Sparse Signal Recovery: Theory, Applications and Algorithms

Bhaskar Rao
Department of Electrical and Computer Engineering
University of California, San Diego

Collaborators: I. Gorodnitsky, S. Cotter, D. Wipf, J. Palmer, K. Kreutz-Delgado

Acknowledgement: Travel support provided by Office of Naval Research Global under Grant Number: N62909-09-1-1024. The research was supported by several grants from the National Science Foundation, most recently CCF - 0830612
Outline

1. Talk Objective
2. Sparse Signal Recovery Problem
3. Applications
4. Computational Algorithms
5. Performance Evaluation
6. Conclusion
Outline

1. Talk Objective
2. Sparse Signal Recovery Problem
3. Applications
4. Computational Algorithms
5. Performance Evaluation
6. Conclusion
Importance of Problem

Organized with Prof. Bresler a Special Session at the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing

SPEC-DSP: SIGNAL PROCESSING WITH SPARSENESS CONSTRAINT

Signal Processing with the Sparseness Constraint
B. Rao (University of California, San Diego, USA)

Application of Basis Pursuit in Spectrum Estimation
S. Chen (IBM, USA); D. Donoho (Stanford University, USA)

Parsimony and Wavelet Method for Denoising
H. Krim (MIT, USA); J. Pesquet (University Paris Sud, France); I. Schick (GTE Internetworking and Harvard Univ., USA)

Parsimonious Side Propagation
P. Bradley, O. Mangasarian (University of Wisconsin-Madison, USA)

Fast Optimal and Suboptimal Algorithms for Sparse Solutions to Linear Inverse Problems
G. Harikumar (Tellabs Research, USA); C. Couvreur, Y. Bresler (University of Illinois, Urbana-Champaign, USA)

Measures and Algorithms for Best Basis Selection
K. Kreutz-Delgado, B. Rao (University of California, San Diego, USA)

Sparse Inverse Solution Methods for Signal and Image Processing Applications
B. Jeffs (Brigham Young University, USA)

Image Denoising Using Multiple Compaction Domains
P. Ishwar, K. Ratakonda, P. Moulin, N. Ahuja (University of Illinois, Urbana-Champaign, USA)
Sparse Signal Recovery is an interesting area with many potential applications.

Tools developed for solving the Sparse Signal Recovery problem are useful for signal processing practitioners to know.
Outline

1 Talk Objective
2 Sparse Signal Recovery Problem
3 Applications
4 Computational Algorithms
5 Performance Evaluation
6 Conclusion
Problem Description

- $t$ is $N \times 1$ measurement vector
- $\Phi$ is $N \times M$ Dictionary matrix. $M \gg N$.
- $w$ is $M \times 1$ desired vector which is sparse with $K$ non-zero entries
- $\epsilon$ is the additive noise modeled as additive white Gaussian
Problem Statement

**Noise Free Case:** Given a target signal $t$ and a dictionary $\Phi$, find the weights $w$ that solve:

$$\min_w \sum_{i=1}^{M} I(w_i \neq 0) \text{ such that } t = \Phi w$$

where $I(.)$ is the indicator function

**Noisy Case:** Given a target signal $t$ and a dictionary $\Phi$, find the weights $w$ that solve:

$$\min_w \sum_{i=1}^{M} I(w_i \neq 0) \text{ such that } \|t - \Phi w\|_2^2 \leq \beta$$
Complexity

- Search over all possible subsets, which would mean a search over a total of \( \binom{M}{K} \) subsets. Problem NP hard. With \( M = 30, N = 20, \) and \( K = 10 \) there are \( 3 \times 10^7 \) subsets (Very Complex)

- A branch and bound algorithm can be used to find the optimal solution. The space of subsets searched is pruned but the search may still be very complex.

- Indicator function not continuous and so not amenable to standard optimization tools.

**Challenge**: Find low complexity methods with acceptable performance
Outline

1. Talk Objective
2. Sparse Signal Recovery Problem
3. Applications
4. Computational Algorithms
5. Performance Evaluation
6. Conclusion
Applications

- Signal Representation (Mallat, Coifman, Wickerhauser, Donoho, ...)
- EEG/MEG (Leahy, Gorodnitsky, Ioannides, ...)
- Functional Approximation and Neural Networks (Chen, Natarajan, Cun, Hassibi, ...)
- Bandlimited extrapolations and spectral estimation (Papoulis, Lee, Cabrera, Parks, ...)
- Speech Coding (Ozawa, Ono, Kroon, Atal, ...)
- Sparse channel equalization (Fevrier, Greenstein, Proakis, )
- Compressive Sampling (Donoho, Candes, Tao..)
DFT Example

$N$ chosen to be 64 in this example.

