is the primal variable, $\mathbf{y} \in \mathbb{R}^{2m\mathcal{E}}$ and $\mathbf{z} \in \mathbb{R}^{2m\mathcal{E}}$ are auxiliary variables, $\mathbf{C} \in \mathbb{R}^{2m\mathcal{E} \times n\mathcal{V}}$ is a constraint matrix constructed with $\mathbf{A}_{i|j}$ for all $(i|j) \in \mathcal{E}_{\text{dir}}$, $\mathbf{P} \in \mathbb{R}^{2m\mathcal{E} \times 2m\mathcal{E}}$ is a symmetric permutation matrix and $\rho > 0$ is a constant that controls the convergence rate. Note that (4) describes the data exchanged amongst neighbouring nodes in the network. Due to the construction of the PDMM algorithm, the update equations are separable across the nodes of the network and thus can be performed in a distributed/parallel manner, see Algorithm 1 for an example of a distributed implementation of PDMM. Furthermore, a complete PDMM iteration can be seen as an operator on the auxiliary variable $\mathbf{z}$, which we denote as $\mathbf{T}_{\mathrm{P},\rho}$. If we define the fixed point set of operator $\mathbf{T}$ as fix $(\mathbf{T}) = \{\mathbf{z} \,|\, \mathbf{T}\mathbf{z} = \mathbf{z}\}$, each $\mathbf{z}^* \in$ fix $(\mathbf{T}_{\mathrm{P},\rho})$ corresponds to a solution of (1). As proved in [3], standard PDMM converges for strongly convex and differentiable cost functions. To guarantee convergence for arbitrary CCP cost functions, we can apply operator averaging. This results in the θ-averaged PDMM operator, which is defined as

$$\mathbf{T}_{\theta\mathrm{P},\rho} = (1-\theta)\mathbf{I} + \theta \mathbf{T}_{\mathrm{P},\rho}, \quad (5)$$

with $\theta \in (0,1)$.

3. STOCHASTIC PDMM

In this section, we define a general stochastic version of PDMM. We formulate a convergence proof inspired by the proof presented in [7] for θ-averaged ADMM, and we show that asynchronous PDMM and PDMM with transmission loss are both specific versions of stochastic PDMM. We will first introduce a stochastic Banach-Picard iteration, which forms the update equation for stochastic PDMM. Once this is defined, we state two assumptions, which are then used to prove the main result of this paper in Theorem 3.1.

3.1. Definitions

Stochastic updates can be defined by assuming that each auxiliary variable $\mathbf{z}_{i|j}$ can be updated based on a Bernoulli random variable $U_{i|j} \in \{0,1\}$, with mean $\mu_{i|j} = \mathbb{E}[U_{i|j}] = \mathbb{P}\{U_{i|j} = 1\} \in (0,1)$, where $\mathbb{E}[\cdot]$ denotes statistical expectation and $\mathbb{P}\{\omega\}$ the probability that event ω occurs. We define the random matrix $\mathbf{U} \in \mathbb{R}^{2m\mathcal{E} \times 2m\mathcal{E}}$ as a block diagonal matrix with diagonal entries $U_{i|j}\mathbf{I}_{m_{i,j}}$, following the same ordering as the entries of $\mathbf{z}$.

Algorithm 1 Asynchronous PDMM.

