ASP Project : Adaptive Least Mean Squares Estimation of Graph Signals

Anoushka Vyas
20171057

About

The aim of this project is to propose a least mean squares (LMS) strategy for adaptive estimation of signals defined over graphs. Assuming the graph signal to be band-limited, over a known bandwidth, the method enables reconstruction, with guaranteed performance in terms of mean-square error, and tracking from a limited number of observations over a subset of vertices.

Furthermore, to cope with the case where the bandwidth is not known beforehand, we propose a method that performs a sparse online estimation of the signal support in the (graph) frequency domain, which enables online adaptation of the graph sampling strategy.

Variable Definitions

- N is number of vertices.
- F is the set of vertices which have non zero signal value in frequency domain.
- The signal y[n] is the observed signal.
- The signal x0 is the original bandlimited signal.
- The signals s0 and s[n] are GFT of x0 and x[n] respectively.
- The set S contains the vertices used to reconstruct the signal.
- D (NxN matrix) is the vertex-limiting operator, which takes nonzero values only in the set S.
- B is $\mathbf{U}\boldsymbol{\Sigma}_{\mathcal{F}}\mathbf{U}^H$ and $\boldsymbol{\Sigma}_{\mathcal{F}}$ is a diagonal matrix defined as $\boldsymbol{\Sigma}_{\mathcal{F}} = \mathrm{diag}\{\mathbf{1}_{\mathcal{F}}\}$.
- Uf is a matrix with |F| eigenvectors of the Laplacian matrix and mu is the learning rate.

Model definition

- The signal is initially assumed to be perfectly band-limited, i.e. its spectral content is different from zero only on a limited set of frequencies F.
- Let us consider partial observations of signal x0 , i.e. observations over only a subset of nodes. The observed signal at time n can be expressed as:

$$\boldsymbol{y}[n] = \mathbf{D}\left(\boldsymbol{x}_0 + \boldsymbol{v}[n]\right) = \mathbf{D}\mathbf{B}\boldsymbol{x}_0 + \mathbf{D}\boldsymbol{v}[n]$$

v[n] is a zero-mean, additive noise with covariance matrix Cv . The second equality comes from the bandlimited assumption, i.e. Bx0 = x0.

Important properties

- **Theorem 1**: There is a vector x, perfectly localized over both vertex set S and frequency set F (i.e. x ∈ B ∩ D) if and only if the operator BDB (or DBD) has an eigenvalue equal to one; in such a case, x is an eigenvector of BDB associated to the unitary eigenvalue.
- **Theorem 2:** Any band-limited signal x0 can be reconstructed from its samples taken in the set S, if and only if ||Dbar B||2 < 1, i.e. if the matrix B Dbar B does not have any eigenvector that is perfectly localized on Sbar and bandlimited on F . Here Dbar= I-D.
- **Theorem 3:** Any bounded initial condition, the LMS strategy defined before asymptotically converges in the mean-square error sense if the sampling operator D and the step-size μ are chosen to satisfy Theorem 2 and $$0 < \mu < \frac{2}{\lambda_{\max}\left(\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}\right)}.$$

  Furthermore, it follows that, for sufficiently small step-sizes: $$\lim_{n\to\infty} \sup_{n} \mathbf{E}\|\widehat{s}[n]\|^2 = O(\mu).$$

  Note: We say that vector x is perfectly localised over subset S if Dx=x and over frequency subset F if Bx=x.

Sampling Strategies

The properties of the LMS algorithm strongly depend on the choice of the sampling set S, i.e. on the vertex limiting operator D. The sampling strategy must be carefully designed in order to:

- enable reconstruction of the signal;
- guarantee stability of the algorithm; and
- impose a desired mean-square error at convergence.

Both the number of samples and their location is fundamental for the performance of the algorithm.

Three algorithms were used to find the samples to be taken for reconstruction algorithm:

- Minimum Mean square deviation
- Maximization of determinant
- Maximization of the minimum eigenvalue

Minimization of MSD

The method iteratively selects the samples from the graph that lead to the largest reduction in terms of steady state MSD.

$$\mathbf{G} = \mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{C}_{v}\mathbf{D}\mathbf{U}_{\mathcal{F}}$$

$$\mathbf{Q} = (\mathbf{I} - \mu\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}) \otimes (\mathbf{I} - \mu\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}).$$

Sampling strategy 1: Minimization of MSD

Input Data : M, the number of samples.
Output Data : $\mathcal{S}$, the sampling set.
Function : initialize $\mathcal{S} \equiv \emptyset$
while $|\mathcal{S}| < M$
$s = \arg\min_{j} \operatorname{vec}(\mathbf{G}(\mathbf{D}_{\mathcal{S}\cup\{j\}}))^T(\mathbf{I} - \mathbf{Q}(\mathbf{D}_{\mathcal{S}\cup\{j\}}))^{\dagger}\operatorname{vec}(\mathbf{I})$;
$\mathcal{S} \leftarrow \mathcal{S} \cup \{s\}$;
end

Maximization of Determinant $\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}$

The rationale underlying this strategy is to design a well suited basis for the graph signal that we want to estimate. This criterion coincides with the maximization of the the pseudo-determinant of the matrix $\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}$ i.e. the product of all non zeros eigenvalues.

Sampling strategy 2: Maximization of $\left|\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}\right|_{+}$

Input Data : M, the number of samples.
Output Data : $\mathcal{S}$, the sampling set.
Function : initialize $\mathcal{S} \equiv \emptyset$
while $|\mathcal{S}| < M$
$s = \arg\max_{j} \left|\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}_{\mathcal{S}\cup\{j\}}\mathbf{U}_{\mathcal{F}}\right|_{+};$
$\mathcal{S} \leftarrow \mathcal{S} \cup \{s\};$
end

Maximization of $\overline{\lambda_{\min}^{+}\left(\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}\right)}$

This algorithm can be formulated as the maximization of the minimum nonzero eigenvalue of the matrix $\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}$.

