SIGNAL RECONSTRUCTION FROM INCOMPLETE AND MIS-PLACED MEASUREMENTS

ABSTRACT

Constrained by practical and economical considerations, one often uses seismic data with missing traces. The use of such data results in image artifacts and poor spatial resolution. Sometimes due to practical limitations, measurements may be available on a perturbed grid, instead of on the designated grid. Due to algorithmic requirements, when such measurements are viewed as those on the designated grid, the recovery procedures may result in additional artifacts. This paper interpolates incomplete data onto regular grid via the Fourier domain, using a recently developed greedy algorithm. The basic objective is to study experimentally as to what could be the size of the perturbation in measurement coordinates that allows for the measurements on the perturbed grid to be considered as on the designated grid for faithful recovery. Our experimental work shows that for compressible signals, a uniformly distributed perturbation can be offset with slightly more number of measurements.

Introduction

The seismic source generates a wavefield. During data acquisition, this wavefield is sampled at discrete locations and at discrete times over a large area. Due to the trade off between geophysical and economical constraints, sampling is rarely regular and complete, which results in artifacts in seismic images when the measurements are processed with simple binning techniques (see e.g. Hennenfent and Herrmann, 2006; Herrmann and Hennenfent, 2007). These artifacts can be minimized by improving upon binning using other techniques. One of these is based on Fourier reconstruction, which is a transform based method.

The recently proposed optimization techniques via *Basis Pursuit (BP)* and *Matching Pursuit (MP)* have promising applications (see e.g. Candes et al., 2006; Hennenfent and Herrmann, 2006; Herrmann and Hennenfent, 2007) for the interpolation of missing data in otherwise regularly sampled grid. The Stagewise Orthogonal Matching Pursuit (StOMP) (see Donoho et al., 2006) assumes easily verifiable recovery conditions and provides a faster algorithm for solving underdetermined system of equations In the present paper, we apply StOMP to the data interpolation problem through a Fourier based technique for naturally occurring compressible signals. It aims primarily at allowing certain misplacement in measurement coordinates.

The discrete Fourier reconstruction of a 1-D (discrete) signal $\mathbf{f}$ can be written in a matrix equation as

$$\mathbf{f}[\mathbf{x}] = \mathbf{F}[\mathbf{x}]\hat{\mathbf{f}}. \tag{1}$$

The notation $\mathbf{f}[\mathbf{x}]$, $\mathbf{F}[\mathbf{x}]$ is used to emphasize the dependence of $\mathbf{f}$, $\mathbf{F}$ on the data grid $\mathbf{x}$. Given measurements $f_j = f(x_j)$ for $x_j \in [-\frac{1}{2}, \frac{1}{2})$ with x_j distributed according to some random distribution, our concern is the evaluation of Fourier coefficients $\hat{f}_k$ from (1). The system (1) represents an underdetermined system for an incomplete set of measurements. In the following section, we describe the signal reconstruction methods proposed in the recent literature (see e.g. Candes et al., 2006; Chen at al., 2001; Donoho et al., 2006) for solving underdetermined systems. After describing the recent recovery algorithms, we present our problem in some detail and analyze our experimental work.

Signal recovery from incomplete measurements

Suppose the vector $\mathbf{y} \in \mathbf{R^n}$ of measurements y_i satisfies

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} \quad \text{for} \quad \mathbf{\Phi} \in I\!R^{n \times N}; \text{ for } n << N, \tag{2}$$

which admits infinitely many solutions. One way of choosing a solution for (2) involves taking the solution that is 'smallest,' (in the l^2 sense), which corresponds to $\tilde{\mathbf{x}} = \mathbf{\Phi^T}(\mathbf{\Phi}\mathbf{\Phi^T})^{-1}\mathbf{y}$, called the *pseudo inverse* of $\mathbf{\Phi}$. Another way of computing a solution of (2) involves taking 'shortest' vector. That is, $\|\mathbf{x}\|_\mathbf{0} := |\#\{\mathbf{i} : \mathbf{x_i} \neq \mathbf{0}\}| < \mathbf{N}$. In this case, *Basis pursuit* (BP) (see e.g. Chen et. al., 2001), a convex programming approach, proposes the solution via l^1 minimization as

$$\min_\alpha \|\alpha\|_\mathbf{1} \quad \textbf{subject} \quad \textbf{to} \quad \mathbf{\Phi}\alpha = \mathbf{y}. \tag{3}$$

One of the important results, for exact recovery from incomplete samples, states that exact recovery is possible as long as the synthesis matrix $\mathbf{\Phi}$ obeys what is called a restricted isometry (see e.g. Candes et al., 2006; Chen et. al., 2001) and certain conditions in terms of the so-called coherence parameter (μ), which is the maximum off-diagonal entry in $\mathbf{\Phi^T}\mathbf{\Phi}$ (see e.g. Candes et al., 2006; Chen et. al., 2001). An important result

by (see e.g. Candes et al.). states that exact recovery of the k nonzero entries in x_0 is possible as long as the number of measurements is roughly proportional to the number of nonzeros, that is, $n \propto \mu^2 k$.

