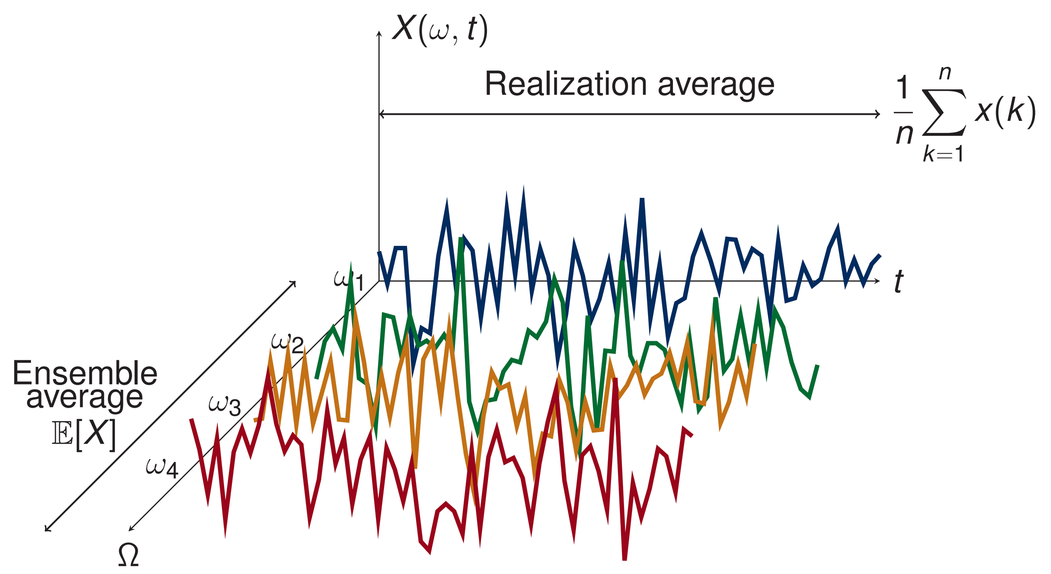
WEAK LAW OF LARGE NUMBERS FOR STATIONARY GRAPH PROCESSES

Fernando Gama[†] and Alejandro Ribeiro[†] E-mail: fgama@seas.upenn.edu

[†]Electrical & Systems Engineering, University of Pennsylvania

Stochastic Graph Processes

- Stochastic processes are essential to model random phenomena
  - $\Rightarrow$ Extract useful information from the available (noisy) data
- Stationarity $\Rightarrow$ Conditions on probability distribution of the process
  - $\Rightarrow$ Strict Sense Stationarity (SSS) $\Rightarrow$ Joint distribution
  - $\Rightarrow$ Wide Sense Stationarity (WSS) $\Rightarrow$ First, second order moments
- Ergodicity $\Rightarrow$ Realization averaging converge to ensemble averaging



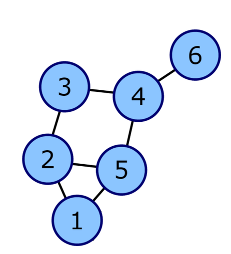
- Regular structure descriptions are insufficient for modern datasets
- Networks $\Rightarrow$ Graphs emerge as more accurate representations
- Random phenomena supported on irregular structures
- Wide sense stationary (WSS) graph processes have been defined
  - $\Rightarrow$ Power spectral density (PSD) estimation methods

Objective

- Extend notion of ergodicity to WSS graph processes
  - $\Rightarrow$ Result reminiscent of weak law of large numbers (WLLN)
- Consistent unbiased estimator by diffusing a single realization
- Optimal design of graph filter that minimizes MSE

Graph signals

- Weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with n nodes $\Rightarrow$ Irregular support
- Graph signal $\mathbf{x} \in \mathbb{R}^n$ $\Rightarrow$ Data value on each node
- Graph shift operator $\mathbf{S} \in \mathbb{R}^{n \times n}$ $\Rightarrow$ Captures local structure in $\mathcal{G}$



$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{32} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix}$$