Sampling strategy 2: Maximization of $\left|\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}\mathbf{U}_{\mathcal{F}}\right|_{+}$

Input Data : M, the number of samples.

Output Data : $\mathcal{S}$, the sampling set.

Function : initialize $\mathcal{S} \equiv \emptyset$

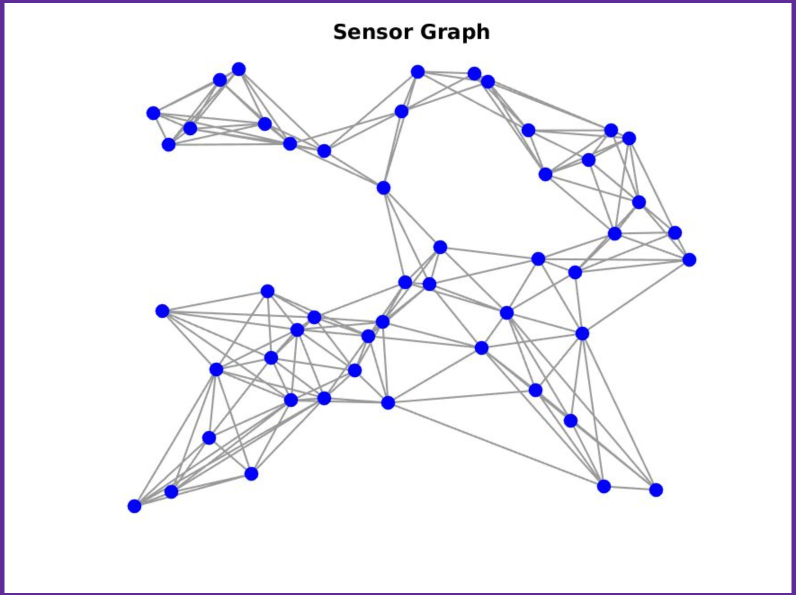
while $|\mathcal{S}| < M$

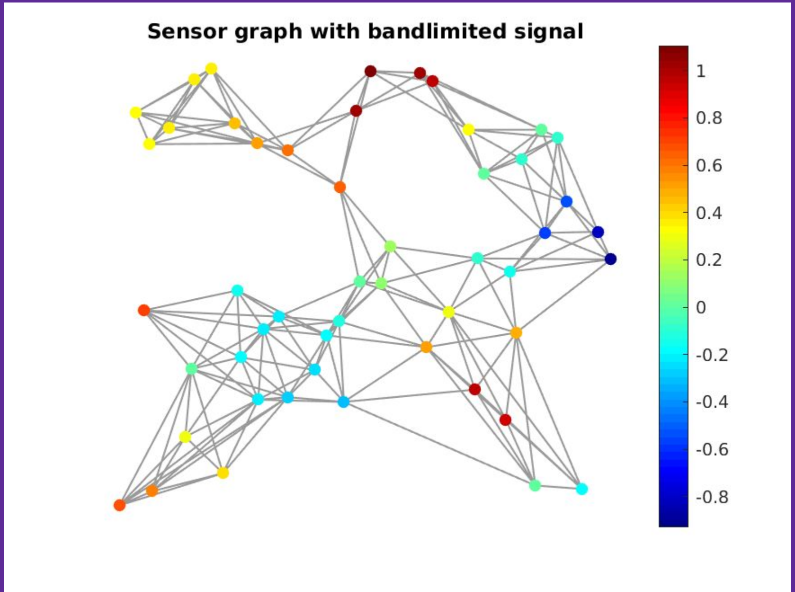
$s = \arg\max_{j} \left|\mathbf{U}_{\mathcal{F}}^{H}\mathbf{D}_{\mathcal{S}\cup\{j\}}\mathbf{U}_{\mathcal{F}}\right|_{+};$

