Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2018-0006

Vol. **24**, No. 1, pp. 286-308
July 2018

A Gradient-Enhanced ℓ_1 Approach for the Recovery of Sparse Trigonometric Polynomials

Zhiqiang Xu[1] and Tao Zhou[1,*]

[1]*LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China.*

Received 9 January 2018; Accepted (in revised version) 7 February 2018

Abstract. In this paper, we discuss a gradient-enhanced ℓ_1 approach for the recovery of sparse Fourier expansions. By *gradient-enhanced* approaches we mean that the directional derivatives along given vectors are utilized to improve the sparse approximations. We first consider the case where both the function values and the directional derivatives at sampling points are known. We show that, under some mild conditions, the inclusion of the derivatives information can indeed decrease the coherence of measurement matrix, and thus leads to the improved the sparse recovery conditions of the ℓ_1 minimization. We also consider the case where either the function values or the directional derivatives are known at the sampling points, in which we present a sufficient condition under which the measurement matrix satisfies RIP, provided that the samples are distributed according to the uniform measure. This result shows that the derivatives information plays a similar role as that of the function values. Several numerical examples are presented to support the theoretical statements. Potential applications to function (Hermite-type) interpolations and uncertainty quantification are also discussed.

AMS subject classifications: 65D15, 41A10, 41A63

Key words: Gradient-enhanced ℓ_1 minimization, compressed sensing, sparse Fourier expansions, restricted isometry property, mutual incoherence.

1 Introduction

Compressed sensing (CS), introduced by Candes, Romberg & Tao [14] and Donoho [25], has been a hot research field in recent years [2,3,7,12,16–19,29,37]. The main motivation for CS is that many real world signals can be well approximated by sparse ones, more precisely, they can be approximated by an expansion in terms of a suitable basis (e.g,

*Corresponding author. *Email addresses:* Xuzq@lsec.cc.ac.cn (Z. Xu), tzhou@lsec.cc.ac.cn (T. Zhou)

Fourier basis), which has only a few non-vanishing terms. CS predicts that sparse vectors in high dimensions can be recovered from few measurements. Recently, CS has been successfully applied in many areas, such as imaging [42,46], radar [24], wireless communication [36], Magnetic Resonance Imaging [15], uncertainty quantification [23,26,32], to name a few.

In this work, we consider the recovery of sparse Fourier expansions by ℓ_1 minimization. Particularly, we propose in this work a *gradient-enhanced* ℓ_1 minimization, which means that the *gradient information* of the Fourier expansions at samples are included in the ℓ_1 minimization. In other words, we consider the ℓ_1 minimization with both function evaluations and the derivative information and we will study the effect of the derivative on the performance of ℓ_1 minimization. We next briefly introduce the motivation for considering the derivative information:

- **Uncertainty quantification (UQ) for random PDEs.** In UQ, one represents the solution as a linear combination of certain bases, such as orthonormal polynomials and Fourier bases, and then uses a few sample evaluations to obtain the sparse approximation of the solutions. Given the sample evaluations, the derivative information can sometime be obtained in a cheaper way, e.g, by solving the adjoint equations [13, 21, 41]. This is a well known approach in the numerical PDE community, and has been used in UQ studies [4,25,28,35]. Naturally, one hopes to use both the function evaluations and the gradient information to enhance the sparse approximation of the solution to PDE.

- **Sparse Hermite-type interpolation.** Recently, one employs the result in CS to study the sparse interpolation [1, 14, 40, 52] which shows many potential applications. To find a function from a finite dimensional function space which interpolates function value and derivatives at some points is called Hermite interpolation [5, 31, 44, 48]). The gradient-enhanced approach here can be viewed as a sparse Hermite-type interpolation which seeks to find a *sparse* interpolation from the function values and derivatives. Hence, the results in this paper also show the connection between compressed sensing and the classical approximation theory.

A simple observation is that the gradient-enhanced ℓ_1 approach uses more information than the standard approach (or in other words, enhance the row size of the sensing matrix), and this opens up the possibility to improve the sparse recovery ability of the ℓ_1 minimization. This work aims at analyzing the *gradient-enhanced* ℓ_1 minimization and provide precise conditions under which the new approach can indeed improve the sparse recovery. To this end, we begin with some preliminaries on CS.