Measurement $t$:

$$t[n] = \cos \omega_0 n, \quad n = 0, 1, 2, \ldots, N - 1$$

$$\omega_0 = \frac{2\pi}{64} \cdot \frac{33}{2}$$

Dictionary Elements:

$$\phi_k = [1, e^{-j\omega_k}, e^{-j2\omega_k}, \ldots, e^{-j(N-1)\omega_k}]^T, \quad \omega_k = \frac{2\pi}{M}$$

Consider $M = 64, 128, 256$ and 512

NOTE: The frequency components are included in the dictionary $\Phi$ for $M = 128, 256$, and 512.
FFT Results with Different $M$

- $M=64$
- $M=128$
- $M=256$
- $M=512$
Magnetoencephalography (MEG)

Given measured magnetic fields outside of the head, the goal is to locate the responsible current sources inside the head.
At any given time, typically only a few sources are active (SPARSE).
MEG Formulation

- Forming the overcomplete dictionary $\Phi$
  - The number of rows equals the number of sensors.
  - The number of columns equals the number of possible source locations.
  - $\Phi_{ij} =$ the magnetic field measured at sensor $i$ produced by a unit current at location $j$.
  - We can compute $\Phi$ using a boundary element brain model and Maxwell's equations.
- Many different combinations of current sources can produce the same observed magnetic field $t$.
- By finding the sparsest signal representation/basis, we find the smallest number of sources capable of producing the observed field.
- Such a representation is of neurophysiological significance.
Compressive Sampling

Transform Coding

\[ \Psi \rightarrow w \rightarrow b \]
Compressive Sampling

Transform Coding

\[ \psi \rightarrow w \rightarrow b \]

Compressive Sensing

\[ A \psi w b = \]
Compressive Sampling

Transform Coding

\[ \psi \quad w \quad b \]

Compressive Sensing

\[ A \quad \psi \quad w \quad A \quad b \quad t \]

Computation: \( t \rightarrow w \rightarrow b \)
Outline

1. Talk Objective
2. Sparse Signal Recovery Problem
3. Applications
4. Computational Algorithms
5. Performance Evaluation
6. Conclusion
Potential Approaches

Problem NP hard and so need alternate strategies

- **Greedy Search Techniques**: Matching Pursuit, Orthogonal Matching Pursuit
- **Minimizing Diversity Measures**: Indicator function not continuous. Define Surrogate Cost functions that are more tractable and whose minimization leads to sparse solutions, e.g. $\ell_1$ minimization
- **Bayesian Methods**: Make appropriate Statistical assumptions on the solution and apply estimation techniques to identify the desired sparse solution
Greedy Search Methods: Matching Pursuit

- Select a column that is most aligned with the current residual

\[
\begin{align*}
\mathbf{t} & \quad \Phi \quad \mathbf{w} \\
0.7 & \quad 4 & \quad \text{...} & \quad -0.3 & \quad \text{...} \\
\end{align*}
\]

- Remove its contribution
- Stop when residual becomes small enough or if we have exceeded some sparsity threshold.

Some Variations
- Matching Pursuit [Mallat & Zhang]
- Orthogonal Matching Pursuit [Pati et al.]
- Order Recursive Matching Pursuit (ORMP)
Inverse techniques

For the systems of equations $\Phi w = t$, the solution set is characterized by \( \{ w_s : w_s = \Phi^+ t + v, \ v \in \mathcal{N}(\Phi) \} \), where $\mathcal{N}(\Phi)$ denotes the null space of $\Phi$ and $\Phi^+ = \Phi^T(\Phi\Phi^T)^{-1}$.