$\mathcal{S} \leftarrow \mathcal{S} \cup \{s\};$

end

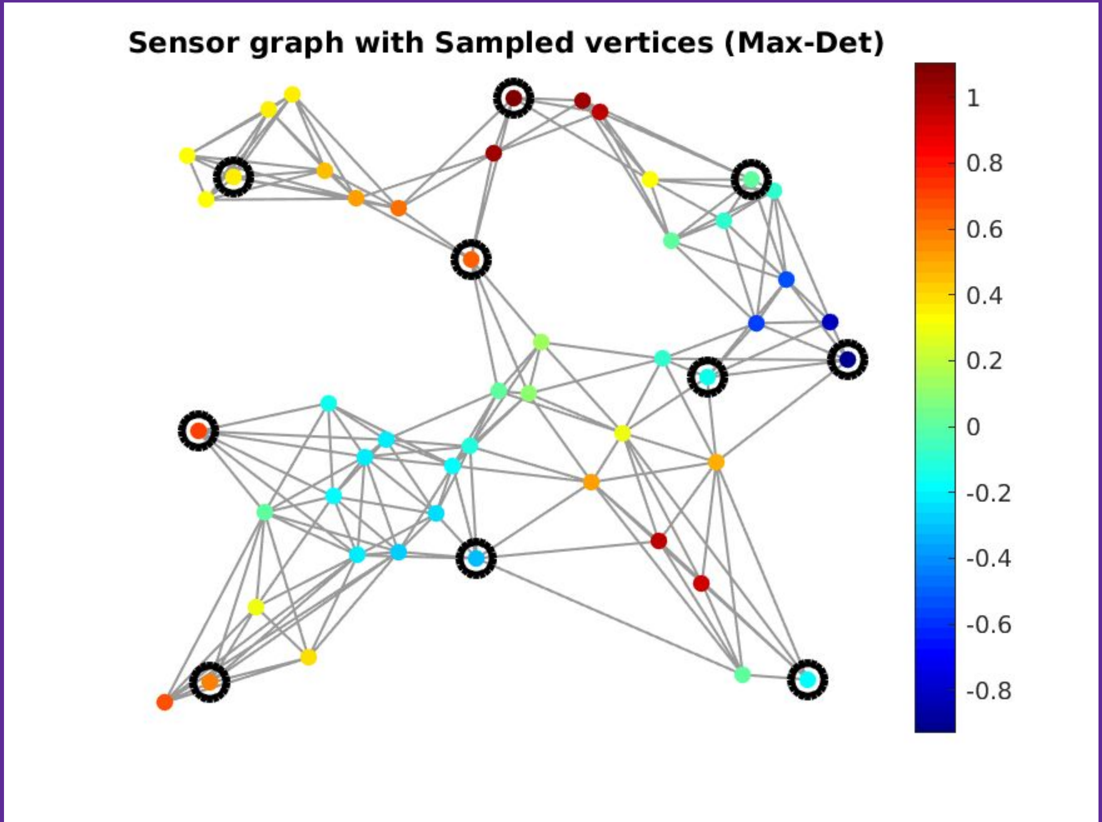
Graph and Graph Signal





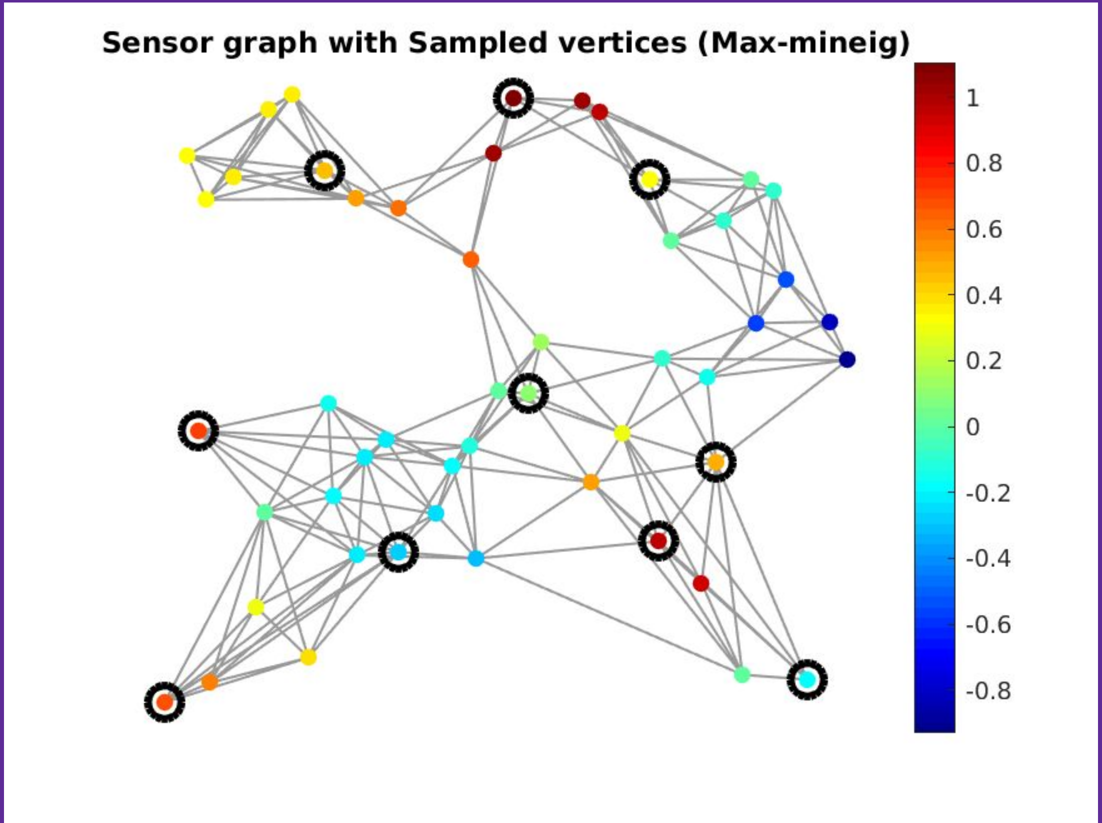
Bandwidth is 10, Nodes=50, Learning rate=0.5