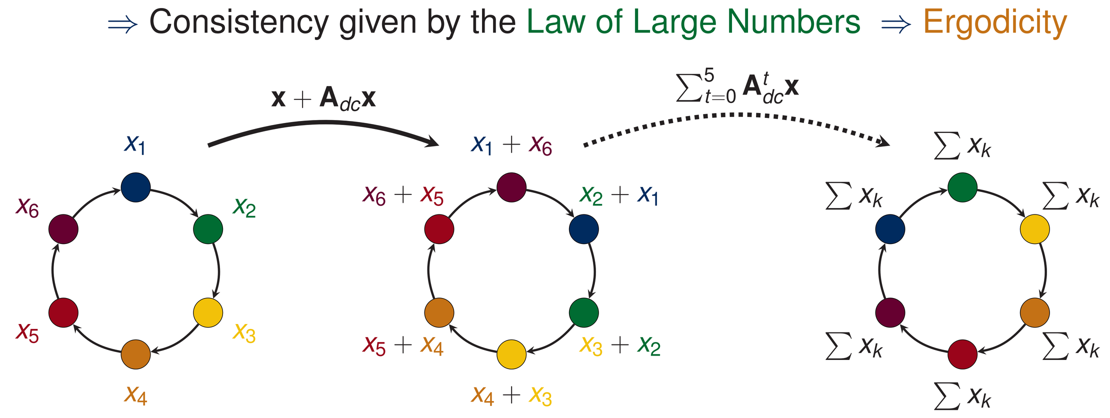
- Interaction between signal and support $\Rightarrow$ $\mathbf{Sx}$ local operation
- Examples re adjacency matrix $\mathbf{A}$ and graph Laplacian $\mathbf{L}$

Discrete-Time Signals

- Sequence of random signal values $\{x_1, \ldots, x_n\}$ $\Rightarrow \mathbb{E}[x_k] = \mu$
- Time average is consistent unbiased estimator of the mean

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k \quad \Rightarrow \quad \mathbb{E}[\hat{\mu}_n] = \mu$$

$\Rightarrow$ Consistency given by the Law of Large Numbers $\Rightarrow$ Ergodicity



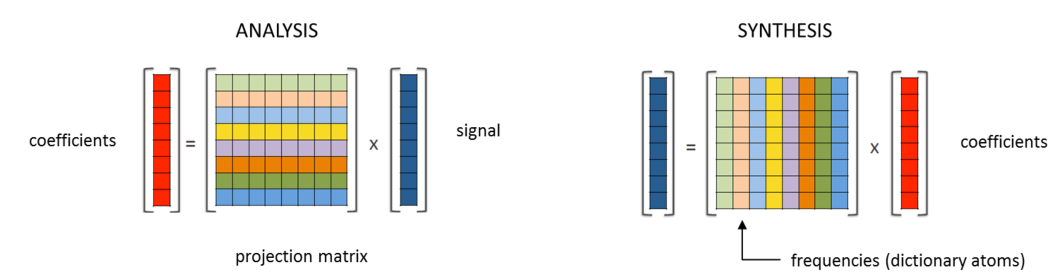
- Model as a graph signal $\Rightarrow \mathcal{G} = \mathcal{G}_{dc} \Rightarrow \mathbf{x} = [x_1, \ldots, x_n]^T \Rightarrow \mathbb{E}[\mathbf{x}] = \mu\mathbf{1}$
- Diffuse signal n times $\Rightarrow$ All nodes contain the estimator

$$\frac{1}{n}\sum_{k=1}^{n} x_k \mathbf{1} = \frac{1}{n}\sum_{\ell=0}^{n-1} \mathbf{A}_{dc}^{\ell}\mathbf{x} = c\sum_{\ell=0}^{n-1} \mathbf{S}^{\ell}\mathbf{x}$$

$\Rightarrow$ Consistent estimator obtained by diffusing a single realization

Graph Fourier Transform

- Assume graph shift operator is normal $\Rightarrow \mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Project graph signal onto eigenbasis $\Rightarrow \tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$
  - $\Rightarrow$ Defined as the graph Fourier transform (GFT)
- Linear combination of eigenvectors weighted by GFT coefficients
  - $\Rightarrow \mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} \Rightarrow$ Inverse graph Fourier transfrom (iGFT)



