Delft University of Technology

Revisiting matching pursuit
Beyond approximate submodularity

Tohidi, Ehsan; Coutino, Mario; Gesbert, David

DOI
10.1016/j.sigpro.2024.109638
Publication date
2025
Document Version
Final published version
Published in
Signal Processing

Citation (APA)
Tohidi, E., Coutino, M., & Gesbert, D. (2025). Revisiting matching pursuit: Beyond approximate submodularity. *Signal Processing*, *226*, Article 109638. https://doi.org/10.1016/j.sigpro.2024.109638

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Signal Processing 226 (2025) 109638

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Revisiting matching pursuit: Beyond approximate submodularity ☆

Ehsan Tohidi [a,b,*], Mario Coutino [c], David Gesbert [d]

[a] Fraunhofer HHI, Berlin, Germany
[b] Technische Universität Berlin, Berlin, Germany
[c] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
[d] EURECOM, Biot, France

ARTICLE INFO

Keywords:
Submodularity
Near-optimality
Matching pursuit

ABSTRACT

We study the problem of selecting a subset of vectors from a large set to obtain the best signal representation over a family of functions. Although greedy methods have been widely used to tackle this problem and many of those have been analyzed under the lens of (weak) submodularity, none of these algorithms are explicitly devised using such a functional property. Here, we revisit the vector-selection problem and introduce a function that is shown to be submodular in expectation. This function not only guarantees near-optimality through a greedy algorithm in expectation but also alleviates the existing deficiencies in commonly used matching pursuit (MP) algorithms. We further show the relation between the single-point-estimate version of the proposed greedy algorithm and MP variants. Moreover, we discuss extending the signal representation problem to instances with knapsack and matroid constraints. Our theoretical findings are supported by numerical experiments on the angle of arrival estimation problem, a typical signal representation task, demonstrating the benefits of our method compared to traditional MP algorithms.

1. Introduction

In this paper, we are interested in revisiting the ubiquitous problem of signal representation: from a given set of vectors, the problem consists in selecting a subset of K of them which *best* represent another vector of interest, i.e., the target signal. This problem has found many applications in signal processing and machine learning [1]. Within the realm of signal processing, the vectors could be delayed versions of a reference signal, and the goal is to represent another signal using a few of them. This problem is encountered, for instance, in the direction of arrival estimation (DOA) [2]. In machine learning, the vectors often relate to features, and we want to make a prediction of a given phenomenon using only a small part of the available features. Typically, the involved vectors are referred to as *atoms*, and the set of atoms is termed *dictionary*.

When dictionaries are composed of orthonormal atoms, the selection problem naturally reduces to representing the target signal as the superposition of basis vectors; see, e.g., [3]. However, in the case of redundant dictionaries, i.e., non-orthonormal atoms, the uniqueness of the representation is not necessarily guaranteed. Due to interpretability motivations or constraints given by a priori information, a parsimonious representation of the target signal is usually preferred, which means that the representation with the least amount of atoms is searched. Although finding the global optimal solution to this problem is computationally intractable, greedy algorithms, e.g., matching pursuit (MP) and orthogonal MP (OMP), efficiently find such a solution under appropriate conditions [4].

Aligned with the goal of the signal representation problem, MP-based methods aim to compute a linear expansion of the target signal in terms of a few atoms of the given dictionary. While MP constructs its solution by successive approximations of the target signal with orthogonal projections on elements of the dictionary [1], OMP builds its solution by ensuring that the residual is orthogonal to the span of previously added atoms [5]. Despite their differences, both methods normalize all the atoms in the beginning, and, at each step, *greedily* select the next atom, among the unselected ones, with the maximum inner product with the residual. Though these methods exhibit good performance in several instances [4], the fact that a part of each unselected atom could lie over the span of the currently selected atoms negatively affects the atom selection step because the remaining parts are not normalized anymore.

☆ The work of Ehsan Tohidi was financially supported by the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification number: 16KISK020K and 16KISK030.

The work of David Gesbert was partially supported by his Chair on AI for wireless networks at the French funded 3IA center on artificial intelligence.

* Corresponding author at: Fraunhofer HHI, Berlin, Germany.
E-mail address: ehsan.tohidi@hhi.fraunhofer.de (E. Tohidi).

https://doi.org/10.1016/j.sigpro.2024.109638
Received 12 May 2023; Received in revised form 4 June 2024; Accepted 25 July 2024
Available online 27 July 2024
0165-1684/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Although there have been many extensions of MP methods, e.g., [6–9], none of these extensions provide guarantees on the optimality of their solution for general problems. In addition, their performance guarantees (if any) are only applicable to special cases, e.g., signals with sparse representation or dictionaries with specific characteristics [1,5–10].

Another family of methods, intended for the same goal, determines the representation using some relaxations/modifications of the original problem [11–15]. For instance, the method of frames (MOF), among all solutions, picks out one whose representation coefficients have the minimum ℓ_2-norm, while the principle of basis pursuit (BP) is to find a representation of the signal whose coefficients have the minimum ℓ_1-norm [11,12]. Yet, no guarantee is presented or the guarantees are valid only for special cases.

Traditional ways to develop theoretical performance bounds for subset selection algorithms mostly involve the spectral conditions of matrices related to the input data, see, e.g., [16]. However, the functional property of submodularity can also be leveraged to analyze and produce performance guarantees. These efforts have led to multiplicative approximation bounds for methods such as forward regression (FR) and OMP improving past known bounds [17].

Submodularity is a characteristic of set functions that exhibit the so-called *diminishing returns* property [18]. Employing greedy algorithms, the maximization of a submodular function subject to cardinality and matroid constraints (which are imposed in many practical signal representation problems) can be performed near-optimally within a $1 - 1/\mathrm{e}$ factor [19–21]. Due to the near-optimality guarantees and the conceptually simple algorithms, submodular optimization has appeared in a wide variety of applications; e.g., antenna selection in multi-input multi-output systems [22], virtual controller placement in 5G software-defined networks [23], resource allocation in wireless communication [24], sparse sampler design for detection [25], intelligent reconfigurable surfaces placement [26], distributed machine learning [27], regression [28], online approximation algorithms [29], multi-objective optimization [30], and active learning [31]. In addition, [32] has proposed dictionary selection algorithms exploiting approximate submodularity where the performance bounds are dependent on the dictionary.

Although submodularity has been used to explain the good performance of greedy algorithms for the signal representation problem, the analyzed algorithms themselves have not been devised explicitly using submodularity as a foundation. In this paper, we revisit the problem of signal representation from a new perspective and introduce a submodular objective function to devise a greedy algorithm, referred to as submodular matching pursuit (SMP), which (i) has probabilistic near-optimality guarantees and (ii) overcomes the shortcomings of the OMP algorithm. Using submodularity as a stepping stone, we discuss the possibility of naturally extending the signal representation problem to instances with constraints, namely knapsack and matroid constraints while still ensuring near-optimality performance. Although other structured constraints, such as block sparsity, have been considered in the past, to the best of our knowledge, this is the first time the aforementioned constraints have been applied in the signal representation setting. Our main contributions can be summarized as follows:

(1) We revisit the problem of signal representation and, motivated by a wide range of applications, introduce a submodular objective set function for this problem.
(2) We propose the SMP algorithm, a greedy algorithm with a $(1 - \mathrm{e}^{-1})$-approximation ratio for the signal representation problem, in expectation.
(3) We demonstrate how the so-called Optimized Orthogonal Matching Pursuit (OOMP) algorithm [6] can be interpreted as the single-point estimate version of the SMP algorithm.
(4) We establish the near-optimality (in probability) of using the finite-sample version of our introduced submodular function, thereby substantiating the near-optimality of the OOMP algorithm.
(5) We discuss extending the signal representation problem to instances with knapsack and matroid constraints, arguing that slightly modified versions of the SMP algorithm return solutions with approximation ratios of $\frac{1}{2}(1 - \mathrm{e}^{-1})$- and $1/2$, respectively.

Notation. Boldface lower(upper)case denote column vectors, e.g., $\mathbf{a}$ (matrices $\mathbf{A}$). Calligraphic letters are used for sets and $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$; $\setminus$ is used as the operator of set subtraction. The Hermitian operator is denoted by $(\cdot)^H$. Given a dictionary matrix $\boldsymbol{\Phi}$, and a subset $\mathcal{S}$ of indices of columns of $\boldsymbol{\Phi}$, $\boldsymbol{\Phi}_{\mathcal{S}}$ denotes a matrix composed of the columns of $\boldsymbol{\Phi}$ indexed by $\mathcal{S}$. $\mathrm{span}\{\mathbf{A}\}$ stands for the span of columns of $\mathbf{A}$; and $\boldsymbol{\Pi}_{\mathcal{S}} = \boldsymbol{\Phi}_{\mathcal{S}}(\boldsymbol{\Phi}_{\mathcal{S}}^H\boldsymbol{\Phi}_{\mathcal{S}})^{-1}\boldsymbol{\Phi}_{\mathcal{S}}^H$ is the projection matrix onto $\mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$. $\mathbf{y}_{\mathcal{S}}$ is the projection of $\mathbf{y}$ over $\mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$. $\langle\cdot,\cdot\rangle$ and $\|\cdot\|_2$ denote the inner product and the ℓ_2-norm, respectively.

