

LMS Estimation of Signals defined over Graphs

Paolo Di Lorenzo¹, Sergio Barbarossa², Paolo Banelli¹, and Stefania Sardellitti²

¹ Department of Engineering, University of Perugia, Via G. Duranti 93, 06125, Perugia, Italy

² Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184, Rome, Italy

Email: paolo.dilorenzo@unipg.it, sergio.barbarossa@uniroma1.it,
paolo.banelli@unipg.it, stefania.sardellitti@uniroma1.it

Abstract—The aim of this paper 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 sampled over a subset of vertices. A detailed mean square analysis provides the performance of the proposed method, and leads to several insights for designing useful sampling strategies for graph signals. Numerical results validate our theoretical findings, and illustrate the advantages achieved by the proposed strategy for online estimation of band-limited graph signals.

Index Terms—Least mean squares estimation, graph signal processing, sampling on graphs.

I. INTRODUCTION

Recent years have witnessed a large interest in developing novel modeling and processing tools for the analysis of signals defined over a graph, or graph signals for short [1]–[3]. Graph signal processing (GSP) extends classical discrete-time signal processing to signals defined over a discrete domain having a very general structure, represented by a graph, which subsumes discrete-time as a very simple case. A central role in GSP is played by spectral analysis of graph signals, which is based on the so called Graph Fourier Transform (GFT). Alternative definitions of GFT have been proposed, see, e.g., [4], [1], [2]. Two basic approaches are available, proposing the projection of the graph signal onto the eigenvectors of either the graph Laplacian, see, e.g., [4], [1], or of the adjacency matrix, see, e.g. [2]. One of the basic problems in GSP is the development of a graph *sampling theory*, whose aim is to recover a band-limited (or approximately band-limited) graph signal from a subset of its samples. A seminal contribution was given in [4], later extended in [5] and, very recently, in [6], [7], [8], [9]. Dealing with graph signals, the recovery problem may easily become ill-conditioned, depending on the location of the samples. Hence, for any given number of samples enabling signal recovery, the identification of the sampling set plays a key role in the conditioning of the recovery problem. It is then particularly important to devise strategies to optimize the selection of the sampling set. Alternative signal reconstruction methods have been proposed, either iterative as in [10], [8], or single shot, as in [6], [7]. Frame-based approaches to reconstruct signals from subsets of samples have also been

proposed in [4], [8], [7]. Finally, in [11], the authors proposed signal recovery methods aimed to recover graph signals that are assumed to be smooth with respect to the underlying graph, from sampled, noisy, missing, or corrupted measurements.

The goal of this paper is to propose LMS strategies for the adaptive estimation of signals defined on graphs. To the best of our knowledge, this is the first attempt to merge the well established theory of adaptive filtering [12] with the emerging field of signal processing on graphs. The proposed method hinges on the graph structure describing the observed signal and, under a band-limited assumption, it enables online reconstruction from a limited number of observations taken over a subset of vertices. A detailed mean square analysis illustrates the role of the sampling strategy on the reconstruction capability, stability, and mean-square performance of the proposed algorithm. Based on these results, we also derive adaptive sampling strategies for LMS estimation of graph signals. Numerical results confirm the theoretical findings, and assess the performance of the proposed strategies.

II. GRAPH SIGNAL PROCESSING TOOLS

We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a set of N nodes $\mathcal{V} = \{1, 2, \dots, N\}$, along with a set of weighted edges $\mathcal{E} = \{a_{ij}\}_{i,j \in \mathcal{V}}$, such that $a_{ij} > 0$, if there is a link from node j to node i , or $a_{ij} = 0$, otherwise. The adjacency matrix \mathbf{A} of a graph is the collection of all the weights a_{ij} , $i, j = 1, \dots, N$. The degree of node i is $k_i := \sum_{j=1}^N a_{ij}$. The degree matrix \mathbf{K} is a diagonal matrix having the node degrees on its diagonal. The Laplacian matrix is defined as $\mathbf{L} = \mathbf{K} - \mathbf{A}$. If the graph is *undirected*, the Laplacian matrix is symmetric and positive semi-definite, and admits the eigendecomposition $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where \mathbf{U} collects all the eigenvectors of \mathbf{L} in its columns, whereas $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of \mathbf{L} . It is well known from spectral graph theory that the eigenvectors of \mathbf{L} are well suited for representing clusters, since they minimize the ℓ_2 norm graph total variation.