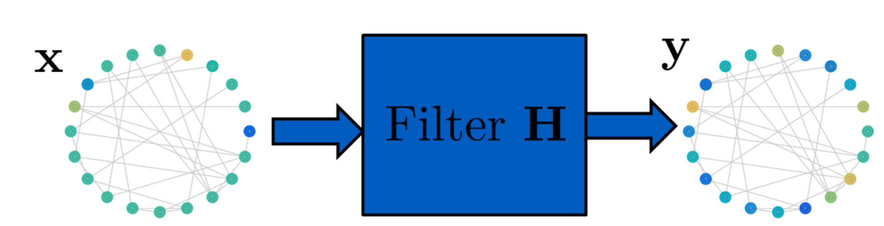
- Change of basis tailored to structure of graph
- If graph is directed cycle (discrete-time signals) $\Rightarrow$ DFT

Graph Filters

- Graph filter $\mathbf{H} : \mathbb{R}^n \to \mathbb{R}^n$ $\Rightarrow$ Map between graph signals
- Consider filters that are linear $\Rightarrow$ $\mathbf{H}$ is a $n \times n$ matrix
- Polynomial in $\mathbf{S}$ of degree $b-1$ with coefficients $\mathbf{h} = [h_0, \ldots, h_{b-1}]^T$

$$\mathbf{H} = h_0\mathbf{I} + h_1\mathbf{S} + \cdots + h_{b-1}\mathbf{S}^{b-1} = \sum_{\ell=0}^{b-1} h_\ell \mathbf{S}^\ell$$

- Linear shift-invariant graph filters (LSI-GF)
  - $\Rightarrow$ Distributed implementation $\Rightarrow$ Only up to b-hop information
- GFT of filter depends on eigenvalues of $\mathbf{S}$ $\Rightarrow \tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h} \in \mathbb{C}^n$
  - $\Rightarrow$ With $[\mathbf{\Psi}]_{k,\ell} = \lambda_k^{\ell-1} \in \mathbb{C}^{n \times b}$ Vandermonde matrix



Wide Sense Stationary Graph Processes

- Probability space $(\Omega, \mathcal{F}, \mathbb{P})$ $\Rightarrow$ Random vector $\mathbf{x} : \Omega \to \mathbb{R}^n$
  - $\Rightarrow [\mathbf{x}]_k$ random variable on each node of $\mathcal{G}$
  - $\Rightarrow$ Mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$ and covariance matrix $\mathbf{C}_x = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H]$
- WSS impose statistical structure related to underlying graph support
  - $\Rightarrow \mathbb{E}[\mathbf{x}] = \mu\mathbf{v}_m$ where $\mathbf{v}_m$ eigenvector of $\mathbf{S}$
  - $\Rightarrow \mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H \Rightarrow \mathbf{p}$: PSD $\Rightarrow \mathbf{C}_{\tilde{x}} = \text{diag}(\mathbf{p})$
  - $\Rightarrow$ Covariance matrix and GSO are simultaneously diagonalizable
- WSS graph process filtered by LSI-GF $\Rightarrow$ Output is WSS
- Definition holds for graph spectra with all different eigenvalues

The Concept of Mean

- Traditional SP $\Rightarrow$ Mean is DC (constant) component of signal
  - $\Rightarrow$ Contribution of zero-frequency $\Rightarrow$ Slowest time-varying
- GSP $\Rightarrow$ Find the slowest node-varying eigenvector $\Rightarrow \mathbf{v}_m$
- Use concept of total variation (TV) to find $\mathbf{v}_m$

$$TV(\mathbf{x}) = \sum_{k=1}^{n}\left| x_k - \sum_{\ell \in \mathcal{N}_k} \frac{w_{\ell,k}}{|\lambda_{\max}|} x_\ell \right| = \left\| \mathbf{x} - \frac{1}{|\lambda_{\max}|}\mathbf{A}^T\mathbf{x} \right\|_1$$

- Ordering $\Rightarrow \lambda_{\max}$ is real and positive for connected graphs
  - $\Rightarrow \mathbf{v}_{\max}$ is the slowest-node varying eigenvector $\Rightarrow \mathbf{v}_m = \mathbf{v}_{\max}$
  - $\Rightarrow$ TV increases as eigenvalues are located further away from $\lambda_{\max}$

$$|\lambda_k - \lambda_{\max}| < |\lambda_\ell - r| \quad \Rightarrow \quad TV(\mathbf{v}_k) < TV(\mathbf{v}_\ell)$$

