Distributed Compressive Spectrum Sensing

Thomas Kealy[1], Oliver Johnson[2] and Robert Piechocki[3]

***Abstract*—Spectrum Sensing is a key technology for Cognitive Radio: the initial task of any cognitive device will be to accurately sense and classify spectral bands. The sampling rates required by many bands (for example TV Whitespaces) makes spectrum sensing costly with current technology. Currently many spectra are sparse and this sparsity can be exploited by using the modern paradigm of Compressive Sampling. A distributed network of sensors can further reduce the sensing load. Should these bands come into use, they will not remain sparse - this can be alleviated by requiring sparsity in the gradient of the spectrum.**

In this paper, we develop a model of the spectral gradient which allows reconstruction from Compressive measurements. Further, we develop a distributed ADMM algorithm for the LASSO with exact, closed form, expressions for the minima per iteration, reducing the computational cost of the algorithm. This allows us to perform spectral reconstruction which is blind to statistics such as the signal sparsity, using only simple statistical operations such as least squares and shrinkage. Numerical experiments show that the algorithm performs to within 10^{-2} of an equivalent centralised algorithm. The algorithm displays the rapid convergence of ADMM, but with a substantially reduced computational burden.

I. Introduction

Spectrum Sensing is a key technology for Cognitive Radio. The initial task of any cognitive device, before any kind of dynamic spectrum management will be to accurately sense and classify spectral bands for availability. Dynamic management holds the promise of satisfying the almost ubiquitous growing demand for mobile and wireless data, with consumers demanding faster speeds and better quality connections in more places. However, there is constrained amount of frequencies over which to transmit this information; and demand for frequencies that provide sufficient bandwidth, good range and in-building penetration is high. Not all spectrum is used in all places and at all times, and judicious spectrum management, by developing approaches to use white spaces where they occur, would likely be beneficial.

Devices seeking to access white spaces need a robust mechanism for learning of the frequencies that can be used at a particular time and location. One approach is to refer to a database, which maps the location of white spaces based on knowledge of existing spectrum users. An alternative approach is for devices to detect white spaces by monitoring spectrum use.

The advantages of spectrum monitoring [1] over persisting a database of space-frequency data are the ability of networks to make use of low-cost low-power devices, only capable of making local (as opposed to national) communications, keeping the cost of the network low and opportunistic channel usage for bursty traffic, reducing channel collisions in dense networks. The main technical difficulty preventing spectrum sensing currently, are the high sampling rates required by wideband spectra such as TV white spaces (TVWS). However such spectra are typically sparse, in that transmissions use far fewer frequencies that the available bandwidth.

Compressive Sensing (CS) [4] has recently emerged as a new sampling paradigm, for acquiring sparse signals. Applying this to wireless communication, we are able to reconstruct sparse signals at sampling rates below what would be required by Nyquist theory, for example the works [7], and [14] detail how this sampling can be achieved.

However, even with CS, spectrum sensing from a single machine will be costly as the proposed TVWS band will be over a large frequency range CS at a single sensor would still require high sampling rates [15]. In this paper we propose a distributed model, which allows a sensing budget at each node far below what is required by centralised CS. The advantages of such a network are that is should be able to average out noise process across some geographic area by making distributed observations, and make use of cheaper sensors due to the lowered sensing budget required per node.

We cannot always guarantee that the frequency spectrum will always be sparse: for example, should TVWS become widely utilised, the spectra will not be sparse. However, even for highly occupied spectra, the gradient of the spectrum will be sparse. This has previously been exploited by [13].

Reconstructing the spectrum from compressive measurements could take place at a fusion centre, but such communications are expensive. It is more efficient therefore to design distributed algorithms where CRs communicate with their neighbours to reach consensus on the reconstruction, given each nodes' private data. However, regularising the reconstruction process would require global co-ordination if Total Variation (the l_1 norm of the gradient of the signal) regularisation was chosen, as.

