

ON THE DUALITY OF GLOBALLY CONSTRAINED SEPARABLE PROBLEMS AND ITS APPLICATION TO DISTRIBUTED SIGNAL PROCESSING

Thomas Sherson¹ Richard Heusdens¹ W. Bastiaan Kleijn^{1,2}

¹ Faculty of EEMCS, Delft University of Technology, Netherlands

² School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

ABSTRACT

In this paper, we focus on the challenge of processing data generated within decentralised wireless sensor networks in a distributed manner. When the desired operations can be expressed as globally constrained separable convex optimisation problems, we show how we can convert these to *extended monotropic programs* and exploit Lagrangian duality to form equivalent distributed consensus problems. Such problems can be embedded in sensor network applications via existing solvers such as the alternating direction method of multipliers or the primal dual method of multipliers. We then demonstrate how this approach can be used to solve specific problems including linearly constrained quadratic problems and the classic Gaussian channel capacity maximisation problem in a distributed manner.

Index Terms— Wireless sensor networks, distributed signal processing, Lagrangian duality, extended monotropic programs.

I. INTRODUCTION

Following the miniaturisation of sensor technologies over the last few decades, there has been a significant increase in the interest in, and deployment of, large scale wireless sensor networks (WSN) [1]. In many applications, including environmental monitoring [1] and distributed power generation [2], the size of such networks makes it impractical to deploy centralised signal processing systems. This stems from the inability of such structures to scale dynamically with changes in network size coupled with the high transmission costs required to communicate data to a central location. In contrast, distributed systems are attractive as they utilise on-node computation and localised communication to achieve the same functionality whilst addressing many of these limitations.

Distributed networks, characterised by their limited connectivity, implicitly restrict the data available at any one node. As data is often generated within the network, this makes many traditional signal processing operations challenging to perform. One approach commonly used in the literature, in applications such as in sensor localisation [3], global averaging of data [4] and network utility optimisation [5], is to embed desired signal processing operations inside

convex optimisation problem, which can then be solved distributedly. Commonly used distributed solvers leverage sub-gradient [6], message passing [7], randomised Gossip [8] and primal-dual based algorithms [9], where each algorithm is chosen based on the specifications of a particular application.

The major challenge faced by this approach is in the transformation of desired problems to equivalent distributable forms. Within the literature, existing approaches have aimed to address this point. In [10], a parallelised proximal based approach was demonstrated, able to exploit the separability of both the objective and constraint functions to distribute part of the computation. However, for global constraints, such methods still require the aggregation of data within the network for some operations and thus can suffer from the same limitations as centralised methods.

In this paper we highlight the observation that *neighbourhood separable problems with global separable constraints* (NSGC) can be transformed to *extended monotropic problems* [11] and from there to fully distributable dual forms. We then demonstrate how this transformation can be used to form distributed versions of particular sensor network optimisation problems, specifically for the cases of linearly constrained quadratic problems as well as the Gaussian channel capacity maximisation problem.

The remainder of this paper is organised as follows. In Section II, we introduce the family of NSGC problems which can be transformed to a distributable dual problem. In Section III we demonstrate how this approach can be used to derive distributed algorithms for solving both linearly constrained quadratic problems as well as the Gaussian channel capacity maximisation problem. Finally in Section IV we draw some conclusions about this approach.

II. DISTRIBUTED DUALITY OF NEIGHBOURHOOD SEPARABLE PROBLEMS

In this section we highlight how particular globally constrained convex optimisation problems can be cast in an equivalent distributed form via Lagrangian duality. Subsection II-A outlines the basic model we will use for a distributed WSN whilst subsection II-B demonstrates the dual-distributability of NSGC problems through their equivalence to extended monotropic programs.

II-A. A Distributed Wireless Sensor Network Model

Consider the problem of processing a set of measured data ($\mathbf{a} \in \mathbb{R}^N$) collected by a WSN comprised on N nodes. We will denote by $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ the associated undirected graph of this network with vertex set \mathbf{V} and edge set \mathbf{E} . It is assumed that the structure of the network may vary with time but does so at a slow rate such that \mathbf{G} describes a dynamic distributed sensor network. A basic example of such a \mathbf{G} is included below in Figure 1.

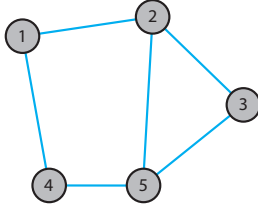


Fig. 1: A simple model of a distributed network topology.