- Specific case of connected graph with positive weights
  - $\Rightarrow$ Eigenvector $\mathbf{v}_{\max}$ has all positive elements
  - $\Rightarrow$ The number of zero-crossings is minimal (none)
- Order eigenvalues from slowest to fastest $\Rightarrow \lambda_1 = \lambda_{\max}, \lambda_2, \ldots, \lambda_n$
- Discrete-time signals $\Rightarrow$ n values used to estimate a scalar μ
- Graph signals $\Rightarrow$ value of n-hop diffusion used to estimate a scalar μ

Unbiased Diffusion Estimator

- Discrete-time estimator $\Rightarrow$ Graph $\mathcal{G} = \mathcal{G}_{dc}$ directed cycle

$$\frac{1}{n}\sum_{k=1}^{n} x_k\mathbf{1} = \frac{1}{n}\sum_{\ell=0}^{n-1}\mathbf{A}_{dc}^{\ell}\mathbf{x} = \frac{1}{n}\sum_{\ell=0}^{n-1}\mathbf{S}^{\ell}\mathbf{x}$$

- Extend to GSP $\Rightarrow$ Estimate mean from diffusing single realization

$$\hat{\mu}_n = \frac{1}{\sum_{\ell=0}^{n-1}\lambda_1^{\ell}}\sum_{\ell=0}^{n-1}\mathbf{S}^{\ell}\mathbf{x}$$

  - $\Rightarrow$ Scaling constant (unbiased) $\Rightarrow$ Eigenvalue associated to mean
- Unbiased $\mathbb{E}[\hat{\mu}_n] = \mu$ $\Rightarrow$ Covariance matrix $\mathbf{C}_{\hat{\mu}} = \mathbf{V}\text{diag}(\mathbf{q})\mathbf{V}^H$
- Estimator $\hat{\mu}_n$ is also a WSS graph process $\Rightarrow$ $\mathbf{q}$: PSD of estimator

$$q_k = p_k\frac{\left|\sum_{\ell=0}^{n-1}\lambda_k^{\ell}\right|^2}{\left|\sum_{\ell=0}^{n-1}\lambda_1^{\ell}\right|^2}, \quad k = 1, \ldots, n$$

- $\mathbf{p}_1$ is associated to the power of the mean
- Since $\Re\{\lambda_1\} > \Re\{\lambda_k\}, |\lambda_1| \geq |\lambda_k|$ $\Rightarrow$ LPF that lowers p_k, $k = 2, \ldots, n$

Weak Law of Large Numbers for WSS Graph Processes

- Prove consistency of unbiased diffusion estimator
  - $\Rightarrow$ Convergence result like the weak law of large numbers
- Bound error of estimating mean at node ℓ

$$\mathbb{P}\left(|[\hat{\mu}_n - \mu]_\ell| > \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} q_k |v_{\ell,k}|$$

  - $\Rightarrow$ Depends on $\mathbf{q}$ and on rows of $\mathbf{V}$ (also orthonormal)
- Behavior of q_k $\Rightarrow$ Assume $|\lambda_k|/\lambda_1 = o(n^{-\delta/2n})$, $\delta > 0$ or $\lambda_1 = 1$

$$q_1 = p_1 \quad , \quad q_k = o(n^{-\delta}) \, , \; k = 2, \ldots, n$$

  - $\Rightarrow$ Directed cycle and Erdős-Rényi graphs satisfy this condition

Weak Law of Large Numbers for WSS graph processes

$$\min_{\ell=1,\ldots,n} \mathbb{P}\left(|[\hat{\mu}_n - \mu]_\ell| > \epsilon\right) \leq \frac{p_1}{n\epsilon^2} + o(n^{-\delta})$$

- Depends on the variance p_1 of mean component
- Graphs satisfying certain spectral conditions $\Rightarrow$ Error $\to 0$ as $n \to \infty$
- Ergodicity $\Rightarrow$ Estimate mean of process from single realization

Optimal Unbiased Graph Filter Estimator

- Diffusion estimator is a LSI graph filter with constant taps

$$h_\ell = \left(\sum_{t=0}^{n-1}\lambda_1^{t}\right)^{-1}, \quad \ell = 0, 1, \ldots, n-1$$

