A General Framework for the Design and Analysis of Sparse FIR Linear Equalizers

Exploiting Sparsity for Interference Management in Broadband Networks: Theory, Application, and Testbeds.

NPRP No. 06-070-2-024

GlobalSIP - Dec. 2015
Outlines

• Background about Sparse Approximation (spA).
• Motivation and Related Work.
• Our Objectives (Contributions).
• Sparse FIR Linear Equalizers.
• Worst-Case Coherence Analysis
• Selected Results.
• Conclusions.
Outlines

• **Background about Sparse Approximation (spA).**
  • Motivation and Related Work.
  • Our Objectives
  • System Model
  • Measurement Matrix Properties.
  • Selected Results.
**What is SpA?**

- SpA is the problem of approximating a signal (vector) with the best linear $m$ combination of elements from a redundant dictionary. *Greedy is Good [Joel A. Tropp, 2006]*

$$
\begin{pmatrix}
0.9593 & 0.2575 & 0.2435 & 0.2511 & 0.8308 \\
0.5472 & 0.8407 & 0.9293 & 0.6160 & 0.5853 \\
0.1386 & 0.2543 & 0.3500 & 0.4733 & 0.5497 \\
0.1493 & 0.8143 & 0.1966 & 0.3517 & 0.9172 \\
\end{pmatrix}
\begin{pmatrix}
0 \\
0.7537 \\
0 \\
0 \\
0 \\
\end{pmatrix}
\approx
\begin{pmatrix}
0.2254 \\
0.7531 \\
0.2366 \\
0.6390 \\
\end{pmatrix}
Sparse Approximation Review

- Consider the linear system of Equations

\[ y = Ax + z \]

Where:
- \( A \) : represents the Sparsifying Dictionary.
- \( y \) : represents the Data (MD).
- \( x \) : represents the Original Signal.
- \( z \) : Additive Noise.

- Two main approaches have been proposed in the literature to solve the spA problems; specifically,
  - \( l_1 \) minimization.
  - Greedy algorithms.
Sparse Approximation Review

1. **Convex optimization**: \( l_1 \)-minimization.

\[
\min_{\tilde{x} \in \mathbb{C}^N} \| \tilde{x} \|_0 \quad \text{s.t.} \quad \| y - A\tilde{x} \|_2^2 \leq \epsilon
\]

2. **Greedy algorithms**: One of the most widely greedy algorithms is the **Orthogonal Matching Pursuit** (OMP)

\[
\tilde{x} = \text{OMP} (y, A, \text{stopping criterion})
\]

**Orthogonal Matching Pursuit Algorithm**

OMP: Selects the most corr. columns of \( A \) with the MD \( y \).
Outlines

• Background about Compressive Sensing (CS).

• **Motivation and Related Work.**
  • Our Objectives
  • System Model
  • Measurement Matrix Properties.
  • Selected Results.
Motivation and Related Work

- Highly dispersive comm. Channels are ch’c by IR’s that span tens-hundreds of symbol periods.
- Very long equalizers have to be employed at the Rx to mitigate the resulting severe ISI. [Bingham, 1990].
- Complexity of finite-impulse-response (FIR) equalizers is proportional to the square of the number of nonzero taps in the filter.

\[
\begin{align*}
\text{FIR filters with non-consecutive non-zero taps typically} \\
\text{referred to as sparse filters}
\end{align*}
\]


1. Design large length dense filter and choose a subset of strongest taps.
2. Design sparse filter on the selected locations.
Outlines

• Background about Compressive Sensing (CS).
• Motivation and Related Work.

• Our Objectives
  • System Model
  • Measurement Matrix Properties.
  • Selected Results.
**Our Contributions,**

- Designing a general framework that transforms the problem of design of sparse linear equalizers (LEs) into the problem of sparsest approximation of a vector in different dictionaries.
Outlines

• Background about Compressive Sensing (CS).
• Motivation and Related Work.
• Our Objectives

• *Sparse FIR LE problem formulation*
  • Measurement Matrix Properties.
  • Selected Results.
Sparse Equalization

- We consider a general case of a digital communication system with $n_i$ inputs and $n_o$ outputs.
- The samples at the $j^{th}$ output ($1 \leq j^{th} \leq n_o$)

$$y_{k}(j) = \sum_{i=1}^{n_i} \sum_{m=0}^{\nu(i,j)} h_{m}^{(i,j)} x_{k-m}^{(i)} + n_{k}^{(j)}$$

