Adaptive Least Mean Squares Estimation of Graph Signals

Paolo Di Lorenzo, *Member, IEEE*, Sergio Barbarossa, *Fellow, IEEE*, Paolo Banelli, *Member, IEEE*, and Stefania Sardellitti, *Member, IEEE*

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 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 performance of the proposed method. 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. Finally, we apply the proposed method to build the power spatial density cartography of a given operational region in a cognitive network environment.

Index Terms—Cognitive networks, graph signal processing, least mean squares estimation, sampling on graphs.

I. INTRODUCTION

IN MANY applications of current interest like social networks, vehicular networks, big data or biological networks, the observed signals lie over the vertices of a graph [1]. This has motivated the development of tools for analyzing signals defined over a graph, or graph signals for short [1]–[3]. Graph signal processing (GSP) aims at extending classical discrete-time signal processing tools to signals defined over a discrete domain whose elementary units (vertices) are related to each other through a graph. This framework subsumes as a very simple special case discrete-time signal processing, where the vertices are associated to time instants and edges link consecutive time instants. A peculiar aspect of GSP is that, since the signal domain is dictated by the graph topology, the analysis tools come to depend on the graph topology as well. This paves the way to a plethora of methods, each emphasizing different aspects of the problem.

P. Di Lorenzo and P. Banelli are with the Department of Engineering, University of Perugia, Perugia 06125, Italy (e-mail: paolo.dilorenzo@unipg.it; paolo.banelli@unipg.it).

S. Barbarossa and S. Sardellitti are with the Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, Rome 00184, Italy (e-mail: sergio.barbarossa@uniroma1.it; stefania.sardellitti@uniroma1.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

An important feature to have in mind about graph signals is that the signal domain is not a metric space, as in the case, for example, of biological networks, where the vertices may be genes, proteins, enzymes, etc, and the presence of an edge between two molecules means that those molecules undergo a chemical reaction. This marks a fundamental difference with respect to time signals where the time domain is inherently a metric space. Processing signals defined over a graph has been considered in [2], [4]–[6]. A central role in GSP is of course played by spectral analysis of graph signals, which passes through the introduction of the so called Graph Fourier Transform (GFT). Alternative definitions of GFT have been proposed, depending on the different perspectives used to extend classical tools [1], [2], [7]–[9]. Two basic approaches are available, proposing the projection of the graph signal onto the eigenvectors of either the graph Laplacian, see, e.g., [1], [7], [9] or of the adjacency matrix, see, e.g. [2], [10]. The first approach applies to *undirected* graphs and builds on the spectral clustering properties of the Laplacian eigenvectors and the minimization of the ℓ_2 norm graph total variation; the second approach was proposed to handle also *directed* graphs and it is based on the interpretation of the adjacency operator as the graph shift operator, which lies at the heart of all linear shift-invariant filtering methods for graph signals [11], [12]. A further very recent contribution proposes to build the graph Forier basis as the set of orthonormal signals that minimize the (directed) graph cut size [13].

After the introduction of the GFT, an uncertainty principle for graph signals was derived in [14]–[18]. with the aim of assessing the link between the spread of a signal on the vertices of the graph and on its dual domain, as defined by the GFT. A simple closed form expressions for the fundamental tradeoff between the concentrations of a signal in the graph and the transformed domains was given in [18].

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 [7], later extended in [19] and, very recently, in [10], [18], [20]–[22]. 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 reconstuction methods have been proposed, either iterative as in [20], [23], [24], or single shot, as in [10], [18]. Frame-based approaches to reconstruct signals from subsets of samples have been proposed in [7], [18], [20].

The theory developed in the last years for GSP was then applied to solve specific learning tasks, such as semi-supervised classification on graphs [25]–[27], graph dictionary learning [28], [29], learning graphs structures [30], smooth graph signal recovery from random samples [31], [32], inpainting [33], denoising [34], and community detection on graphs [35]. Finally, in [36], [37], 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.

Contribution: The goal of this paper is to propose LMS strategies for 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 [38], [39], 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 and tracking from a limited number of observations taken over a subset of vertices. An interesting feature of our proposed strategy is that this subset is allowed to vary over time, in adaptive manner. 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. Several numerical results confirm the theoretical findings, and assess the performance of the proposed strategies. Furthermore, we consider the case where the graph signal is band-limited but the bandwidth is not known beforehand; this case is critical because the selection of the sampling strategy fundamentally depends on such prior information. To cope with this issue, we propose an LMS method with adaptive sampling, which estimates and tracks the signal support in the (graph) frequency domain, while at the same time adapting the graph sampling strategy. Numerical results illustrate the tracking capability of the aforementioned method in the presence of time-varying graph signals. As an example, we apply the proposed strategy to estimate and track the spatial distribution of the electromagnetic power in a cognitive radio framework. The resulting graph signal turns out to be smooth, i.e. the largest part of its energy is concentrated at low frequencies, but it is not perfectly band-limited. As a consequence, recovering the overall signal from a subset of samples is inevitably affected by aliasing [22]. Numerical results show the tradeoff between complexity, i.e. number of samples used for processing, and mean-square performance of the proposed strategy, when applied to such cartography task. Intuitively, processing with a larger bandwidth and a (consequent) larger number of samples, improves the performance of the algorithm, at the price of a larger complexity.

The paper is organized as follows. In Section II, we introduce some basic graph signal processing tools, which will be useful for the following derivations. Section III introduces the proposed LMS algorithm for graph signals, illustrates its mean-square analysis, and derives useful graph sampling strategies. Then, in Section IV we illustrate the proposed LMS strategy with adaptive sampling, while Section V considers the application to power density cartography. Finally, Section VI draws some conclusions.

II. GRAPH SIGNAL PROCESSING TOOLS

We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a set of N nodes $\mathcal{V} = \{1, 2, ..., N\}$, 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 $a_{ij}, i, j = 1, \ldots, N$. 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:

$$\mathbf{L} = \mathbf{K} - \mathbf{A}. \tag{1}$$

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 [40] that the eigenvectors of $\mathbf{L}$ are well suited for representing clusters, since they minimize the ℓ_2 norm graph total variation.

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

$$\boldsymbol{x} = \mathbf{U}\boldsymbol{s} \tag{2}$$

where $\boldsymbol{s}$ 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 allowed to vary arbitrarily from one cluster to the other, the representation in (2) is compact, i.e. the only nonzero (or approximately nonzero) entries of $\boldsymbol{s}$ are the ones associated to the clusters.

The GFT $\boldsymbol{s}$ of a signal $\boldsymbol{x}$ is defined as the projection onto the orthogonal set of vectors $\{\boldsymbol{u}_i\}_{i=1,\ldots,N}$ [1], i.e.

$$\mathbf{GFT}: \quad \boldsymbol{s} = \mathbf{U}^H \boldsymbol{x}. \tag{3}$$

The GFT has been defined in alternative ways, see, e.g., [1], [2], [9], [10]. 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. We denote the support of $\boldsymbol{s}$ in (2) as $\mathcal{F} = \{i \in \{1, \ldots, N\} : s_i \neq 0\}$, and the *bandwidth* of the graph signal $\boldsymbol{x}$ is defined as the cardinality of $\mathcal{F}$, i.e. $|\mathcal{F}|$. Clearly, combining (2) with (3), if the signal $\boldsymbol{x}$ exhibits a clustering behavior, in the sense specified above, computing its GFT is the way to recover the sparse vector $\boldsymbol{s}$ in (2).

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

$$\mathbf{D}_{\mathcal{S}} = \text{diag}\{\mathbf{1}_{\mathcal{S}}\}, \tag{4}$$

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

$$\mathbf{B}_{\mathcal{F}} = \mathbf{U}\mathbf{\Sigma}_{\mathcal{F}}\mathbf{U}^H, \tag{5}$$

where $\mathbf{\Sigma}_{\mathcal{F}}$ is a diagonal matrix defined as $\mathbf{\Sigma}_{\mathcal{F}} = \text{diag}\{\mathbf{1}_{\mathcal{F}}\}$. It is immediate to check that both matrices $\mathbf{D}_{\mathcal{S}}$ and $\mathbf{B}_{\mathcal{F}}$ are self-adjoint and idempotent, and so they represent orthogonal

projectors. The space of all signals whose GFT is exactly supported on the set $\mathcal{F}$ is known as the *Paley-Wiener space* for the set $\mathcal{F}$ [7]. We denote by $\mathcal{B}_{\mathcal{F}} \subseteq L_2(\mathcal{G})$ the set of all finite ℓ_2-norm signals belonging to the Paley-Wiener space associated to the frequency subset $\mathcal{F}$. Similarly, we denote by $\mathcal{D}_{\mathcal{S}} \subseteq L_2(\mathcal{G})$ the set of all finite ℓ_2-norm signals with support on the vertex subset $\mathcal{S}$. In the rest of the paper, whenever there will be no ambiguities 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 $\mathcal{S} \cap \bar{\mathcal{S}} = \emptyset$. Similarly, we define the complement set of $\mathcal{F}$ as $\overline{\mathcal{F}}$. Thus, we define the vertex-projector onto $\overline{\mathcal{S}}$ as $\overline{\mathbf{D}}$ and, similarly, the frequency projector onto the frequency domain $\overline{\mathcal{F}}$ as $\overline{\mathbf{B}}$.

Exploiting the localization operators in (4) and (5), we say that a vector $\boldsymbol{x}$ is perfectly localized over the subset $\mathcal{S} \subseteq \mathcal{V}$ if

$$\mathbf{D}\boldsymbol{x} = \boldsymbol{x}, \tag{6}$$

with $\mathbf{D}$ defined as in (4). Similarly, a vector $\boldsymbol{x}$ is perfectly localized over the frequency set $\mathcal{F}$ (i.e. band-limited on $\mathcal{F}$) if

$$\mathbf{B}\boldsymbol{x} = \boldsymbol{x}, \tag{7}$$

with $\mathbf{B}$ given in (5). The localization properties of graph signals were studied in [22] and later extended in [18] 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. The conditions for having perfect localization are stated in the following theorem, which we report here for completeness of exposition; its proof can be found in [22].