1.1 Compressed sensing

The aim of compressed sensing is to find a sparse solution to linear equations

$$\mathbf{\Phi c} = \mathbf{f},$$

where $\mathbf{f}=(f(\mathbf{z}_1),\cdots,f(\mathbf{z}_N))^T\in\mathbb{C}^N$ is usually the function evaluations, $\mathbf{c}\in\mathbb{C}^M$ is the unknown vector and $\mathbf{\Phi}\in\mathbb{R}^{N\times M}$ is the measurement matrix (Typically, the components here are often evaluations of certain bases on given samples). In CS, one usually uses the following programming to obtain the sparse solution

$$\operatorname{argmin}\|\mathbf{c}\|_0, \quad \text{subject to} \quad \mathbf{\Phi c}=\mathbf{f}. \tag{1.1}$$

The convex relaxation of (1.1) is

$$\operatorname{argmin}\|\mathbf{c}\|_1, \quad \text{subject to} \quad \mathbf{\Phi c}=\mathbf{f}. \tag{1.2}$$

Note that the constraints $\mathbf{\Phi c}=\mathbf{f}$ can be relaxed to $\|\mathbf{\Phi c}-\mathbf{f}\|\le\epsilon$ for some tolerance value ϵ, yielding a regression type *denoising* approach.

Several types of sufficient conditions (for $\mathbf{\Phi}$) have been presented in CS, such as mutual incoherence property (MIP) and restricted isometry property (RIP), under which the solution to (1.2) is the same with the one of (1.1). In the following, we shall give a brief review of these conditions. We first introduce the results about MIP. The mutual incoherence constant (MIC) of $\mathbf{\Phi}$ is defined as

$$\mu:=\mu(\mathbf{\Phi}):=\max_{k\neq j}\frac{|\langle\mathbf{\Phi}_k,\mathbf{\Phi}_j\rangle|}{\|\mathbf{\Phi}_k\|_2\cdot\|\mathbf{\Phi}_j\|_2}, \tag{1.3}$$

where $\{\mathbf{\Phi}_j\}_{j=1}^M$ are the associated column vectors. Assume that $\mathbf{c}_0$ is a s-sparse vector in $\mathbb{C}^M$, i.e., $\|\mathbf{c}_0\|_0\le s$, and if

$$\mu<\frac{1}{2s-1}, \tag{1.4}$$

then the solution to the ℓ_1 minimization (1.2) with $\mathbf{f}=\mathbf{\Phi c}_0$ is exactly $\mathbf{c}_0$, i.e.,

$$\mathbf{c}_0=\operatorname*{argmin}_{\mathbf{c}\in\mathbb{C}^M}\{\|\mathbf{c}\|_1 \text{ subject to } \mathbf{\Phi c}=\mathbf{\Phi c}_0\}.$$

This result was first presented in [16] for the case with $\mathbf{\Phi}$ being the union of two orthogonal matrices, and was later extended to general matrices by Fuchs [20] and Gribonval & Nielsen [22]. In [7], it is also shown that $\mu<\frac{1}{2s-1}$ is sufficient for stably approximating $\mathbf{c}$ in the noisy case.

We next turn to RIP. We say that the matrix $\mathbf{\Phi}$ satisfies s-order RIP with restricted isometry constant (RIC) $\delta_s\in[0,1)$ if there holds

$$(1-\delta_s)\|\mathbf{y}\|_2^2\le\|\mathbf{\Phi y}\|\le(1+\delta_s)\|\mathbf{y}\|_2^2 \tag{1.5}$$

for all $\mathbf{y}\in\mathbb{C}^M$ with $\|\mathbf{y}\|_0\le s$. In fact, δ_s yields a uniform bound on the spectral radius of the sub-matrices of $\mathbf{\Phi}$ formed by selecting any s columns of $\mathbf{\Phi}$. It was shown that one can recover a s-sparse vector by solving the ℓ_1 minimization if $\mathbf{\Phi}$ satisfies RIP with $\delta_{3s}+3\delta_{4s}<2$ [9]. The RIP condition is recently improved to $\delta_s<\frac{1}{3}$ in [8]. More precisely, we have

Theorem 1.1 (Sparse recovery for RIP-matrices [8,9]). *Let $\mathbf{\Phi}\in\mathbb{R}^{N\times M}$ be a matrix satisfying s-order RIP with $\delta_s<\frac{1}{3}$. For any given $\mathbf{c}_0\in\mathbb{R}^M$, let $\mathbf{c}^\#$ be the solution of the ℓ_1-minimization*

$$\operatorname{argmin}\|\mathbf{c}\|_1 \quad s.t. \quad \mathbf{\Phi c}=\mathbf{\Phi c}_0. \tag{1.6}$$

Then we have

$$\|\mathbf{c}^\#-\mathbf{c}_0\|_2\le C\frac{\sigma_{s,1}(\mathbf{c}_0)}{\sqrt{s}} \quad with \quad \sigma_{s,1}(\mathbf{c}_0)=\inf_{\|\mathbf{y}\|_0\le s}\|\mathbf{y}-\mathbf{c}_0\|_1, \tag{1.7}$$

for some constant $C>0$ that depends only on δ_s. In particular, if $\mathbf{c}_0$ is s-sparse then reconstruction is exact, i.e., $\mathbf{c}^\#=\mathbf{c}_0$.