2. Preliminaries

Consider a set of N M-dimensional *atoms*, $\{\boldsymbol{\phi}_i \in \mathbb{C}^M\}_{i=1}^N$, being represented by a matrix $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$. This matrix is usually referred to as *dictionary*. In the signal representation problem, we aim to find an accurate representation of a *target signal* $\mathbf{y} \in \mathbb{C}^{M\times 1}$ over $\mathrm{span}\{\boldsymbol{\Phi}\}$ such that only a few atoms are involved. In other words, we look for a representation $\mathbf{y}_{\mathcal{S}}$ with $|\mathcal{S}|$ as small as possible.

A related problem, but with a different context, is the so-called signal recovery problem. Given a measurement vector $\mathbf{y}$, the goal is to find the atoms and corresponding coefficients that best reconstruct the signal. The recovery problem is modeled as follows

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{M\times 1}$, $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$, $\mathbf{x} \in \mathbb{C}^{N\times 1}$, and $\mathbf{e} \in \mathbb{C}^{M\times 1}$ are the measurement vector, dictionary matrix, unknown coefficients vector, and noise vector, respectively. This problem reduces to estimate of the non-zero coefficients of $\mathbf{x}$, given the measurement vector $\mathbf{y}$, under the assumption that $\boldsymbol{\Phi}$ is known.

As both signal representation and recovery problems have the same goal -selecting a subset of atoms with the minimum cardinality that best represents the signal- though in different contexts, our problem formulation and the proposed algorithms in this work apply to both of them.

2.1. Submodularity

Here, we list a series of definitions needed in this paper and a set of theoretical results, within submodular analysis, which we use to sustain the claims of this work.

First, we introduce the formal definition of submodularity.

Definition 1 (*Submodularity*).A set function $f : 2^{\mathcal{N}} \to \mathbb{R}$ defined over the ground set $\mathcal{N} = \{1, 2, \ldots, N\}$ is submodular if for every $\mathcal{S} \subseteq \mathcal{N}$ and for all $a, b \in \mathcal{N} \setminus \mathcal{S}$ it holds that

$$f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \geqslant f(\mathcal{S} \cup \{a, b\}) - f(\mathcal{S} \cup \{b\}). \tag{2}$$

The submodularity definition offers valuable intuition. Let us consider $\mathcal{S} \subseteq \mathcal{N}$ as a set of elements providing some benefit $f(\mathcal{S})$. According to the definition, for a submodular function f, adding an element b to the set $\mathcal{S}$, denoted as $\mathcal{S} \cup b$, does not increase the marginal benefit of any element a. Consequently, submodular set functions demonstrate a natural property of diminishing returns.

Monotonicity is a common property of set functions appearing in the signal representation problem.

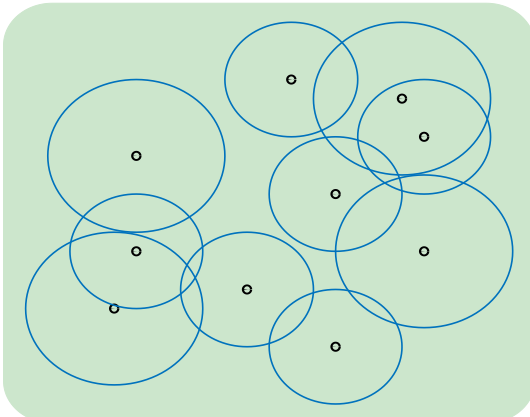


Fig. 1. Illustration of the sensing coverage problem with the coverage of each sensor. Each black dot represents a possible sensor location with the circle depicting the corresponding coverage area. The goal is to select a subset of sensors to maximize the covered area.

Definition 2 (*Monotonicity*).A set function $f : 2^{\mathcal{N}} \to \mathbb{R}$ defined over the ground set $\mathcal{N} = \{1, 2, \dots, N\}$ is monotonic if for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$, $f(\mathcal{A}) \leqslant f(\mathcal{B})$.

Finally, a set function is said to be *normalized* if $f(\emptyset) = 0$.

We now formally define the *constraint subset selection problem*.

Definition 3 (*Constrained Subset Selection*)**.** Given a ground set $\mathcal{N}$, a set function f, and a set of sets $\mathcal{I}$, find a set satisfying $\mathcal{S} \subseteq \mathcal{N}$ and $\mathcal{S} \in \mathcal{I}$, maximizing $f(\mathcal{S})$.

From the above definition, it is clear that the signal representation problem (described before) is an instance of the constrained subset selection problem, where the function f measures how well the set $\mathcal{S}$ of atoms represents the target signal. The set of sets $\mathcal{I}$, in the standard version of the problem, are all subsets of $\mathcal{N}$ whose cardinality is smaller than a prescribed K, i.e., $\mathcal{I}_K := \{\mathcal{A} : \mathcal{A} \subseteq \mathcal{N}, |\mathcal{A}| \leq K\}$.

In order to illustrate constrained subset selection problems in Definition 3 more clearly, here, we briefly discuss the so-called sensing coverage problem. This example application will be used throughout the paper to make concrete constraints discussed later.

Consider a set of potential sensor locations (i.e., the ground set) with corresponding functioning areas that define the regions around each sensor from which useful information can be collected. Given that these sensors incur costs, our objective is to optimize their deployment locations to maximize effectiveness while controlling costs. Consequently, the optimization problem translates into a coverage problem, where the goal is to maximize the covered area of the sensing field while selecting a limited number of sensors. An illustration of such a sensor selection problem is provided in Fig. 1, where each black dot represents a potential sensor location and the circle represents its coverage area. This scenario naturally leads to a *constrained subset selection* [18]. Additionally, the diminishing returns property is evident in this example; namely, the increase in coverage area provided by an additional sensor (i.e., circle) is nonincreasing as the size of the selected sensor set increases.

Though at first glance, the sensing coverage problem seems to be disconnected from the signal representation problem, it is possible to establish a direct relation between these two problems as follows. Consider the correspondence between the sensors and the columns of the dictionary, and the target signal to the target field (i.e., the rectangular field in this case). Clearly, some atoms/sensors are correlated/have coverage overlap and some are uncorrelated/separated. We aim to represent/fill the target signal/field with the atoms/sensors.

Now that the motivation behind the discussed setting is in order, we now provide a well-known result with respect to the solution of this kind of problem using a greedy heuristic. That is, when the set function f is submodular, and $\mathcal{I}$ takes a particular form, we have the following known result.

Theorem 1 ([19]). *Let $f : 2^{\mathcal{N}} \to \mathbb{R}$ be a normalized, monotonic, submodular set function defined on the subsets of a finite ground set $\mathcal{N}$. Let $\mathcal{S}_{\mathrm{G}}$ be a set of K elements selected by Algorithm 1. Then,*

$$f(\mathcal{S}_{\mathrm{G}}) \geqslant (1 - \mathrm{e}^{-1}) f(\mathcal{S}_{\mathrm{OPT}}), \tag{3}$$

where $\mathcal{S}_{\mathrm{OPT}}$ is the solution to the cardinality-constrained subset selection problem with $\mathcal{I} = \mathcal{I}_K$.

2.2. Matching pursuit algorithms

We briefly review the two main variants of MP methods, namely MP and OMP, to put them in context with respect to Algorithm 1.

Matching Pursuit. The original version of MP iteratively constructs a solution by substituting line 2 in Algorithm 1 by

$$i^* = i_k := \underset{i \notin \mathcal{S}_{k-1}}{\arg\max} \, |\langle R_k, \boldsymbol{\phi}_i \rangle|, \tag{4}$$

where $R_{k+1} := R_k - \alpha_k \boldsymbol{\phi}_{i_k}$, with $\alpha_k := |\langle R_k, \boldsymbol{\phi}_{i_k} \rangle|$; and $R_1 := \mathbf{y}$.

Orthogonal Matching Pursuit. In this refined version of MP, the selection rule is the same as that of (4), but with a slightly different definition for the kth residual, i.e.,

$$R_k := \mathbf{y} - \mathbf{y}_{\mathcal{S}_{k-1}}. \tag{5}$$

Although OMP improves on MP, in general, the kth-selected atom cannot be guaranteed to be completely outside of $\mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}_{k-1}}\}$. This characteristic hinders the ability of OMP (and subsequently MP) to find the best representation for the target signal in many instances.

