Signal Processing 142 (2018) 271–280

Contents lists available at ScienceDirect

Signal Processing

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

Fast convex optimization method for frequency estimation with prior knowledge in all dimensions

Zai Yang[a,b,*], Lihua Xie[c]

[a] School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, China
[c] School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Article history:
Received 12 November 2016
Revised 17 July 2017
Accepted 23 July 2017
Available online 24 July 2017

Keywords:
Frequency estimation
Gridless sparse methods
Weighted atomic norm
Prior knowledge

ABSTRACT

This paper investigates the frequency estimation problem in all dimensions within the recent gridless-sparse-method framework. The frequencies of interest are assumed to follow a prior probability distribution. To effectively and efficiently exploit the prior knowledge, a weighted atomic norm approach is proposed in both the 1-D and the multi-dimensional cases. Like the standard atomic norm approach, the resulting optimization problem is formulated as convex programming using the theory of trigonometric polynomials and shares the same computational complexity. Numerical simulations are provided to demonstrate the superior performance of the proposed approach in accuracy and speed compared to the state-of-the-art.

1. Introduction

In this paper we investigate the frequency estimation problem that has wide applications in radar, sonar, wireless communications, seismology, astronomy and so on. Consider the one-dimensional (1-D) case as an example, in which $f_k \in \mathbb{T} := [0, 1]$, $k = 1, \ldots, K$ denote the unknown (normalized) frequencies of interest (K is unknown as well). In the absence of noise, we have the following data model:

$$\boldsymbol{y}^o = \sum_{k=1}^{K} \boldsymbol{a}(f_k) s_k, \tag{1}$$

where $\boldsymbol{a}(f) = [1, e^{i2\pi f}, \ldots, e^{i2\pi (N-1) f}]^T \in \mathbb{C}^N$ denotes a complex sinusoid with frequency f, and $s_k \in \mathbb{C}$ is the amplitude of the kth sinusoid. Suppose that we only observe M out of the N entries of $\boldsymbol{y}^o$ indexed by $\boldsymbol{\Omega} \subset \{1, \ldots, N\}$, which form a subvector $\boldsymbol{y}^o_{\boldsymbol{\Omega}} \in \mathbb{C}^M$. Given $\boldsymbol{y}^o_{\boldsymbol{\Omega}}$, our objective is to recover the frequencies f_k, $k = 1, \ldots, K$. This problem setup is known as the compressive data case in the sense that only part of the full data $\boldsymbol{y}^o$ is observed. In the multi-dimensional (M-D) case, the problem can be stated similarly, which is deferred to the main context.

The difficulty of the frequency estimation problem underlies in the fact that the observed data are highly nonlinear functions of the frequencies. The most prominent conventional approach to 1-D frequency estimation is known as subspace-based methods such as MUSIC and ESPRIT. Readers are referred to [1] for the review. To circumvent the nonlinearity, in these methods the frequencies are estimated from the signal subspace of the data covariance matrix, rather than from the data themselves. They have also been extended to the M-D, especially the 2-D case (see, e.g., [2,3]).[1] However, since the subspace-based methods require a data covariance estimate, their application is challenging in the compressive data case in which this estimate is difficult to obtain. Moreover, the subspace methods require the knowledge of the model order (a.k.a. the number of frequencies).

In the past two decades or so, a class of sparse methods have been developed for frequency estimation thanks to the development of sparse representation and compressed sensing. These method exploit the signal sparsity, which arises from the fact that the number of sinusoids K is small. To overcome the nonlinearity, the continuous frequency domain is gridded/discretized into finite discrete grid points, and the nonlinear parameter estimation problem is then transformed approximately as the recovery of a discrete sparse signal from an underdetermined linear system. Note that gridding in the frequency domain was also necessary since the early sparse representation techniques can only deal with the recovery of discrete signals. The gridding however results in the so-called grid mismatch problem that leads to performance

* Corresponding author.
E-mail addresses: yangzai@njust.edu.cn (Z. Yang), elhxie@ntu.edu.sg (L. Xie).

[1] Note that the subspace-based methods are based on the Vandermonde decomposition of Toeplitz covariance matrices, which dates back to the early 20th century in the 1-D case [4] and was proven in the M-D case in the recent work [5].

http://dx.doi.org/10.1016/j.sigpro.2017.07.028
0165-1684/

degradation and difficulties for theoretical analysis [6,7]. These drawbacks have recently been resolved in *gridless* sparse methods, specifically, those based on the atomic norm—a continuous analog of the ℓ_1 norm [8]—and covariance fitting (see [9–11] on the 1-D case and [5,12,13] on the M-D case). In gridless sparse methods, the frequencies are dealt with directly in the continuous domain, completely resolving the grid mismatch problem; the resulting seemingly nonconvex atomic norm and covariance fitting optimization problems are cast as convex optimization problems; and theoretical guarantees are provided conditioning on that the frequencies are sufficiently separated. Readers are referred to [14] for an overview of both the grid-based and the gridless sparse methods.

This paper is motivated by practical applications in which prior knowledge on the frequencies can be known. For example, due to path loss the delay/range information of a detectable aircraft might be estimated *a priori*. Prior information of the Doppler frequency can be obtained given the characteristic speed. In underwater channel estimation, the frequency parameters of interest can reside in a known small interval [15]. To model various kinds of prior knowledge, in this paper, we assume that the 1-D or the M-D frequencies of interest follow a prior probability distribution (note that K is still unknown).