A signal \mathbf{x} over a graph \mathcal{G} is defined as a mapping from the vertex set to the set of complex numbers, i.e. $\mathbf{x} : \mathcal{V} \rightarrow \mathbb{C}$. In many applications, the signal \mathbf{x} admits a compact representation, i.e., it can be expressed as:

$$\mathbf{x} = \mathbf{U}\mathbf{s} \quad (1)$$

where \mathbf{s} is exactly (or approximately) sparse. As an example, in all cases where the graph signal exhibits clustering features, i.e. it is a smooth function within each cluster, but it is

This work has been supported by TROPIC Project, Nr. ICT-318784. The work of Paolo Di Lorenzo was supported by the ‘‘Fondazione Cassa di Risparmio di Perugia’’.

allowed to vary arbitrarily from one cluster to the other, the representation in (1) is compact, i.e. the only nonzero (or approximately nonzero) entries of \mathbf{s} are the ones associated to the clusters. The GFT \mathbf{s} of a signal \mathbf{x} is defined as the projection onto the set of vectors $\mathbf{U} = \{\mathbf{u}_i\}_{i=1,\dots,N}$ [1], i.e.

$$\text{GFT: } \mathbf{s} = \mathbf{U}^H \mathbf{x}. \quad (2)$$

The GFT has been defined in alternative ways, see, e.g., [1], [2], [6]. In this paper, we basically follow the approach based on the Laplacian matrix, assuming an undirected graph structure, but the theory could be extended to handle directed graphs with minor modifications. From (1) and (2), if the signal \mathbf{x} exhibits a clustering behavior, in the sense specified above, the GFT is the way to recover the sparse vector \mathbf{s} .

Given a subset of vertices $\mathcal{S} \subseteq \mathcal{V}$, we define a vertex-limiting operator as the diagonal matrix

$$\mathbf{D}_{\mathcal{S}} = \text{diag}\{\mathbf{1}_{\mathcal{S}}\}, \quad (3)$$

where $\mathbf{1}_{\mathcal{S}}$ is the set indicator vector, whose i -th entry is equal to one, if $i \in \mathcal{S}$, or zero otherwise. Similarly, given a subset of frequency indices $\mathcal{F} \subseteq \mathcal{V}$, we introduce the filtering operator

$$\mathbf{B}_{\mathcal{F}} = \mathbf{U} \Sigma_{\mathcal{F}} \mathbf{U}^H, \quad (4)$$

where $\Sigma_{\mathcal{F}}$ is a diagonal matrix defined as $\Sigma_{\mathcal{F}} = \text{diag}\{\mathbf{1}_{\mathcal{F}}\}$. It is immediate to check that both matrices $\mathbf{D}_{\mathcal{S}}$ and $\mathbf{B}_{\mathcal{F}}$ are self-adjoint and idempotent, and then they represent orthogonal projectors. The space $\mathcal{B}_{\mathcal{F}}$ of all signals whose GFT is exactly supported on the set \mathcal{F} is known as the *Paley-Wiener space* for the set \mathcal{F} [4]. In the rest of the paper, whenever there will be no ambiguities in the specification of the sets, we will drop the subscripts referring to the sets. Finally, given a set \mathcal{S} , we denote its complement set as $\bar{\mathcal{S}}$, such that $\mathcal{V} = \mathcal{S} \cup \bar{\mathcal{S}}$ and $\mathcal{S} \cap \bar{\mathcal{S}} = \emptyset$. Thus, we define the vertex-projector onto $\bar{\mathcal{S}}$ as $\bar{\mathbf{D}}$.