2.3. Pitfalls of OMP: A motivating toy example

To illustrate the aforementioned problem of OMP, let us study the following toy problem. Consider $M = N = 3$ and $\mathbf{y} = [1000, 10, 1]^H$. Let the dictionary, with normalized columns, be given as

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 0.9959 & 0 \\ 0 & 0.09 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Assume a cardinality constraint that limits the representation to a maximum of $K = 2$ atoms. In the first iteration, OMP selects the first column as it has the maximum inner product with the signal $\mathbf{y}$. Removing the projected part, i.e., $\mathbf{y}_{\{1\}}$, the residual is equal to $R_2 = [0, 10, 1]^T$. A quick calculation of the inner product between R_2 with the two remaining columns gives

$$\langle R_2, \boldsymbol{\phi}_2 \rangle = 0.9, \ \langle R_2, \boldsymbol{\phi}_3 \rangle = 1. \tag{7}$$

Therefore, at the second iteration, OMP selects the third column forming the solution $\mathcal{S}_{\mathrm{OMP}} = \{1, 3\}$. This leaves a final residual of $R_3 = [0, 10, 0]^H$ with norm 10, while, if we select the first two columns, i.e., $\mathcal{S}' = \{1, 2\}$, the final residual would have a norm of 1, i.e., $R_3' = [0, 0, 1]^H$.

The reason for this deficiency in the selection of vectors is that at the second iteration, the second column has a non-zero inner product with the first column which is already selected. This causes the normalization of columns in the initialization phase to be unfair.

In the following, using (5) as a foundation, we introduce a submodular function that (i) naturally tackles the aforementioned problem of OMP and (ii) allows for near-optimal guarantees through Algorithm 1.

3. Sparse representation and submodularity

In this section, we first discuss a closely related result in the intersection of sparse representation and submodularity. We then argue that if the *average performance* of the reconstruction, over the input space, is considered, the resulting function can be shown to be submodular. We use this result to provide an efficient algorithm with probabilistic guarantees in the next section.

3.1. Sparse representation objective

Let us consider a set of m target signals, $\{\mathbf{y}^{(i)}\}_{i=1}^m$. Given that our goal is to best represent the target signals, we can consider the residual [c.f. (5)] for the last step and put forth the following objective for constructing the solution set

$$f_m(\mathcal{S}) := \frac{1}{m}\sum_i \left[\|\mathbf{y}^{(i)}\|_2^2 - \|\mathbf{y}^{(i)} - \mathbf{y}_{\mathcal{S}}^{(i)}\|_2^2\right], \tag{8}$$

where $\mathbf{y}_{\mathcal{S}}^{(i)}$ denotes the projection of the ith target signal onto $\text{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$.

The objective function (8) has been considered before in [32] for the dictionary selection problem and its *approximate submodularity* has been shown. Here, we restate this result for completeness.

Theorem 2 ([32]). *If the dictionary $\boldsymbol{\Phi}$ has incoherence $\mu := \max_{\forall(i,j), i\neq j} |\langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j\rangle|$, then f_m satisfies*

$$f_m(\mathcal{S}\cup\{a\}) - f_m(\mathcal{S}) \geq f_m(\mathcal{S}\cup\{a,b\}) - f_m(\mathcal{S}\cup\{a\}) - \epsilon \tag{9}$$

for all $a, b \in \mathcal{N}\setminus\mathcal{S}$, where $\epsilon \leq 4K\mu$.

Although the above result, in combination with the guarantees provided in [32], portrays an optimistic landscape for finding a *near-optimal* solution to the dictionary learning problem (an instance of the sparse representation problem), the guarantees available for Algorithm 1 degrade with the number of desired atoms, K and does not depend on the number, m, of considered target signals.

In spite of the above issue with the near-optimality guarantee, in practice, the greedy algorithm empirically performs well for large m. Here, we argue that this has to do with the properties of (8) when $m \to \infty$.

3.2. Asymptotics of sparse representation objective function

Let the objective function for $m \to \infty$ be given by

$$f(\mathcal{S}) := \lim_{m\to\infty} f_m(\mathcal{S}) = \mathbb{E}\{\|\mathbf{y}\|_2^2 - \|R_{\mathcal{S}}\|_2^2\}, \tag{10}$$

where $\mathbb{E}\{\cdot\}$ stands for the expectation over the input space; and $R_{\mathcal{S}} := \mathbf{y} - \mathbf{y}_{\mathcal{S}}$.

After some mathematical manipulations (see Appendix A), the marginal gain of (10), i.e., $\Delta(s|\mathcal{S}) := f(s\cup\mathcal{S}) - f(\mathcal{S})$, can be written as

$$\Delta(s|\mathcal{S}) = \text{tr}\left\{\frac{1}{\|\mathbf{v}_s\|_2^2}\mathbf{v}_s^H\mathbf{R}_{\mathbf{y}}\mathbf{v}_s\right\} \tag{11}$$

where $\mathbf{v}_s := (\mathbf{I} - \boldsymbol{\Pi}_{\mathcal{S}})\boldsymbol{\phi}_s$; and $\mathbf{R}_{\mathbf{y}} := \mathbb{E}\{\mathbf{y}\mathbf{y}^H\}$.

Note that when there is no structure in the input space, i.e., $\mathbf{R}_{\mathbf{y}} = \mathbf{I}$, the marginal gain $\Delta(s|\mathcal{S})$ is independent of both $\mathcal{S}$ and s; that is,

$$\Delta(s|\mathcal{S}) = 1 \;\forall\, s \in \mathcal{N}\setminus\mathcal{S},\; \mathcal{S}\subset\mathcal{N}. \tag{12}$$

The assertion in (12) implies that in such a case, (10) is a *modular* set function, i.e., equality holds in (2).

In practice, a sparse representation is usually searched under the assumption that there exists some *structure* in the input space, hence the above results seem irrelevant. However, building on the developed intuition, in the following, we show that function (10) is indeed submodular *independently* of the structure of the input signal space.

Before showing the submodularity of f, we establish its following property.

Algorithm 1 Greedy Algorithm

Input: $f(\cdot)$, K, $\mathcal{N}$.
Output: $\mathcal{S}_K$.
Initialization: $\mathcal{S}_0 = \emptyset$;
for $k = 1$ to K **do**
  $i^* = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}_{k-1}} f(\mathcal{S}\cup\{i\})$
  $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1}\cup\{i^*\}$;
end for

Proposition 1. *The set function f is a normalized and nondecreasing (monotonic) set function.*

Proof. We show that f is normalized by noticing that if $\mathcal{S} = \emptyset$, then $\mathbf{y}_{\mathcal{S}} = \mathbf{0}$, and therefore $f(\emptyset) = \mathbb{E}\{\|\mathbf{y}\|_2^2 - \|\mathbf{y}\|_2^2\} = 0$. To show the monotonic behavior of f, it is sufficient to show that $f(\mathcal{S}) \leqslant f(\mathcal{S}\cup\{a\})$ where $a \notin \mathcal{S}$ is an index of a column of $\boldsymbol{\Phi}$. This fact can be seen by setting the ath coefficient of $\mathbf{y}_{\mathcal{S}\cup\{a\}}$ to 0 (a special case of $\mathbf{y}_{\mathcal{S}\cup\{a\}}$): this selection leads to $f(\mathcal{S}) = f(\mathcal{S}\cup\{a\})$. Thus, in general, $f(\mathcal{S}) \leqslant f(\mathcal{S}\cup\{a\})$. □

Now, to show the submodularity of $f(\cdot)$, we need to introduce the following result.

Lemma 1. *Let $\boldsymbol{y}$ be a vector in a Hilbert space $\mathcal{H}$. If $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{H}$ be two normalized vectors (i.e., unit norm) such that $\text{span}(\boldsymbol{u}, \boldsymbol{v}) = \mathcal{H}$, then, the following inequality holds:*

$$\mathbb{E}\{\|\boldsymbol{y}\|_2^2\} \geqslant \mathbb{E}\{\|\boldsymbol{y} - \langle\boldsymbol{y},\boldsymbol{u}\rangle\boldsymbol{u}\|_2^2 + \|\boldsymbol{y} - \langle\boldsymbol{y},\boldsymbol{v}\rangle\boldsymbol{v}\|_2^2\}. \tag{13}$$

Proof. See Appendix B. □

The above lemma provides a way to guarantee that on average, the energy of the target signal is not over-represented by arbitrary vectors spanning the space wherein it lives. Using the above lemma, we are ready to state the following result.

Theorem 3. *The function $f(\cdot)$ defined in (10) is submodular.*

Proof. See Appendix C. □

The result in Theorem 3 has the following implication.