Stagewise Orthogonal Matching Pursuit (StOMP)

The stronger recovery results for arbitrary systems stated in the above section are very powerful. They, however, tend to be pessimistic requiring many measurements or higher sparsity for the solution vector. Additionally, it is difficult to verify the stronger recovery conditions in general. Nevertheless, there are certain weaker recovery conditions assumed by StOMP, proposed recently by (Donoho et al., 2006). The conditions involve randomness in measurements and Gaussian behaviour of the residual vector, which is defined in the next paragraph. These conditions are relatively easily verifiable. Hence StOMP is attractive for large scale problems such as those in seismic imaging. Suggesting the slogan that "*noiseless underdetermined problems behave like noisy well-determined problems*," StOMP aims at achieving an approximate solution of (2), when the columns of $\boldsymbol{\Phi}$ are Gaussian random variables with unit norm.

Starting with an initial residual $\mathbf{r_0} = \mathbf{y}$, at s^{th} stage, StOMP forms the matched filter $\boldsymbol{\Phi^T r_{s-1}}$ and identifies all coordinates with amplitude exceeding a specially chosen threshold. It then solves a least-squares problem using the selected coordinates and subtracts the least squares fit, which results in a new residual. The method stops after a fixed number of steps. In order to design the threshold parameters, StOMP assumes the Gaussian behaviour of the (*residual*) vector: $\mathbf{z} = \tilde{\mathbf{x}} - \mathbf{x_0} = (\boldsymbol{\Phi^T \Phi} - \mathbf{I})\mathbf{x_0}$. The performance of StOMP is studied in terms of a phase diagram, which provides the information on the reconstruction error in terms of the number of measurements and sparsity of reconstruction vector. Though StOMP is primarily designed for certain matrices of Gaussian random variables, (Donoho et al., 2006) have found that the method applies to partial Fourier ensemble as well. Hence the application of StOMP to the present interpolation problem is justified.

Present work

The optimization techniques described in the previous section are shown to be useful by Hennenfent and Herrmann, 2006; Herrmann and Hennenfent, 2007 for seismic interpolation. As our measurements are missing on otherwise regularly sampled grid, we model them as random variables having uniform distribution in the discrete sense. In seismology, quite often the measurements may not be acquired on the uniformly distributed grid, but instead on some perturbation. Due to the algorithmic requirements, when we use them as on uniformly distributed grid, the recovery algorithms may result in some artifacts. In this article, we empirically study the behaviour of reconstruction error by means of recovery diagram as a function of the number of measurements (n) and the size of perturbation. Mathematically our problem may be described as follows: let the perturbation $\tilde{\mathbf{r}}$ of uniformly distributed measurement coordinate set $\mathbf{r}$ be defined as

$$\tilde{\mathbf{r}} = \mathbf{r} + \lambda \mathbf{p}, \tag{4}$$

where $\mathbf{p}$ is the perturbation vector having some random distribution and λ is control parameter. Let $\mathbf{y}$ and $\mathbf{x}$ (also their 'tilde' counterparts) stand for the signal and its Fourier coefficients such that $\mathbf{y} := \mathbf{y}[\mathbf{r}] = \mathbf{F}[\mathbf{r}]\mathbf{x}$ and $\tilde{\mathbf{y}} := \mathbf{y}[\tilde{\mathbf{r}}] = \mathbf{F}[\tilde{\mathbf{r}}]\mathbf{x}$. When we treat $\tilde{\mathbf{y}}$ to be the measurements over $\mathbf{r}$ and try to recover the underlying signal from $\tilde{\mathbf{y}} = \mathbf{F}[\mathbf{r}]\mathbf{x}$, we get an inaccurate solution (x_1). We study numerically the behaviour of the solution of $\tilde{\mathbf{y}} = \mathbf{F}[\mathbf{r}]\mathbf{x_1}$ in terms of the error $\frac{\|x - x_1\|^2}{\|x\|^2}$. Besides the measurements

being noisy, naturally occurring signals typically are not sparse in strict sense, but rather decay by some power law when sorted in decreasing order of magnitude. Such signals are called compressible signals. That is, for some positive numbers r and C_r, the sorted elements of $\{x\}_{i \in I}$ satisfy

