A User’s Guide to Compressed Sensing for Communications Systems

Kazunori Hayashi
Graduate School of Informatics
Kyoto University
kazunori@i.kyoto-u.ac.jp
Survey paper on compressed sensing:

Agenda

• A Brief Introduction to Compressed Sensing
  • Warming-up
  • Linear Equations Review
  • Problem Setting
  • Algorithms

• Applications to Communications Systems
  • Channel Estimation
  • Wireless Sensor Network
  • Network Tomography
  • Spectrum Sensing
  • RFID
A Brief Introduction to Compressed Sensing
Warming-up Exercise(1)

-Linear simultaneous equations:

\[ x + y + z = 3 \]
\[ x - z = 0 \quad \text{(\# of Eqs. = \# of unknowns)} \]
\[ 2x - y + z = 2 \]

\[ x = 1 \]
\[ y = 1 \]
\[ z = 1 \]

Unique solution
Warming-up Exercise (2)

- Linear simultaneous equations:

\[
\begin{align*}
    x + y + z &= 3 \\
    x - z &= 0 \\
    2x - y + z &= 2 \\
    x + y &= 1
\end{align*}
\]

(# of Eqs. > # of unknowns)

No solution
Warming-up Exercise (3)

- Linear simultaneous equations:

\[
\begin{align*}
x + y + z &= 3 \\
x - z &= 0 \\
2x - y + z &= 2 \\
x + y &= 2
\end{align*}
\]

(# of Eqs. > # of unknowns)

Unique solution

\[
\begin{align*}
x &= 1 \\
y &= 1 \\
z &= 1
\end{align*}
\]
Warming-up Exercise (4)

-Linear simultaneous equations:

\[ x + y + z = 3 \]
\[ x - z = 0 \]  (# of Eqs. < # of unknowns)

\[ x = t \]
\[ y = -2t + 3 \]  Infinitely many solutions
\[ z = t \]
Linear Equations in Engineering Problems

\[ y = Ax \]

- Sensing matrix : \( A \in \mathbb{R}^{m \times n} \)
- Unknown vector : \( x \in \mathbb{R}^{n} \)
- Measurement vector : \( y \in \mathbb{R}^{m} \)

- Answering “No solution” or “Non-unique solutions” is not enough

- Measurement may include noise

\[ y = Ax + v \]

noise
Linear Equations Review: Conventional approach \((n = m, \text{ noise-free})\)

- **Sensing matrix** : \(A \in \mathbb{R}^{m \times n}\)
- **Unknown vector** : \(x \in \mathbb{R}^{n}\)
- **Measurement vector** : \(y \in \mathbb{R}^{m}\)

\[
y = Ax
\]

- **Unique Solution** (if \(A\) is not singular):

\[
x = A^{-1}y
\]

Corresponds to warming-up (1)
Linear Equations Review: Conventional approach ($n < m$, noise-free)

$$y = Ax$$

- Sensing matrix: $A \in \mathbb{R}^{m \times n}$
- Unknown vector: $x \in \mathbb{R}^n$
- Measurement vector: $y \in \mathbb{R}^m$

-Unique Solution (if $A$ is of full column rank):

$$x = (A^T A)^{-1} A^T y$$

Corresponds to warming-up (3)
Linear Equations Review: Conventional approach \((n > m\), noise-free\)

\[ y = Ax \]

Sensing matrix : \( A \in \mathbb{R}^{m \times n} \)

Unknown vector : \( x \in \mathbb{R}^{n} \)

Measurement vector : \( y \in \mathbb{R}^{m} \)

- Non-unique solutions

Select \( x \), which has minimum \( \ell_{2} \)-norm

- Minimum-norm solution :

\[
\hat{x}_{MN} = \arg\min_{x} ||x||_{2}^{2} \quad \text{subject to} \quad Ax = y
\]

\[
= A^{T}(AA^{T})^{-1}y
\]

Corresponds to warming-up (4)
Linear Equations Review: Conventional approach ($n < m$, noisy)

\[
y = Ax
\]

- Sensing matrix: $A \in \mathbb{R}^{m \times n}$
- Unknown vector: $x \in \mathbb{R}^n$
- Measurement vector: $y \in \mathbb{R}^m$

- No solution in general
  (\(y\) is not included in Img. of \(A\) due to noise)

Select $x$, which minimize squared error between $Ax$ and $y$, as a solution

- Least-square (LS) solution:
  \[
  \hat{x}_{LS} = \arg \min_x \|Ax - y\|_2^2 = (A^T A)^{-1} A^T y
  \]