Initialise: $\mathbf{z}^{(0)} \in \mathbb{R}^{2m\mathcal{E}}$ ▷ Initialisation
for $k = 0, ...,$ **do**
  Select a random subset of active nodes: $\xi^{(k)} \in 2^{\mathcal{V}}$
  Select a random subset of active directed edges: $\eta^{(k)} \in 2^{\mathcal{E}_{\text{dir}}}$
  for $i \in \xi^{(k)}$ **do** ▷ Active node updates
    $\mathbf{x}_i^{(k+1)} = \arg\min_{\mathbf{x}_i} \Big[f_i(\mathbf{x}_i) + \sum_{j \in \mathcal{N}_i} \Big((\mathbf{z}_{i|j}^{(k)})^T \mathbf{A}_{i|j}\mathbf{x}_i + \frac{\rho}{2}||\mathbf{x}_i||_2^2 \Big) \Big]$
    for all $j \in \mathcal{N}_i$ **do**
      $\mathbf{y}_{i|j}^{(k+1)} = \mathbf{z}_{i|j}^{(k)} + 2\rho \mathbf{A}_{i|j}\mathbf{x}_i^{(k+1)}$
    end for
  end for
  for $i \in \xi^{(k)}, j \in \mathcal{N}_i$ **do** ▷ Transmit updated variables
    Node$_j$ ← **Node**$_i(\mathbf{y}_{i|j}^{(k+1)})$
  end for
  for $i \in \xi^{(k)}, j \in \mathcal{N}_i \,:\, (i,j) \in \eta^{(k)}$ **do** ▷ Secondary node updates
    $\mathbf{z}_{j|i}^{(k+1)} = \mathbf{y}_{i|j}^{(k+1)}$
  end for
end for

Definition 3.1. *For an operator* $\mathbf{T}$ *and a sequence of realisations of random updating matrices* $\left(\mathbf{U}^{(k)}\right)_{k\in\mathbb{N}}$*, we define the* ***stochastic Banach-Picard iteration*** *as:*

$$\mathbf{z}^{(k+1)} = \left(\mathbf{I} - \mathbf{U}^{(k+1)}\right)\mathbf{z}^{(k)} + \mathbf{U}^{(k+1)}\mathbf{T}\mathbf{z}^{(k)}. \quad (6)$$

In the following, we will make the following assumptions.

Assumption 3.1. $\left(\mathbf{U}^{(k)}\right)_{k\in\mathbb{N}}$ *is a random i.i.d. sequence*[1]

Assumption 3.2. $\bar{\mathbf{U}} = \mathbb{E}[\mathbf{U}] \succ 0$.

Let $\mathcal{F}^{(k)} = \sigma(\mathbf{z}^{(0)}, ..., \mathbf{z}^{(k)})$ denote the sigma-algebra generated by the sequence of random variables $\mathbf{z}^{(0)}, ..., \mathbf{z}^{(k)}$ (see [5, Def. 23.3]). Moreover, let $\|\mathbf{x}\|_{\mathbf{Q}}^2 = \langle \mathbf{Q}\mathbf{x}, \mathbf{x} \rangle$. Assume that $\mathbf{z}^* \in$ fix $(\mathbf{T})$ is deterministic, and that the initialisation $\mathbf{z}^{(0)}$ is known. As shown in Appendix A, we can use the conditional expectation with respect to $\mathcal{F}^{(k)}$ to express

$$\begin{aligned} &\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||_{\bar{\mathbf{U}}^{-1}}^2 | \mathcal{F}^{(k)} \right] \\ &\quad = ||\mathbf{z}^{(k)} - \mathbf{z}^*||_{\bar{\mathbf{U}}^{-1}}^2 - ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2 + ||\mathbf{T}\mathbf{z}^{(k)} - \mathbf{z}^*||^2. \quad (7) \end{aligned}$$

Note that $\mathbb{E}\left[\,\cdot\,|\mathcal{F}^{(k)}\right]$ is a random variable and thus the related expressions include an implicit "almost surely" qualifier.

3.2. Convergence proof

For the remaining part of this proof we will take $\mathbf{z}^* \in$ fix $(\mathbf{T}_{\mathrm{P},\rho})$ and use (2) to define

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left(f(\mathbf{x}) + \left\langle \mathbf{C}^T \mathbf{z}^*, \mathbf{x} \right\rangle + \frac{\rho}{2} ||\mathbf{C}\mathbf{x}||^2 \right). \quad (8)$$

[1] Note that no assumption is made on the dependence between the entries $U_{i|j}$ of $\mathbf{U}^{(k)}$.

Note that $\mathbf{z}^*$, $\mathbf{z}^{(0)}$, and $\mathbf{x}^*$ are deterministic and all other $\mathbf{z}^{(k)}$'s and $\mathbf{x}^{(k)}$'s are random vectors.