In this paper we propose a different model for sensing the gradient of the frequency spectrum to [13] - a model which doesn't require Total Variation regularisation of the objective function.

We also propose a decentralised algorithm to solve the LASSO by consensus optimisation. This allows us to design

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/I028153/1]; Ofcom; and the University of Bristol. The authors would particularly like to thank Gary Clemo of Ofcom for useful discussions.

[1]Thomas Kealy is with CDT in Communications, MVB, School of Engineering University of Bristol, UK tk12098@bristol.ac.uk

[2]Oliver Johnson is with the Department of Mathematics, University Walk, Bristol, University of Bristol, UK. O.Johnson@bristol.ac.uk

[3]Robert Piechocki is with the CSN Group, MVB, School of Engineering, University of Bristol, UK. r.j.piechocki@bristol.ac.uk

an algorithm which requires no global co-ordination whilst reconstructing the gradient of the spectrum. We choose a convex approach, as convex algorithms require no knowledge of the signal statistics (such as sparsity), and are guaranteed to converge. We are able to find exact, closed form, expressions for the Distributed Lasso, reducing the computational load per iteration whilst obviating the need to approximate the objective function [6].

The structure of the paper is as follows: in section III we introduce the sensing model, in section IV we describe the distributed reconstruction algorithm [9], and finally in section V we show some results of the reconstruction quality of this model.

II. Signal Model

Not all signals are sparse in an orthogonal basis: for example, many images are sparse in an over-complete dictionary (set of bases). In particular, frequency spectra for TVWS may no longer be sparse once opportunistic radios begin operating in these frequency bands.

Instead we aim to reconstruct the gradient of the spectrum, as we assume that transitions are constant within a band. Consider the basis defined by the function:

$$l_i(x) = \begin{cases} 1 & \text{if } x \leq i \\ 0 & \text{otherwise} \end{cases} \tag{II-.1}$$

That is, l_i is a left-hand step function.

The basis (II-.1) can be expressed as a matrix in $\mathbb{R}^{n\times n}$ as:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\dots 0 \\ 1 & 1 & 0 & 0 & 0\dots 0 \\ 1 & 1 & 1 & 0 & 0\dots 0 \\ \dots & & & & \\ 1 & 1 & 1 & 1 & 1\dots 1 \end{pmatrix} \tag{II-.2}$$

By direct computation, this inverse of L is:

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\dots 0 \\ -1 & 1 & 0 & 0 & 0\dots 0 \\ 0 & -1 & 0 & 0 & 0\dots 0 \\ \dots & & & & \\ 0 & 0 & 0 & 0\dots -1 & 1 \end{pmatrix} \tag{II-.3}$$

We model our PSD signal g as a linear combination of the basis functions (II-.1):

$$g(x) = \sum_i a_i l_i(x) = L^T a \tag{II-.4}$$

where $a = (a_1, \dots, a_n$ are the coefficients in this basis expansion, and l_i are the rows of L. Note that as defined, g is a column vector.

Proposition II.1.

$$D^T g = a \tag{II-.5}$$

Proof.

$$D^T g = D^T L^T a \tag{II-.6}$$
$$= (LD)^T a \tag{II-.7}$$
$$= a \tag{II-.8}$$

as $LD = I$.

□

III. Sensing Model

We consider a radio environment with a single primary user (PU) and a network of J nodes collaboratively trying to sense and reconstruct the PU signal in a fully distributed manner by local communication and regularisation only.

We try to sense and reconstruct a wideband signal, using a network of J (= 50) nodes placed uniformly at random within the square $[0, 1] \times [0, 1]$.

We consider the frequency domain measurements, formed by each node mixing the signal with a random Gaussian signal $A_j \in \mathbb{R}^n$. The measurements taken at node j are:

$$y_j = A_j H_j g + w_j \tag{III-.9}$$

where $H_j \in \mathbb{R}$ is the scalar channel gain, and $w_j \sim \mathcal{N}(0, \sigma_n^2) \in \mathbb{R}$ is additive white Gaussian noise.

