Distributed Wideband Spectrum Sensing

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

Abstract—We consider the problem of reconstructing wideband frequency spectra from distributed, compressive measurements. The measurements are made by a network of nodes, each independently mixing the ambient spectra with low frequency, random signals. The reconstruction takes place via local transmissions between nodes, each solving simple statistical

I. Introduction

There is an almost ubiquitous need to send more data over wireless networks, with 4G now being rolled out in the UK and US and with 5G being planned for 2020 and beyond [4]. There is constrained frequency set over which to transmit this information. This issue could be alleviated by users dynamically allocating spectrum on a per-transmission basis. However, we (in the UK) currently lack any mechanism by which we are able to gain knowledge of spectral conditions. Frequencies have historically been allocated via auction, organised by OFCOM, so devices are unable to make dynamic decisions about how to transmit this data. This leads to both spectral exhaustion on some frequencies whilst much of the available radio spectrum lies fallow.

Dynamic spectrum access won't become a reality unless spectral holes can be robustly detected. OFCOM requires that secondary users exploit the new spectrum politely, without interference to primary user makes spectrum sensing essential to TV white-space (TVWS) technologies. The realisation of any Cognitive Radio standard (such as IEEE 802.22 [7]), requires the co-existence of primary (TV users) and secondary (everybody else who wants to use TVWS spectrum) users of the frequency spectrum to ensure proper interference mitigation and appropriate network behaviour.

Users of TVWS (Cognitive Radios) must sense whether spectrum is available, and must be able to detect very weak primary user signals. Furthermore they must sense over a wide bandwidth (due to the amount of TVWS spectrum proposed), which challenges traditional Nyquist sampling techniques, because the sampling rates required are not technically feasible with current RF or Analogue-to-Digital conversion technology.

Due to the inherent sparsity of spectral utilisation, Compressive Sensing (CS) [3] is an appropriate formalism within which to tackle this problem. CS has recently emerged as a

new sampling paradigm allowing images to be taken from a single pixel camera for example. Applying this to wireless communication, we are able to reconstruct sparse signals at sampling rates below what would be required by Nyquist theory.

However, even with CS, spectrum sensing from a single machine will be costly as the proposed TVWS band will be over a GHz frequency range. CS at a single sensor would still require GHz sampling rates. 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 main contribution of this paper is that the model can be solved in a fully distributed manner - we do not require a central fusion centre as in [10].

The structure of the paper is as follows: in section (II) we introduce the sensing model, in section (III) we describe the distributed reconstruction algorithm [6], and finally in section (IV) we show some results of the reconstruction quality of this model. [5] [9]

II. Models

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, either in a fully distributed manner (by local communication), or by transmitting measurements to a fusion centre which then solves the linear system.

We try to sense and reconstruct a wideband signal, divided into L channels. We have a (connected) network of J (= 50) nodes placed uniformly at random within the square $[0, 1] \times [0, 1]$. This is the same model as in [10].

The nodes individually take measurements (as in [5]) by mixing the incoming analogue signal $x(t)$ with a mixing function $p_i(t)$ aliasing the spectrum. $x(t)$ is assumed to be bandlimited and composed of up to k uncorrelated transmissions over the L possible narrowband channels - i.e. the signal is k-sparse.

The mixing functions - which are independent for each node - are required to be periodic, with period T_p. Since p_i is periodic it has Fourier expansion:

$$p_i(t) = \sum_{l=-\infty}^{\infty} c_{il} \exp\left(jlt\frac{2\pi}{T_p}\right) \qquad \text{(II-.1)}$$

The c_{il} are the Fourier coefficients of the expansion and are defined in the standard manner. The result of the mixing procedure in channel i is therefore the product xp_i, with Fourier transform:

$$X_i(f) = \int_{-\infty}^{\infty} x(t) p_i(t) \, dt \qquad \text{(II-.2)}$$

$$= \sum_{l=-\infty}^{\infty} c_{il} X(f - l f_p) \qquad \text{(II-.3)}$$

(We insert the Fourier series for p_i, then exchange the sum and integral). The output of this mixing process then, is a linear combination of shifted copies of $X(f)$, with at most $\lceil f_N YQ/f_p \rceil$ terms since $X(f)$ is zero outside its support (we have assumed this Nyquist frequency exists, even though we never sample at that rate).

This process is repeated in parallel at each node so that each band in x appears in baseband.