Theorem 1: There is a vector $\boldsymbol{x}$, perfectly localized over both vertex set $\mathcal{S}$ and frequency set $\mathcal{F}$ (i.e. $\boldsymbol{x} \in \mathcal{B}_{\mathcal{F}} \cap \mathcal{D}_{\mathcal{S}}$) if and only if the operator $\mathbf{BDB}$ (or $\mathbf{DBD}$) has an eigenvalue equal to one; in such a case, $\boldsymbol{x}$ is an eigenvector of $\mathbf{BDB}$ associated to the unitary eigenvalue.

Equivalently, the perfect localization properties can be expressed in terms of the operators $\mathbf{BD}$ and $\mathbf{DB}$. Indeed, since the operators $\mathbf{BD}$ and $\mathbf{DB}$ have the same singular values [18], perfect localization onto the sets $\mathcal{S}$ and $\mathcal{F}$ can be achieved if and only if

$$\|\mathbf{BD}\|_2 = \|\mathbf{DB}\|_2 = 1. \tag{8}$$

Building on these previous results on GSP, in the next section we introduce the proposed LMS strategy for adaptive estimation of graph signals.

III. LMS ESTIMATION OF GRAPH SIGNALS

The least mean square algorithm, introduced by Widrow and Hoff [41], is one of the most popular methods for adaptive filtering. Its applications include echo cancelation, channel equalization, interference cancelation and so forth. Although there exist algorithms with faster convergence rates such as the Recursive Least Square (RLS) methods [38], LMS-type methods are popular because of their ease of implementation, low computational costs and robustness. For these reasons, a huge amount of research was produced in the last decades focusing on improving

the performance of LMS-type methods, exploiting in many cases some prior information that is available on the observed signals. For instance, if the observed signal is known to be sparse in some domain, such prior information can help improve the estimation performance, as demonstrated in many recent efforts in the area of compressed sensing [42], [43]. Some of the early works that mix adaptation with sparsity-aware reconstruction include methods that rely on the heuristic selection of active taps [44], and on sequential partial updating techniques [45]; some other methods assign proportional step-sizes to different taps according to their magnitudes, such as the proportionate normalized LMS (PNLMS) algorithm and its variations [46]. In subsequent studies, motivated by the LASSO technique [47] and by connections with compressed sensing [43], [48], several algorithms for sparse adaptive filtering have been proposed based on LMS [49], RLS [50], and projection-based methods [51]. Finally, sparsity aware distributed methods were proposed in [52]–[56].

In this paper, we aim to exploit the intrinsic sparsity that is present in band-limited graph signals, thus designing proper sampling strategies that guarantee adaptive reconstruction of the signal, with guaranteed mean-square performance, from a limited number of observation sampled from the graph. To this aim, let us consider a signal $\boldsymbol{x}_0 \in \mathbb{C}^N$ defined over the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. 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 $\mathcal{F}$. Later on, we will relax such an assumption. Let us consider partial observations of signal $\boldsymbol{x}_0$, 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:

$$\boldsymbol{y}[n] = \mathbf{D}\left(\boldsymbol{x}_0 + \boldsymbol{v}[n]\right) = \mathbf{DB}\boldsymbol{x}_0 + \mathbf{D}\boldsymbol{v}[n] \tag{9}$$

where $\mathbf{D}$ is the vertex-limiting operator defined in (4), which takes nonzero values only in the set $\mathcal{S}$, and $\boldsymbol{v}[n]$ is a zero-mean, additive noise with covariance matrix $\mathbf{C}_v$. The second equality in (9) comes from the bandlimited assumption, i.e. $\mathbf{B}\boldsymbol{x}_0 = \boldsymbol{x}_0$, with $\mathbf{B}$ denoting the operator in (5) that projects onto the (known) frequency set $\mathcal{F}$. We remark that, differently from linear observation models commonly used in adaptive filtering theory [38], the model in (9) has a free sampling parameter $\mathbf{D}$ that can be properly selected by the designer, with the aim of reducing the computational/memory burden while still guaranteeing theoretical performance, as we will illustrate in the following sections. The estimation task consists in recovering the band-limited graph signal $\boldsymbol{x}_0$ from the noisy, streaming, and partial observations $\boldsymbol{y}[n]$ in (9). Following an LMS approach [41], the optimal estimate for $\boldsymbol{x}_0$ can be found as the vector that solves the following optimization problem:

$$\min_{\boldsymbol{x}} \quad \mathbb{E} \|\boldsymbol{y}[n] - \mathbf{DB}\boldsymbol{x}\|^2 \tag{10}$$

$$\text{s.t.} \quad \mathbf{B}\boldsymbol{x} = \boldsymbol{x},$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator. The solution of problem (10) minimizes the mean-squared error and has a bandwidth limited to the frequency set $\mathcal{F}$. For stationary $\boldsymbol{y}[n]$, the optimal solution of (10) is given by the vector $\hat{\boldsymbol{x}}$ that satisfies

---

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:

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

---

the normal equations [38]:

$$\mathbf{B} \mathbf{D} \mathbf{B} \, \hat{\boldsymbol{x}} = \mathbf{B} \mathbf{D} \, \mathbb{E}\{\boldsymbol{y}[n]\}. \qquad (11)$$

Exploiting (9), this statement can be readily verified noticing that $\hat{\boldsymbol{x}}$ minimizes the objective function of (10) and is bandlimited, i.e. it satisfies $\mathbf{B}\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}$. Nevertheless, in many linear regression applications involving online processing of data, the expectation $\mathbb{E}\{\boldsymbol{y}[n]\}$ may be either unavailable or time-varying, and thus impossible to update continuously. For this reason, adaptive solutions relying on instantaneous information are usually adopted in order to avoid the need to know the signal statistics beforehand. A typical solution proceeds to optimize (10) by means of a steepest-descent procedure. Thus, letting $\boldsymbol{x}[n]$ be the instantaneous estimate of vector $\boldsymbol{x}_0$, 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 $\mathbf{B}\boldsymbol{x}[n] = \boldsymbol{x}[n]$ (i.e., $\boldsymbol{x}[n]$ is band-limited) for all n. Algorithm 1 starts from an initial signal that belongs 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 (12) crucially depend on the choice of the sampling set $\mathcal{S}$, which defines the structure of the operator $\mathbf{D}$ [cf. (4)]. To shed light on the theoretical behavior of Algorithm 1, in the following sections we illustrate how the choice of the operator $\mathbf{D}$ affects the reconstruction capability, mean-square stability, and steady-state performance of the proposed LMS strategy.

A. Reconstruction Properties

It is well known from adaptive filters theory [38] that the LMS algorithm in (12) is a stochastic approximation method for the solution of problem (10), which enables convergence in the mean-sense to the true vector $\boldsymbol{x}_0$ (if the step-size μ is chosen sufficiently small), 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 (12) 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 $\boldsymbol{x}_0$ 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 2: Problem (10) admits a unique solution, i.e. any band-limited signal $\boldsymbol{x}_0$ can be reconstructed from its samples taken in the set $\mathcal{S}$, if and only if

$$\left\| \overline{\mathbf{D}} \mathbf{B} \right\|_2 < 1, \qquad (13)$$

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

Proof: From (11), exploiting the relation $\mathbf{D} = \mathbf{I} - \overline{\mathbf{D}}$, it holds

$$\left(\mathbf{I} - \mathbf{B} \overline{\mathbf{D}} \mathbf{B} \right) \boldsymbol{x}_0 = \mathbf{B} \mathbf{D} \, \mathbb{E}\{\boldsymbol{y}[n]\}. \qquad (14)$$

Hence, it is possible to recover $\boldsymbol{x}_0$ from (14) if $\mathbf{I} - \mathbf{B} \overline{\mathbf{D}} \mathbf{B}$ is invertible. This happens if the sufficient condition (13) holds true. Conversely, if $\|\overline{\mathbf{D}} \mathbf{B}\|_2 = 1$ (or, equivalently, $\|\mathbf{B} \overline{\mathbf{D}}\|_2 = 1$), from (8) we know that there exist band-limited signals that are perfectly localized over $\overline{\mathcal{S}}$. This implies that, if we sample one of such signals over the set $\mathcal{S}$, we get only zero values and then it would be impossible to recover $\boldsymbol{x}_0$ from those samples. This proves that condition (13) is also necessary. ∎

A necessary condition that enables reconstruction, i.e. the non-existence of a non-trivial vector $\boldsymbol{x}$ satisfying $\mathbf{D} \mathbf{B} \boldsymbol{x} = \mathbf{0}$, is that $|\mathcal{S}| \geq |\mathcal{F}|$. However, this condition is not sufficient, because matrix $\mathbf{D} \mathbf{B}$ in (9) may loose rank, or easily become ill-conditioned, depending on the graph topology and sampling strategy defined by $\mathbf{D}$. This suggests that the location of samples plays a key role in the performance of the LMS reconstruction algorithm in (12). For this reason, in Section III-D we will consider a few alternative sampling strategies satisfying different optimization criteria.