For the purposes of comparison in section (V), this corresponds to the concatenated system:

$$y = AHg + w \tag{III-.10}$$

where $H \in \mathbb{R}^{n\times n}$ is a block diagonal matrix of channel gains.

The system III-.10 can then be solved (in the sense of finding the sparse vector a (II-.1) by convex optimisation via minimising the objective function:

$$\hat{a} = \arg\min_a \frac{1}{2}\|AHL^T a - y\|_2^2 + \lambda\|a\|_1 \tag{III-.11}$$

where λ is a parameter chosen to promote sparsity. Larger λ means sparser a.

IV. Constrained Optimisation on Graphs

We model the network of sensors as an undirected graph $G = (V, E)$, where $V = \{1 \dots J\}$ is the set of vertices, and $E = V \times V$ is the set of edges. An edge between nodes i and j implies that the two sensors can communicate. The set of nodes that node i can communicate with is written $\mathcal{N}_i$ and the degree of node i is $D_i = |\mathcal{N}_i|$.

We assume that a proper colouring of the graph is available: that is, each node is assigned a number from a set $C = \{1 \dots c\}$, and no node shares a colour with any neighbour. This is so that nodes may communicate in colour order, as opposed to communicating individually thus reducing the total number of communication rounds required.

Individually nodes make the following measurements (as discussed in section III):

$$\mathbf{y}_j = \mathbf{M}_j \mathbf{x} + \mathbf{n}_j \tag{IV-.12}$$

where $\mathbf{M}_j = \left(\mathbf{AHL^T}\right)_j$ is the p^{th} row of the sensing matrix from (III-.10).

To find the $\mathbf{x}$ we are seeking (the solution to (III-.11), to each node we give a copy of $\mathbf{x}$, $\mathbf{x}_j \in \mathbb{R}^n$, and we constrain the copies to be identical across all edges in the network. To separate the minimisation of the ℓ_2 and ℓ_1 norms, we also introduce a dummy variable $\mathbf{z}_j \in \mathbb{R}^n$ to each node. Each node, thus has a separate optimisation to solve, subject to the constraint that it is consistent with its neighbours.

We write the global optimisation variable as $\bar{x}$, which collects together C copies of a $n \times 1$ vector $\mathbf{x}$:

Definition 1. *We define vectors x_c which represent the subset of nodes with colour c, where $c = 1, \ldots, C$, and write the vector of length nJ:*

$$\bar{x} = \sum_{c=1}^{C} w_c \otimes x_c = \left[x_{c(1)}^T, \ldots, x_{c(J)}^T\right]^T \tag{IV-.13}$$

where $w_{c(i)} = \mathbb{I}(c(i) = c)$, $\mathbb{I}$ is the indicator function, and we have written $c(i)$ for the colour of the ith node.

The problem then is to solve:

$$\begin{aligned} \arg\min_{\bar{x}} \sum_{c=1}^{C} \sum_{j \in c} \|M_j x_j - y_j\|_2^2 + \frac{\lambda}{J}\|z\|_1 \\ \text{and } x_i = x_j \text{ if } \{i,j\} \in E \\ \text{and } x_i - z_i = \ \forall i \in \{1, \ldots, C\} \end{aligned} \tag{IV-.14}$$

That is, at each node we minimise a Lasso functional constrained to be consistent across edges, but that is separable in the ℓ_2 and ℓ_1 norms.

The first set of constraints (edge-agreement) can be written more compactly by introducing the node-arc incidence matrix B: a V by E matrix where each column is associated with an edge $(i,j) \in E$ and has 1 and -1 in the ith and jth entry respectively. We require that $Bx_j = 0$ for all nodes $j = 1, \ldots, J$. The global constraint is simply $(B \otimes I_n)\bar{x} = 0$, and using definition (IV-.13) the constraint $x_i = x_j$ if $\{i,j\} \in E$ can now be written:

$$\sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c = 0 \tag{IV-.15}$$

note that $\left(B^T \otimes I_n\right) \in \mathbb{R}^{nE \times nJ}$.

