QUANTISATION EFFECTS IN PDMM: A FIRST STUDY FOR SYNCHRONOUS DISTRIBUTED AVERAGING

Daan H. M. Schellekens[1,2], Thomas Sherson[1], and Richard Heusdens[1]

[1]Circuits and Systems group, Delft University of Technology, Delft, the Netherlands

[2]Pindrop, Atlanta, United States

dschellekens@pindrop.com, t.w.sherson@tudelft.nl, r.heusdens@tudelft.nl

ABSTRACT

Large-scale networks of computing units, often characterised by the absence of central control, have become commonplace in many applications. To facilitate data processing in these large-scale networks, distributed signal processing is required. The iterative behaviour of distributed processing algorithms combined with energy, computational power, and bandwidth limitations imposed by such networks, place tight constraints on the transmission capacities of the individual nodes. In this paper we investigate the effects of subtractive dithered uniform quantisation in PDMM for the synchronous distributed averaging problem. This is done by deriving expressions for the mean squared error (MSE) that include quantisation noise. Also, the required data rate for quantised PDMM is considered. It was found that for practical applications quantisation in PDMM can be applied with a fixed-rate quantiser, such that significant data rate reduction can be achieved, without compromising the rate of convergence.

Index Terms— PDMM, quantisation, subtractive dithering.

1. INTRODUCTION

Over the past years there has been a considerable growth in the number of *large-scale sensor networks*. These networks consist of a large number of nodes, each having a sensing, data processing, and communication component. Examples of these networks are the 'Internet of Things' (IoT), ad-hoc wireless sensor networks, peer-to-peer networks, mobile networks of vehicles, and social networks [1]. These large-scale networks are characterised by the absence of a central control or processing unit (fusion centre), and as a consequence nodes use their own processing ability to locally carry out simple computations and transmit only the required and partially processed data to neighbouring nodes. The decentralised settings in which algorithms then have to operate are typically dynamic, in the sense that sensors can be added or removed in an unpredictable way. As a consequence, these algorithms must allow for a decentralised implementation, must be easily scalable, must be robust against (small) changes in the network topology, and, in the case of wireless sensor networks, must be robust to possible transmission failures.

There are three popular methods to solve signal processing problems in a distributed manner, namely methods based on distributed consensus algorithms [1], algorithms for probabilistic inference [2], and algorithms based on convex optimisation [3]. In either case, the problem boils down to how to reformulate the problem at hand in a form that allows for an efficient and robust distributable implementation. Over the last decade, the *alternating-direction method of multipliers* (ADMM) has gained considerable attention to solve convex signal processing problems in a distributed manner [4, 5]. Recently, an alternative approach has been introduced, called the *primal-dual method of multipliers* (PDMM) [6, 7, 8] which has the advantage over ADMM that it is completely node-based (which allows for an efficient asynchronous implementation) whilst being robust against transmission failures (packet loss).

Both ADMM and PDMM are iterative algorithms. In applications involving sensor networks, this means that at each iteration the partially processed data needs to be communicated to neighbouring nodes. In order to do so, the data needs to be quantised and represented by a finite number of bits prior to transmission. The more precise the representation, the higher the data rate in the channels and, as a consequence, the higher the total transmission power. Since the algorithm converges for any initialisation of the node variables, we could start the iterations with a course (low-rate) representation of the data, thereby saving transmission power, and gradually increase the accuracy with increasing iterations. On the other hand, the amount of information that needs to be transmitted will decrease with increasing iterations since the difference between successive messages will become (close to) zero when the algorithm converges, which suggests that we can use a low-rate representation throughout the iterations and still end-up with accurate results.

The effect of quantisation on the final accuracy and the convergence rate of some distributed (iterative) algorithms has been investigated [9, 10, 11], including ADMM [12], but no such results are known for PDMM. In this paper we make a first attempt to investigate what the effect of quantisation is for synchronous PDMM. To do so, we restrict ourself to distributable consensus problems, in particular distributed averaging. The main motivation to focus on this particular class of algorithms is its mathematical tractability (it will result in linear update equations) and the numerous applications of consensus algorithms that have been proposed recently, such as power control in smart grids, load balancing, formation of autonomous agents like cars or unmanned aerial vehicles, wireless sensor networks, and coordination of mobile robots [1, 13].