Exploiting the localization operators in (3) and (4), we say that a vector \mathbf{x} is perfectly localized over the subset $\mathcal{S} \subseteq \mathcal{V}$ if

$$\mathbf{D} \mathbf{x} = \mathbf{x}, \quad (5)$$

with \mathbf{D} defined as in (3). Similarly, a vector \mathbf{x} is perfectly localized over the frequency set \mathcal{F} if

$$\mathbf{B} \mathbf{x} = \mathbf{x}, \quad (6)$$

with \mathbf{B} given in (4). The localization properties of graph signals were studied in [7] to derive the fundamental trade-off between the localization of a signal in the graph and on its dual domain. An interesting consequence of that theory is that, differently from continuous-time signals, a graph signal can be perfectly localized in *both* vertex and frequency domains.

III. LMS ESTIMATION OF GRAPH SIGNALS

Let us consider a signal $\mathbf{x}_0 \in \mathbb{C}^N$ defined over the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The signal is assumed to be perfectly band-limited, i.e. its spectral content is different from zero only on a limited set of frequencies \mathcal{F} . Let us consider partial observations of signal \mathbf{x}_0 , i.e. observations over only a subset of nodes. Denoting with \mathcal{S} the sampling set (observation subset), the observed signal at time n can be expressed as:

$$\mathbf{y}[n] = \mathbf{D} (\mathbf{x}_0 + \mathbf{v}[n]) = \mathbf{D} \mathbf{B} \mathbf{x}_0 + \mathbf{D} \mathbf{v}[n] \quad (7)$$

where \mathbf{D} is the vertex-limiting operator defined in (3), which takes nonzero values only in the set \mathcal{S} , and $\mathbf{v}[n]$ is a zero-mean, additive noise with covariance matrix \mathbf{C}_v . The second equality in (7) comes from the bandlimited assumption, i.e. $\mathbf{B} \mathbf{x}_0 = \mathbf{x}_0$, with \mathbf{B} denoting the operator in (4) that projects onto the (known) frequency set \mathcal{F} . The estimation task consists in recovering the band-limited graph signal \mathbf{x}_0 from the noisy, streaming, and partial observations $\mathbf{y}[n]$ in (7). Following an LMS approach [12], the optimal estimate for \mathbf{x}_0 can be found as the vector that solves the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbb{E} \|\mathbf{y}[n] - \mathbf{D} \mathbf{B} \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{B} \mathbf{x} = \mathbf{x} \end{aligned} \quad (8)$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator. The solution of problem (8) minimizes the mean-squared error and has a bandwidth limited to the frequency set \mathcal{F} . A typical LMS-type solution proceeds to optimize (8) by means of a steepest-descent procedure, relying only on instantaneous information. Thus, letting $\mathbf{x}[n]$ be the current estimate of vector \mathbf{x}_0 , the LMS algorithm for graph signals evolves as illustrated in Algorithm 1, where $\mu > 0$ is a (sufficiently small) step-size, and we have exploited the fact that \mathbf{D} is an idempotent operator, and $\mathbf{B} \mathbf{x}[n] = \mathbf{x}[n]$ (i.e., $\mathbf{x}[n]$ is band-limited) for all n . Algorithm 1 starts from an initial signal that belongs

Algorithm 1: LMS algorithm for graph signals

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

$$\mathbf{x}[n+1] = \mathbf{x}[n] + \mu \mathbf{B} \mathbf{D} (\mathbf{y}[n] - \mathbf{x}[n]) \quad (9)$$

to the Paley-Wiener space for the set \mathcal{F} , and then evolves implementing an alternating orthogonal projection onto the vertex set \mathcal{S} (through \mathbf{D}) and the frequency set \mathcal{F} (through \mathbf{B}). The properties of the LMS recursion in (9) crucially depend on the sampling set \mathcal{S} , i.e., on the operator \mathbf{D} [cf. (3)]. Thus, in the sequel we will show how the choice of the operator \mathbf{D} affects the reconstruction capability, the mean-square stability, and the steady-state performance of Algorithm 1.