Together (IV-.13) and (IV-.15), suggests that the problem (IV-.14) can be re-written as:

$$\begin{aligned} \arg\min_{\bar{x}} \sum_{c=1}^{C} \sum_{j \in C_c} \|M_j x_j - y_j\|_2^2 + \beta\|z_j\|_1 \\ \text{s.t. } \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c = 0 \\ \text{and } \bar{x}_c - \bar{z}_c = 0 \end{aligned} \tag{IV-.16}$$

where $\beta = \frac{\lambda}{J}$.

The global Augmented Lagrangian [3] for the problem (IV-.16) can be written down as:

$$\begin{aligned} L_\rho = \sum_{c=1}^{C} \Bigg(\sum_{j \in c} \|M_j x_j - y_j\|_2^2 + \beta\|z_j\|_1 + \\ + \theta^T \left(\bar{x}_j - \bar{z}_j\right) + \frac{\rho}{2} \left|\left|\bar{x}_j - \bar{z}_j\right|\right|_2^2 \Bigg) + \\ + \eta^T \left(B_c^T \otimes I_n\right) \bar{x}_c + \frac{\rho}{2} \left|\left| \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \right|\right|_2^2 \end{aligned} \tag{IV-.17}$$

This is, superficially, similar to the Augmented Lagrangian for the Lasso problem [3][Section 6.4]. That is, the terms indexed by j are a straightforward Lasso problem, constrained by edge-wise variables (indexed by c) forcing consistency across the network. However, the problem (as currently written) is not separable across the edges of the network as the final and penultimate term represent the constraint that the nodes agree on their estimates across edges.

To make it possible that IV-.17 can be posed as a constrained optimisation problem at each node, we introduce the following variable:

Definition 2 (Edge-equality vector)**.**

$$\begin{aligned} u &:= \left(B^T \otimes I_n\right) \bar{x} \\ &= \left(B^T \otimes I_n\right) \sum_{c=1}^{C} w_c \otimes x_c \\ &= \sum_{c=1}^{C} B_c^T \otimes x_c \end{aligned}$$

where we have used the definition (IV-.13) *in the second line, the property of Kronecker products $(A \otimes C)(B \otimes D) = (AB \otimes CD)$ between the second and third lines, and we write $B_c = w_c^T B$.*

The terms $\| \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \|^2$ and $\eta^T \left(B_c^T \otimes I_n\right) \bar{x}_c$ of (IV-.17), can be decomposed across edges, using the following lemma:

Lemma IV.1 (Edge Decomposition)**.**

$$\left|\left| \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \right|\right|^2 = \sum_{j \in C_c} \left(D_j \left|\left|x_j\right|\right|_2^2 - \sum_{k \in \mathcal{N}_j} x_j^T x^k \right) \tag{IV-.18}$$

and

$$\eta^T \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_1 = \sum_{l \in C_c} \sum_{m \in \mathcal{N}_l} \operatorname{sign}(m - l)\, \eta_{ml}^T x_l \tag{IV-.19}$$

where η is decomposed edge-wise: $\eta = (\ldots, \eta_{ij}, \ldots)$, such that $\eta_{i,j} = \eta_{j,i}$, and is associated with the constraint $x_i = x_j$.

Proof. For the first part, note that

$$\begin{aligned} u^T u &= \left|\left| \sum_{c=1}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \right|\right|^2 \\ &= \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} \left(B_{c_1} \otimes x_{c_1}^T\right) \left(B_{c_2}^T \otimes x_c\right) \\ &= \sum_{c_1,c_2} B_{c_1} B_{c_2}^T \otimes x_{c_1}^T x_{c_2} \end{aligned}$$

BB^T is a $J \times J$ matrix, with the degree of the nodes on the main diagonal and -1 in position (i, j) if nodes i and j are neighbours (i.e BB^T is the graph Laplacian). Hence, since we can write $B_{c_1} B_{c_2}^T = w_{c_1}^T B B^T w_{c_2}$, the trace of $B_{c_1} B_{c_1}^T$ is simply the sum of the degrees of nodes with colour 1. Similar reasoning applies to all other colours.