Results of sampling



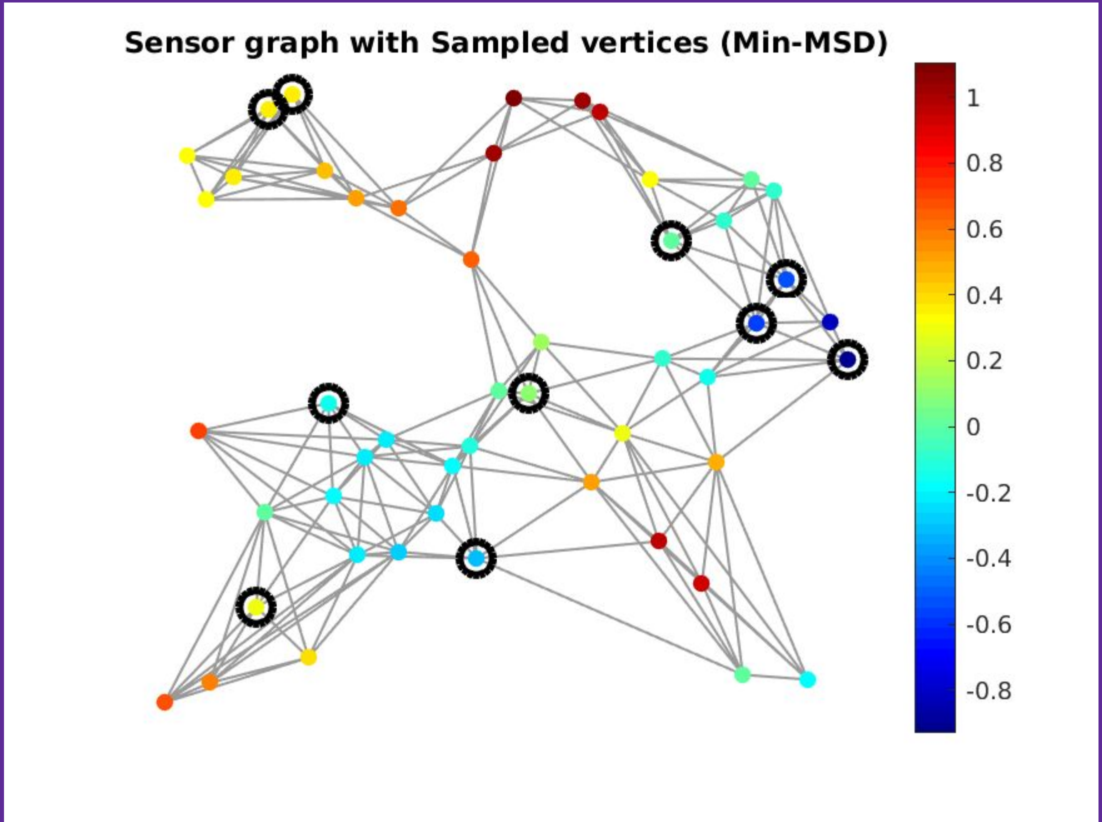
Sampled set of vertices for Max-Det Algorithm
3
48
24
40
50
44
28
5
6
25

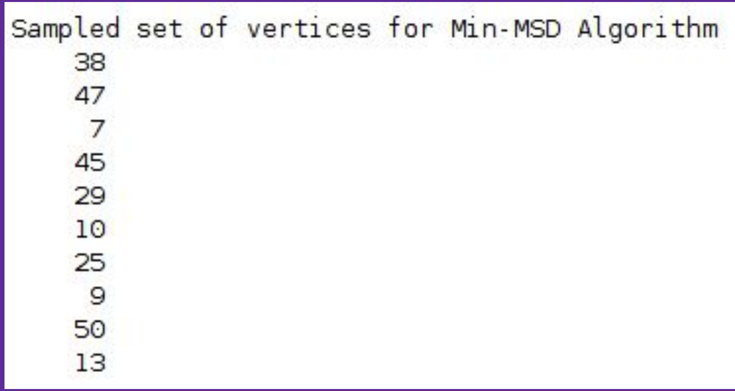
Results of sampling




```
Sampled set of vertices for Max-mineig Algorithm
    12
    19
    48
     1
    28
    35
    41
    36
    29
     3
```

Results of sampling




Sampled set of vertices for Min-MSD Algorithm
38
47
7
45
29
10
25
9
50
13

LMS Strategy

- Following an LMS approach, the optimal estimate for x0 can be found as the vector that solves the following optimization problem:

$$\min_{\boldsymbol{x}} \ \mathbf{E}\,\|\boldsymbol{y}[n] - \mathbf{D}\mathbf{B}\boldsymbol{x}\|^2$$
$$\text{s.t.} \quad \mathbf{B}\boldsymbol{x} = \boldsymbol{x},$$

- For stationary y[n], the optimal solution is given by the vector x̂ that satisfies the normal equations:

$$\mathbf{BDB}\,\hat{\boldsymbol{x}} = \mathbf{BD}\,\mathbf{E}\{\boldsymbol{y}[n]\}.$$

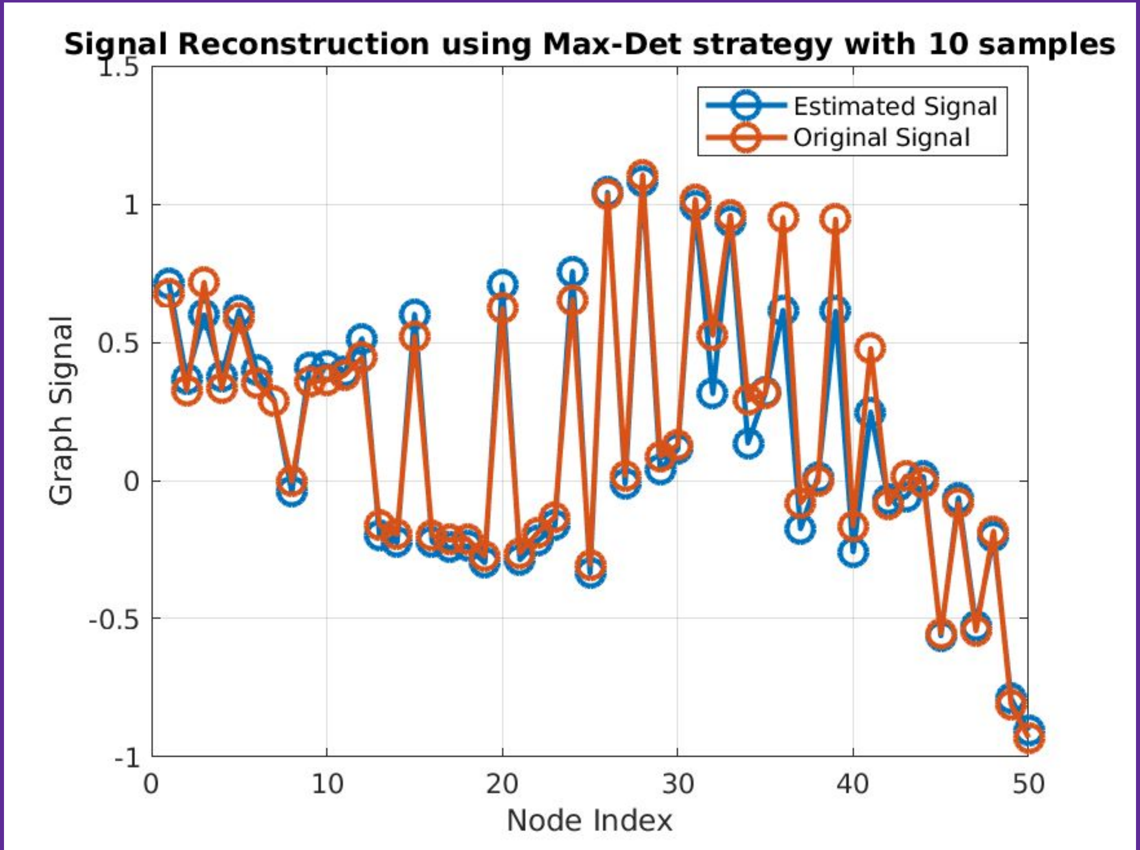
- But we may not have the statistics of y[n] beforehand so adaptive methods are used like steepest descent method like the one defined below.