$$|x_{i \in I}| \leq C_r i^{-r}. \tag{5}$$

Simulation work

Throughout our simulation work, we have taken N, the actual size of the signal, to be 800. In generating the pixel values in the phase diagrams, we have taken averages over 25 iterations. Motivated by the Gaussianity assumption used in StOMP, we have taken the optimum tolerance in StOMP as $\sqrt{var(\tilde{\mathbf{y}} - \mathbf{y})(n + sqrt(2n))}$ for each n. To begin with, we have generated the recovery diagram in Figure 2(a) for the compressible signals with different compression rates . It provides some useful insights about the behaviour of error as a function of decay of compression rate (r) and the number of measurements (n). The bright region in Figure 2(a) corresponds to parameter combinations that favour accurate recovery. As expected, the recovery error decays with both n and r. The transition implies that a small number of measurements are needed for highly compressible signals. In an another experiment, we have implemented the algorithm taking $\mathbf{p}$ in (4) to be (continuously) uniformly distributed in $[-\frac{1}{2N}, \frac{1}{2N}]$ and $\lambda \in [0.05, 1]$. The compressible signal used in generating Figure 2(b) is shown in Figure 1. The phase transition in Figure 2(b) illustrates that with slightly more number of measurements, the effect of perturbation error can be minimized. In Figure 2(a), one may note that for the signal of unit compression rate (that is, $|x_{i \in I}| \leq 1.\frac{1}{i}$) in its frequency domain, the number of random measurements needed for its reliable recovery can be as small as 20-30 % of its length, and in the presence of perturbation as described above, not so many more measurements are needed for different λ, which is demonstrated by Figure 2(b).
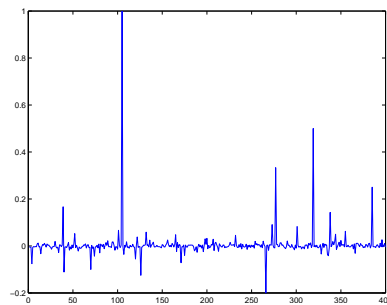

Figure 1: Randomly permuted compressible signal $x_i = \frac{1}{i}$

Conclusion

In the present article, applying StOMP to the Fourier-based interpolation problem, we have demonstrated that the recovery of compressible signals can be made from a small number of its random measurements. Further we have empirically shown that the recovery of compressible signals can be made reliably in spite of error arising in measurements due to moderate amounts of misplacement in measurement coordinates. In typical seismic imaging applications, the seismic traces are neither sparse nor compressible in strict sense, but rather have compressible nature in the transformed domain
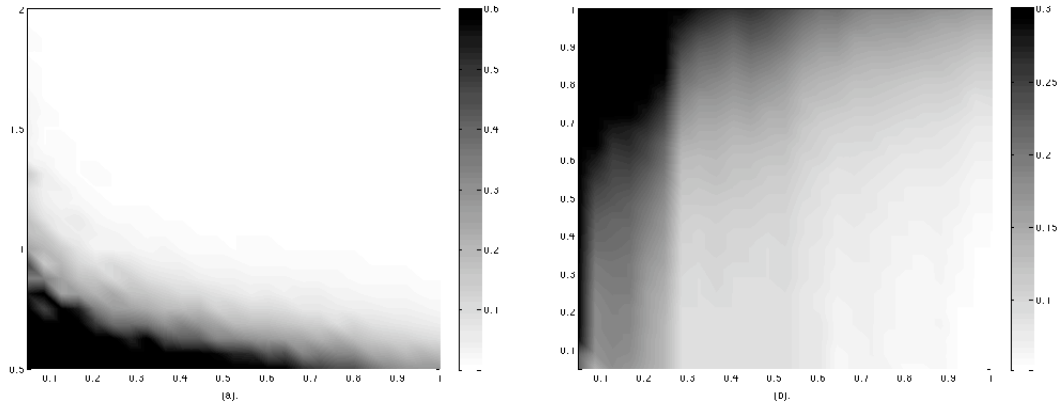
Figure 2: (a). Recovery diagram of compressible signals for different compression rates (y-axis) and $\frac{n}{N}$ (x-axis). This figure depicts that a small number of measurements are needed for the reconstruction of highly compressible signals. (b). Phase diagram of compressible signal of unit compression rate for different sizes of perturbations (y-axis) and $\frac{n}{N}$ (x-axis). This figure illustrates that with slightly more measurements, the effect of perturbation error can be mitigated.

provided by redundant frames such as curvelets . As curvelet transform is Fourier based, the observations made herein are directly relevant for seismic imaging.

References

E. J. Candes, J. Romberg and T. Tao (2006),"Stable Signal Recovery from Incomplete and Inaccurate Measurements," Comm. Pure Appl. Math., 59, 1207-1223.

S.S. Chen, D. L. Donoho, M. A. Saunders (2001), "Atomic decomposition by basis pursuit,", SIAM Review, Vol. 43, No.1 pp: 129-159.

D. L. Donoho, Y. Tsaig, I. Drori, J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Submitted, 2006.

G. Hennenfent, F.J. Herrmann, "Seismic denoising with non-uniformly sampled curvelets," Computing in science and engineering, pp: 50-59, May/June 2006.

F.J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," Submitted, 2006.