2. PRIMAL-DUAL METHOD OF MULTIPLIERS

The primal-dual method of multipliers (PDMM) for iteratively solving a separable convex optimisation problem defined over a graph $G = (V, E)$ has been proposed in [6, 7, 8]. Here, V and E denote the set of *nodes* and *directed edges*, respectively, with $|V| = n$, the total number of nodes in the network. PDMM solves a problem of the form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i \in V} f_i(\mathbf{x}_i) \\ \text{subject to} \quad & A_{ij}\mathbf{x}_i + A_{ji}\mathbf{x}_j = \mathbf{c}_{ij}, \quad \forall (i,j) \in E, \end{aligned} \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$, $\mathbf{x} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T)^T$, with n_i the dimension of $\mathbf{x}_i$. In this paper, PDMM is applied to the distributed scalar averaging problem with local objectives $f_i(x_i) = \frac{1}{2}(x_i - t_i)^2$, where t_i denotes the measurement data at node i [8]. This is a consensus problem with scalar node variables x_i, such that $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, and equality constraints $x_i = x_j$ on every edge $(i,j) \in E$. Hence we have $A_{ij} = 1$ for $i > j$ and $A_{ij} = -1$ otherwise. As shown in [14], the PDMM update steps for this problem can be split in a *primal* update

$$x_i^{(k)} = \frac{t_i + \sum_{j \in N(i)} \left(A_{ij} \mu_{j|i}^{(k-1)} + \rho_{\mathrm{p}} x_j^{(k-1)} \right)}{1 + d_i \rho_{\mathrm{p}}}, \quad \forall i \in V, \quad (2)$$

where $N(i) = \{j \in V | (i,j) \in E\}$ denotes the set of neighbouring nodes and $|N(i)| = d_i$ the degree of node i, and a *dual* update

$$\mu_{i|j}^{(k)} = \mu_{j|i}^{(k-1)} - \rho_{\mathrm{d}} A_{ij} \left(x_i^{(k)} - x_j^{(k-1)} \right), \quad \forall (i,j) \in E, \quad (3)$$

where the superscript $^{(k)}$ refers to iteration k. In this paper it is assumed that the ρ-parameters are chosen as $\rho_{\mathrm{p}} = \rho_{\mathrm{d}}^{-1} = \rho > 0$.

2.1. Linear system of equations

In order to analyze the quantisation effects for PDMM, we consider the vector $\boldsymbol{\mu} = (\mu_{1|2}, \ldots, \mu_{1|n}, \ldots, \mu_{n|1}, \ldots, \mu_{n|n-1})^T \in \mathbb{R}^{n(n-1)}$ of all dual variables in a complete (fully connected) n-node graph. In the case the network is not fully connected, we exclude the dual variables associated to non-existing edges, resulting in the vector $\boldsymbol{\mu} \in \mathbb{R}^d$, with $d = \sum_i d_i$. We combine the primal and dual variables in a vector $\mathbf{y} = (\mathbf{x}^T, \boldsymbol{\mu}^T)^T \in \mathbb{R}^{n+d}$. As a result, the update equations (2) and (3) can be compactly represented as

$$\begin{aligned} \mathbf{y}^{(k)} &= F\mathbf{y}^{(k-1)} + \mathbf{u} \\ &= F^k \mathbf{y}^{(0)} + \sum_{i=0}^{k-1} F^i \mathbf{u}, \end{aligned} \quad (4)$$

where $\mathbf{u} \in \mathbb{R}^{n+d}$ is a vector containing the measurement data.