Algorithm 1: LMS algorithm for graph signals

Start with $\boldsymbol{x}[0] \in \mathcal{B}_{\mathcal{F}}$ chosen at random. Given a sufficiently small step-size $\mu > 0$, for each time $n > 0$, repeat:

$$\boldsymbol{x}[n+1] = \boldsymbol{x}[n] + \mu\,\mathbf{BD}\,(\boldsymbol{y}[n] - \boldsymbol{x}[n]) \qquad (12)$$

Result of LMS algorithm



- Learning rate = 0.5
- Iterations = 100
- Number of samples = 10
- Sampling strategy = Max-Det
- Bandwidth = 10
- Noise is zero mean with diagonal covariance matrix with elements uniformly random between 0 and 0.01.

Observation: Reconstruction is possible with a little error at few node indices.

Mean-Square Analysis

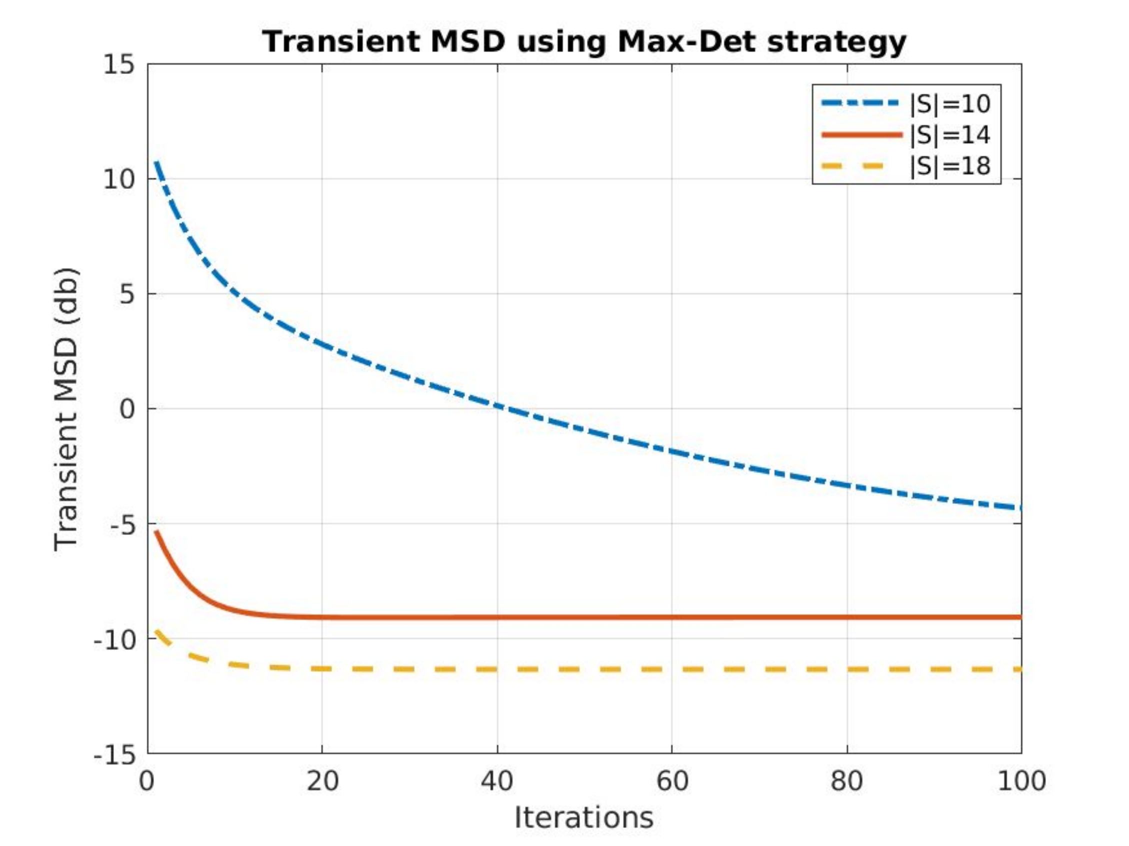
- We view the estimates x[n] as realizations of a random process and analyze the performance of the LMS algorithm in mean-square sense.
- Let $\widetilde{\boldsymbol{x}}[n] = \boldsymbol{x}[n] - \boldsymbol{x}_0$ be the error vector at time n. Thus, the relation is: $\widetilde{\boldsymbol{x}}[n+1] = (\mathbf{I} - \mu\mathbf{BDB})\,\widetilde{\boldsymbol{x}}[n] + \mu\,\mathbf{BD}\boldsymbol{v}[n].$
- Taking the graph fourier transform and solving further: $\mathbf{E}\|\widehat{s}[n+1]\|^2_{\varphi} = \mathbf{E}\|\widehat{s}[n]\|^2_{\mathbf{Q}\varphi} + \mu^2 \mathrm{vec}(\mathbf{G})^T\varphi$ where Q and G are defined before and $\|\widehat{s}[n]\|^2_{\mathbf{\Phi}} = \widehat{s}[n]^H\mathbf{\Phi}\widehat{s}[n].$. Here $\mathbf{\Phi} \in \mathbb{C}^{|\mathcal{F}|\times|\mathcal{F}|}$ is any Hermitian non-negative matrix and where $\varphi = \mathrm{vec}(\mathbf{\Phi})$, where the notation vec(·) stacks the columns of Φ on top of each other.