**Minimum Norm solution** : The minimum $\ell_2$ norm solution $w_{mn} = \Phi^+ t$ is a popular solution.

**Noisy Case**: regularized $\ell_2$ norm solution often employed and is given by

\[
w_{reg} = \Phi^T(\Phi\Phi^T + \lambda I)^{-1} t
\]

**Problem**: Solution is not Sparse
Diversity Measures

- Functionals whose minimization leads to sparse solutions
- Many examples are found in the fields of economics, social science and information theory
- These functionals are concave which leads to difficult optimization problems

Bhaskar Rao
Department of Electrical and Computer Engineering
University of California, San Diego
Examples of Diversity Measures

$\ell_{(p \leq 1)}$ Diversity Measure

$$E^{(p)}(w) = \sum_{l=1}^{M} |w_l|^p, p \leq 1$$

Gaussian Entropy

$$H_G(w) = \sum_{l=1}^{M} \ln |w_l|^2$$

Shannon Entropy

$$H_S(w) = - \sum_{l=1}^{M} \tilde{w}_l \ln \tilde{w}_l. \text{ where } \tilde{w}_l = \frac{w_l^2}{\|w\|^2}$$
Diversity Minimization

Noiseless Case

\[ \min_w E^{(p)}(w) = \sum_{l=1}^{M} |w_l|^p \quad \text{subject to } \Phi w = t \]

Noisy Case

\[ \min_w \left( \|t - \Phi w\|^2 + \lambda \sum_{l=1}^{M} |w_l|^p \right) \]

\(p = 1\) is a very popular choice because of the convex nature of the optimization problem (Basis Pursuit and Lasso).
**Bayesian Methods**

- **Maximum A Posteriori Approach (MAP)**
  - Assume a sparsity inducing prior on the latent variable $w$
  - Develop an appropriate MAP estimation algorithm

- **Empirical Bayes**
  - Assume a parameterized prior on the latent variable $w$ (hyperparameters)
  - Marginalize over the latent variable $w$ and estimate the hyperparameters
  - Determine the posterior distribution of $w$ and obtain a point as the mean, mode or median of this density
Generalized Gaussian Distribution

Density function: Subgaussian: \( p > 2 \) and Supergaussian: \( p < 2 \)

\[
f(x) = \frac{p}{2\sigma \Gamma\left(\frac{1}{p}\right)} \exp \left\{ -\left(\frac{|x|}{\sigma}\right)^p \right\}
\]
**Student t Distribution**

Density function:

\[
f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \nu \Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}
\]
MAP using a Supergaussian prior

Assuming a Gaussian likelihood model for $f(t|w)$, we can find MAP weight estimates

$$w_{MAP} = \arg \max_w \log f(w|t)$$
$$= \arg \max_w (\log f(t|w) + \log f(w))$$
$$= \arg \min_w \left( \| \Phi w - t \|^2 + \lambda \sum_{l=1}^{M} |w_l|^p \right)$$

This is essentially a regularized LS framework. Interesting range for $p$ is $p \leq 1$. 

Bhaskar Rao
Department of Electrical and Computer Engineering
University of California, San Diego

Sparse Signal Recovery: Theory, Applications and Algorithms
28 / 42
MAP Estimate: FOCal Underdetermined System Solver (FOCUSS)

Approach involves solving a sequence of Regularized Weighted Minimum Norm problems

\[ q_{k+1} = \arg \min_q \left( \| \Phi_{k+1} q - t \|^2 + \lambda \| q \|^2 \right) \]

where \( \Phi_{k+1} = \Phi M_{k+1} \), and \( M_{k+1} = \text{diag}(|w_k|^{1-p/2}) \).

\[ w_{k+1} = M_{k+1} q_{k+1}. \]

\( p = 0 \) is the \( \ell_0 \) minimization and \( p = 1 \) is \( \ell_1 \) minimization.
FOCUSS summary

- For $p < 1$, the solution is initial condition dependent
  - Prior knowledge can be incorporated
  - Minimum norm is a suitable choice
  - Can retry with several initial conditions
- Computationally more complex than Matching Pursuit algorithms
- Sparsity versus tolerance tradeoff more involved
- Factor $p$ allows a trade-off between the speed of convergence and the sparsity obtained
Convergence Errors vs. Structural Errors

\begin{align*}
\text{Convergence Error} & \quad \text{Structural Error} \\
-\log p(w | t) & \quad -\log p(w | t) \\
W^* & \quad W^* \\
W_0 & \quad W_0
\end{align*}

\begin{align*}
w^* & = \text{solution we have converged to} \\
W_0 & = \text{maximally sparse solution}
\end{align*}
Shortcomings of these Methods

\( p = 1 \)
- Basis Pursuit/Lasso often suffer from structural errors.
- Therefore, regardless of initialization, we may never find the best solution.