In order to simplify upcoming equations, we will assume that $\mathbf{y}^{(0)} = \mathbf{0}$ although similar results can be obtained for nonzero initialisation. As a consequence, the first term in (4) vanishes. Furthermore, let us assume that F is diagonalisable[1], that is, we can express F as $F = V \Lambda V^{-1}$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n+d})$ is a diagonal matrix containing the eigenvalues of F on the main diagonal and V is a matrix whose columns consist of the eigenvectors of F. With this, (4) can be expressed as

$$\mathbf{y}^{(k)} = V \left(\sum_{i=0}^{k-1} \Lambda^i \right) V^{-1} \mathbf{u}. \quad (5)$$

From the convergence analysis in [8] it follows that the magnitude of the eigenvalues are bounded by one. In addition, in [15] it is shown that the eigenvalues having magnitude one are real. Let $V = (V_1 \; V_2)$, where range(V_1) and range(V_2) are the subspace spanned by the eigenvectors associated to unit magnitude eigenvalues and the subspace spanned by the eigenvectors associated to eigenvalues $|\lambda_i| < 1$, respectively. It then follows that $V_1^T V_2 = O$ and $V_1^+ \mathbf{u} = \mathbf{0}$, where the superscript $^+$ denotes pseudo-inversion [15]. Hence, (5) reduces to

$$\mathbf{y}^{(k)} = V_2 \left(\sum_{i=0}^{k-1} \Lambda_2^i \right) V_2^+ \mathbf{u}, \quad (6)$$

where Λ_2 contains the complex eigenvalues having magnitude strictly less than unity.

2.2. Convergence analysis

Let $\mathbf{y}^*$ denote the fixed point of the iterations. By inspection of (6), $\mathbf{y}^*$ can be written as

$$\mathbf{y}^* = V_2 \left(\sum_{i=0}^{\infty} \Lambda_2^i \right) V_2^+ \mathbf{u} = V_2 \left(I - \Lambda_2 \right)^{-1} V_2^+ \mathbf{u}, \quad (7)$$

which follows from rewriting the infinite geometric series, assuming that $|\lambda_i| < 1$, which holds by the definition of Λ_2. With this, the error $\mathbf{e}^{(k)} = \mathbf{y}^{(k)} - \mathbf{y}^*$ can be expressed as

$$\begin{aligned} \mathbf{e}^{(k)} &= -V_2 \Lambda_2^k \left(I - \Lambda_2 \right)^{-1} V_2^+ \mathbf{u} \\ &= E^{(k)} \mathbf{u}, \end{aligned} \quad (8)$$

which shows that the algorithm converges at a linear rate determined by the magnitude of the largest eigenvalue $|\lambda_{2,\max}|$ of Λ_2.

Some remarks are in place here. Due to our choice $\mathbf{y}^{(0)} = \mathbf{0}$, there will be no information in the subspace spanned by the columns of V_1. If we initialise with arbitrary $\mathbf{y}^{(0)}$, however, this will result in an additional contribution. Since the columns of V_1 are the eigenvectors associated to the unit magnitude eigenvalues, this additional contribution is either present in every iteration (due to the eigenvalues $\lambda_i = 1$) or will alternate between successive iterations (due to eigenvalues $\lambda_i = -1$). However, as shown in [15], all eigenvectors associated to the unit magnitude eigenvalues have its first n entries zero, which implies that the primal variables will converge to a fixed point, while the dual variables could start alternating between two values, depending on the actual choice of $\mathbf{y}^{(0)}$. Since we are primarily interested in the information contained in the primal variables, this is no issue here. However, as we will see later when we consider quantisation noise which will end up in the subspace spanned by the columns of V_1 even though we initialise $\mathbf{y}^{(0)} = \mathbf{0}$, this issue will becomes relevant.