Corollary 1. *Let f be given as (10), and $\mathcal{I} = \mathcal{I}_K$, for some prescribed K. Then, the set $\mathcal{S}_{\text{G}}$ returned by Algorithm 1 provides a $(1-\text{e}^{-1})$-approximation for the cardinality-constrained subset selection problem.*

Despite that the above results guarantee near-optimality for the set constructed by the greedy algorithm, this, in general, requires knowledge of the statistics of the input space, i.e., $\mathbf{R}_{\mathbf{y}}$. Unfortunately, in many cases, this information is not available (or is only known approximately). Thus, Algorithm 1 cannot be directly employed in practice. In the next section, we retake our previous toy example and discuss the single-point-estimate version of the greedy algorithm which deals with the (possible) lack of statistics of the input space. And we provide near-optimality guarantees for the finite-sample case of (10).

4. Submodular matching pursuit

Recall that to maximize f [c.f. (10)], Algorithm 1 selects, at the kth step, the element which maximizes the marginal gain; that is,

$$i^* = \arg\max_{i\notin\mathcal{S}_{k-1}} \Delta(i|\mathcal{S}_{k-1}) = \mathbb{E}\left\{\frac{\langle R_{\mathcal{S}_{k-1}}, \boldsymbol{\phi}_i\rangle^2}{\|(\mathbf{I} - \boldsymbol{\Pi}_{\mathcal{S}_{k-1}})\boldsymbol{\phi}_i\|_2^2}\right\}. \tag{14}$$

Different from OMP (and MP), the greedy heuristic in (14) selects the column of the dictionary with the largest *expected weighted* inner

Algorithm 2 SMP :: Single-point-estimate

Input: $\boldsymbol{\Phi}$, $\mathbf{y}$, K.
Output: $\mathcal{S}$.
Initialization: $\mathcal{S} = \emptyset$; $R = \mathbf{y}$; $\mathbf{Z} = []$; $\mathbf{V} = []$

1: **for** $k = 1$ to K **do**
2: **for** $i \notin \mathcal{S}$ **do**
3: **if** $k = 1$ **then**
4: $\boldsymbol{\Pi}_{\mathcal{S}} = \mathbf{0}$
5: **end if**
6: $\tilde{\boldsymbol{\phi}}_i = (\mathbf{I} - \boldsymbol{\Pi}_{\mathcal{S}})\boldsymbol{\phi}_i$
7: $\tilde{\boldsymbol{\phi}}_i = \tilde{\boldsymbol{\phi}}_i / ||\tilde{\boldsymbol{\phi}}_i||_2$
8: **end for**
9: $i^* = \arg\max_{i \notin \mathcal{S}} |\tilde{\boldsymbol{\phi}}_i^H R|$
10: $\mathcal{S} = \mathcal{S} \cup \{i^*\}$
11: $\mathbf{V} = [\mathbf{V}\ \tilde{\boldsymbol{\phi}}_{i^*}]$
12: Update $\mathbf{Z} = (\mathbf{V}^H\mathbf{V})^{-1}$ using (42)
13: $\boldsymbol{\Pi}_{\mathcal{S}} = \mathbf{V}\mathbf{Z}\mathbf{V}^H$
14: $R = (\mathbf{I} - \boldsymbol{\Pi}_{\mathcal{S}})\mathbf{y}$
15: **end for**

product with respect to the residual. The applied weights in (14) are given by the norm of the component of the column (atom) living in the orthogonal subspace to $\text{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$. From this point on, we refer to Algorithm 1 as Submodular Matching Pursuit (SMP) when the greedy rule (14) is employed.

4.1. Revisiting our motivating example

To illustrate how the weights of (14) work, let us recall our toy example [c.f. (6)] and consider that the expectation in (14) is removed, and only a particular target signal is assumed. After the first step of the greedy algorithm, i.e., $\mathcal{S}_1 = \{1\}$, we have the following *reweighted* atoms

$$\mathcal{V}_1 := \left\{ \tilde{\mathbf{v}}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \tilde{\mathbf{v}}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}, \tag{15}$$

where the atoms are the *decorrelated* versions of the previous columns, i.e., $\tilde{\mathbf{v}}_i = (\mathbf{I} - \boldsymbol{\Pi}_{\{1\}})\boldsymbol{\phi}_i / \|(\mathbf{I} - \boldsymbol{\Pi}_{\{1\}})\boldsymbol{\phi}_i\|_2$. Notice that the last atom is kept unchanged as it has a zero inner product with the first column. Calculating the inner products with the target signal we have

$$\langle \mathbf{y}, \tilde{\mathbf{v}}_2 \rangle = 10, \ \langle \mathbf{y}, \tilde{\mathbf{v}}_3 \rangle = 1. \tag{16}$$

Thus, the second column is selected in the second iteration. This version of the algorithm, referred to as single-point-estimate SMP (see Algorithm 2), is indeed the optimized orthogonal matching pursuit (OOMP) method [6]. Here, we refer to it as *single-point-estimate* as the expectation is approximated with a single sample (unique target signal). Although OOMP was originally derived to alleviate the common issues of OMP, we arrive at it by means of submodularity, taking a completely different route. This further supports the evidence that OOMP can achieve good results in practice as it leverages intrinsically the submodularity of the underlying cost function. It is worth noting that in the case of a dictionary learning problem with $m > 1$, the solution is obtained by substituting line 9 in Algorithm 2 with the following expression:

$$i^* = \arg\max_{i \notin \mathcal{S}} \Delta(i|\mathcal{S}), \tag{17}$$

where $\Delta(i|\mathcal{S})$ is defined based on the objective function f in (8).

4.2. A finite-sample result for submodular matching pursuit

Though we have shown that asymptotically, $m \to \infty$, the set function f_m is a submodular set function, most of the cases of practical interest occur when m is finite; that is, in the *finite-sample regime*. Thus, in the following, we provide a near-optimality guarantee for the case where the finite-sample version [c.f. (8)] (and therefore any variant of SMP) is used to find a solution for (10) using Algorithm 1.

Theorem 4. *Let f_m and f be defined as in (8) and (10), respectively. Further, assume that the target signals $\{\mathbf{y}^{(i)}\}_{i=1}^m$ are i.i.d. Given a prescribed K and tolerance ϵ_m, the set $\mathcal{S}_{\text{G}}$ returned by Algorithm 1, using f_m as input, satisfies*

$$\text{P}(f_m(\mathcal{S}_{\text{G}}) \geq (1 - \text{e}^{-1}) f(\mathcal{S}_{\text{OPT}}) - (2K+1)\epsilon_m) = p_K \tag{18}$$

where

$$\ln p_K = (2K+1)K \ln(1 - \sigma^2/m\epsilon_m^2), \tag{19}$$

where $\sigma^2 := \max_{\mathcal{S} \subseteq \mathcal{V}} \mathbb{E}\{|f_m(\mathcal{S}) - f(\mathcal{S})|^2\}$.

Proof. See Appendix D □

This result is similar in flavor to that of [32] in the sense that it shows the *near-optimality* of employing the cost function (8) to solve the signal representation problem, but differs in the following aspects: (*i*) it is probabilistic; (*ii*) it has a dependency on the number of target signals considered; and (*iii*) it is applicable to Algorithm 2 even though the true objective to maximize is f. Due to (*iii*), we can then guarantee (probabilistically) the near-optimality of the set returned by Algorithm 2 (and hence OOMP) for the signal representation problem as it is a special case of the result of Theorem 4 with $m = 1$.

5. Constrained submodular matching pursuit

Although in the literature there have been efforts to extend classical matching pursuit algorithms to settings with constraints, see, e.g., [33–36], most of this work either focus on linear constraints, i.e., restriction of the solution to a linear subspace, or in the modification altogether of the cost to optimize to promote structural characteristics in the solution. These considerations lead to algorithms, that though based on MP methods, deviate significantly from their original formulation. Thus, in many cases, they do not enjoy the simplicity of MP approaches.

Deviating from the type of constraints addressed in the above-mentioned works, we here shift our attention to constraints that are pervasive in selection problems: *combinatorial constraints*. Also, we show that for these kinds of constraints, under minor modifications to Algorithm 2, near-optimal guarantees can be provided. In this section, we first formally introduce these kinds of constraints, namely knapsack and matroid constraints, and provide examples of applications, within signal processing, wherein they naturally appear. Then, we discuss the modifications to Algorithm 2 that are required to provide near-optimal guarantees by means of a greedy heuristic.

5.1. Knapsack constraints

Instead of enforcing a cardinality constraint for the subset selection problem in Definition 3, in many applications, the elements in $\mathcal{N}$ might have non-uniform costs, i.e., $p_s, s \in \mathcal{N}$. Thus, given a budget B, we are interested in finding a set $\mathcal{S} \subseteq \mathcal{N}$ maximizing $f(\mathcal{S})$, subject to the budget constraint $p_{\mathcal{S}} = \sum_{s \in \mathcal{S}} p_s \leqslant B$. This situation arises, for example, in the sensing coverage problem (Fig. 1) where each of the sensors has a given operative cost. Different sources of energy, variety of accommodation costs, or level of energy efficiency can be the reason behind this difference in operative cost. Subsequently, in such cases, the constraint on the number of sensors, the cardinality constraint, converts to a budget constraint in which the total cost of selected sensors should not exceed the given budget.