B. Mean-Square Analysis

When condition (13) holds true, Algorithm 1 can reconstruct the graph signal from a subset of samples. 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 $\boldsymbol{x}[n]$ as realizations of a random process and analyze the performance of the LMS algorithm in terms of its mean-square behavior. Let $\widetilde{\boldsymbol{x}}[n] = \boldsymbol{x}[n] - \boldsymbol{x}_0$ be the error vector at time n. Subtracting $\boldsymbol{x}_0$ from the left and right hand sides of (12), using (9) and relation $\mathbf{B} \widetilde{\boldsymbol{x}}[n] = \widetilde{\boldsymbol{x}}[n]$, we obtain:

$$\widetilde{\boldsymbol{x}}[n + 1] = \left(\mathbf{I} - \mu \, \mathbf{B} \mathbf{D} \mathbf{B} \right) \widetilde{\boldsymbol{x}}[n] + \mu \, \mathbf{B} \mathbf{D} \boldsymbol{v}[n]. \qquad (15)$$

Applying a GFT to each side of (15) (i.e., multiplying by $\mathbf{U}^H$), and exploiting the structure of matrix $\mathbf{B}$ in (5), we obtain

$$\widetilde{\boldsymbol{s}}[n + 1] = \left(\mathbf{I} - \mu \, \boldsymbol{\Sigma} \mathbf{U}^H \mathbf{D} \mathbf{U} \boldsymbol{\Sigma} \right) \widetilde{\boldsymbol{s}}[n] + \mu \, \boldsymbol{\Sigma} \mathbf{U}^H \mathbf{D} \boldsymbol{v}[n], \quad (16)$$

where $\widetilde{\boldsymbol{s}}[n] = \mathbf{U}^H \widetilde{\boldsymbol{x}}[n]$ is the GFT of the error $\widetilde{\boldsymbol{x}}[n]$. From (16) and the definition of $\boldsymbol{\Sigma}$ in (5), since $\widetilde{s}_i[n] = 0$ for all $i \notin \mathcal{F}$, we can analyze the behavior of the error recursion (16) only on the support of $\widetilde{\boldsymbol{s}}[n]$, i.e. $\widehat{\boldsymbol{s}}[n] = \{\widetilde{s}_i[n], i \in \mathcal{F}\} \in \mathbb{C}^{|\mathcal{F}|}$. 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 $\mathcal{F}$, the error recursion (16) can be rewritten in compact form as:

$$\widehat{\boldsymbol{s}}[n + 1] = \left(\mathbf{I} - \mu \, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \right) \widehat{\boldsymbol{s}}[n] + \mu \, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \boldsymbol{v}[n]. \qquad (17)$$

The evolution of the error $\widehat{\boldsymbol{s}}[n] = \mathbf{U}_{\mathcal{F}}^H \widetilde{\boldsymbol{x}}[n]$ in the compact transformed domain is totally equivalent to the behavior of $\widetilde{\boldsymbol{x}}[n]$ from a mean-square error point of view. Thus, using energy conservation arguments [57], we consider a general weighted squared error sequence $\|\widehat{\boldsymbol{s}}[n]\|_{\boldsymbol{\Phi}}^2 = \widehat{\boldsymbol{s}}[n]^H \boldsymbol{\Phi} \widehat{\boldsymbol{s}}[n]$,

where $\mathbf{\Phi} \in \mathbb{C}^{|\mathcal{F}| \times |\mathcal{F}|}$ 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 $\mathbf{\Phi}$. Then, from (17) we can establish the following variance relation:

$$
\begin{aligned}
\mathbb{E}\|\widehat{s}[n+1]\|_{\mathbf{\Phi}}^2 &= \mathbb{E}\|\widehat{s}[n]\|_{\mathbf{\Phi}'}^2 + \mu^2 \, \mathbb{E}\{\boldsymbol{v}[n]^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \mathbf{\Phi} \mathbf{U}_{\mathcal{F}}^H \mathbf{D}\boldsymbol{v}[n]\} \\
&= \mathbb{E}\|\widehat{s}[n]\|_{\mathbf{\Phi}'}^2 + \mu^2 \, \mathrm{Tr}(\mathbf{\Phi} \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{C}_v \mathbf{D} \mathbf{U}_{\mathcal{F}})
\end{aligned}
\tag{18}
$$

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

$$
\mathbf{\Phi}' = \left(\mathbf{I} - \mu\, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right) \mathbf{\Phi} \left(\mathbf{I} - \mu\, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right). \tag{19}
$$

Let $\boldsymbol{\varphi} = \mathrm{vec}(\mathbf{\Phi})$ and $\boldsymbol{\varphi}' = \mathrm{vec}(\mathbf{\Phi}')$, where the notation $\mathrm{vec}(\cdot)$ stacks the columns of $\mathbf{\Phi}$ on top of each other and $\mathrm{vec}^{-1}(\cdot)$ is the inverse operation. We will use interchangeably the notation $\|\widehat{s}\|_{\mathbf{\Phi}}^2$ and $\|\widehat{s}\|_{\boldsymbol{\varphi}}^2$ to denote the same quantity $\widehat{s}^H \mathbf{\Phi} \widehat{s}$. Exploiting the Kronecker product property

$$
\mathrm{vec}\left(\mathbf{X} \mathbf{\Phi} \mathbf{Y}\right) = (\mathbf{Y}^H \otimes \mathbf{X})\mathrm{vec}(\mathbf{\Phi}),
$$

and the trace property

$$
\mathrm{Tr}\left(\mathbf{\Phi} \mathbf{X}\right) = \mathrm{vec}(\mathbf{X}^H)^T \mathrm{vec}(\mathbf{\Phi}),
$$

in the relation (18), we obtain:

$$
\mathbb{E}\|\widehat{s}[n+1]\|_{\boldsymbol{\varphi}}^2 = \mathbb{E}\|\widehat{s}[n]\|_{\mathbf{Q}\boldsymbol{\varphi}}^2 + \mu^2 \mathrm{vec}(\mathbf{G})^T \boldsymbol{\varphi} \tag{20}
$$

where

$$
\mathbf{G} = \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{C}_v \mathbf{D} \mathbf{U}_{\mathcal{F}} \tag{21}
$$

$$
\mathbf{Q} = \left(\mathbf{I} - \mu\, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right) \otimes \left(\mathbf{I} - \mu\, \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right). \tag{22}
$$

The following theorem guarantees the asymptotic mean-square stability (i.e., convergence in the mean and mean-square error sense) of the LMS algorithm in (12).

Theorem 3: Assume model (9) holds. Then, for any bounded initial condition, the LMS strategy (12) asymptotically converges in the mean-square error sense if the sampling operator $\mathbf{D}$ and the step-size μ are chosen to satisfy (13) and

$$
0 < \mu < \frac{2}{\lambda_{\max}\left(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right)}, \tag{23}
$$

with $\lambda_{\max}(\mathbf{A})$ denoting the maximum eigenvalue of the symmetric matrix $\mathbf{A}$. Furthermore, it follows that, for sufficiently small step-sizes:

$$
\limsup_{n \to \infty} \, \mathbb{E}\|\widehat{s}[n]\|^2 = O(\mu). \tag{24}
$$

Proof: Letting $\boldsymbol{r} = \mathrm{vec}(\mathbf{G})$, recursion (20) can be equivalently recast as:

$$
\mathbb{E}\|\widehat{s}[n]\|_{\boldsymbol{\varphi}}^2 = \mathbb{E}\|\widehat{s}[0]\|_{\mathbf{Q}^n \boldsymbol{\varphi}}^2 + \mu^2 \boldsymbol{r}^T \sum_{l=0}^{n-1} \mathbf{Q}^l \boldsymbol{\varphi} \tag{25}
$$

where $\mathbb{E}\|\widehat{s}[0]\|$ denotes the initial condition. We first note that if $\mathbf{Q}$ is stable, $\mathbf{Q}^n \to \mathbf{0}$ as $n \to \infty$. In this way, the first term on the RHS of (25) vanishes asymptotically. At the same time, the convergence of the second term on the RHS of (25) depends only on the geometric series of matrices $\sum_{l=0}^{n-1} \mathbf{Q}^l$, which is known to be convergent to a finite value if the matrix $\mathbf{Q}$ is a

stable matrix [58]. In summary, if $\mathbf{Q}$ is stable, the RHS of (25) asymptotically converges to a finite value, and we conclude that $\mathbb{E}\|\widehat{s}[n]\|^2$ will converge to a steady-state value. From (22), we deduce that $\mathbf{Q}$ is stable if matrix $\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$ is stable as well. This holds true under the two following conditions. The first condition is that matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} = \mathbf{I} - \mathbf{U}_{\mathcal{F}}^H \overline{\mathbf{D}} \mathbf{U}_{\mathcal{F}}$ must have full rank, i.e. (13) holds true, where we have exploited the relation

$$
\|\overline{\mathbf{D}} \mathbf{U}_{\mathcal{F}}\| = \|\overline{\mathbf{D}} \mathbf{U} \mathbf{\Sigma}\| = \|\overline{\mathbf{D}} \mathbf{B}\|.
$$

Now recalling that, for any Hermitian matrix $\mathbf{X}$, it holds $\|\mathbf{X}\| = \rho(\mathbf{X})$ [58], with $\rho(\mathbf{X})$ denoting the spectral radius of $\mathbf{X}$, the second condition guaranteing the stability of $\mathbf{Q}$ is that $\|\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\| < 1$, which holds true for any step-sizes satisfying (23). This concludes the first part of the Theorem.