The error vector can be split in an error in the primal and dual variables as $\mathbf{e} = (\mathbf{e}_x^T, \mathbf{e}_\mu^T)^T \in \mathbb{R}^{n+d}$. Introducing a selection matrix $S_x = (I_n \; O) \in \mathbb{R}^{n \times (n+d)}$ allows us to write the primal error as $\mathbf{e}_x = S_x \mathbf{e} \in \mathbb{R}^n$. Let the primal normalized squared error (SE)

$$\begin{aligned} \zeta_x^{(k)} &= \frac{1}{n} \mathrm{tr} \left(\mathbf{e}_x^{(k)} \mathbf{e}_x^{(k)H} \right) \\ &= \frac{1}{n} \mathrm{tr} \left(S_x E^{(k)} \mathbf{u}\mathbf{u}^H E^{(k)H} S_x^H \right), \end{aligned} \quad (9)$$

be defined as the sum of the squared error in all the primal variables divided by the number of elements, where the superscript ^{H} denotes the Hermitian transpose and tr$(\cdot)$ denotes the trace of a matrix. By taking the expectation, the following expression for the primal mean squared error (MSE) is obtained

$$\mathbb{E} \left[\zeta_x^{(k)} \right] = \frac{1}{n} \mathrm{tr} \left(S_x E^{(k)} \Sigma_u E^{(k)H} S_x^H \right), \quad (10)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and $\Sigma_u = \mathrm{cov}(\mathbf{u})$, where cov $(\cdot)$ denotes the covariance matrix. Note, that the property

[1] If F is not diagonalisable, we can express F in its Jordan normal form and derive similar results as presented here.

$\mathbb{E}[\mathrm{tr}(\cdot)] = \mathrm{tr}(\mathbb{E}[\cdot])$ is used and that Σ_u is defined by the distribution of the measurements t_i. The primal MSE is of main interest since the estimate of the average value is stored in the primal variables. From (8) it can be seen that the primal MSE, $\mathbb{E}[\zeta_x^{(k)}]$, introduced in (10), converges at a rate $|\lambda_{2,\max}|^{2k}$.

3. UNIFORM SUBTRACTIVE-DITHER QUANTISATION

Let $\hat{\mathbf{y}}^{(k)} = Q(\tilde{\mathbf{y}}^{(k)}) = \tilde{\mathbf{y}}^{(k)} + \mathbf{n}_\mathrm{q}^{(k)}$ be the uniformly quantised version of the vector $\tilde{\mathbf{y}}^{(k)}$, where $\mathbf{n}_\mathrm{q}^{(k)}$ denotes the quantisation noise caused by the uniform quantisation operation and $\tilde{\mathbf{y}}^{(k)}$ is the vector calculated based on previously received quantised vectors. This allows us to rewrite (4) as

$$\tilde{\mathbf{y}}^{(k)} = F\left(\tilde{\mathbf{y}}^{(k-1)} + \mathbf{n}_\mathrm{q}^{(k-1)}\right) + \mathbf{u} = \mathbf{y}^{(k)} + \sum_{i=0}^{k-1} F^{k-i}\mathbf{n}_\mathrm{q}^{(i)}.$$

In this paper, we add a so-called dither signal $\mathbf{n}_\mathrm{d}^{(k)}$ before quantisation, resulting in a dithered output signal $\tilde{\mathbf{y}}_\mathrm{d}^{(k)} = \tilde{\mathbf{y}}^{(k)} + \mathbf{n}_\mathrm{d}^{(k)}$, to give the quantisation noise certain desirable properties to be discussed below. The dither signals are generated by a pseudo-random generator at the transmitting side and subtracted at the receiving side, which has a pseudo-random generator with the same seed. This seed needs to be communicated only once at the start of the iterations. Let $\Delta_{(k)}$ be the distance between the reproduction values of the uniform quantiser at iteration k. If we choose the dither realisations to be i.i.d. uniformly distributed on the interval $[-\Delta_{(k)}, \Delta_{(k)}]$, then the quantisation noise realisations will also be zero-mean i.i.d. uniform distributed with variance $\Delta_{(k)}^2/12$ and they will be independent of the quantiser input [16, Theorem 4]. If, in addition, the dither realisations are statistically independent in time of one another, then so are the quantisation noise realisations [16, Theorem 5].