\( p < 1 \)
- The FOCUSS class of algorithms suffers from numerous suboptimal local minima and therefore convergence errors.
- In the low noise limit, the number of local minima \( K \) satisfies

\[
K \in \left[ \binom{M-1}{N} + 1, \binom{M}{N} \right]
\]
- At most local minima, the number of nonzero coefficients is equal to \( N \), the number of rows in the dictionary.
Empirical Bayesian Method

Main Steps

- Parameterized prior \( f(w|\gamma) \)
- Marginalize

\[
f(t|\gamma) = \int f(t, w|\gamma) dw = \int f(t|w)f(w|\gamma) dw
\]

- Estimate the hyperparameter \( \hat{\gamma} \)
- Determine the posterior density of the latent variable \( f(w|t, \hat{\gamma}) \)
- Obtain point estimate of \( w \)

Example: Sparse Bayesian Learning (SBL by Tipping)
Outline

1 Talk Objective
2 Sparse Signal Recovery Problem
3 Applications
4 Computational Algorithms
5 Performance Evaluation
6 Conclusion
DFT Example

FFT or Minimum Norm Solution

Solution Obtained using Algorithm 1 after 3 iterations
Empirical Tests

- Random overcomplete bases $\Phi$ and sparse weight vectors $w_0$ were generated and used to create target signals $t$, i.e.,
  $$ t = \Phi w_0 + \epsilon $$
- SBL (Empirical Bayes) was compared with Basis Pursuit and FOCUSS (with various $p$ values) in the task of recovering $w_0$. 
Experiment 1: Comparison with Noiseless Data

- Randomized $\Phi$ (20 rows by 40 columns).
- Diversity of the true $w_0$ is 7.
- Results are from 1000 independent trials.

NOTE: An error occurs whenever an algorithm converges to a solution $w$ not equal to $w_0$.

<table>
<thead>
<tr>
<th></th>
<th>FOCUSS ($p = 0.001$)</th>
<th>FOCUSS ($p = 0.9$)</th>
<th>Basis Pursuit ($p = 1.0$)</th>
<th>SBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence Errors</td>
<td>34.1%</td>
<td>18.1%</td>
<td>0.0%</td>
<td>7.4%</td>
</tr>
<tr>
<td>Structural Errors</td>
<td>0.0%</td>
<td>5.7%</td>
<td>22.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total Errors</td>
<td>34.1%</td>
<td>23.8%</td>
<td>22.3%</td>
<td>7.4%</td>
</tr>
</tbody>
</table>
**Experiment II: Comparison with Noisy Data**

- Randomized $\Phi$ (20 rows by 40 columns).
- Diversity of the true $w_0$ is 7.
- 20 db AWGN
- Results are from 1000 independent trials.

*NOTE*: We no longer distinguish convergence errors from structural errors.

<table>
<thead>
<tr>
<th></th>
<th>FOCUSS $(p = 0.001)$</th>
<th>FOCUSS $(p = 0.9)$</th>
<th>Basis Pursuit $(p = 1.0)$</th>
<th>SBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Errors</td>
<td>52.2%</td>
<td>43.1%</td>
<td>45.5%</td>
<td>21.1%</td>
</tr>
</tbody>
</table>
MEG Example

- Data based on CTF MEG system at UCSF with 275 scalp sensors.
- Forward model based on 40,000 points (vertices of a triangular grid) and 3 different scale factors.
- Dimensions of lead field matrix (dictionary): 275 by 120,000.
- Overcompleteness ratio approximately 436.
- Up to 40 unknown dipole components were randomly placed throughout the sample space.
- SBL was able to resolve roughly twice as many dipoles as the next best method (Ramirez 2005).
Outline

1. Talk Objective
2. Sparse Signal Recovery Problem
3. Applications
4. Computational Algorithms
5. Performance Evaluation
6. Conclusion
Summary

- Discussed role of sparseness in linear inverse problems
- Discussed Applications of Sparsity
- Discussed methods for computing sparse solutions
  - Matching Pursuit Algorithms
  - MAP methods (FOCUSS Algorithm and $\ell_1$ minimization)
  - Empirical Bayes (Sparse Bayesian Learning (SBL))

Expectation is that there will be continued growth in the application domain as well as in the algorithm development.