Once the mixing process has been completed the signal in each channel is low-pass filtered and sampled at a rate $f_s \geq f_p$. In the frequency domain this is a ideal rectangle function, so the output of a single channel is:

$$Y_i\left(e^{j2\pi f T_s}\right) = \sum_{l=-L_0}^{+L_0} c_{il} X(f - l f_p) \qquad \text{(II-.4)}$$

since frequencies outside of $[-f_s/2, f_s/2]$ will filtered out. L_0 is the smallest integer number of non-zero contributions in $X(f)$ over $[-f_s/2, f_s/2]$ - at most $\lceil f_N YQ/f_p \rceil$ if we choose $f_s = f_p$. These relations can be written in matrix form as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \qquad \text{(II-.5)}$$

where $\mathbf{y}$ contains the output of the measurement process, and $\mathbf{A}$ is a product matrix of the mixing functions, their Fourier coefficients, a partial Fourier Matrix, and a matrix of channel coefficients. $\mathbf{x}$ is the vector of unknown samples of $x(t)$.

i.e. $\mathbf{A}$ can be written:

$$\mathbf{A}^{m \times L} = \mathbf{S}^{m \times L} \mathbf{F}^{L \times L} \mathbf{D}^{L \times L} \mathbf{H}^{L \times L} \qquad \text{(II-.6)}$$

The measurements $\mathbf{y}$ are transmitted to a Fusion Centre via a control channel. The system II-.5 can then be solved (in the sense of finding the sparse vector $\mathbf{x}$ by convex optimisation via minimising the objective function:

$$\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1 \qquad \text{(II-.7)}$$

where λ is a parameter chosen to promote sparsity.

III. ADMM

The alternating direction method of multipliers [2], (ADMM), algorithm solves problems of the form

$$\min_x f(x) + g(z) \qquad \text{(III-.8)}$$
$$\text{s.t } Ux + Vz = c \qquad \text{(III-.9)}$$

where f and g are assumed to be convex function with range in $\mathbb{R}$, $U \in \mathbb{R}^{p \times n}$ and $V \in \mathbb{R}^{p \times m}$ are matrices (not assumed to have full rank), and $c \in \mathbb{R}^p$.

ADMM consists of iteratively minimising the augmented Lagrangian

$$L_p(x, z, \eta) = f(x) + g(z) + \eta^T (Ux + Vz - c) + \frac{\rho}{2}\|Ux + Vz - c\|_2^2$$

(η is a Lagrange multiplier) with the following iterations:

$$x^{k+1} := \arg\min_x L_\rho\left(x, z^k, y^k\right) \qquad \text{(III-.10)}$$

$$z^{k+1} := \arg\min_z L_\rho\left(x^{k+1}, z, y^k\right) \qquad \text{(III-.11)}$$

$$\eta^{k+1} := \eta^k + \rho\left(Ux^{k+1} + Vz^{k+1} - c\right) \qquad \text{(III-.12)}$$

The alternating minimisation works because of the decomposability of the objective function: the x minimisation step is independent of the z minimisation step and vice versa.

We illustrate an example, relevant to the type of problems encountered in signal processing.

A. Example: ADMM for Centralised LASSO

ADMM can be formulated as an iterative MAP estimation procedure for the problem (which is referred to as LASSO see [8]):

$$\frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \qquad \text{(III-A.13)}$$

This can be cast in constrained form as:

$$\frac{1}{2}\|Ax - b\|_2^2 + \lambda\|z\|_1 \qquad \text{(III-A.14)}$$
$$\text{s.t } z = x \qquad \text{(III-A.15)}$$

i.e this is of the form (III-.10) with $f(x) = \|Ax - y\|_2^2$, $g(z) = \lambda\|z\|_1$, $A = B = I$, and $c = 0$.

and the associated Lagrangian is:

$$L_\rho = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|z\|_1 + \eta(x - z) + \frac{\rho}{2}\|x - z\|^2 \qquad \text{(III-A.16)}$$

Now, given a set of noisy measurements (say of radio spectra) $\mathbf{y}$, and a sensing matrix $\mathbf{A}$ we can use ADMM to find the (sparse) radio spectra.

The ADMM iterations for LASSO, which can be found by alternately differentiating (III-A.16) with respect to x, z and λ, are (in closed form):

$$x^{k+1} := \left(A^T A + \rho I\right)^{-1}\left(A^T b + \rho\left(z^k - y^k\right)\right) \qquad \text{(III-A.17)}$$
$$z^{k+1} := S_{\lambda/\rho}\left(x^{k+1} + y^k/\rho\right) \qquad \text{(III-A.18)}$$
$$y^{k+1} := y^k + \rho\left(x^{k+1} - z^{k+1}\right) \qquad \text{(III-A.19)}$$

where $S_{\lambda/\rho}(\circ)$ is the soft thresholding operator: $S_\gamma(x)_i = sign(x_i)(|x_i| - \gamma)^+$.

This algorithm has a nice statistical interpretation: it iteratively performs ridge regression, followed by shrinkage

towards zero. Exactly like the MAP estimate for LASSO problem under a Laplace prior.