3.1. Convergence analysis

The error $\tilde{\mathbf{e}}^{(k)} = \tilde{\mathbf{y}}^{(k)} - \mathbf{y}^*$ for quantised PDMM is given by

$$\tilde{\mathbf{e}}^{(k)} = \mathbf{e}^{(k)} + \sum_{i=0}^{k-1} F^{k-i}\mathbf{n}_\mathrm{q}^{(i)} = \mathbf{e}^{(k)} + \mathbf{e}_\mathrm{q}^{(k)}.$$

When only considering the primal error $\tilde{\mathbf{e}}_x^{(k)}$, the primal normalised squared error (SE), $\tilde{\zeta}_x^{(k)} = n^{-1}\|\tilde{\mathbf{e}}_x^{(k)}\|_2^2$, can be used to express the primal MSE

$$\mathbb{E}\left[\tilde{\zeta}_x^{(k)}\right] = \mathbb{E}\left[\zeta_x^{(k)}\right] + \mathbb{E}\left[\zeta_{\mathrm{q},x}^{(k)}\right], \tag{11}$$

where cross terms of $\mathbf{e}_x^{(k)}$ and $\mathbf{e}_{\mathrm{q},x}^{(k)}$ dropped out since they are made independently by dithering. Here, the first term $\mathbb{E}[\zeta_x^{(k)}]$ in (11) was given in (10). Let us introduce $F_2^{(k)} = V_2\Lambda_2^k V_2^+$. From [15, Proposition A.6.3] it can be shown that $S_x V_1 = O$, such that $S_x F^k = S_x F_2^{(k)}$. By combining this with earlier introduced properties, it is shown in [15] that the second term $\mathbb{E}[\zeta_{\mathrm{q,x}}^{(k)}]$ in (11) can be expressed as

$$\mathbb{E}\left[\zeta_{\mathrm{q},x}^{(k)}\right] = \frac{1}{12n}\sum_{i=1}^{k}\Delta_{(k-i)}^2\,\mathrm{tr}\left(S_x F_2^{(i)} F_2^{(i)H} S_x^H\right). \tag{12}$$

Note that $\mathrm{tr}(S_x F_2^{(i)} F_2^{(i)H} S_x^H) \to 0$ at a rate $|\lambda_{2,\max}|^{2i}$ as $i \to \infty$. Hence, if we choose

$$\Delta_{(k)} = |\lambda_{2,\max}|^k \Delta_{(0)}, \tag{13}$$

then $\mathbb{E}[\zeta_{\mathrm{q},x}^{(k)}]$ converges at a rate $k \cdot |\lambda_{2,\max}|^{2k}$, while $\mathbb{E}[\zeta_x^{(k)}]$ converges at a rate $|\lambda_{2,\max}|^{2k}$. Since $\mathbb{E}[\zeta_{\mathrm{q},x}]$, in contrast to $\mathbb{E}[\zeta_x]$, is dependent on $\Delta_{(k)}$ and therefore on $\Delta_{(0)}$, we can make $\mathbb{E}[\zeta_{\mathrm{q},x}]$ arbitrarily small by making $\Delta_{(0)}$ sufficiently small. By doing so, the rate of convergence will be unaffected during the first number of iterations. This duration can be made arbitrarily long by making $\Delta_{(0)}$ smaller, such that any required precision for a practical application can be achieved without compromising the rate of convergence. This all comes, however, at the cost of a higher data rate since a more precise quantiser is required for every iteration. This will be evaluated in the next subsection.