We now prove the second part. Selecting $\boldsymbol{\varphi} = \mathrm{vec}(\mathbf{I})$ in (25), we obtain the following bound

$$
\mathbb{E}\|\widehat{s}[n]\|^2 \leq \mathbb{E}\|\widehat{s}[0]\|_{\mathbf{Q}^n \mathrm{vec}(\mathbf{I})}^2 + \mu^2 c \sum_{l=0}^{n} \|\mathbf{Q}\|^l \tag{26}
$$

where $c = \|\boldsymbol{r}\| \|\mathrm{vec}(\mathbf{I})\|$. Taking the limit of (26) as $n \to \infty$, since $\|\mathbf{Q}\| < 1$ if conditions (13) and (23) hold, we obtain

$$
\lim_{n \to \infty} \mathbb{E}\|\widehat{s}[n]\|^2 \leq \frac{\mu^2 c}{1 - \|\mathbf{Q}\|}. \tag{27}
$$

From (22), we have

$$
\begin{aligned}
\|\mathbf{Q}\| &= \|\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\|^2 = \left(\rho\left(\mathbf{I} - \mu \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}\right)\right)^2 \\
&\leq \max\left\{(1 - \mu\delta)^2, (1 - \mu\nu)^2\right\} \\
&\overset{(a)}{\leq} 1 - 2\mu\nu + \mu^2\delta^2
\end{aligned}
\tag{28}
$$

where $\delta = \lambda_{\max}(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}})$, $\nu = \lambda_{\min}(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}})$, and in (a) we have exploited $\delta \geq \nu$. Substituting (28) in (27), we get

$$
\lim_{n \to \infty} \mathbb{E}\|\widehat{s}[n]\|^2 \leq \frac{\mu c}{2\nu - \mu\delta^2}. \tag{29}
$$

It is easy to check that the upper bound (29) does not exceed $\mu c / \nu$ for any stepsize $0 < \mu < \nu/\delta^2$. Thus, we conclude that (24) holds for sufficiently small step-sizes. ∎

C. Steady-State Performance

Taking the limit of (20) as $n \to \infty$ (assuming conditions (13) and (23) hold true), we obtain:

$$
\lim_{n \to \infty} \mathbb{E}\|\widehat{s}[n]\|_{(\mathbf{I} - \mathbf{Q})\boldsymbol{\varphi}}^2 = \mu^2 \mathrm{vec}(\mathbf{G})^T \boldsymbol{\varphi}. \tag{30}
$$

Expression (30) is a useful result: it allows us to derive several performance metrics through the proper selection of the free weighting parameter $\boldsymbol{\varphi}$ (or $\mathbf{\Phi}$). For instance, let us assume that one wants to evaluate the steady-state mean square deviation (MSD) of the LMS strategy in (12). Thus, selecting $\boldsymbol{\varphi} = (\mathbf{I} - \mathbf{Q})^{-1} \mathrm{vec}(\mathbf{I})$ in (30), we obtain

$$
\begin{aligned}
\mathrm{MSD} &= \lim_{n \to \infty} \mathbb{E}\|\widetilde{x}[n]\|^2 = \lim_{n \to \infty} \mathbb{E}\|\widehat{s}[n]\|^2 \\
&= \mu^2 \mathrm{vec}(\mathbf{G})^T (\mathbf{I} - \mathbf{Q})^{-1} \mathrm{vec}(\mathbf{I}).
\end{aligned}
\tag{31}
$$

Sampling strategy 1: Minimization of MSD.

$InputData$: M, the number of samples.
$OutputData$: $\mathcal{S}$, 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

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

$InputData$: M, the number of samples.
$OutputData$: $\mathcal{S}$, 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

If instead one is interested in evaluating the mean square deviation obtained by the LMS algorithm in (12) when reconstructing the value of the signal associated to k-th vertex of the graph, selecting $\boldsymbol{\varphi} = (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}})$ in (30), we obtain

$$\text{MSD}_k = \lim_{n\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{x}}[n]\|_{\mathbf{E}_k}^2 = \lim_{n\to\infty} \mathbb{E}\|\widehat{\boldsymbol{s}}[n]\|_{\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}}}^2$$

$$= \mu^2 \text{vec}(\mathbf{G})^T (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{U}_{\mathcal{F}}^H \mathbf{E}_k \mathbf{U}_{\mathcal{F}}), \qquad (32)$$

where $\mathbf{E}_k = \text{diag}\{\boldsymbol{e}_k\}$, with $\boldsymbol{e}_k \in \mathbb{R}^N$ 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

As illustrated in the previous sections, the properties of the proposed LMS algorithm in (12) 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. In particular, we will see that, when sampling signals defined on graphs, besides choosing the right number of samples, whenever possible it is also fundamental to have a strategy indicating *where* to sample, as the samples' location plays a key role in the performance of the reconstruction algorithm in (12). To select the best sampling strategy, one should optimize some performance criterion, e.g. the MSD in (31), with respect to the sampling set $\mathcal{S}$, or, equivalently, the vertex limiting operator $\mathbf{D}$. However, since this formulation translates inevitably into a selection 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 numerically 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 (31) 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 (31) is inevitably rank deficient. Then, in this case, the criterion builds on the pseudo-inverse of the matrix $\mathbf{I} - \mathbf{Q}$ in (31), 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.

Greedy Selection - Maximum $\left| \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \right|_+$: 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 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 $\left| \mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \right|_+$. In the sequel, we motivate the rationale underlying this strategy. Let us consider the eigendecomposition $\mathbf{Q} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. From (31), we obtain:

$$\text{MSD} = \mu^2 \text{vec}(\mathbf{G})^T (\mathbf{I} - \mathbf{Q})^{-1} \text{vec}(\mathbf{I})$$

$$= \mu^2 \text{vec}(\mathbf{G})^T \mathbf{V} (\mathbf{I} - \boldsymbol{\Lambda})^{-1} \mathbf{V}^H \text{vec}(\mathbf{I})$$

$$= \mu^2 \sum_{i=1}^{|\mathcal{F}|^2} \frac{p_i \cdot q_i}{1 - \lambda_i(\mathbf{Q})} \qquad (33)$$

where $\boldsymbol{p} = \{p_i\} = \mathbf{V}^H \text{vec}(\mathbf{G})$, $\boldsymbol{q} = \{q_i\} = \mathbf{V}^H \text{vec}(\mathbf{I})$. From (33), we notice how the MSD of the LMS algorithm in (12) strongly depends on the values assumed by the eigenvalues $\lambda_i(\mathbf{Q})$, $i = 1, \dots, |\mathcal{F}|^2$. In particular, we would like to design matrix $\mathbf{Q}$ in (22) such that its eigenvalues are as far as possible from 1. From (22), it is easy to understand that

$$\lambda_i(\mathbf{Q}) = \left(1 - \mu\lambda_k(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}})\right) \left(1 - \mu\lambda_l(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}})\right)$$

$k, l = 1, \dots, |\mathcal{F}|$. Thus, requiring $\lambda_i(\mathbf{Q})$, $i = 1, \dots, |\mathcal{F}|^2$, to be as far as possible from 1 translates in designing the matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \in \mathbb{C}^{|\mathcal{F}| \times |\mathcal{F}|}$ such that its eigenvalues are as far as possible from zero. Thus, a possible surrogate criterion for the approximate minimization of (33) can be formulated as the selection of the sampling set $\mathcal{S}$ (i.e. operator $\mathbf{D}$) that maximizes the determinant (i.e. the product of all eigenvalues) of the matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$. When $|\mathcal{S}| < |\mathcal{F}|$, matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$ is inevitably rank deficient, and the strategy builds on the pseudo-determinant of $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$. Of course, when $|\mathcal{S}| \geq |\mathcal{F}|$, the pseudo determinant coincides with the determinant. The resulting algorithm is summarized in the table entitled "Sampling strategy 2". In the

Sampling strategy 3: Maximization of $\lambda_{\min}^+ \left(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \right)$.

$InputData$: M, the number of samples.
$OutputData$: $\mathcal{S}$, the sampling set.
$Function$:   initialize $\mathcal{S} \equiv \emptyset$
            while $|\mathcal{S}| < M$
            $s = \arg\max_j \ \lambda_{\min}^+ \left(\mathbf{U}_{\mathcal{F}}^H \mathbf{D}_{\mathcal{S} \cup \{j\}} \mathbf{U}_{\mathcal{F}} \right);$
            $\mathcal{S} \leftarrow \mathcal{S} \cup \{s\};$
            end
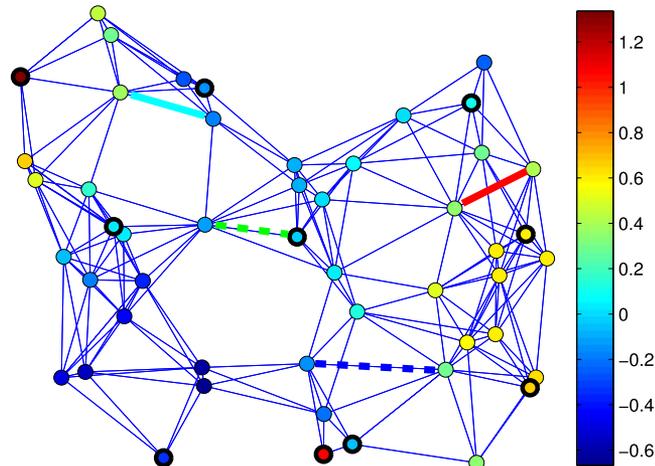
Fig. 1.   Example of graph signal and sampling.

sequel, we will refer to this method as the Max-Det sampling strategy.

Greedy Selection - Maximum $\lambda_{\min}^+ (\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}})$*:* Finally, using similar arguments as before, a further surrogate criterion for the minimization of (33) can be formulated as the maximization of the minimum nonzero eigenvalue of the matrix $\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}}$, which is denoted by $\lambda_{\min}^+ \left(\mathbf{U}_{\mathcal{F}}^H \mathbf{D} \mathbf{U}_{\mathcal{F}} \right)$. This greedy strategy exploits the same idea of the sampling method introduced in [10] in the case of batch signal reconstruction. The resulting algorithm is summarized in the table entitled "Sampling strategy 3". We will refer to this method as the Max-$\lambda_{\min}$ sampling strategy.

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.

E. Numerical Results

In this section, we first illustrate some numerical results aimed at confirming the theoretical results in (31) and (32). Then, we will illustrate how the sampling strategy affects the performance of the proposed LMS algorithm in (12). Finally, we will evaluate the effect of a graph mismatching in the performance of the proposed algorithm.