The soft-thresholding operator can be derived by considering the MAP estimate of the following model:

$$y = x + w \qquad \text{(III-A.20)}$$

where x is some (sparse) signal, and w is additive white Gaussian noise. We seek

$$\hat{x} = \arg\max_x \mathbb{P}_{x|y}(x|y) \qquad \text{(III-A.21)}$$

This can be recast in the following form by using Bayes rule, noting that the denominator is independent of x and taking logarithms:

$$\hat{x} = \arg\max_x \left[\log \mathbb{P}_n(y - x) + \log \mathbb{P}(x) \right] \qquad \text{(III-A.22)}$$

The term $\mathbb{P}_n(y - x)$ arises because we are considering $x + w$ with w zero mean Gaussian, with variance σ_n^2. So, the conditional distribution of y (given x) will be a Gaussian centred at x.

We will take $\mathbb{P}(x)$ to be a Laplacian distribution:

$$\mathbb{P}(x) = \frac{1}{\sqrt{2}\sigma} \exp - \frac{\sqrt{2}}{\sigma}|x| \qquad \text{(III-A.23)}$$

Note that $f(x) = \log \mathbb{P}_x(x) \ - \frac{\sqrt{2}}{\sigma}|x|$, and so by differentiating $f'(x) = -\frac{\sqrt{2}}{\sigma} sign(x)$

Taking the maximum of III-A.22 we obtain:

$$\frac{y - \hat{x}}{\sigma_n^2} - \frac{\sqrt{2}}{\sigma} sign(x) = 0 \qquad \text{(III-A.24)}$$

Which leads the soft thresholding operation defined earlier, with $\gamma = \frac{\sqrt{2}\sigma_n^2}{\sigma}$ as (via rearrangement):

$$y = \hat{x} + \frac{\sqrt{2}\sigma_n^2}{\sigma} sign(x)$$

or

$$\hat{x}(y) = sign(y)\left(y - \frac{\sqrt{2}\sigma_n^2}{\sigma}\right)_+$$

IV. CONSTRAINED OPTIMISATION ON GRAPHS

We model the network as an undirected graph $G = (V, E)$, where $V = \{1 \ldots J\}$ 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 nodes can communicate. The set of nodes 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 (or approximate) colouring of the graph is available: that is each node is assigned a number from a set $C = \{1 \ldots c\}$, and no node shares a colour with any neighbour.

Given the measurement model from the section (II), we can represent the linear system by concatenating the measurements of each node into a single system. Here each row of the matrix represents the measurements of a single node. I.e we are trying to solve the following problem:

$$\min f(x) \text{ subject to } A_p x = y_p \qquad \text{(IV-.25)}$$

where f is some global function we seek to minimise, A_p is the row of the sensing matrix A available to node p, and y_p are the measurements taken by node p.

The problem is coupled at each node by the variable x. To ease this issue, at each node create a local copy, denoted x_j (so there will be J copies of this variable). Then, for consistency, we constrain the problem so that each copy is identical: $x_1 = x_2 = \ldots x_J$. Equivalently, given the graphical structure of the problem, we require that $x_i = x_j$ for each edge of the network.

So now we are solving the problem

$$\min_x \frac{1}{J} \sum_J f_p(x_j) \qquad \text{(IV-.26)}$$

$$\text{s.t } A_p x_p = y_p \text{ and } x_i = x_j \text{ if } \{i,j\} \in E \qquad \text{(IV-.27)}$$

where f_p is a copy of the global function f stored at node p.

This is exactly the kind of problem that can be attacked by the Alternating Direction Method of Multipliers, described previously. We present a short example for bipartite graphs, which can easily be generalised to graphs with many colours.

A. Example: bipartite graphs

We are here considering graphs which are connected, and 2-colourable, that is nodes 1 to c have colour 1 (red say), and nodes $c + 1$ to J have colour 2 (blue). We will demonstrate that we can decompose problem (**??**) into c problems which can be executed in parallel by minimising x at the 1st set of nodes, and $J - c$ problems which also can be executed in parallel by minimising at the second set of nodes.

Firstly, for compactness, we introduced the arc-incidence matrix of the graph B. This is the $J \times E$ matrix where each column corresponds to an edge $\{i, j\} \in E$, with the ith and jth entries equal to 1 and -1 respectively. With this notation, we can rewrite (**??**) as:

$$\min \frac{1}{J} \sum_J f_p(x_j) \text{ subject to } A_p x_p = y_p \text{ and } \left(B^T \circ I\right)\bar{x} = 0 \qquad \text{(IV-A.28)}$$

where $\bar{x} = (x_1, x_2, \ldots x_c)$, and $\circ$ is the Kronecker product.