- Grouping over a block of $N_f$ symbol periods,

$$y_{k:k-N_f+1} = Hx_{k:k-N_f-\nu+1} + n_{k:k-N_f+1}$$

- For SISO-LEs (Error Signal):

$$e_{k} = x_{k} - \hat{x}_{k} = x_{k} - \Delta - w^{H} y_{k:k-N_f+1},$$
Sparse FIR SISO-LEs

MSE = $\xi = E \left[ \left| e_k^2 \right| \right] = E \left[ e_k^H e_k \right] = E \left[ (x_{k-\Delta} - w^H y_{k:k-N_f+1})^H (x_{k-\Delta} - w^H y_{k:k-N_f+1}) \right] = E \left[ x_{k-\Delta}^2 \right] - w^H E \left[ y_{k:k-N_f+1} x_{k-\Delta} \right] - E \left[ y_{k:k-N_f+1} x_{k-\Delta} \right]^H w + w^H \underbrace{y_{k:k-N_f+1} y_{k:k-N_f+1}^H}_{R_{yy}}$ $w$

$\xi(w) = \varepsilon_x - r_{\Delta}^H R_{yy}^{-1} r_{\Delta} + (w - R_{yy}^{-1} r_{\Delta})^H R_{yy} (w - R_{yy}^{-1} r_{\Delta})$

$\hat{w}_S \triangleq \arg\min_{w} \left\| w \right\|_0$ subject to $\xi_e(w) \leq \delta_{eq}$

- $W^H$ : Filter taps.
- $\varepsilon_x$ : Signal Energy.
- $r_{\Delta} = R_{yx}^{-1} \ast 1_\Delta$, where $1_\Delta$ is all zero vector except the $(\Delta+1)$ entry, where this entry is 1.
Proposed sparse approximation framework

- We provide a general framework for designing sparse FIR LEs and equalizers that can be considered as the problem of sparse approximation using different dictionaries. Mathematically, this framework poses the design problem as follows:

\[
\hat{w}_s \triangleq \arg\min_{w \in \mathbb{C}^{N_f}} \|w\|_0 \quad \text{subject to} \quad \|A(\Phi w - b)\|_2^2 \leq \delta_{eq}
\]

where \( \Phi \) is the dictionary that will be used to sparsely approximate \( w \), while \( A \) is a known matrix and \( b \) is a known data vector, both of which change depending upon the sparsifying dictionary \( \Phi \).
Example:

Sparse FIR SISO-LEs

\[
\begin{align*}
\min_{w \in \mathbb{C}^{N_f}} & \quad \|w\|_0 \quad \text{s.t.} \quad \|(L^H w - L^{-1} r_\Delta)\|_2^2 \leq \delta_{eq}, \\
\min_{w \in \mathbb{C}^{N_f}} & \quad \|w\|_0 \quad \text{s.t.} \quad \left\| \left( D^{\frac{1}{2}} U^H w - D^{-\frac{1}{2}} U^H r_\Delta \right) \right\|_2^2 \leq \delta_{eq}, \text{ and} \\
\min_{w \in \mathbb{C}^{N_f}} & \quad \|w\|_0 \quad \text{s.t.} \quad \|L^{-1} (R_{yy} w - r_\Delta)\|_2^2 \leq \delta_{eq}. 
\end{align*}
\]

Table I

EXAMPLES OF DIFFERENT SPARSIFYING DICTIONARIES.

<table>
<thead>
<tr>
<th>Cholesky Factorization</th>
<th>Eigen Decomposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{yy} = LL^H$ or $R_{yy} = P \Lambda P^H$</td>
<td>$R_{yy} = UDU^H$</td>
</tr>
<tr>
<td>$A$</td>
<td>$\Phi$</td>
</tr>
<tr>
<td>$I$</td>
<td>$L^H$</td>
</tr>
<tr>
<td>$L^{-1}$</td>
<td>$R_{yy}$</td>
</tr>
<tr>
<td>$I$</td>
<td>$\Lambda^{\frac{1}{2}} P^H$</td>
</tr>
</tbody>
</table>
To this end ?!

✓ We have shown that the problem of designing sparse FIR equalizers can be cast into one of sparse approximation of a vector by a fixed dictionary. The general form of this problem is given by slide 23.

✓ Which Sparsifying dictionary results in the sparsest design??

Our next challenge is to determine the best sparsifying dictionary for use in our framework.