For $c_1 \neq c_2$, $B_{c_1} B_{c_2}^T$ corresponds to an off diagonal block of the graph Laplacian, and so counts how many neighbours each node with colour 1 has.

For the second part note that $\eta \in \mathbb{R}^{nE}$ and can be written:

$$\eta = \sum_{c=1}^{C} w_c \otimes \eta_c \tag{IV-.20}$$

where η_c is the vector of Lagrange multipliers associated across edges from colour c. Now

$$\eta^T u = \sum_{c_1=1}^{C} \sum_{c_2=1}^{C} w_{c_1}^T B w_{c_2} \otimes \eta_{c_1}^T x_{c_2}$$

where we have repeated the reasoning from the previous part: using the properties of Kronecker products, and the definition of B_c. For $c_1 = c_2$, $\eta^T u$ is zero, as there are no edges between nodes of the same colour by definition. For $c_1 \neq c_2$, $\eta^T u$ counts the edges from c_1 to c_2, with the consideration that the edges from c_2 to c_1 are counted with opposite parity. I.e. for a node l with colour C, $w_{c_1}^T B_{c_2}$ counts the edges to the neighbours of node l, and the edges from the neighbours of node l to node l with opposite parity - $\sum_{l \in C_c} \sum_{m \in N_l} \operatorname{sign}(m - l)$. □

Adding together this with the lemma, lets us write (IV-.17) as:

$$\begin{aligned} L_\rho = \sum_{c=1}^{C} \Bigg(& \sum_{j \in C_c} \|M_j x_j - y_j\|_2^2 + \beta \|z_j\|_1 + \nu^T x_j \\ & + \theta \left(x_j - z_j\right) + \frac{\rho}{2} D_i \left|\left|x_j\right|\right|^2 + \frac{\rho}{2} \left|\left|x_j - z_j\right|\right|^2 \Bigg) \end{aligned} \tag{IV-.21}$$

where we have defined:

$$\nu_i = \left(\sum_{k \in \mathcal{N}_i} sign\left(k - i\right) \eta_{\{i,k\}} - \rho x_k \right) \tag{IV-.22}$$

which is a rescaled version of the Lagrange multiplier, η, which respects the graph structure.

procedure DADMM(y_j, M_j, ε)
  $x^0 = 0$, $z^0 = 0$, $\theta^0 = 0$, $\eta^0 = 0$,
  $Q = \left(M_j^T M_j + (\rho D_J + 1) I\right)^{-1}$, $w_j = M_j^T y_j$
  while $\left|\left|z^{k+1} - z^k\right|\right| \leq \varepsilon$ **do**
    for $c = 1, \ldots, C$ **do**
      $x^{k+1} \leftarrow Q\left(w_j + z^k - \theta^{kT} - \nu^{kT}\right)$
      $z^{k+1} \leftarrow S_{\beta/\rho}\left(x_j^{k+1}\right)$
      $\theta^{k+1} \leftarrow \theta_j^k + \rho\left(x^{k+1} - z^{k+1}\right)$
    end for
  Each node transmits x^{k+1} in $\mathcal{N}_j$ and calculates
    $\nu_j^{k+1} \leftarrow \nu_j^k + \rho\left(\sum_{m \in N_j} z_m^k - z_j^k\right)$
  end while
  return z^{k+1}
end procedure