3.2. Data rate analysis

Using a quantiser with decreasing cell width suggests an increase in data rate with increasing iterations. On the other hand, the amount of information that needs to be transmitted will decrease with increasing iterations. To analyse this, a quantised difference vector $\hat{\mathbf{v}}^{(k)} = Q(\tilde{\mathbf{v}}^{(k)}) = \tilde{\mathbf{v}}^{(k)} + \mathbf{n}_\mathrm{q}^{(k)}$ will be transmitted, where $\tilde{\mathbf{v}}^{(k)} = \tilde{\mathbf{y}}^{(k)} - \hat{\mathbf{y}}^{(k-2)}$, with $\hat{\mathbf{y}}^{(k)} = \mathbf{0}\ \forall k \le 0$. Note that we consider the difference between two successive iterations for reasons mentioned in Section 2.2. If the same previously seen dither signal $\mathbf{n}_\mathrm{d}^{(k)}$ is added before quantisation, we obtain the source vector for the quantiser

$$\begin{aligned}\tilde{\mathbf{v}}_\mathrm{d}^{(k)} = \sum_{i=1}^{2} F_2^{(k-i)}\mathbf{u} + \sum_{i=0}^{k-2}\left(F_2^{(k-i)} - F_2^{(k-2-i)}\right)\mathbf{n}_\mathrm{q}^{(i)} \\ + F\mathbf{n}_\mathrm{q}^{(k-1)} + \mathbf{n}_\mathrm{d}^{(k)},\end{aligned} \tag{14}$$

where the quantisation noise vector $\mathbf{n}_\mathrm{q}^{(k)}$ will be a realisation from the same i.i.d. uniform distribution as before. Therefore, the convergence properties do not change. With (14), the covariance matrix of the quantiser input can be expressed as

$$\begin{aligned}\mathrm{cov}\left(\tilde{\mathbf{v}}_\mathrm{d}^{(k)}\right) &= \left(\sum_{i=1}^{2} F_2^{(k-i)}\right)\Sigma_u\left(\sum_{i=1}^{2} F_2^{(k-i)}\right)^H \\ &+ \sum_{i=0}^{k-2}\frac{\Delta_{(i)}^2}{12}\left(F_2^{(k-i)} - F_2^{(k-2-i)}\right)\left(F_2^{(k-i)} - F_2^{(k-2-i)}\right)^H \\ &+ \frac{\Delta_{(k-1)}^2}{12}FF^H + \frac{\Delta_{(k)}^2}{12}I_{n+d}.\end{aligned} \tag{15}$$

Let the messages $v_i^{(k)}$ with $1 \le i \le n$ be the first n elements of $\tilde{\mathbf{v}}_\mathrm{d}^{(k)}$. Each message is composed of a linear combination of Gaussian and uniformly distributed variables. These messages have a source variance, denoted by $\sigma_{v_i,(k)}^2$, equal to the first n diagonal elements of the covariance matrix in (15), which can be used to upper bound the differential entropy of $v_i^{(k)}$. For any source this upper bound is given by the differential entropy of a Gaussian source with the same variance [17]. With this, the discrete entropy of the quantiser output can be upper bounded at high-rate by

$$H\left(\hat{v_i}^{(k)}\right) \le \frac{1}{2}\log_2\left(\frac{2\pi e\sigma_{v_i,(k)}^2}{\Delta_{(k)}^2}\right), \tag{16}$$

where it is assumed that $\Delta_{(k)}$ is the same for every node i [18].

All but the second term in (15) converge at a rate $|\lambda_{2,\max}|^{2k}$. The second term converges at a rate $k \cdot |\lambda_{2,\max}|^{2k}$, such that the ratio in (16) gradually increases for the choice $\Delta_{(k)} = |\lambda_{2,\max}|^k\Delta_{(0)}$. Therefore, it is not possible to obtain a constant data rate for this specific choice of $\Delta_{(k)}$. However, as can be seen in the next section,

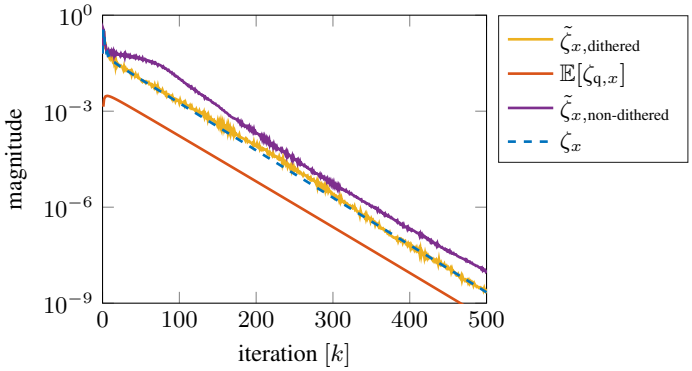

Fig. 1. The SE $\tilde{\zeta}_x$ for quantised PDMM with and without subtractive dithering with decreasing cell width $\Delta_{(k)}$ with $\Delta_{(0)} = 10^{-1}$.

