Space-Shift Sampling of Graph Signals

Santiago Segarra, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro

Dept. of Electrical and Systems Engineering
University of Pennsylvania
ssegarra@seas.upenn.edu
http://www.seas.upenn.edu/~ssegarra/

ICASSP, March 23, 2016
Network science analytics

Desiderata: Process, analyze and learn from network data [Kolaczyk’09]
Network science analytics

Desiderata: Process, analyze and learn from network data [Kolaczyk’09]

Network as graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships

Interest here not in $G$ itself, but in data associated with nodes in $\mathcal{V}$

⇒ The object of study is a graph signal

Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

Online social media | Internet | Clean energy and grid analytics
Motivating examples – Graph signals

Graph SP: broaden classical SP to graph signals [Shuman etal’13]
⇒ Our view: GSP well suited to study network processes

As.: Signal properties related to topology of $G$ (e.g., smoothness)
⇒ Algorithms that fruitfully leverage this relational structure
Graph signals

Consider a graph \( G(\mathcal{V}, \mathcal{E}) \). Graph signals are mappings \( x : \mathcal{V} \rightarrow \mathbb{R} \)

\[ \Rightarrow \text{Defined on the vertices of the graph (data tied to nodes)} \]

May be represented as a vector \( x \in \mathbb{R}^N \)

\[ \Rightarrow x_n \text{ denotes the signal value at the } n\text{-th vertex in } \mathcal{V} \]

\[ \Rightarrow \text{Implicit ordering of vertices} \]

\[ x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.7 \\ 0.3 \\ \vdots \\ 0.7 \end{bmatrix} \]
To understand and analyze $\mathbf{x}$, useful to account for $G$'s structure

- Graph $G$ is endowed with a graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$
  \[ S_{ij} = 0 \text{ for } i \neq j \text{ and } (i,j) \notin \mathcal{E} \] (captures local structure in $G$)

- $\mathbf{S}$ can take nonzero values in the edges of $G$ or in its diagonal

- **Ex:** Adjacency $\mathbf{A}$, degree $\mathbf{D}$, and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices
Relevance of the graph-shift operator

» Q: Why is $S$ called shift?
Relevance of the graph-shift operator

Q: Why is $S$ called shift? A: Resemblance to time shifts

Set $S = A_{dc}$

What is $Sx$?

$$
\begin{pmatrix}
0 \\
0 \\
x_2 \\
x_3 \\
0 \\
0
\end{pmatrix}
= \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
x_2 \\
x_3 \\
0 \\
0 \\
0
\end{pmatrix}
$$
Relevance of the graph-shift operator

Q: Why is $S$ called shift? A: Resemblance to time shifts

Set $S = A_{dc}$

What is $Sx$?

$S$ will be building block for GSP algorithms

⇒ Same is true in the time domain (filters and delay)
Locality of the graph-shift operator

- **S** is a linear operator that can be computed locally at the nodes in \( \mathcal{V} \)

- Consider the graph signal \( \mathbf{y} = \mathbf{Sx} \) and node \( i \)'s neighborhood \( \mathcal{N}_i \)
  \[ \Rightarrow \text{Node } i \text{ can compute } y_i \text{ if it has access to } x_j \text{ at } j \in \mathcal{N}_i \]
  \[ y_i = \sum_{j \in \mathcal{N}_i} S_{ij} x_j, \quad i \in \mathcal{V} \]

- Recall \( S_{ij} \neq 0 \) only if \( i = j \) or \( (j, i) \in \mathcal{E} \)

- If \( \mathbf{y} = \mathbf{S}^2 \mathbf{x} \Rightarrow y_i \text{ found using values } x_j \text{ within 2 hops} \)
Graph Fourier transform (GFT)

- As.: $S$ related to generation (description) of the signals of interest
  $\Rightarrow$ Spectrum of $S = \mathbf{V}\Lambda\mathbf{V}^{-1}$ will be especially useful to analyze $\mathbf{x}$

- The Graph Fourier Transform (GFT) of $\mathbf{x}$ is defined as
  $$\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$$

- While the inverse GFT (iGFT) of $\tilde{\mathbf{x}}$ is defined as
  $$\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$$
  $\Rightarrow$ Eigenvectors $V = [v_1, ..., v_N]$ are the frequency basis (atoms)

- Ex: For the directed cycle (temporal signal) $\Rightarrow$ GFT $\equiv$ DFT
  $\Rightarrow$ DFT matrix diagonalizes circulant matrices like $S = A_{dc}$
Sampling of graph signals

- **Sampling** is a cornerstone inverse problem in classical SP
  - How to find $x \in \mathbb{R}^N$ using $P < N$ observations?