Steady-state Analysis

- Taking the limit as n tends to infinity we obtain: $\lim_{n\to\infty} \mathbf{E}\|\widehat{s}[n]\|^2_{(\mathbf{I}-\mathbf{Q})\varphi} = \mu^2\mathrm{vec}(\mathbf{G})^T\varphi.$
- Selecting vec(ϕ) = pinv(I-Q) vec(I) we get:

$$\begin{aligned}\mathrm{MSD} = \lim_{n\to\infty} \mathbf{E}\|\widetilde{\boldsymbol{x}}[n]\|^2 &= \lim_{n\to\infty} \mathbf{E}\|\widehat{s}[n]\|^2 \\ &= \mu^2\mathrm{vec}(\mathbf{G})^T(\mathbf{I}-\mathbf{Q})^{-1}\mathrm{vec}(\mathbf{I}).\end{aligned}$$

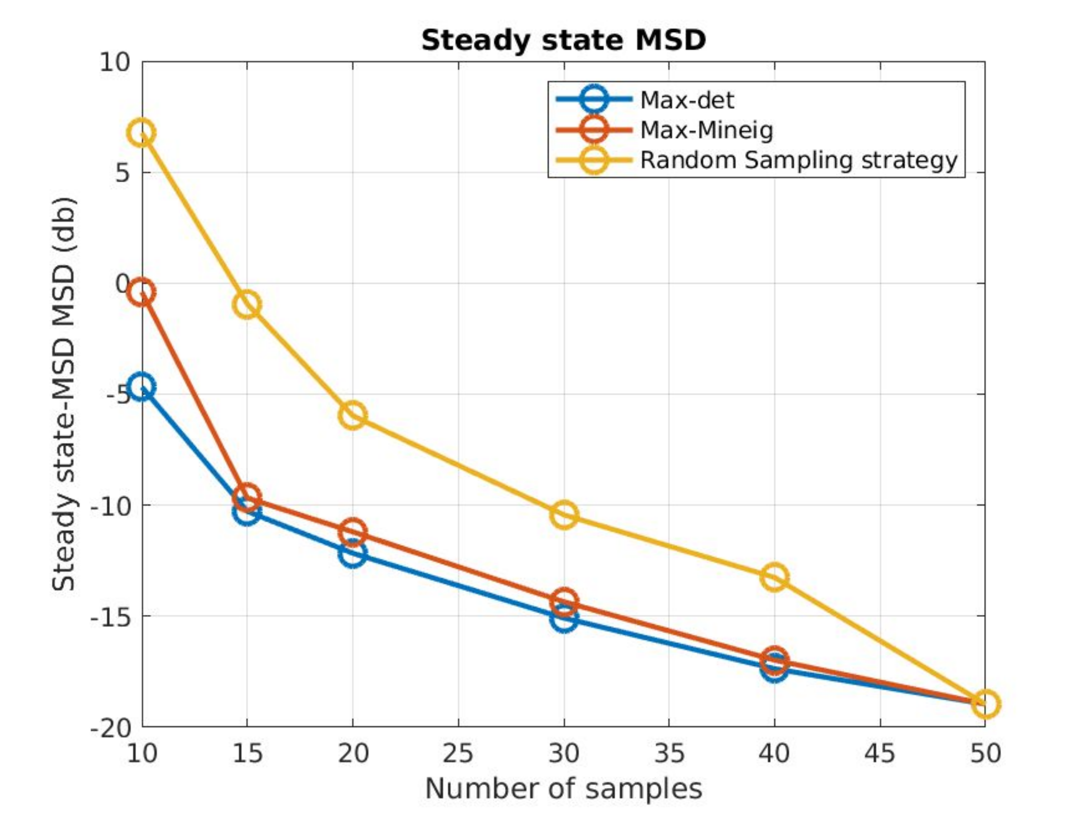
Results of Transient MSD



- Learning rate = 0.5
- Iterations = 100 over 200 independent iterations
- Number of samples = 10 , 14, 18
- Sampling strategy = Max-Det
- Bandwidth = 10
- Noise is zero mean with diagonal covariance matrix with elements uniformly random between 0 and 0.01.

Observation: The error in reducing with increase in number of samples as more information is available to reconstruct the signal.

Results of steady state MSD



- Learning rate = 0.5
- Iterations = 100 over 200 independent iterations
- Number of samples = varying
- Sampling strategy = Max-Det, Max-Mineig and Random sampling
- Bandwidth = 10
- Noise is zero mean with diagonal covariance matrix with elements uniformly random between 0 and 0.01.

Observation:

- Random sampling performs worst especially at low number of samples. Thus, when sampling a graph signal, both the number of samples and their location matter.
- The Max-Det strategy outperforms the Max-Mineig strategy as it considers all the eigenvalues and not just one.
- The Min-MSD strategy is highly costly to compute but will give the most minimum value for some samples as it takes into account information from both graph topology and spatial distribution of the observation noise.