- Consider a general LSI graph filter $\Rightarrow$ Unbiased estimator

$$\left(\sum_{\ell=0}^{n-1} h_\ell \mathbf{S}^\ell \mathbf{x}\right) \quad \Rightarrow \quad \mathbf{z}_n = \frac{1}{\sum_{\ell=0}^{n-1} h_\ell\lambda_1^\ell}\sum_{\ell=0}^{n-1} h_\ell\mathbf{S}^\ell\mathbf{x}$$

- $\mathbf{z}_n$ is a WSS graph process $\Rightarrow$ Covariance matrix $\mathbf{C}_z = \mathbf{V}\text{diag}(\mathbf{r})\mathbf{V}^H$

$$r_k = p_k\frac{\left|\sum_{\ell=0}^{n-1} h_\ell\lambda_k^\ell\right|^2}{\left|\sum_{\ell=0}^{n-1} h_\ell\lambda_1^\ell\right|^2}$$

- Recall the GFT of filter taps $\Rightarrow \tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$ (depends on eigenvalues)

$$\tilde{h}_k = \sum_{\ell=0}^{n-1} h_\ell\lambda_k^\ell \quad \Rightarrow \quad r_k = p_k\frac{|\tilde{h}_k|^2}{|\tilde{h}_1|^2}, \quad k = 1, \ldots, n$$

- Select filter taps that minimize mean squared error (MSE)

$$\min_{\tilde{\mathbf{h}}} \text{tr}[\mathbf{C}_z] \quad \Rightarrow \quad \tilde{h}_1 \neq 0, \tilde{h}_k = 0 \, , \quad k = 2, \ldots, n$$

- The PSD of the optimal unbiased graph filter estimator

$$r_1 = p_1 \, , \; r_k = 0 \, , \quad k = 2, \ldots, n$$

  - $\Rightarrow$ Attenuates all frequencies except for the DC component

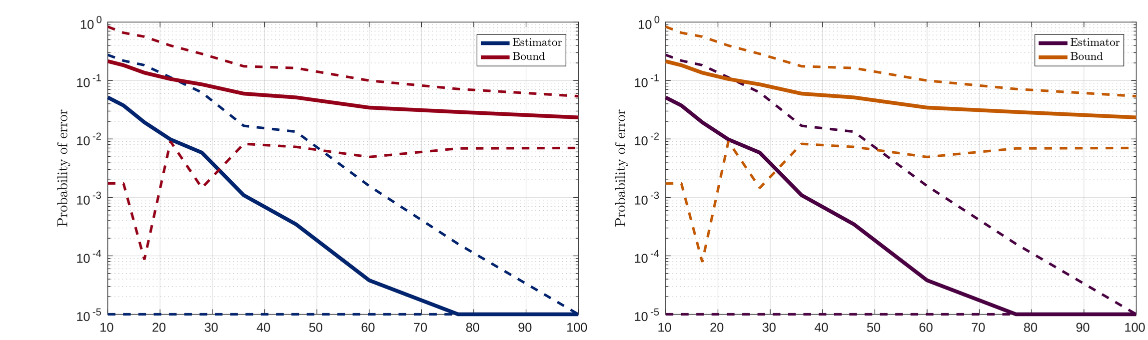
Consistency for optimal unbiased graph filter estimator

$$\min_{\ell=1,\ldots,n} \mathbb{P}\left(|[\mathbf{z}_n - \mu]_\ell| > \epsilon\right) \leq \frac{p_1}{n\epsilon^2}$$

- This result holds for any underlying graph support
- PSD of $\mathbf{z}_n$ is nonzero only for component associated to mean
  - $\Rightarrow$ Perfect low pass filter applied to the realization
  - $\Rightarrow$ No need for other eigenvalues to satisfy any conditions

Erdős-Rényi Graphs

- Simulate 100 ER graphs for each n with $p = 0.2$
- Generate realization of WSS process $\Rightarrow$ Compute bound
- Estimate probability of error out of 10,000 realizations