Figure IV.1. The algorithm at Node j

Then by differentiating (IV-.21) with respect to x_j and z_j we can find closed forms for the updates as:

$$x_j^{k+1} := \left(M_j^T M_j + (\rho D_J + 1) I\right)^{-1} \left(M_j^T y_j + z^k - \theta^{kT} - \nu^{kT}\right) \tag{IV-.23}$$

$$z_j^{k+1} := S_{\beta/\rho}\left(x_j^{k+1}\right) \tag{IV-.24}$$

$$\theta_j^{k+1} := \theta_j^k + \rho\left(x^{k+1} - z^{k+1}\right) \tag{IV-.25}$$

$$\nu_j^{k+1} := \nu_j^k + \rho\left(\sum_{m \in \mathcal{N}_j} z_m^k - z_j^k\right) \tag{IV-.26}$$

Where we have defined

Definition 3 (Soft-thresholding)**.**

$$S_\tau\left(y\right) := \operatorname{sign}(y) \max(y - |\tau|, 0) \tag{IV-.27}$$

Remark IV.2. *This algorithm can be thought of as a distributed EM algorithm with memory: each node places a Gaussian prior with variance proportional to its degree on its private data, and solves a posterior least-squares problem. Each node then soft thresholds and then exchanges the result of this computation with its one-hop neighbours.*

This explains the inclusion of an extra Lagrange multiplier: the multiplier θ controls how far each node moves from its previous estimate in each iteration, whilst the multiplier η enforces consistency between nodes by integrating past disagreements between neighbouring nodes. Note that there is no communication of data between the nodes - only the result the computation in each round.

V. Results

The model described in section (III), equation (III-.10) was simulated. The signal $g \in \mathbb{R}^{300}$ was composed of 3 rectangular pulses, mimicking primary user signals in TVWS, as shown in figure (V.2) (a). The signal was put through a Rayleigh channel, before being sensed by the nodes. The network was generated as a random geometric graph in $[0, 1] \times [0, 1]$, with

50 nodes. If the network wasn't connected, it was redrawn. 200 mixing patters were drawn i.i.d from a $\mathcal{N}\left(0, \sigma^2 I_{300}\right)$ distribution, with $\sigma^2 = 1/200$, to from the matrix $A \in \mathbb{R}^{200\times 300}$.

Monte Carlo simulations were performed at 18 σ_n^2 values ranging from 1 to 10 and the expected Mean Squared Error (MSE) of solutions of a centralised ADMM solver and a our distributed solver were calculated over 500 repetitions with 1200 iterations (k) per repetition.

The MSE was calculated as follows:

$$\frac{\left|\left|L^t z^k - g^*\right|\right|}{||g^*||} \tag{V-.28}$$

where z^k is the result of the algorithm at iteration k, and g^* is the optimal solution.

The SNR for each repetition was calculated as

$$\frac{||g^*||}{||w||} \tag{V-.29}$$

and averaged over the 500 repetitions. The results are shown in figure (V.3). Following [**?**], for each repetition we chose

$$\lambda = \sqrt{2\sigma_n^2 \log n} \tag{V-.30}$$

The error bars indicate the empirical variance across the 500 repetitions.

These results indicate that for both the centralised and distributed solvers, their performance degrades as the noise power increases in a roughly log-linear fashion. The performance of the distributed algorithm is consistently worse than the centralised version, this contrasts with results from [2]; this is due to the differing sparsity models: [2] use a joint space and frequency model for the sparsity, and as such observe an spatial averaging out of noise when using a distributed solver. The performance of DADMM is within the error bars of the centralised version at low SNR, and gap in performance between the two versions is no more than 10^{-2}. Even at relatively lower SNRs both solvers reach a solution within 10^{-1} of the optimal (as measured by normalised MSE), which will be adequate for the task of spectrum sensing. For example the reconstructions in figures (V.2) (c) and (d) show realisations of the reconstruction from DADMM with $\sigma_n^2 = 5$ and $\sigma_n^2 = 20$ respectively. It is still possible to distinguish the occupied bands from unoccupied frequencies for both reconstructi ons.

The distributed algorithm has consistently larger variance, than the centralised solver at all SNRs. This is due to individual nodes only having access to a subset of the data to perform calculations on: the variance will be proportional to the square-root of number of data samples at each node, which are fewer than the total number of samples available to the centralised solver.