In order to maximize the knapsack-constrained representation problem, we build a solution using two variants of the proposed SMP

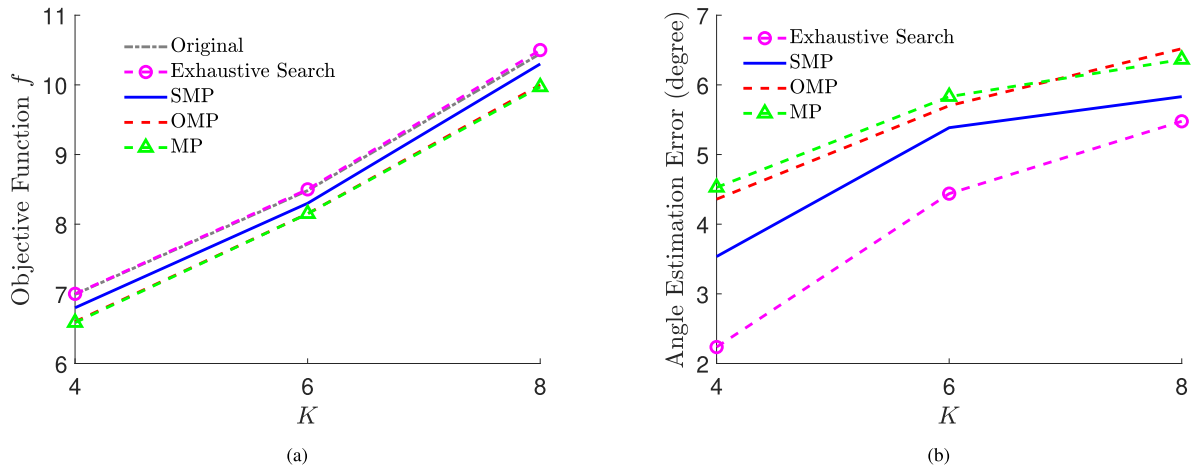


Fig. 2. Performance comparison of different algorithms for a small scenario with a cardinality constraint (a) Objective function versus K, (b) Estimation error versus K.

algorithm. The first variant substitutes the selection rule (line 9) in Algorithm 2 by

$$i^* = \arg\max \ |\tilde{\boldsymbol{\phi}}_i^H R|, \quad \text{s.t. } i \in \{j : j \notin \mathcal{S}, p_{\mathcal{S}\cup\{j\}} \leq B\}, \tag{20}$$

and in the second variant, the selection rule is substituted by the following rule

$$i^* = \arg\max \ |\tilde{\boldsymbol{\phi}}_i^H R|/p_i \quad \text{s.t. } i \in \{j : j \notin \mathcal{S}, p_{\mathcal{S}\cup\{j\}} \leq B\}. \tag{21}$$

Both algorithms stop when there is no element satisfying the constraints. It is shown that if f is a normalized, monotonic, submodular set function, executing both variants of the proposed algorithm and selecting the best of the two results, achieves the optimal solution with the factor $\frac{1}{2}(1-\mathrm{e}^{-1})$ [37]. Further, with a partial enumeration of all feasible sets of cardinality one or two, and substituting the selection rule (21) in Algorithm 2, a $(1-1/\mathrm{e})$-approximation guarantee is achieved [38].

5.2. Matroid constraints

A matroid is defined as a pair $(\mathcal{N}, \mathcal{I})$ in which $\mathcal{N}$ is a finite set and $\mathcal{I} \subseteq 2^{\mathcal{N}}$ comprises any subset of $\mathcal{N}$ which satisfies the following properties:

- $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and $\mathcal{B} \in \mathcal{I}$ implies $\mathcal{A} \in \mathcal{I}$.
- $\mathcal{A}, \mathcal{B} \in \mathcal{I}$ and $|\mathcal{B}| > |\mathcal{A}|$ implies that $\exists a \in \mathcal{B} \setminus \mathcal{A}$ such that $\mathcal{A} \cup \{a\} \in \mathcal{I}$.

Recalling the sensing coverage problem, in some applications, the existence of some of the sensors can be mutually exclusive. For example, in case the locations of candidate sensors are physically overlapping, only one would fit in the considered position. Also, one can consider the case that different subgroups of sensors are fed from particular sources of energy with limited capacities. This could lead, for instance, to constraints with respect to the number of selected sensors from the different subgroups. To provide a more concrete example, let us consider $\{\mathcal{N}_i, i = 1, \ldots, M\}$ to be a partition of the ground set $\mathcal{N}$, i.e., $\mathcal{N} = \bigcup_{i=1}^{M} \mathcal{N}_i : \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \ \forall \ i, j$. Defining the matroid $(\mathcal{N}, \mathcal{I})$ with $\mathcal{I} = \{\mathcal{S} | \mathcal{S} \subseteq \mathcal{N}, |\mathcal{S} \cap \mathcal{N}_i| \leqslant K_i, i = 1, \ldots, M\}$, we obtain a so-called *partition matroid*. This kind of matroids has applications, for instance, in the problem of joint transmit/receive antenna selection, e.g., [22]. In this setting, the partition matroid, $(\mathcal{N}, \mathcal{I})$, is composed of the sets of transmitters and receivers, and K_1 and K_2 denote the maximum allowed number of transmitters and receivers, respectively, that can be selected.

Recalling the subset selection problem in Definition 3, to maximize the objective function subject to a given matroid $(\mathcal{N}, \mathcal{I})$, we employ the proposed SMP algorithm by substituting the selection rule (line 9) in Algorithm 2 by

$$i^* = \arg\max \ |\tilde{\boldsymbol{\phi}}_i^H R|, \quad \text{s.t. } i \in \{j : j \notin \mathcal{S}, \mathcal{S} \cup \{j\} \in \mathcal{I}\}. \tag{22}$$

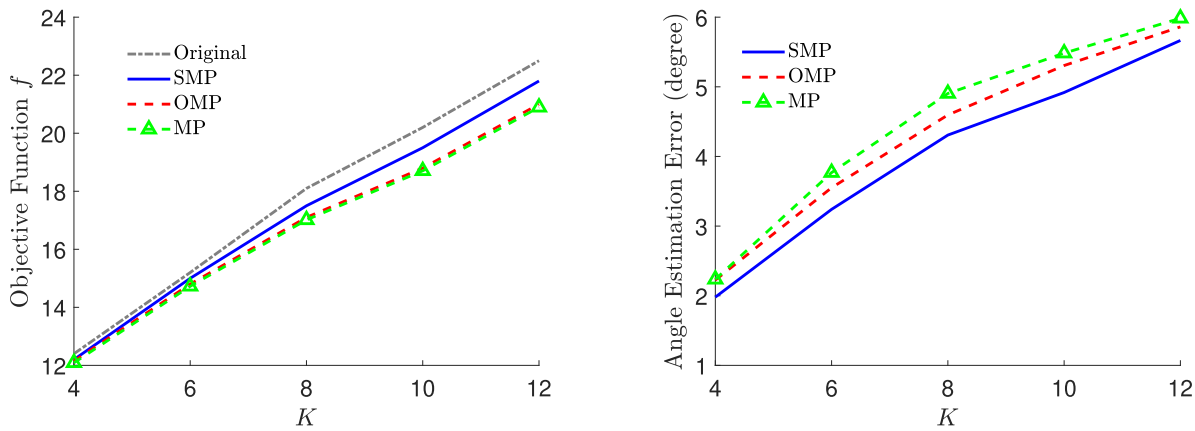
and the algorithm stops when there is no such element. It is shown that if f is a normalized, monotonic, submodular set function, the proposed algorithm achieves the optimal solution with the factor $\frac{1}{2}$ [39]. Alternatively, a $(1-1/\mathrm{e})$-approximation guarantee can be achieved if the continuous greedy algorithm in [21] is used instead.

6. Experiments

In this section, we make a comparison between the SMP algorithm and the MP and OMP algorithms. First of all, it is worth noting that the MP and OMP algorithms have computational complexities of $O(KMN)$ and $O(KMN + K^3)$, respectively. When comparing computational complexities, the proposed SMP algorithm requires approximately K times more computations. In the sequel, we compare the algorithms from their performance perspective.