Performance: 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 (9) is zero-mean, Gaussian, with a diagonal
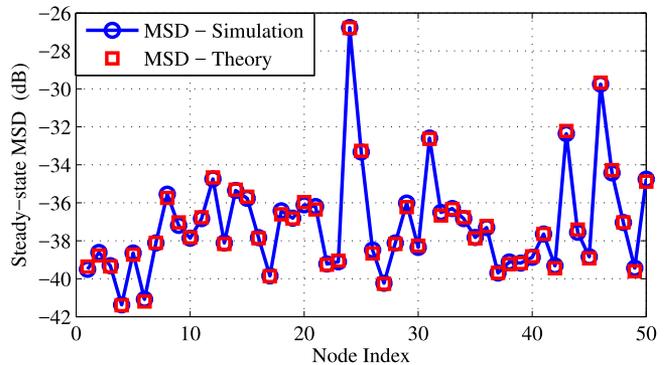
Fig. 2.   Comparison between theoretical MSD in (30) and simulation results, at each vertex of the graph. The theoretical expressions match well with the numerical results.
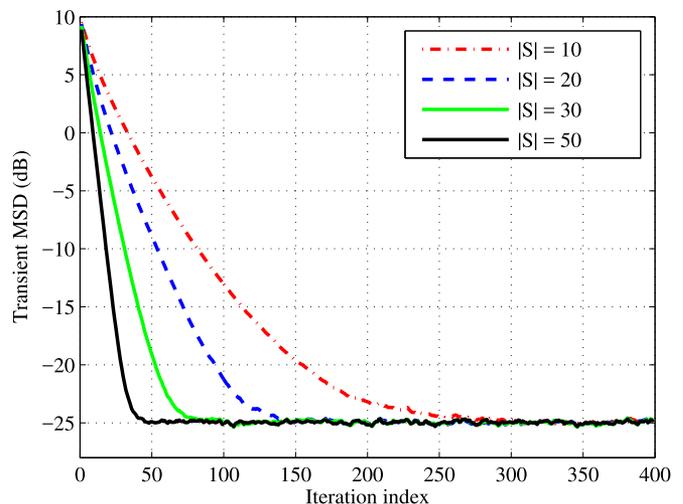
Fig. 3.   Transient MSD, for different number of samples $|\mathcal{S}|$. Increasing the number of samples, the learning rate improves.

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 (32), 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 (13) and (23). As we can notice from Fig. 2, the theoretical predictions match well the simulation results.

Effect of sampling strategies: It is fundamental to assess the performance of the LMS algorithm in (12) 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,
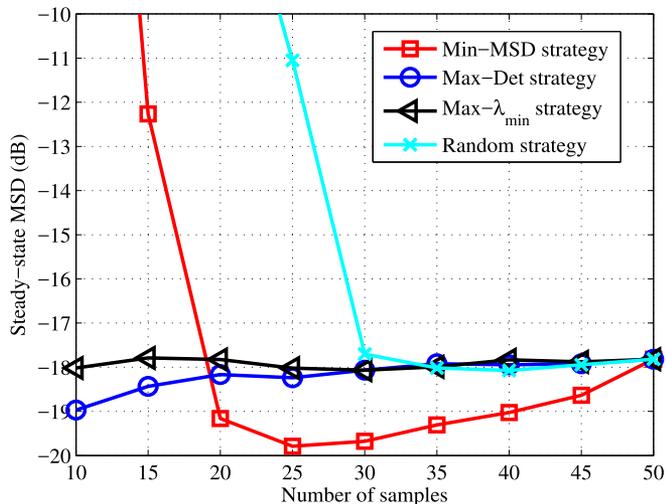
Fig. 4.   Steady-state MSD versus number of samples, for different sampling strategies.

Fig. 5.   Transient MSD versus iteration index, for different links removed from the original graph in Fig. 1.

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 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 (12) comparing the performance obtained by four different sampling strategies, namely: a) the Max-Det strategy; b) the Max-$\lambda_{\min}$ strategy; c) the Min-MSD strategy; and d) 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. Comparing the other sampling strategies, we notice from Fig. 4 that the Max-Det and Max-$\lambda_{\min}$ strategies perform well also at low number of samples ($|\mathcal{S}| = 10$ is the minimum number of samples that allows signal reconstruction). As expected, the Max-Det strategy outperforms the Max-$\lambda_{\min}$ strategy, because it considers all the modes of the MSD in (33), as opposed to the single mode associated to the minimum eigenvalue considered by the Max-$\lambda_{\min}$ strategy. 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 above the limit $|\mathcal{S}| = |\mathcal{F}| = 10$, the Min-MSD strategy outperforms all other methods. This happens because the Min-MSD strategy takes into account information from *both* graph topology and spatial distribution of the observation noise (cf. (31)). Thus, when the number of samples is large enough to have sufficient degrees of freedom in selecting the samples' location, the Min-MSD strategy has the capability of selecting the vertexes in a good position to enable a well-conditioned signal recovery,
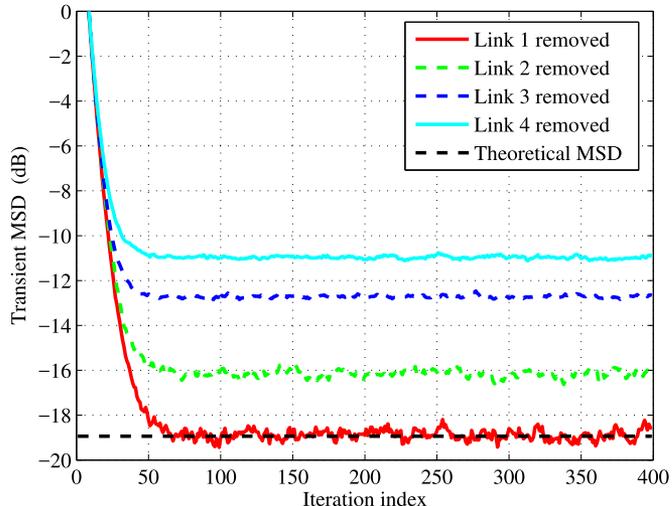
with possibly low additive noise, thus improving the overall performance of the LMS algorithm in (12). Conversely, when the number of samples is very close to its minimum value, the Min-MSD criterion may give rise to ill-conditioning of the signal recovery strategy because the low noise samples may be in sub-optimal positions with respect to signal recovery. This explains its losses with respect to Max-Det and Max-λ_{min} strategies, for low values of the number of samples. 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.

Effect of graph mismatching: In this last example, we aim at illustrating how the performance of the proposed method is affected by a graph mismatching during the processing. To this aim, we take as a benchmark the graph signal in Fig. 1, where the signal bandwidth is set equal to $|\mathcal{F}| = 10$. The bandwidth defines also the sampling operator $\mathbf{D}$, which is selected through the Max-Det strategy, introduced in Section III-D, using $|\mathcal{S}| = 10$ samples. Now, we assume that the LMS processing in (12) is performed keeping fixed the sampling operator $\mathbf{D}$, while adopting an operator $\mathbf{B}$ in (5) that uses the same bandwidth as for the benchmark case (i.e., the same matrix $\mathbf{\Sigma}_{\mathcal{F}}$), but different GFT operators $\mathbf{U}$, which are generated as the eigenvectors of Laplacian matrices associated to graphs that differs from the benchmark in Fig. 1 for one (removed) link. The aim of this simulation is to quantify the effect of a single link removal on the performance of the LMS strategy in (12). Thus, in Fig. 5, we report the transient MSD versus the iteration index of the proposed LMS strategy, considering four different links that are removed from the original graph. The four removed links are those shown in Fig. 1 using thicker lines; the colors and line styles are associated to the four behaviors of the transient MSD in Fig. 5. The results are averaged over 100 independent simulations, using a step-size $\mu = 0.5$. The theoretical performance in (31) achieved by the ideal LMS, i.e. the one perfectly matched to the graph, are also reported as a benchmark. As we can see from

---

Algorithm 2: LMS with Adaptive Graph Sampling.

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

1) $s[n+1] = \boldsymbol{T}_{\lambda\mu}\left(s[n] + \mu\,\mathbf{U}^H\mathbf{D}[n]\left(\boldsymbol{y}[n] - \mathbf{U}s[n]\right)\right)$;

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 Section III-D;

Fig. 5, the removal of different links from the graph leads to very different performance obtained by the algorithm. Indeed, while removing Link 1 (i.e., the red one), the algorithm performs as in the ideal case, the removal of links 2, 3, and 4, progressively determine a worse performance loss. This happens because the structure of the eigenvectors of the Laplacian of the benchmark graph is more or less preserved by the removal of specific links. Some links have almost no effects (e.g., Link 1), whereas some others (e.g., Link 4) may lead to deep modification of the structure of such eigenvectors, thus determining the mismatching of the LMS strategy in (12) and, consequently, its performance degradation. This example opens new theoretical questions that aim at understanding which links affect more the graph signals' estimation performance in situations where both the signal and the graph are jointly time-varying. We plan to tackle this exciting case in future work.