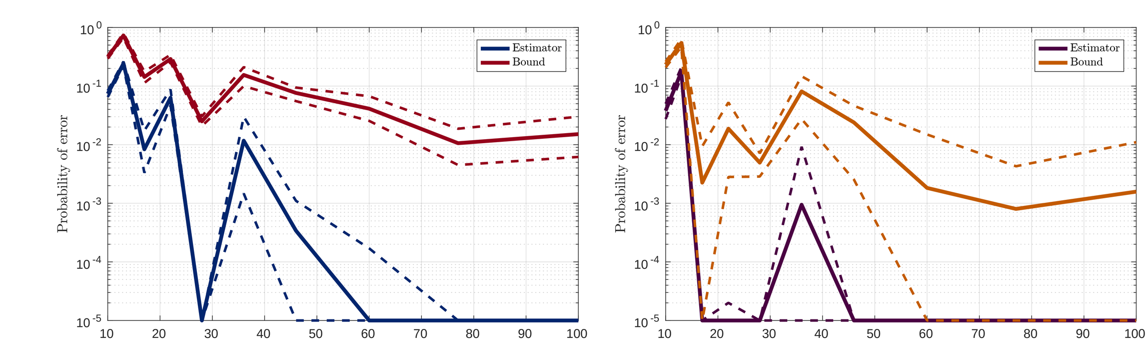


(a) Diffusion Estimator (b) Optimal Estimator

- ER satisfies conditions $\Rightarrow$ Both estimators yield same result

Covariance graph

- Simulate 100 covariance graphs for each n $\Rightarrow$ Gaussian r. v.



(a) Diffusion Estimator (b) Optimal Estimator

- Bound and probability of error for optimal estimator are lower

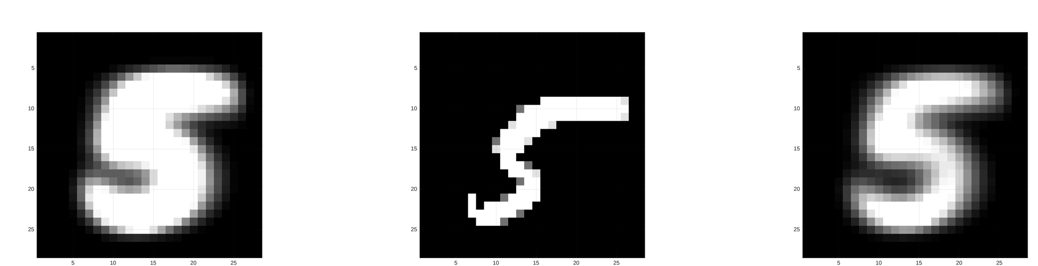
MNIST Handwritten digits

- Build covariance graph from a set of 5,000 training samples
- Sample mean obtained from the training set $\{\mathbf{x}_i\}_{i=1}^N$

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \approx \mathbb{E}[\mathbf{x}]$$

- Single realization $\mathbf{x}$ $\Rightarrow$ Compute unbiased diffusion estimator

$$\frac{1}{\sum_{\ell=0}^{n-1}\lambda_1^\ell}\sum_{\ell=0}^{n-1}\mathbf{S}^\ell\mathbf{x} \approx \mathbb{E}[\mathbf{x}]$$



(a) Sample average (b) Single realization (c) Diffusion estimator

- Consider the sample average to be the true mean
- Diffusing single image $\Rightarrow$ Resulting image respresents true mean

Conclusions and future work

- Obtained consistent unbiased estimator of the mean
  - $\Rightarrow$ Reminiscent of weak law of large numbers
- Based on diffusion of a single realization $\Rightarrow$ Ergodicity
  - $\Rightarrow$ Graph ergodicity $\Rightarrow$ Information about $\mathbb{E}[\mathbf{x}]$ is in the graph
- Designed optimal graph filter that minimizes MSE
- Obtain a Strong Law of Large Numbers result
- Characterize ergodicity for all moments $\Rightarrow$ Strict Sense Stationarity

References

- D. I. Shuman et al., "The emerging field of signal processing on graphs," *IEEE SPM*, May 2013.
- A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency Analysis," *IEEE TSP*, June 2014.
- B. Girault et al., "Translation and stationarity for graph signals," Research Report RR-8719, École Normale Supérieure de Lyon, Inria Rhône-Alpes, Apr. 2015.
- A. G. Marques et al., "Stationary graph processes and spectral estimation," *arXiv:1603.04667v1 [cs.SY]*, 14 March 2016.
- N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *arXiv:1601.02522v4 [cs.DS]*, 8 July 2016.