1.2 The recovery of sparse trigonometric polynomials

We denote by $\Pi(\Gamma)$ the space of all trigonometric polynomials which are of the following form

$$f(\mathbf{x})=\sum_{\mathbf{k}\in\Gamma}c_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \mathbf{x}=(x_1,\cdots,x_d)^\top\in[-\pi,\pi)^d, \tag{1.8}$$

where $c_{\mathbf{k}}\in\mathbb{C}$ and $\Gamma\subset\mathbb{Z}^d$ is a finite index set. The dimension of $\Pi(\Gamma)$ (or the cardinality of the index set Γ) is denoted by $M=\#\Gamma$. To state conveniently, we also impose an order on these bases and re-write the expansion into the following single index version

$$f(\mathbf{x})=\sum_{\mathbf{k}\in\Gamma}c_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}=\sum_{j=1}^{M}c_j\phi_j(\mathbf{x}), \tag{1.9}$$

where we denoted by $\{\phi_j\}_{j=1}^M$ the re-ordered trigonometric bases. Suppose that $\Xi=\{\mathbf{z}_j\}_{j=1}^N\subset[-\pi,\pi)^d$. Throughout this paper, we define the *interpolation matrix* corresponding to $\{\phi_t\}_{t=1}^M$ and $\{\mathbf{z}_j\}_{j=1}^N$ as

$$\mathbf{\Phi}=\left[\phi_t(\mathbf{z}_j)\right]_{j=1,\cdots,N,t=1,\cdots,M}\in\mathbb{C}^{N\times M}.$$

We denote the support of the coefficient vector $\{c_{\mathbf{k}}\}_{\mathbf{k}\in\Gamma}$ by T, i.e.,

$$T:=\{\mathbf{k}:c_{\mathbf{k}}\neq 0\}.$$

In this work, we are concerned with the recovery of sparse trigonometric polynomials, namely, we assume that $s=\#T$ is much smaller than the dimension of $M=\#\Gamma$, i.e., $s\ll M$. We shall also set

$$\Pi_s(\Gamma):=\bigcup_{\substack{T\subset\Gamma\\ \#T\le s}}\Pi(T),$$

where $\Pi(T)$ denotes the space of all trigonometric polynomials whose coefficients are supported on T. Note that the set $\Pi_s(\Gamma)$ is the union of linear spaces and consists of all trigonometric polynomials in $\Pi(\Gamma)$ with the number of nonzero Fourier coefficients less than or equal to s.

An important topic in CS is to design a sampling set $\Xi=\{\mathbf{z}_j\}_{j=1}^N\subset[-\pi,\pi)^d$ to recover $f\in\Pi_s(\Gamma)$, and this is equivalent to find a s-sparse solution to

$$\mathbf{\Phi}\mathbf{c}=\mathbf{\Phi}\mathbf{c}_0,$$

where $\mathbf{\Phi}:=(\phi_t(\mathbf{z}_j))_{j=1,\cdots,N,t=1,\cdots,M}\in\mathbb{C}^{N\times M}$ and $\mathbf{c}_0\in\mathbb{C}^M$ satisfying $\|\mathbf{c}_0\|_0\le s$. Hence, it is possible to use the ℓ_1 minimization to recover $f\in\Pi_s(\Gamma)$ from $f(\mathbf{z}_j)$, $j=1,\cdots,N$. According to Theorem 1.1, to guarantee

$$\operatorname{argmin}\{\|\mathbf{c}\|_1:\mathbf{\Phi}\mathbf{c}=\mathbf{\Phi}\mathbf{c}_0\}=\mathbf{c}_0, \tag{1.10}$$