LMS estimation with adaptive graph sampling- Model Definition

- Suppose that there is no prior information about the frequency set F which must then be inferred directly from the streaming data y[n].
- Here, we consider the important case where the graph is fixed, and the spectral content of the signal can vary over time in an unknown manner.
- The signal observation can be recast as: $\boldsymbol{y}[n] = \mathbf{D}\mathbf{U}\boldsymbol{s}_0 + \mathbf{D}\boldsymbol{v}[n].$
- Thus, the overall problem can be formulated as the joint estimation of sparse representation s and sampling strategy D from the observations y[n] as: $\min_{\boldsymbol{s}, \mathbf{D}\in\mathcal{D}} \mathbf{E}\|\boldsymbol{y}[n] - \mathbf{D}\mathbf{U}\boldsymbol{s}\|^2 + \lambda f(\boldsymbol{s}),$ where D(italic) is the (discrete) set that constraints the selection of the sampling strategy D, f (·) is a sparsifying penalty function (typically, l0 or l1 norms), and λ > 0 is a parameter that regulates how sparse we want the optimal GFT vector s.
- The rationale behind this choice is that, given an estimate for the support of vector s, i.e. F , we can select the sampling operator D in a very efficient manner through one of the sampling strategies.

LMS estimation with adaptive graph sampling- Algorithm

- The aim of algorithm is to estimate the GFT s0 of the graph signal x0, while selectively shrinking to zero all the components of s0 that are outside its support, i.e., which do not belong to the bandwidth of the graph signal. Then, the online identification of the support of the GFT s0 enables the adaptation of the sampling strategy, which can be updated using one of the strategies.
- Intuitively, the algorithm will increase (reduce) the number of samples used for the estimation, depending on the increment (reduction) of the current signal bandwidth.
- Thresholding functions are used which depend on the sparsity-inducing penalty f(.).

Algorithm 2: LMS with Adaptive Graph Sampling

Start with $\boldsymbol{s}[0]$ chosen at random, $\mathbf{D}[0] = \mathbf{I}$, and $\mathcal{F}[0] = \mathcal{V}$.
Given $\mu > 0$, for each time $n > 0$, repeat:

1) $\boldsymbol{s}[n+1] = T_{\lambda\mu}\left(\boldsymbol{s}[n] + \mu \mathbf{U}^H \mathbf{D}[n] \left(\boldsymbol{y}[n] - \mathbf{U}\boldsymbol{s}[n]\right)\right)$;
2) Set $\mathcal{F}[n+1] = \{i \in \{1, \ldots, N\} : s_i[n+1] \neq 0\}$;
3) Given $\mathbf{U}_{\mathcal{F}[n+1]}$, select $\mathbf{D}[n+1]$ according to one of the criteria proposed in Sec. III.D;

LMS estimation with adaptive graph sampling- Thresholding functions

- **Lasso constraint:** Here γ = μλ, $$T_\gamma(s_m) = \begin{cases} s_m - \gamma, & s_m > \gamma; \\ 0, & -\gamma \le s_m \le \gamma; \\ s_m + \gamma, & s_m < -\gamma. \end{cases}$$, sets to zero the components whose magnitude are within the threshold γ. Since the Lasso constraint is known for introducing a large bias in the estimate, the performance would deteriorate for vectors that are not sufficiently sparse, i.e. graph signals with large bandwidth.