A. Reconstruction Properties

The LMS algorithm in (9) is a stochastic approximation method for the solution of problem (8), which enables convergence in the mean-sense to the true vector \mathbf{x}_0 , while guaranteeing a bounded mean-square error (as we will see in the sequel). However, since the existence of a unique band-limited solution for problem (9) depends on the adopted sampling strategy, the first natural question to address is: What conditions must be satisfied by the sampling operator \mathbf{D} to guarantee reconstruction of signal \mathbf{x}_0 from the selected samples? The answer is given in the following theorem, which gives a necessary and sufficient condition to reconstruct graph signals from partial observations using Algorithm 1.

Theorem 1: Problem (8) admits a unique solution, i.e. any band-limited signal \mathbf{x}_0 can be reconstructed from its samples

taken in the set \mathcal{S} , if and only if

$$\|\overline{\mathbf{D}}\mathbf{B}\|_2 < 1, \quad (10)$$

i.e. if the matrix $\overline{\mathbf{D}}\mathbf{B}$ does not have any eigenvector that is perfectly localized on $\overline{\mathcal{S}}$ and bandlimited on \mathcal{F} .

Proof. See [13]. ■

A necessary condition that enables reconstruction, i.e. the non-existence of a non-trivial vector \mathbf{x} satisfying $\mathbf{D}\mathbf{B}\mathbf{x} = \mathbf{0}$, is that $|\mathcal{S}| \geq |\mathcal{F}|$. However, this condition is not sufficient, because matrix $\mathbf{D}\mathbf{B}$ in (7) may loose rank, or easily become ill-conditioned, depending on the graph topology and sampling strategy. This suggests that the location of samples plays a key role in the performance of Algorithm 1.

B. Mean-Square Analysis

In this section, we study the mean-square behavior of the proposed LMS strategy, illustrating how the sampling operator \mathbf{D} affects its stability and steady-state performance. From now on, we view the estimates $\mathbf{x}[n]$ as realizations of a random process and analyze the performance of the LMS algorithm in terms of its mean-square behavior. Let $\tilde{\mathbf{x}}[n] = \mathbf{x}[n] - \mathbf{x}_0$ be the error vector at time n . Subtracting \mathbf{x}_0 from the left and right hand sides of (9), using (7) and $\mathbf{B}\tilde{\mathbf{x}}[n] = \tilde{\mathbf{x}}[n]$, we get:

$$\tilde{\mathbf{x}}[n+1] = (\mathbf{I} - \mu \mathbf{B}\mathbf{D}\mathbf{B})\tilde{\mathbf{x}}[n] + \mu \mathbf{B}\mathbf{D}\mathbf{v}[n]. \quad (11)$$

Letting $\tilde{\mathbf{s}}[n]$ be the GFT of $\tilde{\mathbf{x}}[n]$, we will analyze the mean-square behavior of the error recursion (11) only on the support of $\tilde{\mathbf{s}}[n]$, i.e. $\hat{\mathbf{s}}[n] = \{\tilde{s}_i[n], i \in \mathcal{F}\} \in \mathbb{C}^{|\mathcal{F}|}$. Thus, letting $\mathbf{U}_{\mathcal{F}} \in \mathbb{C}^{N \times |\mathcal{F}|}$ be the matrix having as columns the eigenvectors of the Laplacian matrix associated to the frequency indices \mathcal{F} , and multiplying each side of (11) by $\mathbf{U}_{\mathcal{F}}^H$, the error recursion (11) can be rewritten in compact form as:

$$\hat{\mathbf{s}}[n+1] = (\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}})\hat{\mathbf{s}}[n] + \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{v}[n]. \quad (12)$$

Using energy conservation arguments [14], we consider a general weighted squared error sequence $\hat{\mathbf{s}}[n]^H \Phi \hat{\mathbf{s}}[n]$, where $\Phi \in \mathbb{C}^{|\mathcal{F}| \times |\mathcal{F}|}$ is any Hermitian nonnegative-definite matrix that we are free to choose. In the sequel, it will be clear the role played by a proper selection of the matrix Φ . Then, from (12) we can establish the following variance relation:

$$\begin{aligned} \mathbb{E}\|\hat{\mathbf{s}}[n+1]\|_{\Phi}^2 &= \mathbb{E}\|\hat{\mathbf{s}}[n]\|_{\Phi'}^2 + \mu^2 \mathbb{E}\{\mathbf{v}[n]^H \mathbf{D}\mathbf{U}_{\mathcal{F}} \Phi \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{v}[n]\} \\ &= \mathbb{E}\|\hat{\mathbf{s}}[n]\|_{\Phi'}^2 + \mu^2 \text{Tr}(\Phi \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{C}_v \mathbf{D}\mathbf{U}_{\mathcal{F}}) \end{aligned} \quad (13)$$

where $\text{Tr}(\cdot)$ denotes the trace operator, and

$$\Phi' = (\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}}) \Phi (\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}}). \quad (14)$$

Let $\varphi = \text{vec}(\Phi)$ and $\varphi' = \text{vec}(\Phi')$, where the notation $\text{vec}(\cdot)$ stacks the columns of Φ on top of each other and $\text{vec}^{-1}(\cdot)$ is the inverse operation. We will use interchangeably the notation $\|\hat{\mathbf{s}}\|_{\Phi}^2$ and $\|\hat{\mathbf{s}}\|_{\varphi}^2$ to denote the same quantity $\hat{\mathbf{s}}^H \Phi \hat{\mathbf{s}}$. Exploiting the Kronecker product property $\text{vec}(\mathbf{X}\Phi\mathbf{Y}) = (\mathbf{Y}^H \otimes \mathbf{X})\text{vec}(\Phi)$, and the trace property $\text{Tr}(\Phi\mathbf{X}) = \text{vec}(\mathbf{X}^H)^T \text{vec}(\Phi)$, in the relation (13), we obtain:

$$\mathbb{E}\|\hat{\mathbf{s}}[n+1]\|_{\varphi}^2 = \mathbb{E}\|\hat{\mathbf{s}}[n]\|_{\varphi'}^2 + \mu^2 \text{vec}(\mathbf{G})^T \varphi \quad (15)$$

where

$$\mathbf{G} = \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{C}_v \mathbf{D}\mathbf{U}_{\mathcal{F}} \quad (16)$$

$$\mathbf{Q} = (\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}}) \otimes (\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}}). \quad (17)$$

The following theorem guarantees the asymptotic mean-square stability of the LMS algorithm in (9).

Theorem 2: Assume model (7) holds. Then, for any bounded initial condition, the LMS strategy (9) asymptotically converges in the mean-square error sense if the sampling operator \mathbf{D} and the step-size μ are chosen to satisfy (10) and

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{U}_{\mathcal{F}}^H \mathbf{D}\mathbf{U}_{\mathcal{F}})}, \quad (18)$$

where $\lambda_{\max}(\mathbf{A})$ is the maximum eigenvalue of matrix \mathbf{A} .

Proof. See [13]. ■

C. Steady-State Performance

Taking the limit of (15) as $n \rightarrow \infty$ (assuming condition (18) holds true), we obtain:

$$\lim_{n \rightarrow \infty} \mathbb{E}\|\hat{\mathbf{s}}[n]\|_{(\mathbf{I}-\mathbf{Q})\varphi}^2 = \mu^2 \text{vec}(\mathbf{G})^T \varphi. \quad (19)$$

Expression (19) is a useful result: it allows us to derive several performance metrics through the proper selection of the free weighting parameter φ (or Φ). For instance, let us assume that one wants to evaluate the steady-state mean square deviation (MSD) of the LMS strategy (9). Thus, selecting $\varphi = (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{I})$ in (19), we obtain

$$\begin{aligned} \text{MSD} &= \lim_{n \rightarrow \infty} \mathbb{E}\|\tilde{\mathbf{x}}[n]\|^2 = \lim_{n \rightarrow \infty} \mathbb{E}\|\hat{\mathbf{s}}[n]\|^2 \\ &= \mu^2 \text{vec}(\mathbf{G})^T (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{I}). \end{aligned} \quad (20)$$