it is enough to require that $\mathbf{\Phi}$ satisfies the s-order RIP with $\delta_s<1/3$ (see also [50]). Variants choices of sampling strategies have been proposed recently to guarantee the RIP condition. In [11], Candès and Tao consider the case where $\mathbf{z}_1,\cdots,\mathbf{z}_N$ have the uniform distribution on the grid $\{0,\frac{2\pi}{N},\cdots,\frac{2\pi(N-1)}{N}\}^d$ and the result is improved by Rudelson and Vershynin [43]. In [37,39], Rauhut consider the case where $\mathbf{z}_1,\cdots,\mathbf{z}_N$ have uniform distribution on $[-\pi,\pi)^d$ and show that $\mathbf{\Phi}$ satisfy RIP $\delta_s\le\delta$ with probability at least $1-\epsilon$ provided

$$N/\log(N)\ge\frac{C}{\delta^2}s\log^2(s)\log(M)\log(\epsilon^{-1}).$$

Beyond RIP, one can also use MIP to study the performance of the ℓ_1 minimization for the recovery of $f\in\Pi_s(\Gamma)$. As said before, $\mu<\frac{1}{2s-1}$ is a sufficient condition for (1.10) holding. In [38], Kunis and Rauhut proved that $\mu<\frac{1}{2s-1}$ holds with probability at least $1-\epsilon$ provided $N\ge C(2s-1)^2\log(4M^2/\epsilon)$ and $\mathbf{z}_1,\cdots,\mathbf{z}_N$ are uniformly distributed on $[-\pi,\pi)^d$. In [51], for the case where $\Gamma=[-q,q]^d$ and $d\ge2$, Xu introduce the deterministic sampling points

$$\mathbf{z}_j=2\pi(j,j^2,\cdots,j^d)/N,\qquad j=1,\cdots,N, \tag{1.11}$$

and show that $\mu(\mathbf{\Phi})<\frac{1}{2s-1}$ provided $N>\max\{(2s-1)^2(d-1)^2,2q+1\}$ is a prime number. We would like to remark that for the case $d=1$, a deterministic sampling strategy has been introduced in [49].

1.3 Our contribution

In this work, we shall analyze a gradient-enhanced ℓ_1 approach for the recovery of sparse trigonometric polynomials. More precisely, we denote the sampling set by

$$\Xi=\{\mathbf{z}_j\}_{j=1}^N\subset[-\pi,\pi)^d\quad\text{ with }\quad\#\Xi=N.$$

The aim is to reconstruct $f\in\Pi_s(\Gamma)$ from its sample evaluations and/or the gradient evaluations. We shall consider the following two problems:

- **Problem** 1: Recover $f\in\Pi_s(\Gamma)$ from

$$f(\mathbf{z}_j)=f_j, \qquad \mathbf{z}_j\in\Xi, \tag{1.12}$$

$$D_{\mathbf{v}_t}f(\mathbf{z}_j)=f'_{j,t}, \qquad t=1,\cdots,k,\ \mathbf{z}_j\in\Xi, \tag{1.13}$$

  where $D_{\mathbf{v}_t}f(\mathbf{z}_j):=\langle\nabla f(\mathbf{x}),\mathbf{v}_t\rangle|_{\mathbf{x}=\mathbf{z}_j}$ and $\mathbf{v}_t\in\mathbb{R}^d,t=1,\cdots,k$ ($k\le d$). Namely, we assume that both function values and the directional derivative information at the sampling points are known.

- **Problem** 2: Recover $f\in\Pi_s(\Gamma)$ from

$$D^{\tau_j}_{\mathbf{v}_j}f(\mathbf{z}_j)=y_j, \qquad \mathbf{z}_j\in\Xi, \tag{1.14}$$

  where $\mathbf{v}_j\in\mathbb{R}^d,\tau_j\in\mathbb{Z}_{\ge0},j=1,\cdots,N$. Here, it is supposed that one knows either the τ_j-order directional derivative of f at $\mathbf{z}_j$ or the function value $f(\mathbf{z}_j)$. If $\tau_j=0$, then (1.14) means that we know only the function value of f at $\mathbf{z}_j$, i.e., $y_j=f(\mathbf{z}_j)$. We consider this problem as an purely approximation problem, as its connection to UQ is up to now unclear.