IV. LMS ESTIMATION WITH ADAPTIVE GRAPH SAMPLING

The LMS strategy in (12) assumes perfect knowledge of the support where the signal is defined in the graph frequency domain, i.e. $\mathcal{F}$. Indeed, this prior knowledge allows to define the projector operator $\mathbf{B}$ in (5) in a unique manner, and to implement the sampling strategies introduced in Section III-D. However, in many practical situations, this prior knowledge is unrealistic, due to the possible variability of the graph signal over time at various levels: the signal can be time varying according to a given model; the signal model may vary over time, for a given graph topology; the graph topology may vary as well over time. In all these situations, we cannot always assume that we have prior information about the frequency support $\mathcal{F}$, which must then be inferred directly from the streaming data $\boldsymbol{y}[n]$ in (9). 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. Exploiting the definition of GFT in (3), the signal observations in (9) can be recast as:

$$\boldsymbol{y}[n] = \mathbf{D}\mathbf{U}s_0 + \mathbf{D}\boldsymbol{v}[n]. \tag{34}$$

The problem then translates in estimating the coefficients of the GFT s_0, while identifying its support, i.e. the set of indexes where s_0 is different from zero. The support identification is deeply related to the selection of the sampling set. Thus, the overall problem can be formulated as the joint estimation of sparse representation s and sampling strategy $\mathbf{D}$ from the observations $\boldsymbol{y}[n]$ in (34), i.e.,

$$\min_{s,\mathbf{D}\in\mathcal{D}} \mathbb{E}\|\boldsymbol{y}[n] - \mathbf{D}\mathbf{U}s\|^2 + \lambda\,f(s), \tag{35}$$

where $\mathcal{D}$ is the (discrete) set that constraints the selection of the sampling strategy $\mathbf{D}$, $f(\cdot)$ is a sparsifying penalty function (typically, ℓ_0 or ℓ_1 norms), and $\lambda > 0$ is a parameter that regulates how sparse we want the optimal GFT vector s. Problem (35) is a mixed integer nonconvex program, which is very complicated to solve, especially in the adaptive context considered in this paper. Thus, to favor low complexity online solutions for (35), we propose an algorithm that alternates between the optimization of the vector s and the selection of the sampling operator $\mathbf{D}$. The rationale behind this choice is that, given an estimate for the support of vector s, i.e. $\mathcal{F}$, we can select the sampling operator $\mathbf{D}$ in a very efficient manner through one of the sampling strategies illustrated in Section III-D. Then, starting from a random initialization for s and a full sampling for $\mathbf{D}$ (i.e., $\mathbf{D} = \mathbf{I}$), the algorithm iteratively proceeds as follows. First, fixing the value of the sampling operator $\mathbf{D}[n]$ at time n, we update the estimate of the GFT vector s using an online version of the celebrated ISTA algorithm [59], [60], which proceeds as:

$$s[n+1] = \boldsymbol{T}_{\lambda\mu}\left(s[n] + \mu\,\mathbf{U}^H\mathbf{D}[n]\left(\boldsymbol{y}[n] - \mathbf{U}s[n]\right)\right), \tag{36}$$

$n \geq 0$, where $\mu > 0$ is a (sufficiently small) step-size, and $\boldsymbol{T}_\gamma(s)$ is a thresholding function that depends on the sparsity-inducing penalty $f(\cdot)$ in (35). Several choices are possible, as we will illustrate in the sequel. The aim of recursion (36) is to estimate the GFT s_0 of the graph signal x_0 in (9), while selectively shrinking to zero all the components of s_0 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 s_0 enables the adaptation of the sampling strategy, which can be updated using one of the strategies illustrated in Section III-D. Intuitively, the algorithm will increase (reduce) the number of samples used for the estimation, depending on the increment (reduction) of the current signal bandwidth. The main steps of the LMS algorithm with adaptive graph sampling are listed in Algorithm 2.

Thresholding functions : Several different functions can be used to enforce sparsity. A commonly used thresholding function comes directly by imposing an ℓ_1 norm constraint in (35), which is commonly known as the Lasso [47]. In this case, the vector threshold function $\boldsymbol{T}_\gamma(s)$ is the component-wise thresholding function $T_\gamma(s_m)$ applied to each element of vector s, with

$$T_\gamma(s_m) = \begin{cases} s_m - \gamma, & s_m > \gamma; \\ 0, & -\gamma \leq s_m \leq \gamma; \\ s_m + \gamma, & s_m < -\gamma. \end{cases} \tag{37}$$

The function $\boldsymbol{T}_\gamma(s)$ in (37) tends to shrink all the components of the vector s and, in particular, 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. To reduce
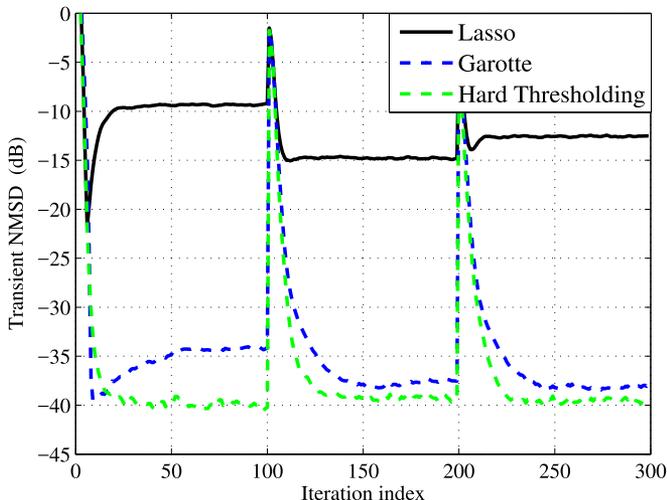
Fig. 6.   LMS with Adaptive Sampling: NMSD versus iteration index, for different thresholding functions.

Fig. 7.   LMS with Adaptive Sampling: $|\mathcal{F}|$ versus iteration index, for different thresholding functions.

the bias introduced by the Lasso constraint, several other thresholding functions can be adopted to improve the performance also in the case of less sparse systems. A potential improvement can be made by considering the non-negative Garotte estimator as in [61], whose thresholding function is defined as a vector whose entries are derived applying the threshold

$$T_\gamma(s_m) = \begin{cases} s_m \ (1 - \gamma^2/s_m^2), & |s_m| > \gamma; \\ 0, & |s_m| \le \gamma; \end{cases} \quad (38)$$

$m = 1, \ldots, M$. Finally, to completely remove the bias over the large components, we can implement a hard thresholding mechanism, whose function is defined as a vector whose entries are derived applying the threshold

$$T_\gamma(s_m) = \begin{cases} s_m, & |s_m| > \gamma; \\ 0, & |s_m| \le \gamma; \end{cases} \quad (39)$$

In the sequel, numerical results will illustrate how different thresholding functions such as (37), (38), and (39), affect the performance of Algorithm 2.

IV.   Numerical Results

In this section, we illustrate some numerical results aimed at assessing the performance of the proposed LMS method with adaptive graph sampling, i.e. Algorithm 2. In particular, to illustrate the adaptation capabilities of the algorithm, we simulate a scenario with a time-varying graph signal with $N = 50$ nodes, which has the same topology shown in Fig. 1, and spectral content that switches between the first 5, 15, and 10 eigenvectors of the Laplacian matrix of the graph. The elements of the GFT s_0 inside the support are chosen to be equal to 1. The observation noise in (9) is zero-mean, Gaussian, with a diagonal covariance matrix $\mathbf{C}_v = \sigma_v^2 \mathbf{I}$, with $\sigma_v^2 = 4 \times 10^{-4}$. In Fig. 6 we report the transient behavior of the normalized Mean-Square Deviation

(NMSD), i.e.

$$\mathrm{NMSD}[n] = \frac{\|\boldsymbol{s}[n] - \boldsymbol{s}_0\|^2}{\|\boldsymbol{s}_0\|^2},$$

versus the iteration index, considering the evolution of Algorithm 2 with three different thresholding functions, namely: a) the Lasso threshold in (37), the Garotte threshold in (38), and the hard threshold in (39). Also, in Fig. 7, we illustrate the behavior of the estimate of the cardinality of $\mathcal{F}$ versus the iteration index (cf. Step 2 of Algorithm 2), obtained by the three aforementioned strategies at each iteration. The value of the cardinality of $\mathcal{F}$ of the true underlying graph signal is also reported as a benchmark. The curves are averaged over 100 independent simulations. The step-size is chosen to be $\mu = 0.5$, the sparsity parameter $\lambda = 0.1$, and thus the threshold is equal to $\gamma = \mu\lambda = 0.05$ for all strategies. The sampling strategy used in Step 3 of Algorithm 2 is the Max-Det method introduced in Section III-D, where the number of samples $M[n]$ to be selected at each iteration is chosen to be equal to the current estimate of cardinality of the set $F[n]$. As we can notice from Fig. 6, the LMS algorithm with adaptive graph sampling is able to track time-varying scenarios, and its performance is affected by the adopted thresholding function. In particular, from Fig. 6, we notice how the algorithm based on the hard thresholding function in (39) 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 in (38). Finally, we can notice how the LMS algorithm based on Lasso may lead to very poor performance, due to misidentifications of the true graph bandwidth. This can be noticed from Fig. 7 where, while the Garotte and hard thresholding strategies are able to learn exactly the true bandwidth of the graph signal (thus leading to very good performance in terms of NMSD, see Fig. 6), the Lasso strategy overestimates the bandwidth of the signal, i.e. the cardinality of the set $\mathcal{F}$ (thus leading to poor estimation
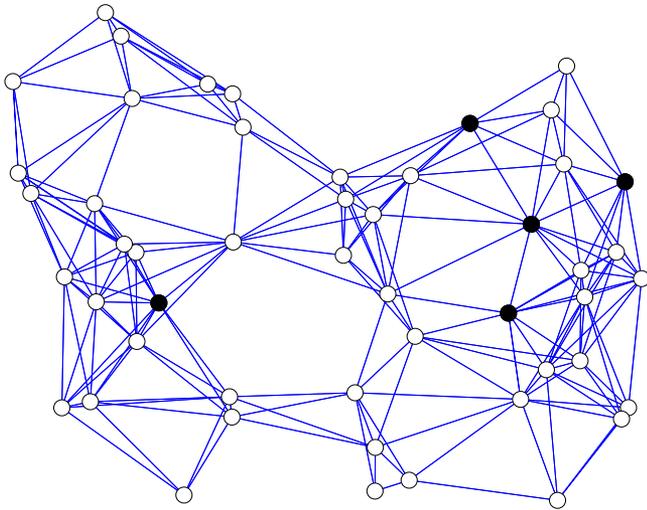
Fig. 8. Optimal Sampling at iteration $n = 80$.
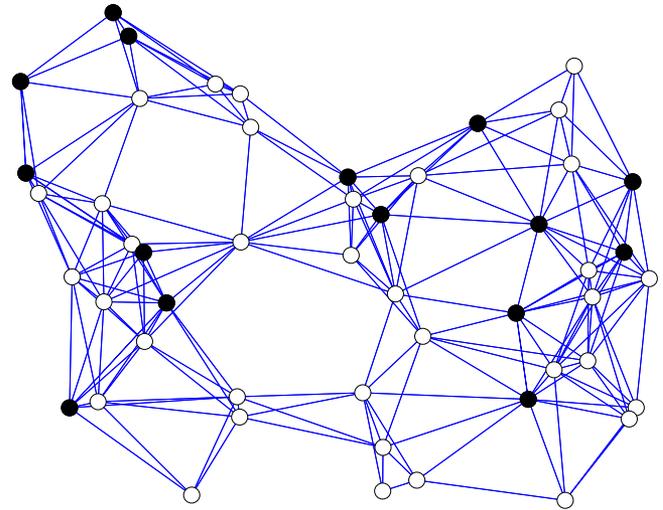
Fig. 10. Optimal Sampling at iteration $n = 280$.
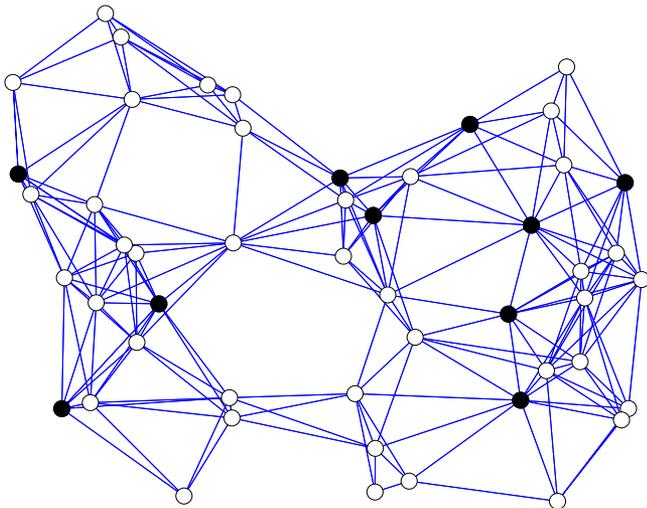
Fig. 9. Optimal Sampling at iteration $n = 180$.