- Our focus on bandlimited signals, but other models possible
  - $\tilde{x} = V^{-1}x$ sparse
    - $x = \sum_{k \in K} \tilde{x}_k v_k$, with $|K| = K < N$
  - $S$ involved in generation of $x$
    - Agnostic to the particular form of $S$

- Two sampling schemes were introduced in the literature
  - Selection [Anis14, Chen15, Tsitsvero15, Puy15, Wang15]
  - Aggregation [Marques15, Segarra15]

- We combine both to create a hybrid scheme
  - Space-shift sampling
Sampling of graph signals

- **Sampling** is a cornerstone inverse problem in classical SP
  - How to find $x \in \mathbb{R}^N$ using $P < N$ observations?

- Our focus on **bandlimited** signals, but other models possible

  $\tilde{x} = V^{-1}x$ sparse
  
  $x = \sum_{k \in \mathcal{K}} \tilde{x}_k v_k$, with $|\mathcal{K}| = K < N$
  
  $S$ involved in generation of $x$

  Agnostic to the particular form of $S$

- Two sampling schemes were introduced in the literature
  - **Selection** [Anis14, Chen15, Tsitsvero15, Puy15, Wang15]
  - **Aggregation** [Marques15, Segarra15]

- We combine both to create a **hybrid** scheme ⇒ **Space-shift** sampling
There are two ways of interpreting sampling of time signals:

1. We can either freeze the signal and sample values at different times.
2. We can fix a point (present) and sample the evolution of the signal.

Both strategies coincide for time signals but not for general graphs.

⇒ Give rise to selection and aggregation sampling.
Selection sampling

- **Intuitive extension** of sampling to graph signals
  - Select a subset of the nodes and observe the signal value
  - Let $C \in \{0, 1\}^{P \times N}$ be a selection matrix ($P$ rows of $I_N$)
    $$\tilde{x} = Cx$$

- Goal: recover $x$ based on $\tilde{x}$
  - Assume that the support of $\mathcal{K}$ is known (w.l.o.g. $\mathcal{K} = \{k\}_{k=1}^K$)
  - Since $\tilde{x}_k = 0$ for $k > K$, define $\tilde{x}_K$ as (with $E_K := [e_1, \ldots, e_K]$)
    $$\tilde{x}_K := [\tilde{x}_1, \ldots, \tilde{x}_K]^T = E_K^T \tilde{x}$$

- Use $\tilde{x}$ to find $\tilde{x}_K$, and then recover $x$ as
  $$x = V_K \tilde{x}_K = V_K (CV_K)^{-1} \tilde{x}$$
Aggregation sampling

- Idea: incorporating $S$ to the **sampling** procedure
- Consider shifted (aggregated) signals $y^{(l)} = S^l x$
  - $y^{(l)} = Sy^{(l-1)}$ ⇒ they can be found sequentially
  - $S_{ij} = 0$ if $i \notin N_j$ ⇒ Only local exchanges are required
- Form signal $y_i = [y_i^{(0)}, y_i^{(1)}, ..., y_i^{(N-1)}]^T$

![Graph examples](image-url)
Aggregation sampling

- Idea: incorporating $S$ to the sampling procedure

- Consider shifted (aggregated) signals $y^{(l)} = S'x$
  \[ y^{(l)} = Sy^{(l-1)} \Rightarrow \text{they can be found sequentially} \]
  \[ S_{ij} = 0 \text{ if } i \not\in \mathcal{N}_j \Rightarrow \text{Only local exchanges are required} \]

- Form signal $y_i = [y_i^{(0)}, y_i^{(1)}, \ldots, y_i^{(N-1)}]^T$

- Sampled signal $\bar{y}_i = Cy_i \Rightarrow \bar{y}_i$ can be obtained locally by node $i$

- Goal: recover $x$ based on $\bar{y}_i \Rightarrow$ Find $\tilde{x}_K$ and recover $x$ as $x = V_K\tilde{x}_K$

- Define $\bar{u}_i := V^T_K e_i$ and the Vandermonde matrix $\Psi$ s.t. $\psi_{kl} = \lambda_k^{l-1}$

\[
x = V_K\tilde{x}_K = V_K\text{diag}^{-1}(\bar{u}_i)(C\Psi^T E_K)^{-1}\bar{y}_i
\]
Space-shift sampling

- **Hybrid** scheme combining selection and aggregation sampling
  - **Selection** ⇒ sampling the dimension of nodes
  - **Aggregation** ⇒ sampling the dimension of shift applications
  - **Space-shift** ⇒ sampling the 2D space spanned by the above

**Selection:** 4 nodes, 1 sample  
**Space-shift:** 2 nodes, 2 samples  
**Aggregat.:** 1 node, 4 samples

Define the matrix $Y := [y^{(0)}, \ldots, y^{(N-1)}] = [x, Sx, \ldots, S^{N-1}x]$
  - **Selection** samples the first column of $Y$
  - **Aggregation** samples the $i$-th row of $Y$
  - **Space-shift** samples the whole matrix $Y$
Define the matrix \( \bar{\Upsilon} := [\text{diag}(\bar{u}_1), \ldots, \text{diag}(\bar{u}_N)]^T \) and \( \gamma := \text{vec}(Y^T) \).