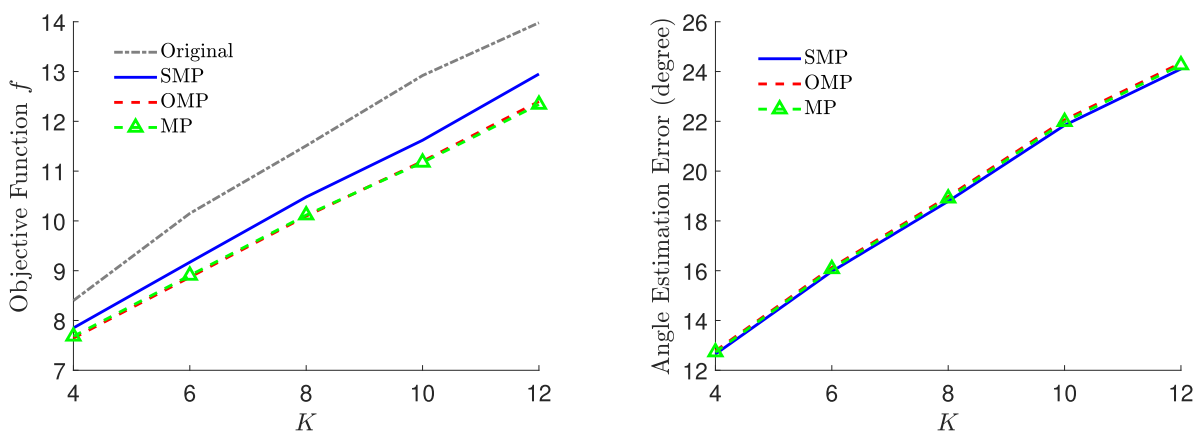
For the following set of simulations, we consider the direction of arrival (DOA) estimation problem [2]. We assume an array of M sensors placed along a line with $\lambda/2$ spacing (λ is the wavelength). Moreover, we grid the total angle span of interest, for these simulations $[-80, 80]$ degree, into a set of N discrete angle cells. Using the formulation in (1), we consider an additive white Gaussian noise (AWGN) with a signal-to-noise ratio (SNR) equal to 20dB. Also, we consider there exist K sources in the assumed angle span. To evaluate the performance of the algorithms for both signal representation and recovery problems, we consider two metrics. One is the objective function introduced in (10) (without the expectation) which determines the quality of representation. The other metric is the estimation error which is defined as $\|\boldsymbol{\theta}(\mathcal{S}) - \boldsymbol{\theta}_{Original}\|_2$ where $\boldsymbol{\theta}(\mathcal{S})$ and $\boldsymbol{\theta}_{Original}$ are the estimate and real DOAs, respectively.

For the first case, we compare SMP, OMP, and MP algorithms for a scenario with $M = 10$ and $N = 15$ for the problem with cardinality constraint. Since it is a small scenario, the optimal solution (the result of an exhaustive search) is also compared. It should be noted that since we are working with a noise-contaminated signal, the exhaustive search may result in a higher objective value in comparison with the original set of columns. Fig. 2(a) depicts the objective function versus the cardinality constraint K. As observed in the figure, SMP outperforms OMP and MP and approaches the optimal solution. Using the solution of these algorithms for signal representation, in Fig. 2(b), the estimation



(a) (b)

Fig. 3. Performance comparison of different algorithms for a large scenario with a cardinality constraint (a) Objective function versus K, (b) Estimation error versus K.



(a) (b)

Fig. 4. Performance comparison of different algorithms for a large scenario with a matroid constraint (a) Objective function versus K, (b) Estimation error versus K.

error versus K is plotted (clearly the estimation error for the original set of columns is zero and is not plotted). It is worth pointing out that a higher value of the objective function does not necessarily result in a lower estimation error. As shown in Fig. 2(b), the SMP has a lower estimation error in comparison with OMP and MP.

For the second case, in Fig. 3, we compare different algorithms for a scenario with $M = 30$ and $N = 100$ where we have the cardinality constraint. Again, the superiority of SMP over OMP and MP is clear in Figs. 3(a) and 3(b).

For the final case, we consider the representation problem subject to a matroid constraint. In this scenario, we have $M = 30$ and $N = 100$. Moreover, the matroid constraint is constructed as follows: we split the total 100 angle cells into 50 groups of non-overlapping angle cells each consisting of 2 consecutive angle cells. We put the limitation that at most one angle cell from each group can be selected. This defines a partition matroid. Fig. 4 depicts the results of different algorithms. In Fig. 4(a), although the gap between the original solution and the result of SMP is increased, SMP is still better than OMP and MP algorithms. In addition, the estimation error plotted in Fig. 4(b) presents similar results for all three algorithms.

Even though the results look quite good for cardinality constraints, the gains seem to be less impressive for matroid constraints, especially in Fig. 4(b). An intuitive explanation is that in the matroid example, we have two possibilities in each partition (consecutive angle cells). Even if the selected cell within the two cells of a partition is wrong, its contribution to the estimation error, i.e., $\|\boldsymbol{\theta}(S) - \boldsymbol{\theta}_{Original}\|_2$, is small, while the superiority of the SMP is still evident in Fig. 4(a) due to applying a different objective function.

7. Conclusion

Investigating the problem of signal representation from a new perspective, we introduced a new formulation and proved that the objective function is submodular. We discussed how the problems of signal representation and signal recovery are related, and thus, the same model and algorithm work for both. Besides the conventional cardinality constraint, for the first time, we posed this problem with a matroid constraint which enables using the proposed SMP algorithms for a wider variety of applications. Leveraging submodularity, we proposed the SMP algorithm that not only (for the first time) provides a valid near-optimal performance bound for any signal representation problem, but also resolves the existing deficiency in the OMP algorithm. In addition, we show the connection of the single-point-estimate version of SMP with OOMP. Finally, as shown through the simulations, the SMP algorithm outperforms the MP and OMP algorithms.

CRediT authorship contribution statement

Ehsan Tohidi: Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Mario Coutino:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **David Gesbert:** Supervision, Conceptualization.

Data availability

No data was used for the research described in the article.

Appendix A. Marginal gain expression

Let us first expand the set function and rewrite it as

$$\begin{aligned} f(\mathcal{S}) &= \mathbb{E}\{\|\mathbf{y}\|_2^2 - \|\mathbf{y}-\mathbf{y}_{\mathcal{S}}\|_2^2\} = 2\mathbb{E}\{\langle \mathbf{y},\mathbf{y}_{\mathcal{S}}\rangle\} - \mathbb{E}\{\|\mathbf{y}_{\mathcal{S}}\|_2^2\} \\ &= 2\mathbb{E}\{\mathbf{y}^H \boldsymbol{\Pi}_{\mathcal{S}}\mathbf{y}\} - \mathbb{E}\{(\mathbf{y}^H\boldsymbol{\Pi}_{\mathcal{S}})(\boldsymbol{\Pi}_{\mathcal{S}}\mathbf{y})\} = \mathbb{E}\{\mathbf{y}^H\boldsymbol{\Pi}_{\mathcal{S}}\mathbf{y}\}, \end{aligned} \tag{23}$$

where the fact $\boldsymbol{\Pi}_{\mathcal{S}}\boldsymbol{\Pi}_{\mathcal{S}} = \boldsymbol{\Pi}_{\mathcal{S}}$ has been used. Now, using the above expression, the marginal gain $\Delta(s|\mathcal{S})$ can be expressed as

$$\begin{aligned} \Delta(s|\mathcal{S}) &= f(\mathcal{S}\cup\{s\}) - f(\mathcal{S}) = \mathbb{E}\{\mathbf{y}^H(\boldsymbol{\Pi}_{\mathcal{S}} + \boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s})\mathbf{y}\} - \mathbb{E}\{\mathbf{y}^H\boldsymbol{\Pi}_{\mathcal{S}}\mathbf{y}\} \\ &= \mathbb{E}\{\mathbf{y}^H\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s}\mathbf{y}\}. \end{aligned} \tag{24}$$

Recalling that $\tilde{\mathbf{v}}_s := \mathbf{v}_s/\|\mathbf{v}_s\|_2$, with $\mathbf{v}_s := (\mathbf{I}-\boldsymbol{\Pi}_{\mathcal{S}})\boldsymbol{\phi}_s$, the projection matrix $\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s}$ can be written as

$$\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s} = \tilde{\mathbf{v}}_s\tilde{\mathbf{v}}_s^H = \frac{1}{\|\mathbf{v}_s\|_2^2}\mathbf{v}_s\mathbf{v}_s^H. \tag{25}$$

Hence, applying trace properties to the terms in (24), i.e.,

$$\Delta(s|\mathcal{S}) = \mathbb{E}\{\mathbf{y}^H\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s}\mathbf{y}\} = \mathbb{E}\{\mathrm{tr}\big(\mathbf{y}^H\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s}\mathbf{y}\big)\} = \mathrm{tr}\big(\boldsymbol{\Pi}_{\tilde{\mathbf{v}}_s}\mathbb{E}\{\mathbf{y}\mathbf{y}^H\}\big), \tag{26}$$

and setting $\mathbf{R}_{\mathbf{y}} := \mathbb{E}\{\mathbf{y}\mathbf{y}^H\}$, the expression (11) for the marginal gain is obtained.