Corresponds to warming-up (2)
Linear Equations Review: Conventional approach ($n > m$, noisy)

\[ y = Ax \]

**Sensing matrix:** \( A \in \mathbb{R}^{m \times n} \)

**Unknown vector:** \( x \in \mathbb{R}^n \)

**Measurement vector:** \( y \in \mathbb{R}^m \)

- Non-unique solutions

- Regularized LS solution:

\[
\hat{x}_{rLS} = \arg \min_x \left( \|Ax - y\|_2^2 + \lambda \|x\|_2^2 \right)
\]

\[
= (\lambda I + A^T A)^{-1} A^T y
\]

\( \lambda \): regularization parameter
Problem Setting of Compressed Sensing

Linear equations of \( n > m \) with a prior information that the true solution is sparse

\[
\begin{align*}
\mathbf{y} &= \mathbf{A}\mathbf{x} \\
\text{Sensing matrix (non-adaptive):} &\quad \mathbf{A} \in \mathbb{R}^{m \times n} \\
\text{Unknown vector:} &\quad \mathbf{x} \in \mathbb{R}^{n} \\
\text{Measurement vector:} &\quad \mathbf{y} \in \mathbb{R}^{m} \\
\text{Underdetermined:} &\quad n > m \\
\text{Sparse unknown vector:} &\quad \|\mathbf{x}\|_0 \ll n
\end{align*}
\]

\( \|\mathbf{x}\|_0 \): \# of nonzero elements of \( \mathbf{x} \)
Signal Recovery via $\ell_0$ Optimization

-Natural and straightforward approach:

$$\hat{x}_{\ell_0} = \arg\min_x ||x||_0 \text{ subject to } Ax = y$$

- if $m > ||x||_0$, almost always true solution is obtained
- NP hard in general
Signal Recovery via $\ell_1$ Optimization

Replace $\|x\|_0$ with $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

$\hat{x}_{\ell_1} = \arg\min_x \|x\|_1 \quad \text{subject to} \quad Ax = y$

- It can be posed as linear programing (LP) problem
- True solution can be obtained even when $n > m$ under some conditions

If $A$ satisfies RIP (restricted isometry property), only $O(k \log(n/k))$ measurement will be enough for exact recovery
Intuitive Illustrations of Sparse Signal Recovery

\( \ell_1 \) recovery:
\[
\hat{x}_{\ell_1} = \arg \min_x \|x\|_1 \quad \text{subject to} \quad Ax = y
\]

MN solution:
\[
\hat{x}_{MN} = \arg \min_x \|x\|_2^2 \quad \text{subject to} \quad Ax = y
\]
Reconstruction with Noisy Measurement (1/2)

Constrained $\ell_1$ reconstruction:

$$\hat{x}_{\ell_1} = \arg\min_x \|x\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \epsilon$$

- Natural constraint for bounded noise
- A certain guarantee of signal recovery can be given for Gaussian noise
Reconstruction with Noisy Measurement (2/2)

\[ \ell_1-\ell_2 \text{ optimization:} \]

\[ \hat{x}_{\ell_1-\ell_2} = \arg \min_x \left( \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1 \right) \]

- Equivalent to constrained \( \ell_1 \) reconstruction for a certain choice of \( \lambda \)
- Analogy to regularized LS problem:
  \[ \hat{x}_{rLS} = \arg \min_x \left( \|Ax - y\|_2^2 + \lambda \|x\|_2^2 \right) \]
- LARS (least angle regression) algorithm
Equivalent approaches to $\ell_1-\ell_2$ optimization

- **Lasso** (Least absolute shrinkage and selection operator): 
  \[
  \hat{x}_{\text{Lasso}} = \arg\min_x \|Ax - y\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq t \quad (t > 0)
  \]

- Maximum a posteriori (MAP) estimator for AWGN with Laplacian prior $\propto \exp(-\lambda\|x\|_1)$
# Algorithms for Compressed Sensing

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solver</th>
<th>Subsec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\min_x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>$\min_x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>SPGL1 [49]</td>
<td>5.1.5</td>
</tr>
<tr>
<td>$\min_x \left( \frac{1}{2}</td>
<td></td>
<td>Ax - y</td>
</tr>
<tr>
<td></td>
<td>FPC [51]</td>
<td>5.1.5</td>
</tr>
<tr>
<td></td>
<td>Bregman [52]</td>
<td>5.1.5</td>
</tr>
<tr>
<td></td>
<td>GPSR [53]</td>
<td>5.1.5</td>
</tr>
<tr>
<td></td>
<td>TwIST [54]</td>
<td>5.1.5</td>
</tr>
<tr>
<td></td>
<td>SpaRSA [55]</td>
<td>5.1.5</td>
</tr>
<tr>
<td>$\min_x</td>
<td></td>
<td>Ax - y</td>
</tr>
</tbody>
</table>

Sample program (Scilab):

http://www.msys.sys.i.kyoto-u.ac.jp/csdemo.html
Applications to Communications Systems
Key issues in applications

- Sparsity of signals of interest
  - Is it natural to assume the sparsity of them?
  - In which domain?