In figure (V.4), we plot the progress of DADMM along the solution path for a variety of regularisation parameters λ. The y-axis is the relative (unormalised) MSE between the optimal solution and the current iteration, and the x-axis is the iteration number. We note that for a fixed λ there is a single unique optimal solution, which DADMM converges to (in the sense of stationary error between consecutive iterations). This solution may not be attained in the allotted number of iterations, as the rate of convergence is determined by λ, ρ and the eigenvalues of the Laplacian of G. The paper [11], proves linear convergence for DADMM, with explicit expressions for the rate. In particular the rate convergence of DADMM is affected by the choice of λ: smaller λ corresponds to slower convergence - this is intuitive as solutions with fewer non-zero components should require fewer iterations to fully specify. Notice that for some λs the solution path exhibits phenomenological behaviour similar to damped oscillations: this phenomena has been explored in [10] and [12].

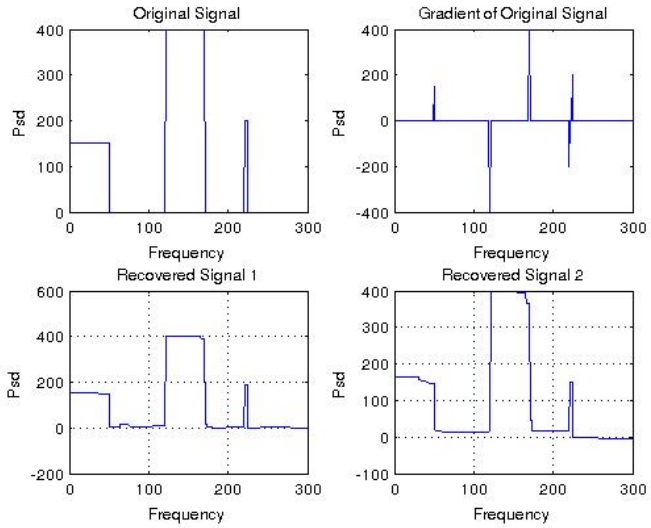

Figure V.2. Left to right: (a) The original signal. (b) The gradient (II.1) of the original signal. (c) Recovery using DADMM, 1000 iterations, $\sigma_n^2 = 5$. (d) Recovery using DADMM, 1000 iterations, $\sigma_n^2 = 20$

VI. Conclusions

We have demonstrated an alternating direction algorithm for distributed optimisation with exact (as opposed to linear or quadratic approximations to the objective as in [8] and [6]) closed form expressions for the computation at each iteration, and discussed the statistical properties of the estimation.

We have simulated the performance of this distributed algorithm for the distributed estimation of frequency spectra, in the presence of additive (white, Gaussian) and multiplicative noise. We have shown that the algorithm is robust to a variety of SNRs and converges to a similar solution as an equivalent centralised algorithm (in relative mean-squared-error).

We plan to work on larger, more detailed, models for the frequency spectra, to extend our regression framework to solve the MMV problem, to accelerate the convergence via Nesterov type methods to smooth the convergence of the distributed algorithm [5], and to incorporate spatial variation into our model to further promote sparsity. We also plan to automate the choice of λ via continuation methods, and study how the choice of λ and ρ affect the rate of convergence.