The decentralised nature of \mathbf{G} imposes that no one node has access to the entire set of node-generated data. For even simple operations, such as computing the inner product between α and a known vector $\tilde{\alpha} \in \mathbb{R}^N$, this limited access restricts the use of traditional signal processing techniques without the use of additional data aggregation methods [12]. The problem with such approaches is that the memory required by each node to store α during this aggregation scales with the dimension of the network thereby increasing the cost of node based hardware. Furthermore, aggregation techniques require each node's data to be uniquely identifiable where the allocation of these identifies requires network wide consensus to ensure their uniqueness. We are therefore interested in the transformation of centralised signal processing algorithms to distributed forms to directly exploit the nature of data generation in such sensor networks.

II-B. A Distributable Class of Optimisation Problems

To develop distributed algorithms for use with G , we can embed a desired signal processing operation within a convex optimisation problem and then transform this to an equivalent *distributed form*. By *distributed form* we specifically refer to optimisation problems with convex node-separable objective functions and edge-based constraints such that

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i \in \mathbf{V}} f_i(\mathbf{x}_i) \\ \text{s.t.} \quad & g_i(\mathbf{x}_i) + g_j(\mathbf{x}_j) \leq 0 \quad \forall (i, j) \in \mathbf{E} \\ & \mathbf{A}_{i,j} \mathbf{x}_i + \mathbf{A}_{j,i} \mathbf{x}_j = \mathbf{c}_{i,j} \quad \forall (i, j) \in \mathbf{E}, \end{aligned}$$

where x_i denotes the local variables (potentially vector valued) at node i . Unfortunately, the transformation of centralised optimisation problems to a distributed form is not always feasible. However, for a specific subclass of problems, we will show how this is possible.

We consider what we term as *neighbourhood separable problems with global constraints* which have the following form:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i \in \mathbf{V}} \left(f_i(\mathbf{x}_i, \{\mathbf{x}_j \mid j \in \mathcal{N}(i)\}) \right) \\ \text{s.t.} \quad & \sum_{i \in \mathbf{V}} g_{i,k}(\mathbf{x}_i, \{\mathbf{x}_j \mid j \in \mathcal{N}(i)\}) \leq 0 \quad \forall k = 1, \dots, K \\ & \mathbf{A}_l \mathbf{x} = \mathbf{b}_l \quad \forall l = 1, \dots, L \end{aligned}$$

where f_i and $g_{i,k}$ are closed proper convex (CCP) functions and $j \in \mathcal{N}(i)$ denotes the set of nodes in the neighbourhood of node i . We will assume that this problem is strictly feasible such that strong duality holds.

By introducing local versions of neighbouring variables at each node, denoted by $\mathbf{z}_{i,j} \forall j \in \mathcal{N}(i)$, and imposing the constraint that at consensus each $\mathbf{z}_{i,j} = \mathbf{x}_j$, we can transform any NSGC to the following form:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \sum_{i \in \mathbf{V}} \left(f_i(\mathbf{x}_i, \{\mathbf{z}_{i,j} \mid j \in \mathcal{N}(i)\}) \right) \\ \text{s.t.} \quad & \sum_{i \in \mathbf{V}} g_{i,k}(\mathbf{x}_i, \{\mathbf{z}_{i,j} \mid j \in \mathcal{N}(i)\}) \leq 0 \quad \forall k = 1, \dots, K \\ & \mathbf{A}_l \mathbf{x} = \mathbf{b}_l \quad \forall l = 1, \dots, L \\ & \mathbf{x}_j - \mathbf{z}_{i,j} = \mathbf{0} \quad \forall (i, j) \in \mathbf{E} \end{aligned} \quad (1)$$

The reformulated problem, reduces the broader class of NSGC problems, to equivalent *extended monotropic programs* [11] which are a generalisation of *monotropic programs*, first proposed by Rockafellar [13]. *Extended monotropic programs* are separable objective problems with separable convex constraints and contain as a subset all distributed optimisation problems. However, due to the global constraints noted in (1), this problem is not yet distributable.