Lemma 3.1. *Consider the PDMM operator $\mathbf{T}_{P,\rho}$, then*

$$||\mathbf{T}_{P,\rho}(\mathbf{z}^{(k)}) - \mathbf{z}^*||^2 = ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2 - 4\rho\left\langle \partial f(\mathbf{x}^{(k+1)}) - \partial f(\mathbf{x}^*), \mathbf{x}^{(k+1)} - \mathbf{x}^*\right\rangle.$$

Proof. See [3, Appendix C]. □

Using Lemma 3.1, (7) becomes

$$\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}}|\mathcal{F}^{(k)}\right] = ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}} - 4\rho\left\langle \partial f(\mathbf{x}^{(k+1)}) - \partial f(\mathbf{x}^*), \mathbf{x}^{(k+1)} - \mathbf{x}^*\right\rangle. \tag{9}$$

Note that since the subdifferential of a CCP function is monotone, the sequence $\left(||\mathbf{z}^{(k)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}}\right)_{k\in\mathbb{N}}$ is a nonnegative supermartingale and therefore converges almost surely. However, in general $\mathbf{z}^{(k)} - \mathbf{T}_{\mathrm{P},\rho}(\mathbf{z}^{(k)}) \not\to 0$ since PDMM is at best nonexpansive for arbitrary CCP functions. To guarantee almost sure convergence of the primal variable $\mathbf{x}^{(k)}$, we need additional assumptions on f, like f being differentiable and β-strongly convex.

Theorem 3.1. *For differentiable and β-strongly convex cost functions, stochastic standard PDMM converges almost surely to a primal optimal point $\mathbf{x}^*$.*

Proof. Assume that f is differentiable and β-strongly convex. In that case, the gradient ∇f is β-strongly monotone so that

$$\left\langle \nabla f(\mathbf{x}^{(k)}) - \nabla f(\mathbf{x}^*), \mathbf{x}^{(k)} - \mathbf{x}^*\right\rangle \geq \beta||\mathbf{x}^{(k)} - \mathbf{x}^*||^2.$$

With this, (9) can be reformulated as

$$\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}}|\mathcal{F}^{(k)}\right] \leq ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}} - 4\rho\beta||\mathbf{x}^{(k+1)} - \mathbf{x}^*||^2. \tag{10}$$

Taking expectations on both sides of (10) and iterating over k we obtain

$$\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}}\right] \leq ||\mathbf{z}^{(0)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}} - 4\rho\beta\sum_{t=1}^{k+1}\mathbb{E}\left[||\mathbf{x}^{(t)} - \mathbf{x}^*||^2\right]. \tag{11}$$

Since $\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}}\right] \geq 0$, we have that

$$\sum_{t=1}^{\infty}\mathbb{E}\left[||\mathbf{x}^{(t)} - \mathbf{x}^*||^2\right] \leq \frac{1}{4\rho\beta}||\mathbf{z}^{(0)} - \mathbf{z}^*||^2_{\bar{\mathbf{U}}^{-1}} < \infty,$$

which shows that the sum of the expected values of the primal error is bounded. Hence, using Markov's inequality, we conclude that

$$\sum_{t=1}^{\infty}\mathbb{P}\left\{||\mathbf{x}^{(t)} - \mathbf{x}^*||^2 \geq \epsilon\right\} \leq \frac{1}{\epsilon}\sum_{t=1}^{\infty}\mathbb{E}\left[||\mathbf{x}^{(t)} - \mathbf{x}^*||^2\right] < \infty,$$

for all $\epsilon > 0$, so that by Borel Cantelli's lemma [5, Theorem 10.5]

$$\mathbb{P}\left\{\limsup_{t\to\infty}\left(||\mathbf{x}^{(t)} - \mathbf{x}^*||^2 \geq \epsilon\right)\right\} = 0.$$