performance, see Fig. 6). Finally, to illustrate an example of adaptive sampling, in Figs. 8, 9, and 10 we report the samples (depicted as black nodes) chosen by the proposed LMS algorithm based on hard thresholding at iterations $n = 80$, $n = 180$, and $n = 280$. As we can notice from Figs. 6, 7 and 8, 9, and 10, the algorithm always selects a number of samples equal to the current value of the signal bandwidth, while guaranteeing good reconstruction performance.

V. APPLICATION TO POWER SPATIAL DENSITY ESTIMATION IN COGNITIVE NETWORKS

The advent of intelligent networking of heterogeneous devices such as those deployed to monitor the 5G networks, power grid, transportation networks, and the Internet, will have a strong impact on the underlying systems. Situational awareness provided by such tools will be the key enabler for effective information dissemination, routing and congestion control, network health management, risk analysis, and security assurance. The vision is for ubiquitous smart network devices to enable data-driven statistical learning algorithms for distributed, robust, and online network operation and management, adaptable to the dynamically evolving network landscape with minimal need for human intervention. In this context, the unceasing demand for continuous situational awareness in cognitive radio (CR) networks calls for innovative signal processing algorithms, complemented by sensing platforms to accomplish the objectives of layered sensing and control. These challenges are embraced in the study of power cartography, where CRs collect data to estimate the distribution of power across space, namely the power spatial density (PSD). Knowing the PSD at any location allows CRs to dynamically implement a spatial reuse of idle bands. The estimated PSD map need not be extremely accurate, but precise enough to identify idle spatial regions.

In this section, we apply the proposed framework for LMS estimation of graph signals to spectrum cartography in cognitive networks. We consider a 5G scenario, where a dense deployment of radio access points (RAPs) is envisioned to provide a service environment characterized by very low latency and high rate access. Each RAP collects streaming data related to the spectrum utilization of primary users (PU's) at its geographical position. This information can then be sent to a processing center, which collects data from the entire system, through high speed wired links. The aim of the center is then to build a spatial map of the spectrum usage, while processing the received data on the fly and envisaging proper sampling techniques that enable a proactive sensing of the system from only a limited number of RAP's measurements. As we will see in the sequel, the proposed approach hinges on the graph structure of the signal received from the RAP's, thus enabling real-time PSD estimation from a small set of observations that are smartly sampled from the graph.

Numerical examples: Let us consider an operating region where 100 RAPs are randomly deployed to produce a map of the spatial distribution of power generated by the transmissions of two active primary users. The PU's emit electromagnetic radiation with power equal to 1 Watt. For simplicity, the propagation medium is supposed to introduce a free-space path loss attenuation on the PU's transmissions. The graph among RAPs is built from a distance based model, i.e. stations that are
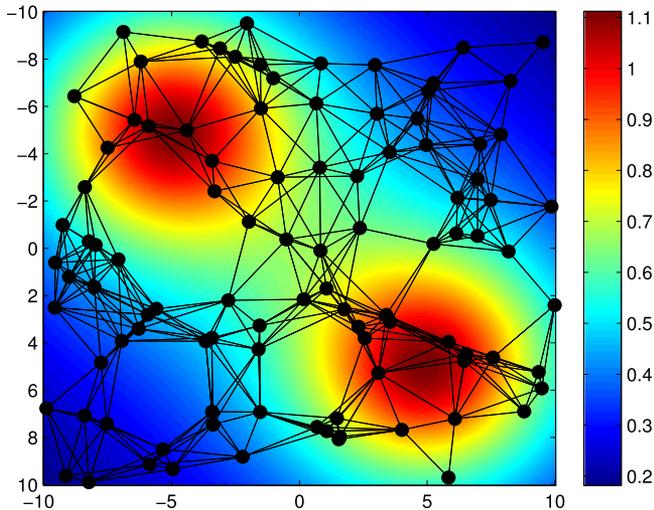
Fig. 11.   PSD cartography: spatial distribution of primary users' power, small cell base stations deployment, graph topology, and graph signal.

Fig. 13.   PSD cartography: Transient NMSD versus iteration index, for different number of samples and bandwidths used for processing.

Fig. 12.   PSD cartography: Steady-state NMSD versus number of samples taken from the graph, for different bandwidths used for processing.

sufficiently close to each other are connected through a link (i.e. $a_{ij} = 1$, if nodes i and j are neighbors). In Fig. 11, we illustrate a pictorial description of the scenario, and of the resulting graph signal. We assume that each RAP is equipped with an energy detector, which estimates the received signal using 100 samples, considering an additive white Gaussian noise with variance $\sigma_v^2 = 10^{-4}$. The resulting signal is not perfectly band-limited, but it turns out to be smooth over the graph, i.e. neighbor nodes observe similar values. This implies that sampling such signals inevitably introduces aliasing during the reconstruction process. However, even if we cannot find a limited (lower than N) set of frequencies where the signal is completely localized, the greatest part of the signal energy is concentrated at low frequencies. This means that if we process the data using a sufficient number of observations and (low) frequencies, we should still be able to reconstruct the signal with a satisfactory performance.

To illustrate an example of cartography based on the LMS algorithm in (12), in Fig. 12 we report the behavior of the steady-state NMSD versus the number of samples taken from the graph, for different bandwidths used for processing. The

step-size is chosen equal to 0.5, while the adopted sampling strategy is the Max-Det method introduced in Section III-D. The results are averaged over 200 independent simulations. As expected, from Fig. 12, we notice that the steady-state NMSD of the LMS algorithm in (12) improves by increasing the number of samples and bandwidths used for processing. Interestingly, in Fig. 12 we can see a sort of threshold behavior: the NMSD is large for $|\mathcal{S}| < |\mathcal{F}|$, when the signal is undersampled, whereas the values become lower and stable as soon as $|\mathcal{S}| > |\mathcal{F}|$. Finally, we illustrate an example that shows the tracking capability of the proposed method in time-varying scenarios. In particular, we simulate a situation the two PU's switch between idle and active modes: for $0 \leq n < 133$ only the first PU transmits; for $133 \leq n < 266$ both PU's transmit; for $266 \leq n \leq 400$ only the second PU's transmits. In Fig. 13 we show the behavior of the transient NMSD versus iteration index, for different number of samples and bandwidths used for processing. The results are averaged over 200 independent simulations. From Fig. 13, we can see how the proposed technique can track time-varying scenarios. Furthermore, its steady-state performance improves with increase in the number of samples and bandwidths used for processing. These results, together with those achieved in Fig. 12, illustrate an existing tradeoff between complexity, i.e. number of samples used for processing, and mean-square performance of the proposed LMS strategy. In particular, using a larger bandwidth and a (consequent) larger number of samples for processing, the performance of the algorithm improves, at the price of a larger computational complexity.

VI. CONCLUSION

In this paper we have proposed LMS strategies for adaptive 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 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. Furthermore, to cope with time-varying scenarios, we also propose an LMS method with adaptive graph sampling, which estimates and tracks the signal support in the (graph)frequency domain, while at the same time adapting the graph sampling strategy. Several numerical simulations confirm the theoretical findings, and illustrate the potential advantages achieved by these strategies for adaptive estimation of band-limited graph signals. Finally, we apply the proposed method to estimate and track the spatial distribution of power transmitted by primary users in a cognitive network environment, thus illustrating the existing tradeoff between complexity and mean-square performance of the proposed strategy.

We expect that such processing tools will represent a key technology for the design and proactive sensing of Cyber Physical Systems, where a proper adaptive control mechanism requires the availability of data driven sampling strategies able to control the overall system by only checking a limited number of nodes, in order to collect correct information at the right time, in the right place, and for the right purpose.

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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, Apr. 2013.

[2] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 1644–1656, Apr. 2013.

[3] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[4] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[5] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.

[6] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

[7] I. Z. Pesenson, "Sampling in Paley–Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

[8] I. Z. Pesenson and M. Z. Pesenson, "Sampling, filtering and sparse approximations on combinatorial graphs," *J. Fourier Anal. Appl.*, vol. 16, no. 6, pp. 921–942, 2010.

[9] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 3921–3924.

[10] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[11] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-D time," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3572–3585, Aug. 2008.