- Linear measurement of the signals
  - Can we formulate with linear equation?
  - Is it happy to reduce the number of equations?
Channel Estimation

Multipath channel

Transmitter

Receiver

Tx. signal $S_n$  Convolution channel $h_0, \ldots, h_M$  Rx. signal $y_n$

$$y_n = \sum_{m=0}^{M} h_m s_{n-m} + \nu_n$$

Estimation of channel impulse response using Rx. signals
Channel Estimation

- Received signal vector:
  \[ y = [y_1, \ldots, y_{P-M}]^T \]
  \[ = Hs + v \]

Channel matrix:
\[
H = \begin{bmatrix}
  h_M & \cdots & h_0 & 0 & \cdots & 0 \\
  0 & h_M & h_0 & \cdots & \vdots \\
  \vdots & \vdots & \ddots & \ddots & \vdots \\
  0 & \cdots & 0 & h_M & \cdots & h_0 \\
\end{bmatrix} \in (P-M) \times P
\]

Known pilot signal:
\[ s = [s_1, \ldots, s_P]^T \in \mathbb{R}^P \]

White noise:
\[ v = [v_1, \ldots, v_{P-M}]^T \]
Channel Estimation

- Received signal vector (rewritten) :

\[ y = Sh + v \]

\[
S = \begin{bmatrix}
    s_{M+1} & s_M & \cdots & s_1 \\
    s_{M+2} & s_{M+1} & \cdots & s_2 \\
    \vdots & \vdots & \ddots & \vdots \\
    s_P & s_{P-1} & \cdots & s_{P-M}
\end{bmatrix} \quad (P - M) \times (M + 1)
\]

\[
h = [h_0, \ldots, h_M]^T \in \mathbb{R}^{M+1}
\]

- Conventional approaches :

  - LS solution \((P - M \geq M + 1)\) :
    \[ \hat{h} = (S^T S)^{-1} S^T y \]

  - MN solution \((P - M < M + 1)\) :
    \[ \hat{h} = S^T (S S^T)^{-1} y \]
Channel Estimation

- Sparsity of wireless channel impulse response:

  Wireless channel is discrete in nature, and the feature appears as the bandwidth increases.

- Estimation via compressed sensing:

  \[ \hat{h} = \arg \min_{h} ||h||_1 \quad \text{subject to} \quad ||Sh - y||_2 \leq \epsilon \]

  \( S \) : Sensing matrix (Random, Toeplitz)

  \( h \) : Unknown sparse channel impulse response

Merit: Reduction of pilot signals
(Improvement of frequency efficiency)
Wireless Sensor Network (Single-Hop)

- **Tx. signal** @ time \(i\) @ \(j\)-th node:

\[
A_{i,j} s_j
\]

\(s_j\) : measurement data

\(A_{i,j}\) : random coefficients (generated from node id)

- **Rx. signal** @ time \(i\) @ fusion center

\[
y_i = \sum_{j=1}^{n} A_{i,j} s_j + v_i
\]

\[
y = [y_1, \ldots, y_m]^T
\]

\[
= As + v
\]
Wireless Sensor Network (Single-Hop)

- Sparsity of measurement data:
  \[ s = \Phi x \quad (\Phi : \text{invertible transformation matrix, typically denoting DFT, DCT, wavelet, etc.}) \]

  \( s \) might be sparse in a certain representation with basis \( \Phi \)
  (due to spatial correlation of data)

- Data recovery via compressed sensing:
  \[ \hat{x} = \arg \min_x \| x \|_1 \quad \text{subject to} \quad \| A\Phi x - y \|_2 \leq \epsilon \]

  \( A\Phi \): sensing matrix
  \( x \): unknown sparse vector

Merit: Reduction of data collection time
Wireless Sensor Network (Multi-Hop)