If instead one is interested in evaluating the mean square deviation obtained by the LMS algorithm (9) when reconstructing the value of the signal associated to k -th vertex of the graph, selecting $\varphi = (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}})$ in (19), we obtain

$$\begin{aligned} \text{MSD}_k &= \lim_{n \rightarrow \infty} \mathbb{E}\|\tilde{x}[n]\|_{\mathbf{E}_k}^2 = \lim_{n \rightarrow \infty} \mathbb{E}\|\hat{\mathbf{s}}[n]\|_{\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}}}^2 \\ &= \mu^2 \text{vec}(\mathbf{G})^T (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}}), \end{aligned} \quad (21)$$

where $\mathbf{E}_k = \text{diag}\{e_k\}$, with $e_k \in \mathbb{R}^N$ denoting the k -th canonical vector. In the sequel, we will confirm the validity of these theoretical expressions by comparing them with numerical simulations.

D. Sampling Strategies

The properties of the proposed LMS algorithm in (9) strongly depend on the choice of the sampling set \mathcal{S} , i.e. on the vertex limiting operator \mathbf{D} . Indeed, building on the previous analysis, it is clear that the sampling strategy must be carefully designed in order to: a) enable reconstruction of the signal; b) guarantee stability of the algorithm; and c) impose a desired mean-square error at convergence. To select the best sampling strategy, one should optimize some performance criterion, e.g. the MSD in (20), with respect to the sampling set \mathcal{S} . However, since this formulation translates inevitably into a combinatorial problem, whose solution in general requires an exhaustive

search over all the possible combinations, the complexity of such procedure becomes intractable also for graph signals of moderate dimensions. Thus, in the sequel we will provide some efficient, albeit sub-optimal, greedy algorithms to tackle the problem of selecting the sampling set.

Greedy Selection - Minimum MSD: This strategy aims at minimizing the MSD in (20) via a greedy approach: the method iteratively selects the samples from the graph that lead to the largest reduction in terms of MSD. Since the proposed greedy approach starts from an initially empty sampling set, when $|\mathcal{S}| < |\mathcal{F}|$, matrix $\mathbf{I} - \mathbf{Q}$ in (20) is inevitably rank deficient. Then, in this case, the criterion builds on the pseudo-inverse of the matrix $\mathbf{I} - \mathbf{Q}$ in (20), denoted by $(\mathbf{I} - \mathbf{Q})^\dagger$, which coincides with the inverse as soon as $|\mathcal{S}| \geq |\mathcal{F}|$. The resulting algorithm is summarized in the table entitled ‘‘Sampling strategy 1’’, where we made explicit the dependence of matrices \mathbf{G} and \mathbf{Q} on the sampling operator \mathbf{D} . In the sequel, we will refer to this method as the Min-MSD strategy.

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 $|\mathcal{S}| < M$

$s = \arg \min_j \text{vec}(\mathbf{G}(\mathbf{D}_{\mathcal{S} \cup \{j\}}))^T (\mathbf{I} - \mathbf{Q}(\mathbf{D}_{\mathcal{S} \cup \{j\}}))^\dagger \text{vec}(\mathbf{I});$

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

end

Greedy Selection - Maximum $|\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}|_+$: In this case, the strategy aims at maximizing the volume of the parallelepiped build with the selected rows of matrix $\mathbf{U}_{\mathcal{F}}$. The algorithm starts including the row with the largest norm in $\mathbf{U}_{\mathcal{F}}$, and then it adds, iteratively, the rows having the largest norm and, at the same time, are as orthogonal as possible to the vectors already in \mathcal{S} . 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 pseudo determinant of the matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$ (i.e. the product of all nonzero eigenvalues), which is denoted by $|\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}|_+$. The resulting algorithm is summarized in the table entitled ‘‘Sampling strategy 2’’. We will refer to this method as the Max-Det sampling strategy.

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

Input Data : M , the number of samples.

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

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

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

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

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

end

In the sequel, we will illustrate some numerical results aimed at comparing the performance achieved by the proposed LMS algorithm using the aforementioned sampling strategies.