[12] M. Puschel and J. M. F. Moura, "Algebraic signal processing theory: 1-D space," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3586–3599, Aug. 2008.

[13] S. Sardellitti and S. Barbarossa, "On the graph fourier transform for directed graphs," 2016. [Online]. Available: http://arxiv.org/abs/1601.05972

[14] A. Agaskar and Y. M. Lu, "A spectral graph uncertainty principle," *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4338–4356, Aug. 2013.

[15] B. Pasdeloup, R. Alami, V. Gripon, and M. Rabbat, "Toward an uncertainty principle for weighted graphs," in *Proc. 23rd Eur. Signal Process. Conf.*, 2015, pp. 1496–1500, arXiv:1503.03291.

[16] J. J. Benedetto and P. J. Koprowski, "Graph theoretic uncertainty principles," in *Proc. 2015 Int. Conf. Sampling Theory Appl.*, 2015. [Online]. Available: http://www.math.umd.edu/jjb/graph_theoretic_UP_April_14.pdf

[17] P. J. Koprowski, "Finite frames and graph theoretic uncertainty principles," Ph.D. dissertation, Digit. Repository, Univ.Maryland, College Park, MD, USA, 2015. [Online]. Available: http://hdl.handle.net/1903/16666

[18] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Jan. 2016.

[19] S. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 5445–5449.

[20] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2432–2444, May 2015.

[21] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.

[22] M. Tsitsvero and S. Barbarossa, "On the degrees of freedom of signals on graphs," in *Proc. 2015 Eur. Signal Proc. Conf.*, Sep. 2015, pp. 1521–1525.

[23] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 491–494.

[24] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 728–740, Jun. 2015.

[25] S. Chen, F. Cerda, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovacevic, "Semi-supervised multiresolution classification using adaptive graph filtering with application to indirect bridge structural health monitoring," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2879–2893, Jun. 2014.

[26] A. Sandryhaila and J. M. Moura, "Classification via regularization on graphs." in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 495–498.

[27] V. N. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, "Wavelet-regularized graph semi-supervised learning," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 423–426.

[28] X. Zhang, X. Dong, and P. Frossard, "Learning of structured graph dictionaries," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 3373–3376.

[29] D. Thanou, D. I. Shuman, and P. Frossard, "Parametric dictionary learning for graph signals," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 487–490.

[30] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 1, 2016.

[31] D. Zhou and B. Schölkopf, "A regularization framework for learning from graph data," in *Proc. ICML Workshop Statist. Relational Learn. Connections Fields*, 2004, vol. 15, pp. 67–68.

[32] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, 2006.

[33] S. Chen *et al.*, "Signal inpainting on graphs via total variation minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 8267–8271.

[34] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal denoising on graphs via graph filtering," in *Proc. IEEE Global Conf. Signal Inform. Process.*, 2014, pp. 872–876.

[35] P.-Y. Chen and A. O. Hero, "Local fiedler vector centrality for detection of deep and overlapping communities in networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 1120–1124.

[36] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal recovery on graphs: Variation minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4609–4624, Sep. 2015.

[37] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Trans. Signal Inf. Process. over Networks*, 2016.

[38] A. H. Sayed, *Adaptive Filters*. New York, NY, USA: Wiley, 2011.

[39] A. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends® Mach. Learn.*, vol. 7, no. 4/5, pp. 311–801, 2014.

[40] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society, 1997.

[41] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.

[42] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[43] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[44] J. Homer, I. Mareels, R. R. Bitmead, B. Wahlberg, and F. Gustafsson, "LMS estimation via structural detection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2651–2663, Oct. 1998.

[45] M. Godavarti and A. O. Hero III, "Partial update LMS algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2382–2399, Jul. 2005.

[46] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[47] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Series B*, vol. 58, pp. 267–288, 1996.

[48] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted L1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5/6, pp. 877–905, 2008.

[49] Y. Chen, Y. Gu, and A. O. Hero III, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3125–3128.

[50] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, Jul. 2010.

[51] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted balls," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 936–952, Mar. 2011.

[52] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, "A sparsity promoting adaptive algorithm for distributed learning," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5412–5425, Oct. 2012.

[53] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1419–1433, Mar. 2013.

[54] P. Di Lorenzo, S. Barbarossa, and A. H. Sayed, "Distributed spectrum estimation for small cell networks based on sparse diffusion adaptation," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1261–1265, Oct. 2013.

[55] P. Di Lorenzo, "Diffusion adaptation strategies for distributed estimation over gaussian Markov random fields," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5748–5760, Jul. 2014.

[56] S. Barbarossa and S. Sardellitti, and P. Di Lorenzo, *Distributed Detection and Estimation in Wireless Sensor Networks*, vol. 2. New York, NY, USA: Academic, 2014, pp. 329–408.

[57] A. H. Sayed and V. H. Nascimento, *Energy Conservation and the Learning Ability of LMS Adaptive Filters*. New York, NY, USA: Wiley, 2003.

[58] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1986.

[59] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[60] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[61] M. Yuan and Y. Lin, "On the non-negative garrotte estimator," *J. Roy. Statist. Soc.: Series B*, vol. 69, no. 2, pp. 143–161, 2007.

Sergio Barbarossa (S'84–M'88–F'12) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Rome "La Sapienza," Rome, Italy, in 1984 and 1988, respectively. He held positions as a Research Engineer with Selenia SpA during 1984–1986 and with the Environmental Institute of Michigan in 1988, as a Visiting Professor at the University of Virginia, Charlottesville, VA, USA in 1995 and 1997, and with the University of Minnesota, Minneapolis, MN, USA, in 1999. He is currently a Full Professor with the University of Rome "La Sapienza." He was an IEEE Distinguished Lecturer from the Signal Processing Society in 2012–2013. He is the author of a research monograph titled Multiantenna Wireless Communication Systems. He has been the Scientific Coordinator of various European projects on wireless sensor networks, femtocell networks, and mobile cloud computing. His research interests include signal processing for self-organizing networks, mobile-edge computing, signal processing over graphs, and distributed optimization algorithms. Dr. Barbarossa is an EURASIP Fellow. From 1997 to 2003, he was a Member of the IEEE Technical Committee for Signal Processing in Communications. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (during 1998–2000 and 2004–2006) and the IEEE SIGNAL PROCESSING MAGAZINE. He is currently an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. He has been the General Chairman of the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2003 and the Technical Cochair of SPAWC, 2013. He has been the Guest Editor for Special Issues on the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *the EURASIP Journal of Applied Signal Processing*, *the EURASIP Journal on Wireless Communications and Networking*, the IEEE SIGNAL PROCESSING MAGAZINE, and the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He received the 2010 EURASIP Technical Achievements Award and the 2000 and 2014 IEEE Best Paper Awards from the IEEE Signal Processing Society. He is the co-author of papers that received the Best Student Paper Award at ICASSP 2006, SPAWC 2010, EUSIPCO 2011, and CAMSAP 2011.

Paolo Banelli (S'90–M'99) received the Laurea degree (*cum laude*) in electronics engineering and the Ph.D. degree in telecommunications from the University of Perugia, Perugia, Italy, in 1993 and 1998, respectively. In 2005, he was appointed as Associate Professor in the Department of Electronic and Information Engineering, University of Perugia, where he has been an Assistant Professor since 1998. In 2001, he joined the SpinComm Group, as a Visiting Researcher, in the ECE Department, University of Minnesota, Minneapolis. His research interests include signal processing for wireless communications, with emphasis on multicarrier transmissions, signal processing for biomedical applications, spectrum sensing for cognitive radio, waveform design for 5G communications, and recently graph signal processing. He was a Member (2011–2013) of the IEEE Signal Processing Society's Signal Processing for Communications and Networking Technical Committee. In 2009, he was a General Cochair of the IEEE International Symposium on Signal Processing Advances for Wireless Communications. He currently serves as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the *EURASIP Journal on Advances in Signal Processing*.

Stefania Sardellitti (M'12) received the M.Sc. degree in electronic engineering from the University of Rome "La Sapienza," Rome, Italy, in 1998 and the Ph.D. degree in electrical and information engineering from the University of Cassino, Cassino, Italy, in 2005. Since 2005, she has been an appointed Professor of digital communications with the University of Cassino, Italy. She is currently a Research Assistant at the Department of Information, Electronics and Telecommunications, University of Rome, Sapienza, Italy. Her research interests include the area of statistical signal processing, mobile cloud computing, femtocell networks and wireless sensor networks, with emphasis on distributed optimization. She received the 2014 IEEE Best Paper Award from the IEEE Signal Processing Society. She has participated in the European project WINSOC (on wireless sensor networks) and in the European projects FREEDOM (on femtocell networks) and TROPIC, on distributed computing, storage and radio resource allocation over cooperative femtocells.

Paolo Di Lorenzo (S'10–M'13) received the M.Sc. degree in 2008 and the Ph.D. degree in electrical engineering in 2012, both from University of Rome "La Sapienza," Rome, Italy. He is currently an Assistant Professor in the Department of Engineering, University of Perugia, Perugia, Italy. During 2010, he held a Visiting Research appointment in the Department of Electrical Engineering, University of California, Los Angeles (UCLA). He has participated in the European research project FREEDOM, on femtocell networks, SIMTISYS, on moving target detection through satellite constellations, and TROPIC, on distributed computing, storage and radio resource allocation over cooperative femtocells. His research interests include signal processing theory and methods, distributed optimization and learning over networks, graph theory, and adaptive filtering. He is currently an Associate Editor of the *Eurasip Journal on Advances in Signal Processing*. He received three Best Student Paper Awards, respectively, at the IEEE SPAWC'10, the EURASIP EUSIPCO'11, and the IEEE CAMSAP'11, for works in the area of signal processing for communications and synthetic aperture radar systems. He also received the 2012 GTTI (Italian national group on telecommunications and information theory) Award for the Best Ph.D. Thesis in information technologies and communications.