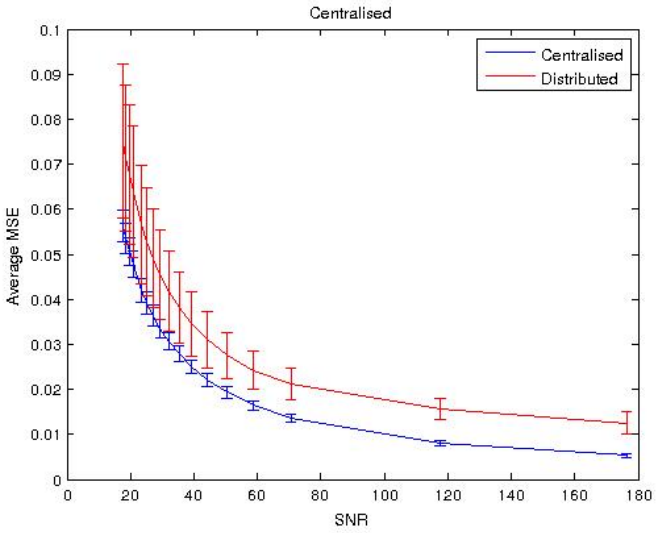

Figure V.3. MSE vs SNR for the sensing model showing the performance of distributed and centralised solvers. The performance of DADMM is consistently within 10^{-2} of ADMM, and within the error bars of ADMM at low SNRs. The variance of estimates produced by DADMM is larger than ADMM, due to nodes performing computations on a subset of data. Both estimates are consistently within 10^{-1} of the optimal solution, which is sufficient to classify occupied bands.

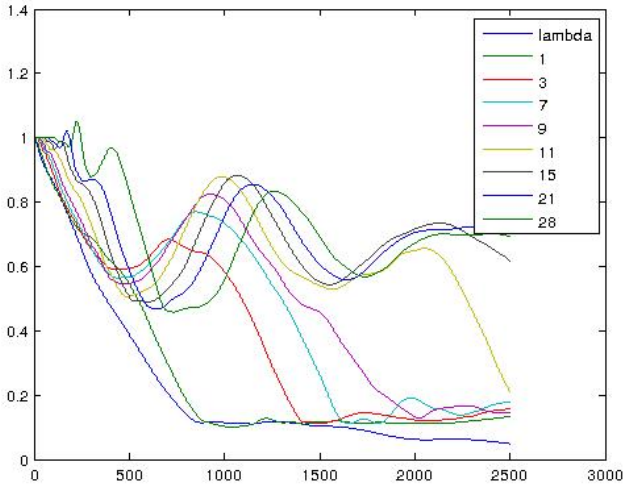

Figure V.4. The progress of the distributed solver as a function of the number of iterations, with different values of the regression parameter λ. For a fixed λ there is a single unique optimal solution, with higher λ favouring sparser solutions. The convergence of DADMM is slowed by smaller λ. This is intuitive: solutions with fewer non-zero components should be identified in fewer iterations.

REFERENCES

[1] O. B. Akan, O. Karli, and O. Ergul, "Cognitive Radio Sensor Networks," *Network, IEEE*, vol. 23, no. 4, pp. 34–40, 2009.

[2] G. B. Bazerque, J.A Giannakis, "Distributed spectrum sensing for cognitive radios by exploiting sparsity," *Proc. of 42nd Asilomar Conf. on Signals, Systems, and Computers*, 2008.

[3] S. Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, no. 1, pp. 1–122.

[4] E. J. Candes, J. Romberg, and T. Tao, "Robust Uncertainty Principles : Exact Signal Frequency Information," vol. 52, no. 2, pp. 489–509, 2006.

[5] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.

[6] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," 2014.

[7] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 375–391, 2010.

[8] A. Mokhtari, W. Shi, Q. Ling, and A. Ribeiro, "DQM: Decentralized quadratically approximated alternating direction method of multipliers," *arXiv preprint arXiv:1508.02073*, 2015.

[9] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Puschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2718–2723.

[10] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan, "A General Analysis of the Convergence of ADMM," *arXiv preprint arXiv:1502.02009*, 2015.

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

[12] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling nesterov's accelerated gradient method: Theory and insights," in *Advances in Neural Information Processing Systems*, 2014, pp. 2510–2518.

[13] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *Cognitive radio oriented wireless networks and communications, 2006. 1st international conference on*. IEEE, 2006, pp. 1–5.

[14] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520–544, 2010.

[15] H. Zhang, Z. Zhang, and Y. Chau, "Distributed compressed wideband sensing in Cognitive Radio Sensor Networks," in *2011 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2011*, 2011, pp. 13–17.