The associated Lagrangian of Eq. (1) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu}) = & \sum_{i \in \mathbf{V}} \left(f_i(\mathbf{x}_i, \{\mathbf{z}_{i,j} \mid j \in \mathcal{N}(i)\}) \right) \\ & + \sum_{k=1}^K \lambda_k g_{i,k}(\mathbf{x}_i, \{\mathbf{z}_{i,j} \mid j \in \mathcal{N}(i)\}) + \sum_{j \in \mathcal{N}(i)} \boldsymbol{\nu}_{i,j}^T \mathbf{z}_{i,j} \\ & - \sum_{j \in \mathcal{N}(i)} \boldsymbol{\nu}_{j,i}^T \mathbf{x}_i - \sum_{l=1}^L \boldsymbol{\mu}_l^T (\mathbf{a}_{l,i} x_i - \mathbf{b}_l) \end{aligned}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ denote the introduced dual variables and $\mathbf{a}_{i,l}$ denotes the i th column vector of \mathbf{A}_l . It can then be shown that the dual problem also exhibits the same separability as the Lagrangian such that

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu}) &= \inf_{\mathbf{x}, \mathbf{z}} \mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu}) \\ &= \sum_{i \in \mathbf{V}} \inf_{\mathbf{x}_i, \mathbf{z}_i} \mathcal{L}_i(\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu}) \\ &= \sum_{i \in \mathbf{V}} q_i(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu}). \end{aligned}$$

Introducing local versions of the dual variables, λ_i, ν_i and $\mu_i \forall i \in V$, and imposing that at consensus each are equal to λ, ν and μ , respectively, we can form a final distributed dual problem given by

$$\begin{aligned} \min_{\lambda, \nu, \mu} & - \sum_{i \in V} q_i(\lambda_i, \nu_i, \mu_i) \\ \text{s.t.} & [\lambda_i^T, \nu_i^T, \mu_i^T] - [\lambda_j^T, \nu_j^T, \mu_j^T] = \mathbf{0} \quad \forall (i, j) \in \mathbf{E} \\ & (\lambda_i, \nu_i, \mu_i) \in \mathcal{D}_i \quad \forall i \in V \end{aligned} \quad (2)$$

Here \mathcal{D}_i represent the dual feasible domains relative to each node. Importantly, Eq. (2) is distributable, noted in the node-separable objective and edge-based constraints assuming that each q_i is closed proper concave. Utilising existing solvers, such as the alternating direction method of multipliers (ADMM) [14] or the primal dual method of multipliers (PDMM) [15], and assuming that the negative of the dual functions are CCP, the equivalent dual problem can be solved in general network topologies.

To form distributed signal processing algorithms we therefore only need to show that our desired operation can be rephrased as a NSGC problem and to then exploit duality to construct an equivalent distributable form.

III. SEPARABLE PROBLEMS WITH GLOBAL CONSTRAINTS

In this section, given the distributed duality of NSGC problems, we demonstrate how we can develop distributed algorithms for specific problems instances. In particular, in subsection III-A we apply this approach to a linearly constrained quadratic problems whilst in subsection III-B we demonstrate the distributability of the classic Gaussian channel capacity maximisation problem.

III-A. Example: Constrained Quadratic Programming

Consider a linearly constrained problem given by

$$\begin{aligned} \min_{\mathbf{x}} & \frac{1}{2} \mathbf{x}^T (\mathbf{Q} + \Phi) \mathbf{x} + \mathbf{q}^T \mathbf{x} \\ \text{s.t.} & \mathbf{A} \mathbf{x} = \mathbf{b} \end{aligned} \quad (3)$$

where $\mathbf{Q} \in S_+^{N \times N}$, $\Phi \in S_{++}^{N \times N}$ is a diagonal positive definite matrix and $\mathbf{q} \in \mathbb{R}^N$. Additionally, $\mathbf{A} \in \mathbb{R}^{N \times P}$ whilst $\mathbf{b} \in \mathbb{R}^P$. In general, the entries of \mathbf{A} , \mathbf{b} and \mathbf{q} may be dependent on node-based measurements of the network. The eigenvalue decomposition of \mathbf{Q} is given by

$$\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ denotes the unitary matrix of eigenvectors of \mathbf{Q} and $\mathbf{\Lambda} \in \mathbb{R}^{N \times N}$ denotes the diagonal matrix of eigenvalues. Furthermore $M < N$ denotes the number of non-zero entries of $\mathbf{\Lambda}$. Traditionally, to solve such problems in a centralised context, \mathbf{U} and $\mathbf{\Lambda}$ would require the storage of $M(N+1)$ data entries, a memory requirement that scales with the network size. In contrast, if each node were to store the i th elements of the columns of \mathbf{U} , $u_{i,k} \forall k = 1, \dots, M$,

and the non-zero diagonal entries of $\mathbf{\Lambda}$, then only $2M$ data values would need to be stored at each node. This motivates us to rewrite Eq. (3) as

$$\begin{aligned} \min_{\mathbf{x}} & \frac{1}{2} \sum_{k=1}^M \left(\lambda_k \left(\sum_{i \in V} u_{i,k} x_i \right)^2 \right) + \frac{1}{2} \mathbf{x}^T \Phi \mathbf{x} + \mathbf{q}^T \mathbf{x} \\ \text{s.t.} & \mathbf{A} \mathbf{x} = \mathbf{b} \end{aligned} \quad (4)$$