The use of prior distribution for improving estimation performance is common in the literature on statistical estimation and is typically accomplished based on statistical inference, which however needs the value of K and often requires nonconvex optimization due to the nonlinear nature of the problem [16]. Differently from the existing methods, we try to propose a deterministic convex optimization method. Concretely, we propose a *weighted* atomic norm approach in which the prior knowledge is encoded in a sophisticatedly chosen weighting function that is further used to specify the preference of frequencies. We show that a careful choice of the weighting function enables us to cast the resulting weighted approach as convex programming and to solve it using off-the-shelf solvers. While the proposed method in the 1-D case was presented in our conference publication [17], we show in this extended journal paper that similar techniques can be utilized to propose a method in the M-D case.

1.1. Related work

The paper [18] studied the 1-D case and assumed that the model order K is known and each frequency follows a prior distribution. The maximum a posterior (MAP) estimator was derived by using nonconvex optimization, which suffers from local convergence.

Weighted optimization is common in the literature on compressed sensing (in the discrete setting) for exploiting prior knowledge of the support of the sparse signal. To do so, a subset of the support is usually assumed known [19]. But results are rare in case when the prior knowledge is described in a probabilistic manner. Related papers include [20,21], in which weighted ℓ_1 norm methods were studied given the probability of each entry of the sparse signal being nonzero (note that the number of nonzeros, or K, is thus approximately known).

To date, the weighted atomic norm has been studied by two research groups for 1-D frequency estimation. In our previous work [22], no prior knowledge but sparsity was assumed and a majorizaiton-minimization algorithm was implemented based on a nonconvex objective function, resulting in a reweighted atomic norm method that iteratively enhances sparsity and resolution compared to the atomic norm. While the weighting function in [22] is automatically determined in the iterative algorithm, it is sophisticatedly designed in this paper to efficiently exploit the prior knowledge.

In [23], the so-called constrained atomic norm method, which is interpreted as the weighted atomic norm in this paper, was proposed to deal with a class of piecewise-constant prior distributions. Inspired by [22,23], a new reweighted atomic norm method was introduced in [24] in the absence of prior knowledge. Note that, compared to [23], the approach of this paper can deal with more general priors. When working with the specialized priors in [23], the proposed convex optimization method has significantly fewer constraints and thus lower computational complexity. In fact, it shares the same computational complexity as the standard atomic norm method in which prior knowledge is not considered. It is shown via numerical simulations that the proposed method can be an order of magnitude faster than the method in [23], with comparable accuracy. Finally, the presented solution can be applied to the M-D case. To the best of our knowledge, this is the only convex optimization method for M-D frequency estimation which can exploit the prior knowledge and work in the continuous domain.

1.2. Paper organization

The rest of this paper is organized as follows. Section 2 introduces the problem statement and presents the proposed solution in the context of general priors in the 1-D case. The proposed solution is also specialized there to the case of block priors and compared to the method in [23]. Section 3 extends the presented solution from the 1-D to the M-D case. Section 4 provides numerical simulations to demonstrate the performance of the proposed solution. Section 5 concludes this paper.

2. Proposed solution: the 1-D case

2.1. Problem statement

Let $F \in [0, 1]$ be a random variable that describes the 1-D frequencies $\{f_k\}$ and $p(f)$ be the probability density function (pdf) of F. The objective is to recover the frequencies $\{f_k\}$ given the data model in (1), the observed data $\boldsymbol{y}^o_{\boldsymbol{\Omega}}$ and the prior distribution $p(f)$.

2.2. Preliminaries

We first recall the atomic norm method for frequency estimation in which only the signal sparsity is exploited. Note that $\boldsymbol{y}^o$ in (1) can be written as a linear combination of K atoms in the following set of atoms:

$$\mathcal{A} = \{\boldsymbol{a}(f, \phi) = \boldsymbol{a}(f)\phi : \ f \in \mathbb{T}, \phi \in \mathbb{S}^1\}, \tag{2}$$

where $\mathbb{S}^1 = \{\phi \in \mathbb{C} : \ |\phi| = 1\}$ denotes the unit circle. Following from the literature on sparse representation, we attempt to solve the following optimization problem:

$$\min_{\boldsymbol{y}} \|\boldsymbol{y}\|_{\mathcal{A},0}, \ \text{subject to } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}, \tag{3}$$

where $\|\boldsymbol{y}\|_{\mathcal{A},0}$ denotes the atomic ℓ_0 norm that is defined as the smallest number of atoms composing $\boldsymbol{y}$—a continuous analog of the ℓ_0 norm. Note that, by using the ℓ_0 norm, the signal sparsity is exploited to the greatest extent possible. But unfortunately, (3) is a rank minimization problem that cannot be easily solved [10]. Therefore, its convex relaxation—the following atomic norm problem is considered instead [9,10]:

$$\min_{\boldsymbol{y}} \|\boldsymbol{y}\|_{\mathcal{A}}, \ \text{subject to } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}, \tag{4}$$

where $\|\boldsymbol{y}\|_{\mathcal{A}}$ denotes the atomic norm that is defined as:

$$\|\boldsymbol{y}\|_{\mathcal{A}} := \inf_{f_k, \phi_k, c_k > 0} \left\{ \sum_k c_k : \ \boldsymbol{y} = \sum_k c_k \boldsymbol{a}(f_k, \phi_k) \right\}. \tag{5}$$