In this form, the updates for each node (within a colour) of the problem (III-C.28) can be written in the following way:

$$\bar{x}_p^{k+1} = \frac{1}{J} \sum_{j \in C_i} f_p(x_j) + \phi_1(\bar{x}_i, \lambda)$$

where

$$\phi_i\left(\bar{x}_i, \lambda\right) = \lambda^T \left(B_i^T \circ I_n\right) \bar{x}_i + \frac{\rho}{2} \left\| \left(B_i^T \circ I_n\right) \bar{x}_i \right\|^2$$
$$= \left(\left(B_i^T \circ I_n\right) \lambda\right)^T \bar{x}_i + \frac{\rho}{2} \bar{x}_i^T \left(B_1 B_2^T \circ I_n\right) \bar{x}_i$$

Since, no nodes within C_i are neighbours $B_i B_i^T$ is a diagonal matrix (with the degree of node i as the ith entry). So, we can finally re-write

$$\phi_i\left(\bar{x}_i, \lambda\right) = \sum_{j \in C_i} \left(\left(\sum_{k \in \mathcal{N}_i} sign\left(k - i\right) \lambda_{\{i,k\}} \right)^T x_i + \frac{\rho}{2} D_i \left\| x_i \right\|^2 \right)$$

where the dual variable λ has been decomposed edge-wise so that $\lambda_{\{i,j\}} = \lambda_{\{j,i\}}$ and is associated with edge $\{i,j\}$ and the constraint $x_i = x_j$. This function $\sum_{p \in C_c} f_p\left(x_p\right) + \phi_i\left(\bar{x}_i, \lambda\right)$ is what is minimised at each node within each colour (in parallel). For LASSO take $f_p := \|Ax_p - b_p\|_2^2 + \frac{\beta}{J}\|x_p\|_1$.

V. Results

The model described in section (II) was simulated, with a wideband signal of 201 channels and a network of 50 nodes (i.e. the signal will be sampled at a 1/4 of rate predicted by Nyquist theory). The mixing patterns were generated from iid Gaussian sources (i.e the matrix S had each entry drawn from an iid Gaussian source). Monte Carlo simulations were performed at EbN0db values ranging from 5 to -5 in steps of 1, and the expected Mean Squared Error (MSE) of solutions of a centralised solver (spgl1) and a distributed solver (ADMM). The results can be seen in fig (IV.1).

These results indicate that for both centralised and distributed solvers, adding noise to the system results in a degrading of performance. Interestingly note, that the distributed solver seems to (slightly) outperform the centralised solver at all SNRs. This is counter-intuitive, as it would be expected that centralised solvers knowing *all* the available information would outperform distributed solutions. We conjecture that the updates described in section **??**, take into account differences in noise across the network. The distributed averaging steps, which form the new prior for each node, then penalise updates from relatively more noisy observations. This corroborates observations from [1].

This observation is (partially) confirmed in figure (IV.2), which plots the progress of the centralised and distributed solvers (as a function of iterations) towards the optimum solution. The SNR is 0.5 (i.e the signal is twice as strong as the noise).
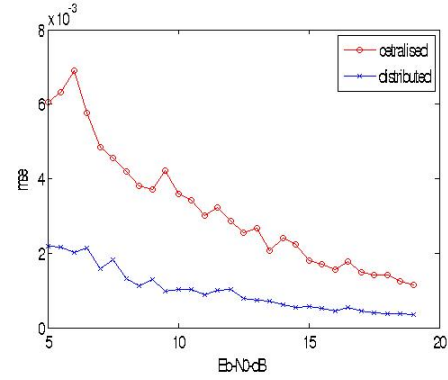

Figure V.1. Mse vs SNR for the 'frequency only sensing model, showing the performance of distributed and centralised solvers
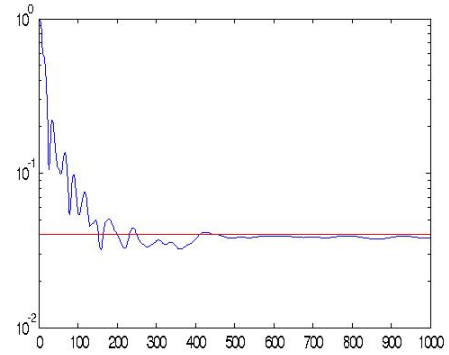

Figure V.2. The progress of a distributed and a centralised solver as a function of the number of iterations

References

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

[2] S. Boyd, "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, 2010. [Online]. Available: http://www.nowpublishers.com/product.aspx?product=MAL&doi=2200000016

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

[4] E. e. a. Dahlman, "5g wireless acces: Requirements and realization," *IEEE Communications Magazine*, no. December, pp. 42–47, 2014.

[5] 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.

[6] 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.

[7] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "Ieee 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, 2009.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[9] 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.

[10] 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.