Hence, there is a set of events with probability 1 where $\forall\epsilon > 0\ \exists t_0$ such that $||\mathbf{x}^{(t)} - \mathbf{x}^*||^2 < \epsilon, \forall t \geq t_0$. Hence $||\mathbf{x}^{(t)} - \mathbf{x}^*||^2 \to 0$ almost surely. Because $\mathbf{z}^* \in \mathrm{fix}\,(\mathbf{T}_{\mathrm{P},\rho})$, $\mathbf{x}^*$ is a primal optimal solution by construction, see (8). □

3.3. Averaged PDMM

Stochastic θ-averaged PDMM can be seen as a stochastic version of a θ-averaged operator on the auxiliary variable $\mathbf{z}$. Thus, we can directly apply Theorem 3 from [7], to prove the auxiliary convergence of $\mathbf{z}^{(k)}$ and in turn the convergence of stochastic θ-averaged PDMM. This convergence holds for arbitrary CCP cost functions.

3.4. Asynchronous Schemes with Transmission Losses

In practice, synchronous algorithm operation implies the presence of a global clocking system between nodes. In many practical situations this requirement represents an additional and undesirable overhead. Asynchronous algorithm operation, on the other hand, alleviates this problem as individual nodes can update their variables according to a local clock. That is, at each iteration, a single node, or possibly a subset of nodes chosen at random, are activated. More formally, let $(\xi^{(k)})_{k\in\mathbb{N}}$ denote an i.i.d. random process defined on a common probability space such that $\xi^{(k)} \in 2^{\mathcal{V}}$ denotes a set of indices indicating which node will be updated at iteration k. Hence, $\xi^{(k)}$ denotes the set of *active nodes* at iteration k. Asynchronous PDMM can be seen as a specific case of stochastic PDMM where the entries of $\mathbf{U}^{(k)}$ are defined as

$$U_{i|j}^{(k)} = \begin{cases} 1, & \text{if } i \in \xi^{(k)}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that by Assumption 3.2, we have $\forall i \in \mathcal{V}$, $\forall k \in \mathbb{N}$, $\mathbb{P}\{i \in \xi^{(k)}\} > 0$. Since $(\xi^{(k)})_{k\in\mathbb{N}}$ is i.i.d., this guarantees that at every iteration, node $i \in \mathcal{V}$ has nonzero probability to be updated.

PDMM with transmission losses can also be seen as a special case of stochastic PDMM. Let $(\eta^{(k)})_{k\in\mathbb{N}}$ denote an i.i.d. random process defined on a common probability space such that $\eta^{(k)} \in 2^{\mathcal{E}_{\mathrm{dir}}}$ denotes a set of ordered pairs of nodes indicating which entries of $\mathbf{z}^{(k+1)}$ will be updated at iteration k. Hence, $\eta^{(k)}$ denotes the set of *active directed edges* at iteration k; updating $\mathbf{z}_{j|i}^{(k)}$ implies that $\mathbf{z}_{j|i}^{(k+1)} = \mathbf{y}_{i|j}^{(k+1)}$ so that there has been a successful transmission from node i to node j, but we could have a transmission failure from node j to i. PDMM with transmission losses can thus be seen as a specific case of stochastic PDMM where the entries of $\mathbf{U}^{(k)}$ are defined as

$$U_{i|j}^{(k)} = \begin{cases} 1, & \text{if } (i,j) \in \eta^{(k)}, \\ 0, & \text{otherwise.} \end{cases}$$

Obviously, a combination of asynchronous updating and transmission loss can be modelled by defining

$$U_{i|j}^{(k)} = \begin{cases} 1, & \text{if } i \in \xi^{(k)} \text{ and } (i,j) \in \eta^{(k)}, \\ 0, & \text{otherwise.} \end{cases}$$

The pseudocode for lossy asynchronous PDMM is given in Algorithm 1.

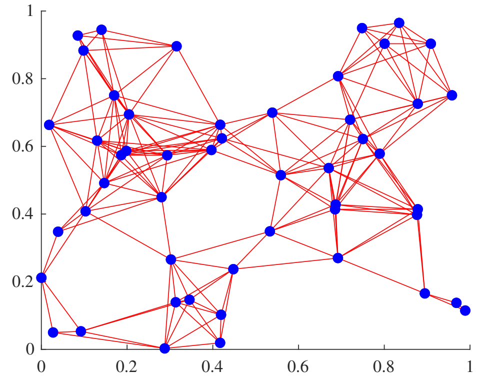

Fig. 1. Random geometric graph. Blue dots represent nodes whilst red lines are the edges connecting them.