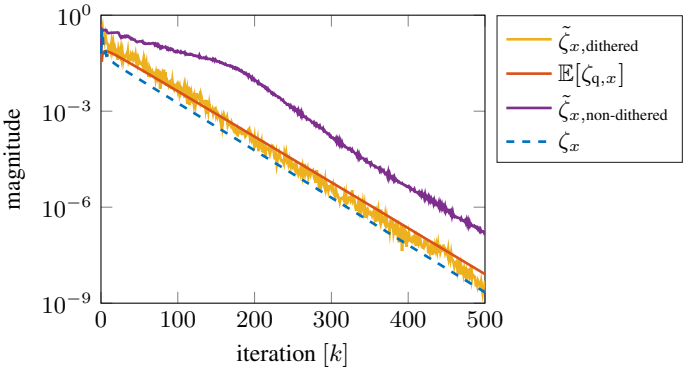

Fig. 2. The SE $\tilde{\zeta}_x$ for quantised PDMM with and without subtractive dithering with decreasing cell width $\Delta_{(k)}$ with $\Delta_{(0)} = 5 \cdot 10^{-1}$.

the second term in (15) is negligible small compared to the other terms, such that for practical applications a fixed-rate quantiser can be used.

4. SIMULATION RESULTS

In this section we present results obtained by computer simulations. The simulations were generated using the same random $n = 10$ node network, with each comprised of 500 iterations. The penalty parameter was chosen as $\rho = 1/n = 0.1$. Only the primal MSE was considered.

To ensure the convergence of the primal MSE in the case of quantised transmission, a variable $\Delta_{(k)}$ with a linear decreasing rate $|\lambda_{2,\max}|^k$, similar to the non-quantised MSE, was introduced. In Section 3 it was shown that $\Delta_{(0)}$ plays an important role in the performance in terms of the convergence rate. Therefore, results for two different values of $\Delta_{(0)}$ are demonstrated.

Fig. 1 and 2 contain the non-quantised reference ζ_x, the SE $\tilde{\zeta}_x$ for the same realisations of the measurements t_i, and the MSE of the noise contribution $\mathbb{E}[\zeta_{q,x}]$ when subtractive dither is used. By comparing Fig. 1 and 2, it can be seen that a smaller $\Delta_{(0)}$ in Fig. 1 causes $\mathbb{E}[\zeta_{q,x}]$ to reduce, which minimises the influence of the quantisation noise. For a small enough $\Delta_{(0)}$, $\tilde{\zeta}_x$ will be identical to ζ_x, however this comes with the cost of a higher transmission data rate as can be seen from Fig. 3. The SE $\tilde{\zeta}_x$ without subtractive dither is plotted in purple to show the benefits of subtractive dither.

In Section 3 it was shown that for an exponentially decreasing cell width $\Delta_{(k)}$, $\mathbb{E}[\zeta_{q,x}]$ from (12) has a different rate of convergence than $\mathbb{E}[\zeta_x]$ from (10). This different rate of convergence is most apparent in Fig. 2, where the red and the blue dashed line have distinctly different slopes. However, regardless the initial selection of $\Delta_{(0)}$, asymptotically $\mathbb{E}[\zeta_{q,x}]$ will always dominate $\mathbb{E}[\zeta_x]$ after a certain number of iterations. However, for practical applications, with small enough $\Delta_{(0)}$, this point can be made to occur after the required precision has been achieved. As such, an exponentially decreasing $\Delta_{(k)}$ is sufficient to maintain the convergence rate of unquantised PDMM.