If the number of nodes in the network were to increase, assuming the rank of \mathbf{Q} remains constant, then the on-node memory requirements of the network would remain fixed. Thus, the question is, to take advantage of the distributed storage of \mathbf{Q} and node based generation of \mathbf{A} , \mathbf{b} or \mathbf{q} , is it possible to solve Eq. (4) in a fully distributed manner?

We will first show that (4) can be transformed to a NSGC problem by utilising a tight convex relaxation, as previously demonstrated in [16]. To do so, we introduce local variables $y_{i,k}$ and impose the constraint that, at consensus

$$\sum_{i \in V} y_{i,k} = N \sum_{i \in V} u_{i,k} x_i$$

Thus (4) can be rewritten as

$$\begin{aligned} \min_{\mathbf{x}} & \sum_{i \in V} \left(\sum_{k=1}^M \left(\frac{\lambda_k |y_{i,k}|^2}{2N} \right) + \frac{\phi_i |x_i|^2}{2} + q_i^T x_i \right) \\ \text{s.t.} & \sum_{i \in V} \mathbf{a}_i x_i = \mathbf{b} \\ & \sum_{i \in V} y_{i,k} = N \sum_{i \in V} u_{i,k}^T x_i \quad \forall k = 1, \dots, M \end{aligned} \quad (5)$$

where \mathbf{a}_i denotes the i th column of \mathbf{A} .

By considering the Lagrangian of (5), with dual variables ν and ω ,

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{y}, \nu, \omega) &= \sum_{i \in V} \left(\sum_{k=1}^M \frac{\lambda_k |y_{i,k}|^2}{2N} + \frac{\phi_i |x_i|^2}{2} \right. \\ &\quad \left. - \left(\mathbf{a}_i^T \omega - q_i - \sum_{k=1}^M N u_{i,k} \nu_k \right) x_i \right. \\ &\quad \left. + \frac{\omega^T \mathbf{b}}{N} - \sum_{k=1}^M \nu_k y_{i,k} \right) \end{aligned} \quad (6)$$

it can be shown that problems (5) and (3) are equivalent [16]. As the stationary points of (6) occur when

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= \phi_i x_i - \mathbf{a}_i^T \omega + q_i + \sum_{k=1}^M N u_{i,k} \nu_k = 0 \\ \frac{\partial \mathcal{L}}{\partial y_{i,k}} &= \frac{\lambda_k}{N} y_{i,k} - \nu_k = 0 \\ \therefore x_i &= \frac{1}{\phi_i} \left(\mathbf{a}_i^T \omega - q_i - \sum_{k=1}^M N u_{i,k} \nu_k \right) \\ \therefore y_{i,k} &= \frac{N}{\lambda_k} \nu_k, \end{aligned} \quad (7)$$

one can note that $y_{i,k} = y_k \forall k = 1, \dots, M, i \in V$. By inspection this means that (4) and (5) have the same minimum and optimal \mathbf{x} .

The dual function of (6) is given by

$$q(\boldsymbol{\nu}, \boldsymbol{\omega}) = \sum_{i \in \mathbf{V}} \left(\frac{\boldsymbol{\omega}^T \mathbf{b}}{N} + \sum_{k=1}^M \frac{-N}{2\lambda_k} |\nu_k|^2 - \frac{1}{2\phi_i} \left| \mathbf{a}_i^T \boldsymbol{\omega} - q_i - \sum_{k=1}^M N u_{i,k} \nu_k \right|^2 \right),$$