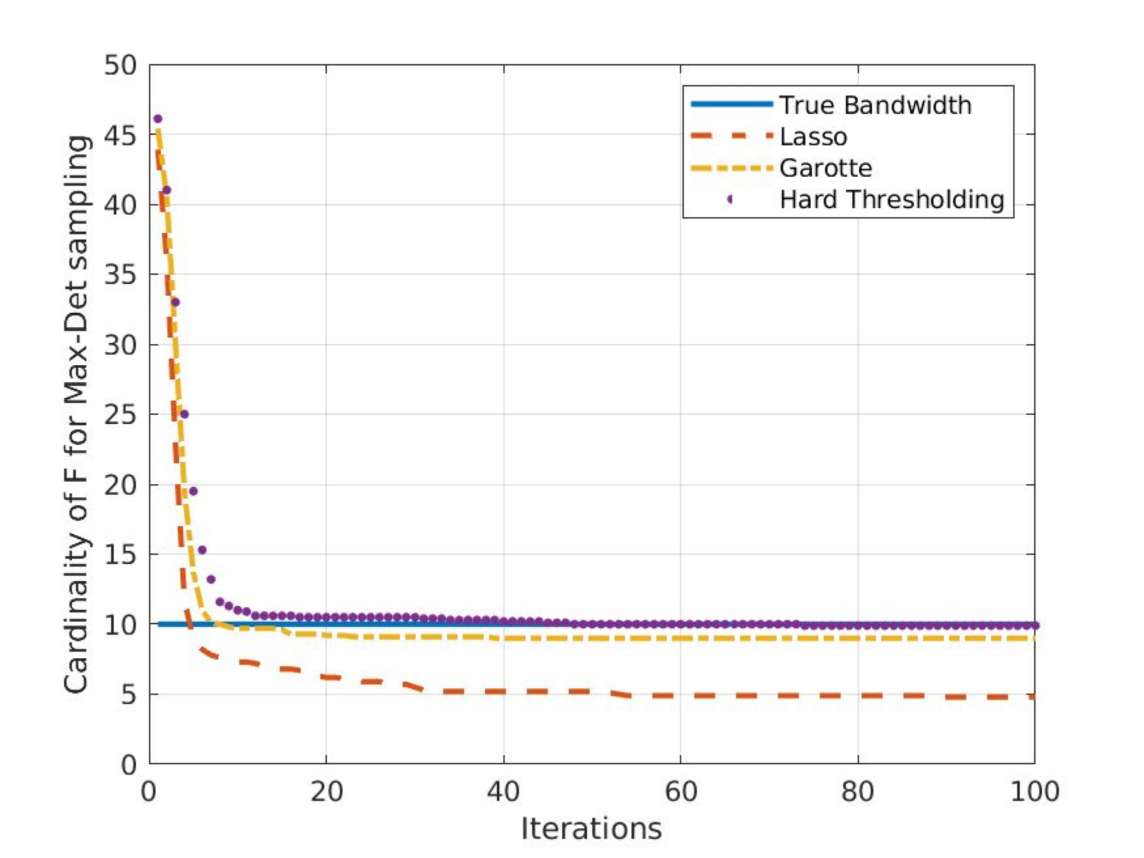
- **Garotte constraint:**

$$T_\gamma(s_m) = \begin{cases} s_m\left(1 - \gamma^2/s_m^2\right), & |s_m| > \gamma; \\ 0, & |s_m| \le \gamma; \end{cases}$$

- **Hard Thresholding:** Completely removing the bias.

$$T_\gamma(s_m) = \begin{cases} s_m, & |s_m| > \gamma; \\ 0, & |s_m| \le \gamma; \end{cases}$$

Results - Cardinality of F

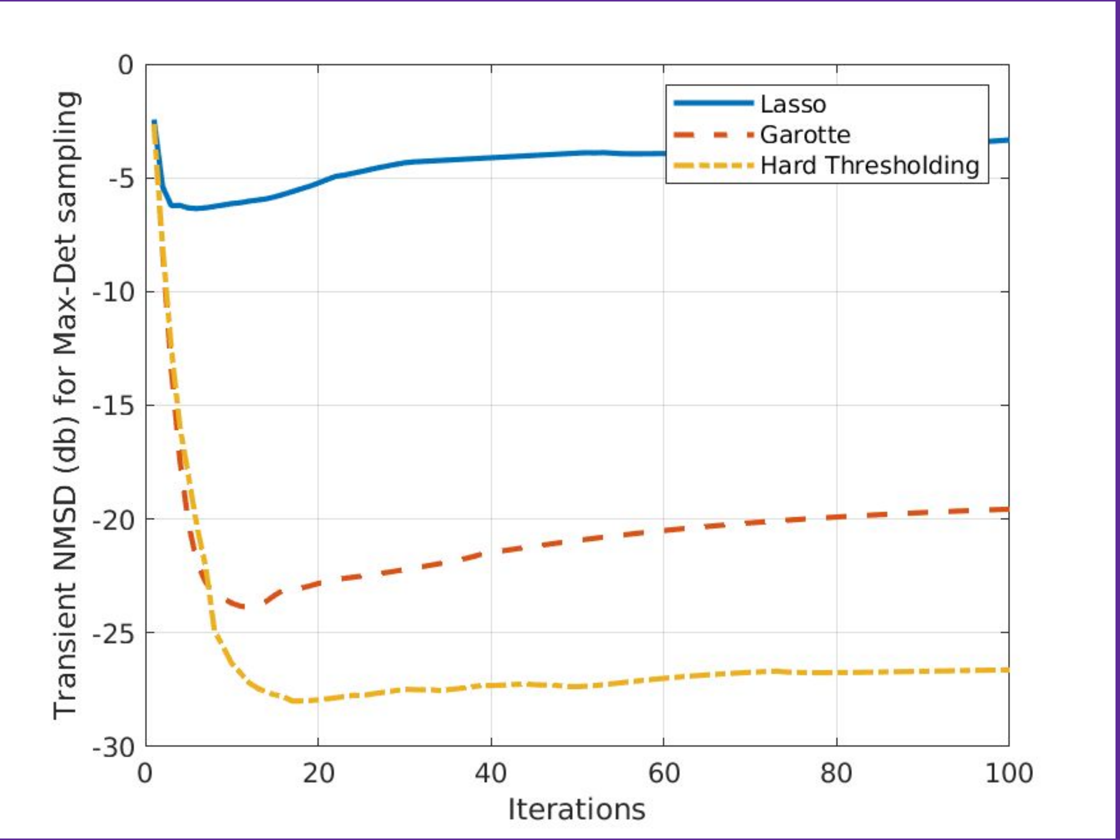


- Learning rate = 0.5, Lambda= 0.1
- Iterations = 100 over 10 independent iterations
- Number of samples = varying
- Sampling strategy = Max-Det
- Noise is zero mean with diagonal covariance matrix with elements 0.0004.

Observation:

- The Garotte and hard thresholding strategies are able to learn the true bandwidth of the graph signal better than Lasso.
- Lasso strategy underestimates the bandwidth of the signal, i.e. the cardinality of the set F

Results- Normalised MSD



$$\mathrm{NMSD}[n] = \frac{\|s[n] - s_0\|^2}{\|s_0\|^2},$$

- Learning rate = 0.5, Lambda= 0.1
- Iterations = 100 over 10 independent iterations
- Number of samples = varying
- Sampling strategy = Max-Det
- Noise is zero mean with diagonal covariance matrix with elements 0.0004.

Observation:

- The algorithm based on the hard thresholding function outperforms the other strategies in terms of steady-state NMSD, while having the same learning rate.
- The Garotte based algorithm has slightly worse performance with respect to the method exploiting hard thresholding, due to the residual bias introduced at large values by the thresholding function.
- The LMS algorithm based on Lasso may lead to very poor performance, due to misidentifications of the true graph bandwidth.

Challenges

- Optimization of the Min-MSD sampling method can be done to get better results for the sampled set to get better reconstruction.
- The Algorithm can be extended to reconstruct signal when the graph is also changing.