Appendix B. Proof of Lemma 1

Proof. Expanding right side of (13) leads to

$$\begin{aligned} &\mathbb{E}\{\|\mathbf{y}\|_2^2\} \geqslant \mathbb{E}\{\|\mathbf{y}\|_2^2 + \langle\mathbf{y},\mathbf{u}\rangle^2 - 2\langle\mathbf{y},\mathbf{u}\rangle^2 + \|\mathbf{y}\|_2^2 + \langle\mathbf{y},\mathbf{v}\rangle^2 - 2\langle\mathbf{y},\mathbf{v}\rangle^2\} \\ \Rightarrow\ &\mathbb{E}\{\langle\mathbf{y},\mathbf{u}\rangle^2 + \langle\mathbf{y},\mathbf{v}\rangle^2\} \geqslant \mathbb{E}\{\|\mathbf{y}\|_2^2\} \ \Rightarrow\ \mathbb{E}\{\frac{\|\mathbf{y}\|_2^2}{2} + \frac{\|\mathbf{y}\|_2^2}{2}\} \geqslant \mathbb{E}\{\|\mathbf{y}\|_2^2\}, \end{aligned} \tag{27}$$

where the last inequality is obtained by the following integration:

$$\begin{aligned} \mathbb{E}\{\langle\mathbf{y},\mathbf{u}\rangle^2\} &= \frac{\mathbb{E}\{\|\mathbf{y}\|_2^2\}}{2\pi}\int_0^{2\pi}\cos^2\theta d\theta = \frac{\mathbb{E}\{\|\mathbf{y}\|_2^2\}}{2\pi}\int_0^{2\pi}\frac{1+\cos 2\theta}{2}d\theta \\ &= \frac{\mathbb{E}\{\|\mathbf{y}\|_2^2\}}{2}, \end{aligned} \tag{28}$$

where a uniform distribution is considered for the angle between $\mathbf{y}$ and $\mathbf{u}$. This assumption is due to the fact that $\mathbf{y}$ in Lemma 1 is not the original signal but a projection of the original signal into the span of two vectors as derived in Theorem 3 and particularly in (34). As there is no prior information on these two vectors, it is reasonable to consider a uniform distribution for the angle between the projected part and $\mathbf{u}$.

It is worth mentioning that the uniform distribution is not the only distribution that brings the result of (28). Based on (28), it is straightforward to see that (27) is met for any probability density function (pdf) of the angle between $\mathbf{y}$ and $\mathbf{u}$, denoted by $f(\theta)$, as long as the inequality $\int_0^{2\pi} f(\theta)\cos 2\theta d\theta \geqslant 0$ is satisfied. Apart from the pdfs that hold the constraint in an equality form (e.g., the uniform distribution as shown in (28)), for any pdf $f_1(\theta)$ that strictly satisfies the inequality, there is a corresponding pdf $f_2(\theta) = f_1((\theta+\pi/2) \mod 2\pi)$ that violates the constraint, and vice versa, i.e.,

$$\begin{aligned} \int_0^{2\pi} f_2(\theta)\cos 2\theta d\theta &= \int_0^{2\pi} f_1(\theta)\cos(2\theta-\pi)d\theta \\ &= -\int_0^{2\pi} f_1(\theta)\cos 2\theta d\theta. \end{aligned} \tag{29}$$

Consequently, more than half of the possible pdfs (i.e., including the ones that satisfy the equality form plus half of the rest) lead to the result of Lemma 1. □

Appendix C. Proof of Theorem 3

Proof. Substituting (10) in (2) leads to

$$\begin{aligned} &\mathbb{E}\{\|\mathbf{y}\|_2^2 - \|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{a\}}\|_2^2 - \|\mathbf{y}\|_2^2 + \|\mathbf{y}-\mathbf{y}_{\mathcal{S}}\|_2^2\} \\ &\quad\geqslant \mathbb{E}\{\|\mathbf{y}\|_2^2 - \|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{a,b\}}\|_2^2 - \|\mathbf{y}\|_2^2 + \|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{b\}}\|_2^2\}. \end{aligned} \tag{30}$$

Simplifying equal terms from both sides, (30) reduces to

$$\mathbb{E}\{\|\mathbf{y}-\mathbf{y}_{\mathcal{S}}\|_2^2 - \|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{a\}}\|_2^2\} \geqslant \mathbb{E}\{\|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{b\}}\|_2^2 - \|\mathbf{y}-\mathbf{y}_{\mathcal{S}\cup\{a,b\}}\|_2^2\}. \tag{31}$$

Here, we introduce $\boldsymbol{\phi}_{\{\tilde{a}\}}$ and $\boldsymbol{\phi}_{\{\tilde{b}\}}$ in order to present the part of $\boldsymbol{\phi}_{\{a\}}$ and $\boldsymbol{\phi}_{\{b\}}$, respectively, that are not in $\mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$, i.e., $\boldsymbol{\phi}_{\{\tilde{a}\}}, \boldsymbol{\phi}_{\{\tilde{b}\}} \perp \mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}}\}$. Moreover, we restate $\mathbf{y}$ as

$$\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{y}_{\mathcal{S}} + \mathbf{y}_{\{\tilde{a},\tilde{b}\}}, \tag{32}$$

where $\tilde{\mathbf{y}}$ is defined as the residual of projecting $\mathbf{y}$ over $\mathrm{span}\{\boldsymbol{\Phi}_{\mathcal{S}\cup\{\tilde{a},\tilde{b}\}}\}$, i.e., $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{y}_{\mathcal{S}\cup\{\tilde{a},\tilde{b}\}}$. Expanding $\mathbf{y}$ using (32) and following some simplifications we obtain

$$\mathbb{E}\{\|\tilde{\mathbf{y}} + \mathbf{y}_{\{\tilde{a},\tilde{b}\}}\|_2^2 - \|\tilde{\mathbf{y}} + \mathbf{y}_{\{\tilde{a},\tilde{b}\}} - \mathbf{y}_{\{\tilde{a}\}}\|_2^2\} \geqslant \mathbb{E}\{\|\tilde{\mathbf{y}} + \mathbf{y}_{\{\tilde{a},\tilde{b}\}} - \mathbf{y}_{\{\tilde{b}\}}\|_2^2 - \|\tilde{\mathbf{y}}\|_2^2\}. \tag{33}$$

Since $\tilde{\mathbf{y}}$ is orthogonal to the other terms in (33), the following inequality is obtained:

$$\mathbb{E}\{\|\mathbf{y}_{\{\tilde{a},\tilde{b}\}}\|_2^2\} \geqslant \mathbb{E}\{\|\mathbf{y}_{\{\tilde{a},\tilde{b}\}} - \mathbf{y}_{\{\tilde{a}\}}\|_2^2 + \|\mathbf{y}_{\{\tilde{a},\tilde{b}\}} - \mathbf{y}_{\{\tilde{b}\}}\|_2^2\}, \tag{34}$$

which is correct based on Lemma 1. □

Appendix D. Proof of Theorem 4

Proof. To show the result, consider the following chain of inequalities, i.e.,

$$\begin{aligned} f(\mathcal{S}_{\mathrm{OPT}}) &\overset{(a)}{\leq} f(\mathcal{S}_{\mathrm{OPT}}\cup\mathcal{S}_i) \overset{(b)}{=} f(\mathcal{S}_i) + \sum_{j=1}^{k}\Delta(v_j^*|\mathcal{S}_i\cup v_1,\dots,v_{j-1}) \\ &\overset{(c)}{\leq} f(\mathcal{S}_i) + \sum_{v\in\mathcal{S}^*}\Delta(v|\mathcal{S}_i) \end{aligned} \tag{35}$$

$$\overset{(d)}{\leq} f_m(\mathcal{S}_i) + \epsilon_m + \sum_{v\in\mathcal{S}_{\mathrm{OPT}}}[\Delta_{i+1} + 2\epsilon_m] = f_m(\mathcal{S}_i) + K\Delta_{i+1} + \alpha\epsilon_m, \tag{36}$$

where (a) is due to monotonicity; (b) is established by a telescopic sum; (c) is result of submodularity; and (d) holds with probability $(1-\sigma^2/m\epsilon_m^2)^{2K+1}$ by a Chernoff's bound-type inequality. $\Delta_{i+1} := f_m(\mathcal{S}_{i+1}) - f_m(\mathcal{S}_i)$ and $\alpha = 2K+1$.

Now, we can define the following quantity.

$$\delta_i := f(\mathcal{S}_{\mathrm{OPT}}) - f_m(\mathcal{S}_i) \overset{(a)}{\leq} K\Delta_{i+1} + \alpha\epsilon_m = K(\delta_i - \delta_{i+1}) + \alpha\epsilon_m, \tag{37}$$

where (a) is established using the previous inequalities. Rearranging terms from the above expression, we get

$$\delta_{i+1} \leq \frac{(K-1)}{K}\delta_i + \epsilon_m', \tag{38}$$

where $\epsilon_m' := (2+K^{-1})\epsilon_m$.