Fig. 3 shows the required data rate for quantised PDMM for three different values $\Delta_{(0)}$ when difference messages are transmitted. The lines represent $n^{-1}\sum_i H(\hat{v}_i^{(k)})$, the mean over ten nodes of the upper bound on the data rate from (16). Based on (15), (16) and Fig. 3, it can be noted that a smaller $\Delta_{(0)}$ requires a higher data rate, since there are more quantisation cells to represent with a smaller cell width. Furthermore, it can be seen that during the first iterations the data rate is relatively high. This corresponds with the effect of a very fast decrease in the SE during the first iterations, which requires more information to be transmitted. Lastly, the near horizontal lines for the later iterations indicate that a fixed-rate quantiser will be sufficient for practical applications that stop transmitting data after the required precision has been achieved.

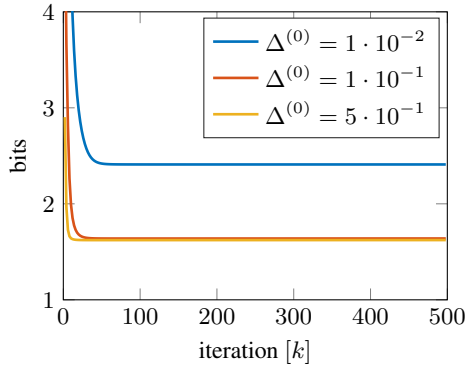

Fig. 3. The mean required data rate to quantise $v_i^{(k)}$ for a variable cell width for three different values of $\Delta_{(0)}$.

5. CONCLUSIONS

In this paper we examined the influence of quantisation on the convergence of PDMM for the synchronous distributed averaging problem. We showed that it is possible to apply quantisation in PDMM and thereby reduce the data rate while maintaining the same convergence rate by exponentially decreasing the quantiser cell width at every iteration with the same rate as the linear convergence bound. The expressions for the MSE and the figures in Section 4 emphasise the role that the initial quantiser cell width plays in this. The presented conclusions, however, only hold for practical applications which demand a finite precision such that the algorithm stops after a certain number of iterations. This is because the rate of convergence of quantised PDMM asymptotically becomes $k \cdot |\lambda_{2,\max}|^{2k}$ for an exponentially decreasing cell width, due to the part of the MSE caused by the quantisation noise. This means, that for the proposed decreasing cell width scheme, an infinite precision can never be obtained without compromising convergence performance.

6. REFERENCES

[1] D. Shah, *Gossip algorithms*, Now Publishers Inc, 2009.

[2] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, New York, corr. second printing edition edition, 2007.

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

[4] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, January 2011.

[6] G. Zhang and R. Heusdens, "Bi-alternating direction method of multipliers," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3317–3321.

[7] G. Zhang, R. Heusdens, and W. B. Kleijn, "On the convergence rate of the bi-alternating direction method of multipliers," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3869–3873.

[8] G. Zhang and R. Heusdens, "Bi-alternating direction method of multipliers over graphs," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 3571–3575.

[9] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4905–4918, Oct 2008.

[10] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, March 2010.

[11] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.

[12] S. Zhu and B. Chen, "Quantized consensus by the ADMM: Probabilistic versus deterministic quantizers," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1700–1713, April 2016.

[13] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc wsns with noisy links; part i: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, Jan 2008.

[14] G. Zhang and R. Heusdens, "On simplifying the primal-dual method of multipliers," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4826–4830.

[15] D. H. M. Schellekens, "Quantization effects in PDMM: A first study for synchronous distributed averaging," M.S. thesis, Delft University of Technology, Delft, the Netherlands, 2016.

[16] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: A theoretical survey," *J. Audio Eng. Soc*, vol. 40, no. 5, pp. 355–375, May 1992.

[17] D. J. C. MacKay, *Information theory, inference and learning algorithms*, Cambridge university press, 2003.

[18] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, September 1968.