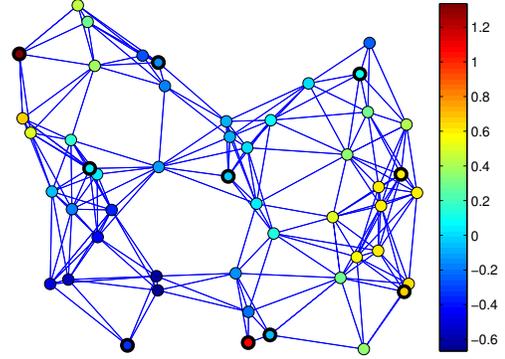


Fig. 1: Example of graph signal and sampling.

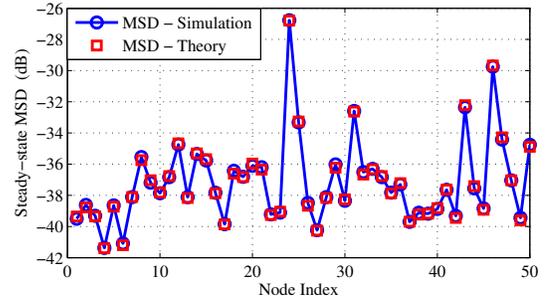


Fig. 2: Comparison between theoretical MSD in (19) and simulation results, at each vertex of the graph.

IV. NUMERICAL RESULTS

Let us consider the graph signal shown in Fig. 1 and composed of $N = 50$ nodes, where the color of each vertex denotes the value of the signal associated to it. The signal has a spectral content limited to the first ten eigenvectors of the Laplacian matrix of the graph in Fig. 1, i.e. $|\mathcal{F}| = 10$. The observation noise in (7) is zero-mean, Gaussian, with a diagonal covariance matrix, where each element is chosen uniformly random between 0 and 0.01. An example of graph sampling, obtained selecting $|\mathcal{S}| = 10$ vertexes using the Max-Det sampling strategy, is also illustrated in Fig. 1, where the sampled vertexes have thicker marker edge. To validate the theoretical results in (21), in Fig. 2 we report the behavior of the theoretical MSD values achieved at each vertex of the graph, comparing them with simulation results, obtained averaging over 200 independent simulations and 100 samples of squared error after convergence of the algorithm. The step-size is chosen equal to $\mu = 0.5$ and, together with the selected sampling strategy \mathbf{D} , they satisfy the reconstruction and stability conditions in (10) and (18). As we can notice from Fig. 2, the theoretical predictions match well the simulation results.

It is fundamental to assess the performance of the LMS algorithm in (9) with respect to the adopted sampling set \mathcal{S} . As a first example, using the Max-Det sampling strategy, in Fig. 3 we report the transient behavior of the MSD, considering different number of samples taken from the graph, i.e. different cardinalities $|\mathcal{S}|$ of the sampling set. The results are averaged over 200 independent simulations, and the step-sizes are tuned in order to have the same steady-state MSD for each value of $|\mathcal{S}|$. As expected, from Fig. 3 we notice how the learning

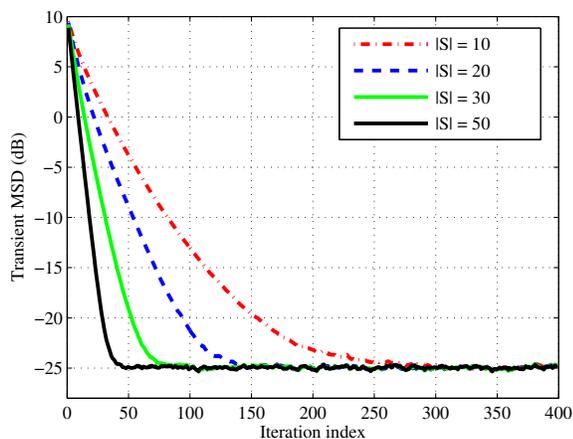


Fig. 3: Transient MSD, for different number of samples $|\mathcal{S}|$. Increasing the number of samples, the learning rate improves.