4. SIMULATIONS AND RESULTS

In this section we present simulation results that illustrate the convergence of asynchronous PDMM and its robustness against transmission loss. We consider the straightforward problem of distributed averaging, where each node $i \in \mathcal{V}$ collects a scalar measurement w_i and the nodes in the network aim to calculate the average of all of these measurements. The related cost function is $f(\mathbf{x}) = \sum_{i\in\mathcal{V}} \left(\frac{1}{2}||x_i - w_i||^2\right)$. Due to its resemblance to wireless networks, we perform distributed averaging using PDMM over a connected random geometric graph [10] with $n = 50$ nodes and a communication radius given by $r = \sqrt{\frac{\log n}{n}}$, as shown in Figure 1. The measurement values w_i are drawn from a unit variance, zero mean, Gaussian distribution. For the asynchronous simulations, at each iteration a single random node is activated according to a uniform distribution and the transmission losses are modelled using a Bernoulli random variable for each directed edge. In Figure 2, the convergence of the primal error of synchronous PDMM and asynchronous PDMM are shown with respect to the number of transmissions. A number of different transmission error probabilities were simulated using asynchronous PDMM and asynchronous ADMM (1/2-averaged PDMM [3]) is added as a comparison. We can see that transmission loss does not prevent the convergence of PDMM; the convergence rate decreases proportional to the loss rate. Furthermore, we can see that asynchronous PDMM shows similar convergence properties to synchronous PDMM and faster convergence than ADMM.

5. CONCLUSIONS

In this paper a formal convergence proof is derived for the convergence of stochastic PDMM. Stochastic PDMM is a general framework that can be used to model PDMM variations like asynchronous updating and PDMM with transmission losses. As proved in Section 3, stochastic PDMM converges almost surely to a primal optimal solution. The only assumption required is that the updating probability of each entry of the auxiliary variable $\mathbf{z}$ is nonzero. The convergence proof for stochastic standard PDMM holds for differentiable and strongly convex cost functions and the convergence proof of stochastic θ-averaged PDMM holds for arbitrary CCP cost functions.

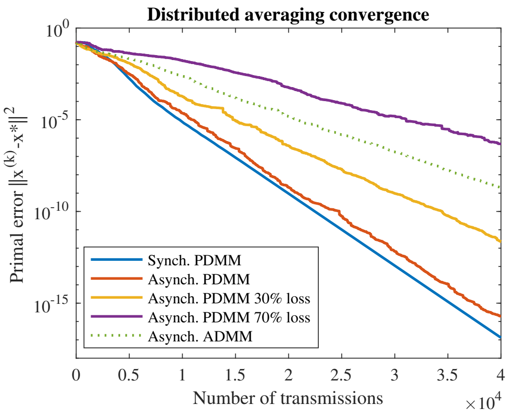

Fig. 2. Convergence of distributed averaging with different levels of transmission loss.

Appendix A: Derivation of (7)

Using the assumptions listed in Section 3.1 and the conditional expectation with respect to the sigma-algebra $\mathcal{F}^{(k)} = \sigma\left(\mathbf{z}^{(0)}, ..., \mathbf{z}^{(k)}\right)$, we can derive