- Rx. signal@sink node in the $i$-th multi-hop communication:

$$y_i = \sum_{j \in P_i} A_{i,j} s_j$$

$P_i$: index set of nodes included in the $i$-th multi-hop communication

$A_{i,j}$: random coefficient of the $j$-th node in the $i$-th multi-hop communication ($A_{i,j} = 0$ if $j \notin P_i$)

$$y = A\Phi x$$

Sensing matrix depends on routing algorithm

Merit: Reduction of # of communications (power consumption)
Network Tomography

- Link-level parameter estimation based on end-end measurements
- End-End performance evaluation based on link-level measurements
Delay Tomography (Link Delay Estimation)

- delay @ link \( e_j : x_j \)
- total delay in the \( i \)-th path:

\[
y_i = \sum_{j \in \mathcal{P}_i} A_{i,j} x_j
\]

\( \mathcal{P}_i \): set of link index included in the \( i \)-th path

\[
A_{i,j} = \begin{cases} 
  1 & j \in \mathcal{P}_i \\
  0 & \text{otherwise}
\end{cases}
\]

\[
y = Ax \quad \text{A: Routing Matrix}
\]
Delay Tomography (Link Delay Estimation)

- Sparsity of link delay:
  Only limited number of links have large delay, due to node failure or some other reasons in typical network

- Estimation via compressed sensing:

\[
\hat{x} = \arg \min_x \|x\|_1 \quad \text{subject to} \quad Ax = y
\]

\(A\) : sensing matrix (routing matrix)
\(x\) : (approximately) sparse delay vector

Merit: Reduction of # of prove packets
Loss Tomography
(Link Loss Probability Estimation)

- packet loss probability @ link $e_j : p_j$
- packet loss probability @ $i$-th path: $q_i$
- packet success probability @ $i$-th path: $1 - q_i = \prod_{j \in P_i} (1 - p_j)$

$$-\log(1 - q_i) = -\log \left( \prod_{j \in P_i} (1 - p_j) \right)$$

$$= -\sum_{j \in P_i} \log(1 - p_j)$$

$y_i = -\log(1 - q_i)$
$x_j = -\log(1 - p_j)$

$$y_i = \sum_{j \in P_i} A_{i,j} x_j \quad \Rightarrow \quad y = Ax$$
Spectrum Sensing

- Rx. signal (continuous time): \( r(t) \quad t \in [0, nT_s] \)
  \( T_s \): inverse of Nyquist rate

- Rx. signal with Nyquist rate sampling:
  \[ r_t = [r(T_s), \ldots, r(nT_s)]^T \]

- Rx. signal with (discrete time) linear measurement:
  \[ y_t = \Psi r_t \]
  \( \Psi \): \( m \times n \) sensing matrix

( \( m < n \Rightarrow \) sub-Nyquist rate sampling)
Spectrum Sensing

- Sparsity of occupied spectrum:
  Primary system uses

\[ r_f = D r_t \]

\( D \): DFT matrix

- Spectrum sensing via compressed sensing:

\[ y_t = \Psi D^H r_f \]

\( \Psi D^H \): Sensing matrix

\( r_f \): Sparse spectrum vector

Merit: Reduction of sampling rate
Spectrum Sensing

- Edge spectrum

\[ z_s = \Gamma D \Phi_s r_t \]

\( \Phi_s \): smoothing function

\[ \Gamma = \begin{bmatrix}
1 & 0 & \cdots & \cdots & 0 \\
-1 & 1 & \ddots & \ddots & \ddots \\
0 & \ddots & \ddots & \ddots & \ddots \\
\cdots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -1 & 1
\end{bmatrix} \]

\[ y_t = \Psi (\Gamma D \Phi_s)^{-1} z_s \]
RFID

- Each tag sends its own ID by reflecting wireless signal from reader
- Conventionally, collision avoiding protocols are used
RFID Tag Detection

- ID of tag $j$: $a_j$ ($j = 1, \ldots, n$)

- Received signal @ reader:

\[
\begin{align*}
    y &= \sum_{j \in \mathcal{P}} a_j \\
    &= [a_1, \ldots, a_n] x \\
    &= Ax
\end{align*}
\]

$\mathcal{P}$: Index set of tags in reader’s range

\[
    x_j = \begin{cases} 
        1 & j \in \mathcal{P} \\
        0 & \text{otherwise}
    \end{cases}
\]
RFID Tag Detection

- Sparsity of RFID tags:

  Number of tags in the reader’s range is much smaller than that of whole ID space (similar to CS-based MUD in the previous talk)

- Tag detection via compressed sensing:

  \[ \hat{x} = \arg\min_{x} \|x\|_1 \quad \text{subject to} \quad Ax = y \]

  \( A \) : Sensing matrix (ID matrix)
  \( x \) : Unknown sparse vector

  Merit: Reduction of detection time