➢ We know from the sparse approximation literature that the sparsity of the OMP solution tends to be inversely proportional to the worst-case coherence $\mu(\Phi)$. 
Worst-Case Coherence $\mu(\Phi)$ Analysis

• Once again, the sparsity of the OMP solution tends to be inversely proportional to the worst-case coherence.

$$\mu(\Phi) \triangleq \max_{i \neq j} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|_2 \|\phi_j\|_2}$$

• Orthogonal Matching Pursuit (OMP) and like methods will estimate the optimum representation if $\mu(\emptyset) < 1$. [Tropp, 2006]

• Our Approach in investigating $\mu(\Phi)$?

$$R_{yy} = HH^H + \frac{1}{SNR} I.$$
Worst-Case Coherence of $R_{yy}$, $\mu(R_{yy})$:

$$R_{yy} = HH^H + \frac{1}{SNR} I.$$

$$= \varepsilon_x$$

$$= \phi_1^H$$

$$R_{yy} = \text{Toeplitz} \left( \begin{bmatrix} r_0 & r_1 & \ldots & r_v & 0 & \ldots & 0 \end{bmatrix} \right),$$
Worst-Case Coherence of $R_{yy}$, $\mu(R_{yy})$:

$$R_{yy} = HH^H + \frac{1}{SNR} I.$$

$$R_{yy} = \text{Toeplitz} \left( \begin{bmatrix} r_0 & r_1 & \ldots & r_v & 0 & \ldots & 0 \end{bmatrix} \right).$$

- To get the worst CIR taps which result in the worst-case coherence:

$$\max \sum_{i=1}^{v} |h_i h_{i-1}^*| \quad \text{s.t.} \quad \sum_{i=0}^{v} |h_i|^2 = 1.$$

$$\max |h^H R h| \quad \text{s.t.} \quad h^H h = 1,$$

- The solution of the above optimization problem is:

$$\lambda_s = 2 \cos \left( \frac{\pi s}{v + 2} \right), \quad h_{j}^{(s)} = \sqrt{\frac{2}{v + 2}} \sin \left( \frac{j \pi s}{v + 2} \right),$$
Worst-Case Coherence of $R_{yy}$ Factors

- It is important to note here that the other dictionaries, which result from decomposing $R_{yy}$ and $R_{x/y}$, can be considered as square roots of them in the spectral-norm sense. For example:

$$
\| R_{yy} \|_2 = \| L_y L_y^H \|_2 \leq \| L_y^H \|^2_2
$$

$$
\| R_\delta \|_2 = \| U_\delta D_\delta U_\delta^H \|^2_2 \leq \| D_\delta^{1/2} U_\delta^H \|^2_2
$$

- Through simulation, we show that the coherence of the factors of $R_{yy}$ is less than that of $\mu(R_{yy})$. 
Can we do better?

WE CAN DO BETTER
Reduced-Complexity Design

- The proposed designs for Sparse FIR LEs involve Cholesky factorization and/or eigen decomposition, whose computational costs could be large for channels with large delay spreads.
- For a Toeplitz matrix, the most efficient algorithms for Cholesky factorization are Levinson or Schur algorithms, which involve $O(M^2)$ computations, where $M$ is the matrix dimension.
- In contrast, since a circulant matrix is asymptotically equivalent to a Toeplitz matrix, for reasonably large dimension, the eigen decomposition of a circulant matrix can be computed efficiently using the fast Fourier transform (FFT) and its inverse with only $O(M \log(M))$ operations.
- By using this asymptotic equivalence between Toeplitz and circulant matrices, all computations needed for $R_{yy}$ factorization can be done efficiently with the FFT and inverse FFT. In addition, direct matrix inversion can be avoided.
Reduced-Complexity Design

- A circulant matrix, \( C \), has the discrete Fourier transform (DFT) basis vectors as its eigenvectors and the DFT of its first column as its eigenvalues.
- An \( M \times M \) circulant matrix \( C \) can be decomposed as
  \[
  C = \frac{1}{M} (F_M^H \Gamma_c F_M)
  \]
- The autocorrelation matrix \( R_{yy} \) is computed as:
  \[
  \overline{R_{yy}} = E[\tilde{y}_k \tilde{y}_k] + \frac{1}{SNR} I_{N_f}
  \]
  \[
  \overline{R_{\tilde{y}\tilde{y}}} + \frac{1}{\sigma_n^2}
  \]
- To approximate \( R_{yy} \) as a circulant matrix, we assume that \( \{y_k\} \) is cyclic. Hence, \( E\{y_k y_k\} \) can be approximated as a time-averaged autocorrelation function as follows:
Reduced-Complexity Design:  $R_{yy}$