Introducing local estimates of $\boldsymbol{\omega}$ and $\boldsymbol{\nu}$ we define the equivalent distributed dual problem given by

$$\begin{aligned} \min_{\boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega}} \quad & \sum_{i \in \mathbf{V}} \left(-\frac{\boldsymbol{\omega}_i^T \mathbf{b}}{N} + \sum_{k=1}^M \frac{N}{2\lambda_k} |\nu_{i,k}|^2 + \frac{1}{2\phi_i} \left| \mathbf{a}_i^T \boldsymbol{\omega}_i - q_i - \sum_{k=1}^M N u_{i,k} \nu_{i,k} \right|^2 \right) \\ \text{s.t.} \quad & \boldsymbol{\omega}_i - \boldsymbol{\omega}_j = \mathbf{0} \quad \forall (i, j) \in \mathbf{E}, \\ & \nu_{i,k} - \nu_{j,k} = 0 \quad \forall k = 1, \dots, M. \end{aligned} \quad (8)$$

As desired, (8) has a fully distributable form. Figure 2 demonstrates the convergence rates of two different distributed solvers, ADMM and PDMM, in optimising (8) for a randomly generated 25 node network with approximately 25% connectivity. Once consensus is reached, the optimal \mathbf{x}

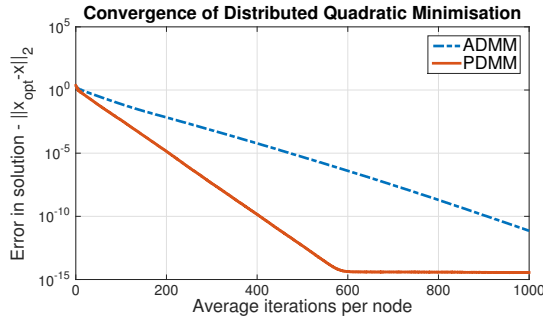


Fig. 2: Convergence of quadratic minimisation using empirically derived optimal step size for each algorithm.

can be recovered via (7) allowing both the optimal solution and optimisers of the original non-separable quadratic problem to be computed in a fully distributed manner. The linear convergence of both algorithms in this figure stems from the strong convexity of (8).

III-B. Example: Channel Capacity Maximisation

As a more practical example, consider the use of a WSN of N antennas to cooperatively transmit a signal back to a target location over a set of additive white Gaussian channels (AWGNs). Given that each channel has equal bandwidth B , the objective is to optimally configure the transmission power of the antennas (\mathbf{x}) to maximise channel capacity

whilst restricting the total output power. From the Shannon-Hartley theorem [17], the capacity for each channel (C_i) is given by

$$C_i = B \log_2 \left(1 + \frac{x_i}{\sigma_i^2} \right) = \frac{B (\ln(\sigma_i^2 + x_i) - \ln(\sigma_i^2))}{\ln(2)},$$

where σ_i^2 is the noise variance of the i th channel. In practice, each σ^2 will most likely be estimated locally at each node. Thus, where traditional water-filling algorithms would require the aggregation of these estimates to a central point for processing, we are instead interested in computing the optimal \mathbf{x} in a distributed manner. The maximisation of the total channel capacity of our system is equivalent to solving

$$\begin{aligned} \min_{\mathbf{x}} \quad & - \sum_{i \in \mathbf{V}} \ln(\sigma_i^2 + x_i) \\ \text{s.t.} \quad & \mathbf{0} \leq \mathbf{x} \\ & \mathbf{1}^T \mathbf{x} = 1, \end{aligned} \quad (9)$$

where, unlike the previous quadratic programming example we have an additional separable inequality constraint. By inspection, we can note that this optimisation problem takes the form of a *monotropic program* and thus has a distributed dual form. The Lagrangian of (9) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mu) = \sum_{i \in \mathbf{V}} (-\ln(\sigma_i^2 + x_i)) - \boldsymbol{\lambda}^T \mathbf{x} - \mu (\mathbf{1}^T \mathbf{x} - 1) \\ \text{s.t.} \quad \mathbf{0} \leq \boldsymbol{\lambda}, \end{aligned} \quad (10)$$

where $\boldsymbol{\lambda}$ and μ are the introduced dual variables.