Finally, applying the above inequality repetitively (k times), we obtain

$$\delta_k \leq \left(\frac{K-1}{K}\right)^k\delta_0 + \epsilon_m'\sum_{l=0}^{k-1}\left(\frac{K-1}{K}\right)^l \overset{(a)}{\leq} \mathrm{e}^{-k/K}f(\mathcal{S}_{\mathrm{OPT}}) + k\epsilon_m', \tag{39}$$

which leads to the desired result

$$f_m(\mathcal{S}_{\mathrm{G}}) \geq (1-\mathrm{e}^{-1})f(\mathcal{S}_{\mathrm{OPT}}) - K\epsilon_m'. \ \square \tag{40}$$

Appendix E. Computational aspects: Matrix update

Given a matrix $\mathbf{X}$ with inverse $\mathbf{Y} = (\mathbf{X}^H\mathbf{X})^{-1}$, the procedure for updating $\mathbf{Y}$, when a column is added to the matrix $\mathbf{X}$, is given by the following procedure.

Consider $\mathbf{X}_{\text{new}} = \begin{bmatrix}\mathbf{X} & \mathbf{v}\end{bmatrix}$, thus

$$\mathbf{Y}_{\text{new}}^{-1} = \mathbf{X}_{\text{new}}^H\mathbf{X}_{\text{new}} = \begin{bmatrix}\mathbf{X}^H \\ \mathbf{v}^H\end{bmatrix}\begin{bmatrix}\mathbf{X} & \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^H\mathbf{X} & \mathbf{X}^H\mathbf{v} \\ \mathbf{v}^H\mathbf{X} & \mathbf{v}^H\mathbf{v}\end{bmatrix} \tag{41}$$

Using the inverse of a partitioned matrix, the updated inverse is achieved as follows [40]

$$\mathbf{Y} = \begin{bmatrix}\mathbf{X}^H\mathbf{X} & \mathbf{X}^H\mathbf{v} \\ \mathbf{v}^H\mathbf{X} & \mathbf{v}^H\mathbf{v}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{F} & -\alpha\mathbf{Y}\mathbf{X}^H\mathbf{v} \\ -\alpha\mathbf{v}^H\mathbf{X}\mathbf{Y}^H & \alpha\end{bmatrix} \tag{42}$$

where

$$\alpha = \frac{1}{\mathbf{v}^H\mathbf{v} - \mathbf{v}^H\mathbf{X}\mathbf{Y}\mathbf{X}^H\mathbf{v}}, \quad \mathbf{F} = \mathbf{Y} + \alpha\mathbf{Y}\mathbf{X}^H\mathbf{v}\mathbf{v}^H\mathbf{X}\mathbf{Y}^H. \tag{43}$$

Eq. (42) is then used in Algorithm 2 to alleviate the computational complexity of taking the full inverse.

References

[1] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.

[2] H.L. Van Trees, Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory, John Wiley & Sons, 2004.

[3] A.N. Akansu, Y. Liu, On-signal decomposition techniques, Opt. Eng., Bellingham 30 (7) (1991) 912–921.

[4] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.

[5] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, vol. 1, 1993, pp. 40–44.

[6] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Process. Lett. 9 (4) (2002) 137–140.

[7] M. Andrle, L. Rebollo-Neira, E. Sagianos, Backward-optimized orthogonal matching pursuit approach, IEEE Signal Process. Lett. 11 (9) (2004) 705–708.

[8] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, Found. Comput. Math. 9 (3) (2009) 317–334.

[9] J. Wang, S. Kwon, B. Shim, Generalized orthogonal matching pursuit, IEEE Trans. Signal Process. 60 (12) (2012) 6202–6216.

[10] M.A. Davenport, M.B. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, IEEE Trans. Inform. Theory 56 (9) (2010) 4395–4401.

[11] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.

[12] S. Chen, D. Donoho, Basis pursuit, in: Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers, vol. 1, 1994, pp. 41–44.

[13] J.D. Blanchard, M. Cermak, D. Hanle, Y. Jing, Greedy algorithms for joint sparse recovery, IEEE Trans. Signal Process. 62 (7) (2014) 1694–1704.

[14] Y.C. Eldar, H. Rauhut, Average case analysis of multichannel sparse recovery using convex relaxation, IEEE Trans. Inform. Theory 56 (1) (2010) 505–519.

[15] F. Guerrini, A. Gnutti, R. Leonardi, Iterative mirror decomposition for signal representation, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2019, pp. 5541–5545.

[16] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory 50 (10) (2004) 2231–2242.

[17] A. Das, D. Kempe, Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection, 2011, arXiv preprint arXiv:1102.3975.

[18] E. Tohidi, R. Amiri, M. Coutino, D. Gesbert, G. Leus, A. Karbasi, Submodularity in action: from machine learning to signal processing applications, IEEE Signal Process. Mag. (2020).

[19] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions—I, Math. Program. 14 (1) (1978) 265–294.

[20] G. Calinescu, C. Chekuri, M. Pál, J. Vondrák, Maximizing a submodular set function subject to a matroid constraint, in: International Conference on Integer Programming and Combinatorial Optimization, Springer, 2007, pp. 182–196.

[21] G. Calinescu, C. Chekuri, M. Pal, J. Vondrák, Maximizing a monotone submodular function subject to a matroid constraint, SIAM J. Comput. 40 (6) (2011) 1740–1766.

[22] E. Tohidi, M. Coutino, S.P. Chepuri, H. Behroozi, M.M. Nayebi, G. Leus, Sparse antenna and pulse placement for colocated MIMO radar, IEEE Trans. Signal Process. 67 (3) (2019) 579–593.

[23] E. Tohidi, S. Parsaeefard, M.A. Maddah-Ali, B.H. Khalaj, A. Leon-Garcia, Near-optimal robust virtual controller placement in 5G software defined networks, IEEE Trans. Netw. Sci. Eng. 8 (2) (2021) 1687–1697.

[24] K. Thekumparampil, A. Thangaraj, R. Vaze, Combinatorial resource allocation using submodularity of waterfilling, IEEE Trans. Wireless Commun. 15 (1) (2016) 206–216.

[25] M. Coutino, S.P. Chepuri, G. Leus, Submodular sparse sensing for Gaussian detection with correlated observations, IEEE Trans. Signal Process. 66 (15) (2018) 4025–4039.

[26] E. Tohidi, S. Haesloop, L. Thiele, S. Stańczak, Near-optimal LOS and orientation aware intelligent reflecting surface placement, in: ICC 2023 - IEEE International Conference on Communications, 2023, pp. 498–504.

[27] B. Mirzasoleiman, A. Karbasi, R. Sarkar, A. Krause, Distributed submodular maximization, J. Mach. Learn. Res. 17 (235) (2016) 1–44.

[28] F. Bach, Efficient algorithms for non-convex isotonic regression through submodular optimization, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018, pp. 1–10.

[29] S. Jegelka, J.A. Bilmes, Online submodular minimization for combinatorial structures, in: ICML, Citeseer, 2011, pp. 345–352.

[30] R. Udwani, Multi-objective maximization of monotone submodular functions with cardinality constraint, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018, pp. 9493–9504.

[31] K. Wei, R. Iyer, J. Bilmes, Submodularity in data subset selection and active learning, in: International Conference on Machine Learning, 2015, pp. 1954–1963.

[32] A. Krause, V. Cevher, Submodular dictionary selection for sparse representation, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML '10, Omni Press, USA, 2010, pp. 567–574.

[33] A. Adler, V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, M.D. Plumbley, A constrained matching pursuit approach to audio declipping, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2011, pp. 329–332.

[34] T. Huang, Y. Liu, H. Meng, X. Wang, Adaptive matching pursuit with constrained total least squares, EURASIP J. Adv. Signal Process. 2012 (1) (2012) 76.

[35] M. Yaghoobi, D. Wu, M.E. Davies, Fast non-negative orthogonal matching pursuit, IEEE Signal Process. Lett. 22 (9) (2015) 1229–1233.

[36] R. Baraldi, R. Kumar, A. Aravkin, Basis pursuit denoise with nonsmooth constraints, IEEE Trans. Signal Process. 67 (22) (2019) 5811–5823.

[37] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007, pp. 420–429.

[38] M. Sviridenko, A note on maximizing a submodular set function subject to a knapsack constraint, Oper. Res. Lett. 32 (1) (2004) 41–43.

[39] M.L. Fisher, G.L. Nemhauser, L.A. Wolsey, An analysis of approximations for maximizing submodular set functions—II, in: Polyhedral Combinatorics, Springer, 1978, pp. 73–87.

[40] M. Khan, Updating Inverse of a Matrix When a Column is Added, Tech. Rep., University of British Columbia, Vancouver, 2008.