$$
\bar{R}_{\tilde{y}\tilde{y}} = \frac{1}{N_f} \sum_{k=0}^{N_f-1} \tilde{y}_k \tilde{y}_k^H = \frac{1}{N_f} C_y C_y^H \\
= \frac{1}{N_f} \left( \frac{1}{N_f} F_{N_f}^H \Lambda \tilde{Y} F_{N_f} \right) \left( \frac{1}{N_f} F_{N_f}^H \Lambda^H Y \tilde{Y} F_{N_f} \right) \\
= \frac{1}{N_f^2} F_{N_f}^H \Lambda \tilde{Y} \Lambda^H Y \tilde{Y} F_{N_f} \\
= \frac{1}{N_f^2} \left( F_{N_f}^H \Lambda |\tilde{Y}|^2 F_{N_f} \right). \\
$$

Note that $F_{N_f}^H F_{N_f} = F_{N_f} F_{N_f}^H = N_f I_{N_f}$ and

$$
C_y = \begin{bmatrix}
\tilde{y}_{N_f-1} & \tilde{y}_0 & \tilde{y}_1 & \cdots & \tilde{y}_{N_f-2} \\
\tilde{y}_{N_f-2} & \tilde{y}_{N_f-1} & \tilde{y}_0 & \cdots \\
\tilde{y}_{N_f-3} & \tilde{y}_{N_f-2} & \tilde{y}_{N_f-1} & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\tilde{y}_0 & \cdots & \cdots & \cdots & \tilde{y}_{N_f-1}
\end{bmatrix}
$$
Reduced-Complexity Design: $R_{yy}$

$$\tilde{Y} = H^H \odot P_\Delta \odot X$$

$$P_\Delta = \begin{bmatrix} 1 & e^{-j2\pi \Delta/N_f} & \ldots & e^{-j2\pi (N_f-1)\Delta/N_f} \end{bmatrix}^T$$

$$|\tilde{Y}|^2 = N_f |H_i|^2 = N_f \begin{bmatrix} |H_0|^2 & \ldots & |H_{N_f-1}|^2 \end{bmatrix}^H$$

where $\odot$ denotes the element-wise multiplication. Then,

$$R_{yy} = R_{\tilde{y}\tilde{y}} + \sigma_n^2 I_{N_f}$$

$$= \frac{1}{N_f^2} F_{N_f}^H \left( \Lambda_{N_f} |H|^2 + N_f \sigma_n^2 1_{N_f} \right) F_{N_f}$$

$$= F_{N_f}^H \left( \frac{1}{N_f} \Lambda_{q} |H|^2 + \sigma_n^2 1_{N_f} \right) F_{N_f}$$

$$= F_{N_f}^H \left( \frac{1}{N_f} \Lambda_{q} \right) F_{N_f} = QQ^H.$$
Outlines

• Background about Compressive Sensing (CS).
• Motivation and Related Work.
• Our Objectives

• Measurement Matrix Properties.

• Selected Results.
Performance of circulant approximation based approach for UPDP channel with $v = 5$ and input $SNR = 30dB$. 

Feed–Forward Filter taps weights of single realizations of the MMSE and the equivalent circulant–approx. based solution for SISO–LE with $SNR = 30dB$. 

![Graph showing tap weights](image-url)
Worst-case coherence for $R_{yy}$ and its factors $L_y^H$ and $\frac{1}{2} D_y^2 U_y^H$

- Dashed lines represent the coherence of the corresponding circulant approximation for $D_y^2 U_y^H$ (i.e., $Q^H$) and $R_{yy}$ (i.e., $R_{yy} = QQ^H$).
- UPDP with $v = 8$ and $N_f = 80$. 
Active taps Percentage versus Performance Loss: Vech-A-PDP (LEs)

Allowing for 0.25 dB performance loss results in a significant reduction in the number of the total filter taps (65% reduction) 😊😊
65% reduction (0.25dB loss) in FIR taps with almost same performance 😊 😊
Outlines

• Conclusions
Conclusions

- Dramatic complexity reduction is achieved at a small performance loss compared to the conventional MMSE non-sparse FIR equalizer design.
- Sparse FIR equalizers enjoy more rapid taps weight adaptation to changing channel condition which support higher mobile velocities.
- Based on the asymptotic equivalence of Toeplitz and circulant matrices, we provide reduced-complexity designs, where matrix factorizations can be carried out efficiently using the FFT and inverse FFT.
- The dictionary with the smallest coherence gives the sparsest filter design.
Thank you

a.o.alabbasi@ieee.org