Let \( C \in \{0, 1\}^{K \times N^2} \) be a selection matrix \( \Rightarrow \bar{\gamma} = C\gamma \).

### Recovery of space-shift sampling

Signal \( x \) can be recovered from \( K \) space-shift samples as

\[
x = V_K \tilde{x}_K = V_K (C(I \otimes (\Psi E_K)) \bar{\Upsilon})^{-1} \bar{\gamma}
\]

provided that the inverse exists.

If \( C(I \otimes (\Psi E_K)) \bar{\Upsilon} \) is not invertible \( \Rightarrow \) additional samples required.

In general, invertibility is not easy to check a priori \( \Rightarrow \) Selection.

For some forms of \( C \), invertibility can be ensured \( \Rightarrow \) Aggregation.
Space-shift sampling: Discussion

Appealing features of Space-shift Sampling

- Natural scheme when $S$ encodes an underlying network dynamics
- Appropriate for inference based on a few access nodes
- Includes cases where node observes neighboring signal values
- Consistent with sampling in DSP
- Recovery error is reduced by combining selection and aggregation

Extensions

- Sampling in the presence of noise
  - Design of optimal sampling schemes
  - Aggregating nodes and $C$ play a key role in minimizing error
- Unknown frequency support ⇒ Sparse recovery
We have assumed the first $K$ frequencies of $x$ to be active.

A more challenging problem $\Rightarrow$ Frequency support $\mathcal{K}$ is unknown.

Defining $\Upsilon := [\text{diag}(u_1), \ldots, \text{diag}(u_N)]^T$, reformulate the problem

$$\tilde{x}^* = \arg \min_{\tilde{x}} \| \tilde{x} \|_0 \quad \text{s.t.} \quad \tilde{\gamma} = C(I \otimes \Psi) \Upsilon \tilde{x}$$

Identifiable when $C(I \otimes \Psi) \Upsilon$ is full spark and has at least $2K$ rows.

For some $C$, full-spark can be assessed by inspecting $\{\lambda_i\}_{i=1}^N$ and $V$.

Computationally, the $\ell_0$ norm renders the optimization non-convex $\Rightarrow$ Convexify it by replacing the $\ell_0$ with an $\ell_1$ norm.

Recoverability based on the coherence and the RIP of $C(I \otimes \Psi) \Upsilon$.

With noise, the constraint can be replaced by $\| \tilde{\gamma} - C(I \otimes \Psi) \Upsilon \tilde{x} \|_2^2 < \epsilon$. 
Comparing sampling schemes

- **62 economic sectors** in USA + 2 synthetic sectors
  - Graph: average flows of production in 2007-2010, \( S = A \)
  - Signal \( x \): Production of sectors in 2011 (approx. bandlimited)

- Comparable **minimum errors**
- **Median** errors reduced via space-shift sampling

<table>
<thead>
<tr>
<th>Sampling strategy</th>
<th>Error</th>
<th>Min.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>([x]_i)</td>
<td>([x]_j)</td>
<td>([x]_k)</td>
<td>([x]_l)</td>
</tr>
<tr>
<td>([x]_i)</td>
<td>([Sx]_i)</td>
<td>([S^2x]_i)</td>
<td>([S^3x]_i)</td>
</tr>
<tr>
<td>([Sx]_i)</td>
<td>([Sx]_j)</td>
<td>([Sx]_k)</td>
<td>([Sx]_l)</td>
</tr>
<tr>
<td>([S^2x]_i)</td>
<td>([S^2x]_j)</td>
<td>([S^2x]_k)</td>
<td>([S^2x]_l)</td>
</tr>
<tr>
<td>([x]_i)</td>
<td>([Sx]_i)</td>
<td>([x]_j)</td>
<td>([Sx]_j)</td>
</tr>
</tbody>
</table>
Signals of bandwidth $K \in \{1, 2, \ldots, 5\}$ on the economic network

$\Rightarrow$ Value of $K$ known but not the specific support

Nr. of observations $M$, i.e. rows of $C$, where $M \in \{5, 10, \ldots, 40\}$

$\Rightarrow$ Chosen among values in original signal $x$ and first shift $Sx$

Solve iterative randomized version of convex relaxation

For large $M$ and small $K$

$\Rightarrow$ perfect recovery

Gradual detriment for more adverse configurations
Conclusion

- Presented basic building blocks of GSP
  \[ S = V \Lambda V^{-1}, \text{ GFT } V^{-1} \]

- Discussed differences between selection and aggregation sampling
- Selection and aggregation can be combined in space-shift sampling
  \[ \Rightarrow \text{ All of them reduce to traditional sampling in DSP} \]

- Natural scheme for network processes
  \[ \Rightarrow \text{ Appropriate for inference with few access nodes} \]

- Conditions for perfect recovery and joint support identification

- Illustrated concepts via the U.S. economic network