The stationary points of (10) occur when

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= \frac{-1}{\sigma_i^2 + x_i} - \lambda_i - \mu \\ \therefore x_i &= \frac{-1}{\lambda_i + \mu} - \sigma_i^2, \end{aligned} \quad (11)$$

and thus, by substituting (11) into (10) it can be shown that the dual function is given by

$$\begin{aligned} q(\boldsymbol{\lambda}, \mu) = \sum_{i \in \mathbf{V}} \left(-\ln \left(\frac{-1}{\lambda_i + \mu} \right) + 1 + (\lambda_i + \mu) \sigma_i^2 \right) + \mu \\ \text{s.t.} \quad \mathbf{0} \leq \boldsymbol{\lambda}. \end{aligned}$$

The distributed dual problem of (9) can then be found by introducing local versions of μ at each node and constraining these μ_i along each edge such that

$$\begin{aligned} \min_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \sum_{i \in \mathbf{V}} \left(\ln \left(\frac{-1}{\lambda_i + \mu_i} \right) - (\lambda_i + \mu_i) \sigma_i^2 - \frac{\mu_i}{N} \right) \\ \text{s.t.} \quad & 0 \leq \lambda_i \quad \forall i \in \mathbf{V} \\ & \mu_i - \mu_j = 0 \quad \forall (i, j) \in \mathbf{E}, \end{aligned} \quad (12)$$

where implicitly it is imposed that $\lambda_i + \mu_i \leq 0$ such that $\ln \left(\frac{-1}{\lambda_i + \mu_i} \right)$ is real valued. It is worth noting that,

due to the inequality constraints, (12) has no analytical solution but fortunately, as it is distributable, each local optimisation problem will only require solving a constrained convex problem of two variables at each iteration. Echoing III-A, Figure 3 serves as an example to demonstrate the convergence rate of (12) when solved via both ADMM and PDMM for a randomly generated network of 25 nodes.

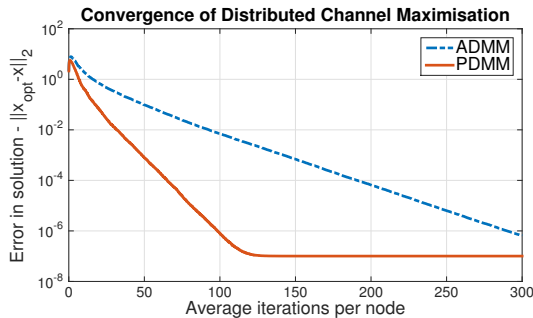


Fig. 3: Convergence of channel capacity maximisation using empirically derived optimal step size for each algorithm.

Again, at consensus, the optimal \mathbf{x} can then be recovered via (11) therefore solving (9) in a fully distributed manner.

IV. CONCLUSIONS

In this paper we have demonstrated how the class of non-distributable NSGC convex optimisation problems can be transformed to a distributed form via Lagrangian duality. For WSN problems, where data is generated locally at each node, we can use this transformation to develop distributed signal processing algorithms by first embedding desired operations within convex optimisation problems and then casting these into an NSGC form. Furthermore we have shown that by combining this approach with existing distributed solvers, such as ADMM and PDMM, we can solve specific problem classes including linearly constrained quadratic problems as well as the Gaussian channel capacity maximisation in a fully distributable manner.

V. REFERENCES

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4. IEEE, 2001, pp. 2033–2036.
- [2] 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.
- [3] A. Simonetto and G. Leus, "Distributed maximum likelihood sensor network localization," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1424–1437, 2014.
- [4] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proceedings ISIT'10*, no. EPFL-CONF-148711. IEEE, 2010.
- [5] J. Mota, J. Xavier, P. Aguiar, and M. Puschel, "Distributed admm for model predictive control and congestion control," in *IEEE Conf. Decision and Control (CDC)*. IEEE, 2012, pp. 5110–5115.
- [6] S. Ram, A. Nedić, and V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of optimization theory and applications*, vol. 147, no. 3, pp. 516–545, 2010.
- [7] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.
- [8] 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.
- [9] D. Yuan, S. Xu, and H. Zhao, "Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 6, pp. 1715–1724, 2011.
- [10] N. Parikh and S. P. Boyd, "Proximal algorithms." *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [11] D. Bertsekas, "Extended monotropic programming and duality," *Journal of optimization theory and applications*, vol. 139, no. 2, pp. 209–225, 2008.
- [12] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *Int. Conf. Distributed Computing Systems Workshops*. IEEE, 2002, pp. 575–578.
- [13] R. Rockafellar, *Network flows and monotropic optimization*. Belmont, Mass: Athena Scientific, 1998.
- [14] 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.
- [15] G. Zhang and R. Heusdens, "Bi-alternating direction method of multipliers over graphs," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, 2015, pp. 3571–3575.
- [16] T. Sherson, W. B. Kleijn, and R. Heusdens, "A distributed algorithm for robust LCMV beamforming," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, March 2016.
- [17] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.