rate of the algorithm improves by increasing the number of samples. Finally, in Fig. 4 we illustrate the steady-state MSD of the LMS algorithm in (9) comparing the performance obtained by three different sampling strategies, namely: a) the Min-MSD strategy; b) the Max-Det strategy; and c) the random sampling strategy, which simply picks at random $|\mathcal{S}|$ nodes. We consider the same parameter setting of the previous simulation. The results are averaged over 200 independent simulations. As we can notice from Fig. 4, the LMS algorithm with random sampling can perform quite poorly, especially at low number of samples. This poor result of random sampling emphasizes that, when sampling a graph signal, what matters is not only the number of samples, but also (and most important) where the samples are taken. We also notice from Fig. 4 that the Max-det strategy performs well also at low number of samples ($|\mathcal{S}| = 10$ is the minimum number of samples that allows signal reconstruction), where the other methods fail. It is indeed remarkable that, for low number of samples, Max-Det outperforms also Min-MSD, even if the performance metric is MSD. There is no contradiction here because we need to remember that all the proposed methods are greedy strategies, so that there is no claim of optimality in all of them. However, as the number of samples increases, the Min-MSD strategy outperforms all other methods. This is a consequence of the fact that Min-MSD exploits information about the spatial distribution of the observation noise (cf. (20)). Thus, increasing the number of samples, this strategy avoids to select the most noisy vertexes of the graph, thus improving the overall performance of the LMS algorithm in (9). This analysis suggests that an optimal design of the sampling strategy for graph signals should take into account processing complexity (in terms of number of samples), prior knowledge (e.g., graph structure, noise distribution), and achievable performance.

V. CONCLUSIONS

In this paper we have proposed LMS strategies for online estimation of signals defined over graphs. The proposed strategies are able to exploit the underlying structure of the graph signal, which can be reconstructed from a limited number of observations properly sampled from a subset of vertexes, under

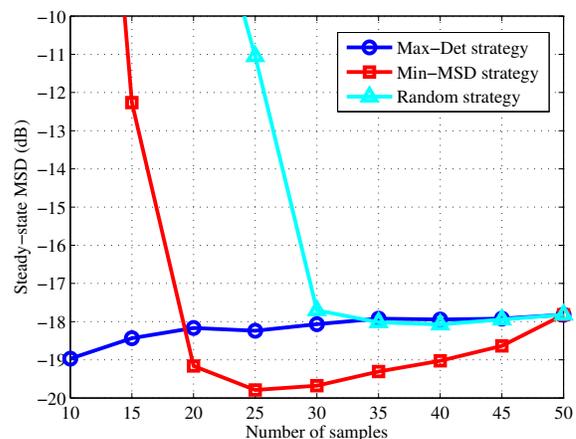


Fig. 4: Steady-state MSD versus number of samples, for different sampling strategies.

a band-limited assumption. A detailed mean square analysis illustrates the deep connection between sampling strategy and the properties of the proposed LMS algorithm in terms of reconstruction capability, stability, and mean-square error performance. From this analysis, some sampling strategies for adaptive estimation of graph signals are also derived. Several numerical simulations confirm the theoretical findings, and illustrate the potential advantages achieved by these strategies for online estimation of band-limited graph signals.

REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Proc. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.
- [2] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. on Signal Proc.*, vol. 61, pp. 1644–1656, 2013.
- [3] —, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Proc. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.
- [4] I. Z. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.
- [5] S. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 5445–5449.
- [6] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. on Signal Proc.*, vol. 63, pp. 6510–6523, Dec. 2015.
- [7] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," to appear in *IEEE Trans. on Signal Processing*; available at <http://arxiv.org/abs/1507.08822>, 2015.
- [8] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Trans. on Signal Proc.*, vol. 63, no. 9, pp. 2432–2444, 2015.
- [9] A. G. Marquez, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. on Signal Process.*, vol. 64, no. 7, pp. 1832–1843.
- [10] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *IEEE Global Conference on Signal and Information Processing*, 2013.
- [11] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal recovery on graphs: Variation minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, 2015.
- [12] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.
- [13] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Least mean squares estimation of graph signals," submitted to *IEEE Trans. on Signal and Inform. Proc. over Networks*, preprint arXiv:1602.05703, 2016.
- [14] A. H. Sayed and V. H. Nascimento, *Energy conservation and the learning ability of LMS adaptive filters*. John Wiley, New York, 2003.