$$
\begin{aligned}
&\mathbb{E}\left[||\mathbf{z}^{(k+1)} - \mathbf{z}^*||^2_{\mathbf{U}^{-1}}|\mathcal{F}^{(k)}\right] \\
&= \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1}{\mu_{i|j}}\mathbb{E}\left[\left\|\mathbf{z}^{(k)}_{i|j} - U^{(k+1)}_{i|j}\mathbf{z}^{(k)}_{i|j}\right.\right. \\
&\qquad\qquad \left.\left. + U^{(k+1)}_{i|j}\left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} - \mathbf{z}^*_{i|j}\right\|^2 |\mathcal{F}^{(k)}\right] \\
&= \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1}{\mu_{i|j}}\mathbb{E}\left[||\mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}||^2\right. \\
&\qquad + \left(U^{(k+1)}_{i|j}\right)^2\left\|\left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} - \mathbf{z}^{(k)}_{i|j}\right\|^2 \\
&\qquad \left. + 2U^{(k+1)}_{i|j}\left\langle \mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}, \left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} - \mathbf{z}^{(k)}_{i|j}\right\rangle |\mathcal{F}^{(k)}\right] \\
&\overset{(a)}{=} \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1}{\mu_{i|j}}\left(||\mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}||^2 + \mu_{i|j}\left[\left[\mathbf{T}\mathbf{z}^{(k)}\right]^2_{i|j} - \left(\mathbf{z}^{(k)}_{i|j}\right)^2\right.\right. \\
&\qquad \left.\left. - 2\mathbf{z}^*_{i|j}\left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} + 2\mathbf{z}^*_{i|j}\mathbf{z}^{(k)}_{i|j}\right]\right) \\
&= \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1}{\mu_{i|j}}\left(||\mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}||^2 + \mu_{i|j}\left[\left\|\left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} - \mathbf{z}^*_{i|j}\right\|^2\right.\right. \\
&\qquad \left.\left. - ||\mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}||^2\right]\right) \\
&= \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\frac{1-\mu_{i|j}}{\mu_{i|j}}||\mathbf{z}^{(k)}_{i|j} - \mathbf{z}^*_{i|j}||^2 + \left\|\left[\mathbf{T}\mathbf{z}^{(k)}\right]_{i|j} - \mathbf{z}^*_{i|j}\right\|^2 \\
&= ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2_{\mathbf{U}^{-1}} - ||\mathbf{z}^{(k)} - \mathbf{z}^*||^2 + ||\mathbf{T}\mathbf{z}^{(k)} - \mathbf{z}^*||^2,
\end{aligned}
\tag{12}
$$

where (a) results from the conditioning on $\mathcal{F}^{(k)}$ and the fact that $U^2_{i|j} = U_{i|j}, \forall i \in \mathcal{V}, j \in \mathcal{N}_i$. □

References

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[2] Guoqiang Zhang and Richard Heusdens, "Distributed Optimization Using the Primal-Dual Method of Multipliers," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 173–187, Mar. 2018.

[3] Thomas William Sherson, Richard Heusdens, and W. Bastiaan Kleijn, "Derivation and Analysis of the Primal-Dual Method of Multipliers Based on Monotone Operator Theory," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 2, pp. 334–347, June 2019.

[4] Heinz H. Bauschke and Patrick L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics. Springer International Publishing, Cham, 2017.

[5] Jean Jacod and Philip E Protter, *Probability essentials*, Springer, Berlin; New York, 2004, OCLC: 780176454.

[6] Franck Iutzeler, Pascal Bianchi, Philippe Ciblat, and Walid Hachem, "Asynchronous distributed optimization using a randomized alternating direction method of multipliers," in *52nd IEEE Conference on Decision and Control*, Firenze, Dec. 2013, pp. 3671–3676, IEEE.

[7] Pascal Bianchi, Walid Hachem, and Franck Iutzeler, "A Coordinate Descent Primal-Dual Algorithm and Application to Distributed Asynchronous Optimization," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2947–2957, Oct. 2016.

[8] Patrick L. Combettes and Jean-Christophe Pesquet, "Stochastic Quasi-Fejér Block-Coordinate Fixed Point Iterations with Random Sweeping," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 1221–1248, Jan. 2015.

[9] Nicola Bastianello, Ruggero Carli, Luca Schenato, and Marco Todescato, "Asynchronous Distributed Optimization Over Lossy Networks via Relaxed ADMM: Stability and Linear Convergence," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2620–2635, June 2021.

[10] Jesper Dall and Michael Christensen, "Random geometric graphs," *Physical Review E*, vol. 66, no. 1, pp. 016121, July 2002.