We shall adopt the MIP framework and the RIP framework to analyze the above two problems respectively. We first consider Problem 1. We denote by $\mathbf{\Phi}$ the interpolation matrix corresponding to $\{\phi_j\}_{j=1}^M$ and Ξ. Similarly, for $t=1,\cdots,k$, we denote by $\mathbf{\Phi}_t$ the interpolation matrix corresponding to $\{D_{\mathbf{v}_t}\phi_j\}_{j=1}^M$ and Ξ. Then, the gradient-enhanced ℓ_1 minimization yields:

$$\operatorname{argmin}\|\mathbf{c}\|_1 \quad \text{subject to} \quad \tilde{\mathbf{\Phi}}\mathbf{c}=\tilde{\mathbf{f}}, \tag{1.15}$$

where

$$\tilde{\mathbf{f}}=\begin{pmatrix}\mathbf{f}\\ \mathbf{f}_1\\ \vdots\\ \mathbf{f}_k\end{pmatrix}, \quad \tilde{\mathbf{\Phi}}=\begin{pmatrix}\mathbf{\Phi}\\ \mathbf{\Phi}_1\\ \vdots\\ \mathbf{\Phi}_k\end{pmatrix}.$$

Notice that by combining gradient information, we have increased the row size of the design matrix, i.e., $(k+1)N$ instead of N, while the length of the unknown vector remains the same. Naturally, one is interested in whether $\mathbf{\Phi}_t$ is helpful in improving the MIP condition of $\tilde{\mathbf{\Phi}}$. One of our main purpose is to show $\mu(\tilde{\mathbf{\Phi}})\le\lambda\mu(\mathbf{\Phi})$ where $\lambda<1$ is a positive constant depending on Γ and $\mathbf{v}_1,\cdots,\mathbf{v}_k$ which implies that the gradient information is indeed helpful for improving the recovery guarantee of ℓ_1 minimization.

For Problem 2, we consider the following ℓ_1 minimization

$$\operatorname{argmin}\|\tilde{\mathbf{c}}\|_1 \quad \text{subject to} \quad \mathbf{\Psi}\tilde{\mathbf{W}}\tilde{\mathbf{c}}=\mathbf{f}, \tag{1.16}$$

where $\mathbf{\Psi} := \left[D_{\mathbf{v}_j}^{\tau_j}\phi_t(\mathbf{z}_j)\right]_{t=1,\cdots,M,j=1,\cdots,N}$, $\mathbf{f} := (y_1,\cdots,y_N)^T$ and $\tilde{\mathbf{W}}$ is a diagonal normalizing matrix so that the column norm of $\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\tilde{\mathbf{W}}$ is normalized. If the solution to (1.16) is $\tilde{\mathbf{c}}^{\#}$ then the coefficient vector of $f \in \Pi_s(\Gamma)$ is $\tilde{\mathbf{W}}\tilde{\mathbf{c}}^{\#}$. We shall study the RIP of $\tilde{\mathbf{\Psi}}$ under the setting of $\mathbf{z}_1,\cdots,\mathbf{z}_N$ being uniformly distributed samples on $[-\pi,\pi)^d$. In particular, we show that $\tilde{\mathbf{\Psi}}$ satisfies RIP with RIC $\delta_s \le \delta$ with probability at least $1-\epsilon$ provided

$$N \ge 2\frac{(C_0R_0)^2}{\delta^2} s(\ln 100s)^2 \ln(4M)\ln(10N)\ln\frac{\beta}{\epsilon} + \kappa,$$

where κ and R_0 are defined in Section 3. This indicates that the gradient information plays a similar role as that of the function values.

1.4 Related work

We are not only one investigating the performance of the gradient-enhance ℓ_1 approach. We would like to mention that many numerical experiments were reported in [4, 25, 28, 35]. In [28], the new bases (Sobleve orthogonal bases) are constructed which are suitable for the gradient-enhanced approach for the least-squares regression. A special case of Problem 1 is studied in [35] under the setting of the bases being Hermite polynomial and the sampling points being random. The results in [35] show that $\mu(\tilde{\mathbf{\Phi}}) \le \mu(\mathbf{\Phi})$ and this inequality is *almost-surely* strict. Our analysis is a little bit more general and it is shown that $\mu(\tilde{\mathbf{\Phi}}) \le \lambda\mu(\mathbf{\Phi})$ where $\lambda < 1$ is a positive constant and our result does not suffer from probabilistic qualifiers (e.g., "almost surely"). We also present an estimation for the constant λ (see Theorem 2.1). However, we would like to remark that in UQ applications, Hermite polynomials are more important than that of the trigonometric polynomials.

1.5 Organization

The rest of the paper is organized as follows. In Section 2, we focus on Problem 1 and we shall show that $\mu(\tilde{\mathbf{\Phi}}) \le \lambda\mu(\mathbf{\Phi})$ with $\lambda < 1$ being a constant. We also shown that multiplying a diagonal matrix $\mathbf{W}$ to the interpolation matrix is helpful to decrease the coherence of $\tilde{\mathbf{\Phi}}$. We turn to Problem 2 in Section 3. Particularly, we present the RIP of $\tilde{\mathbf{\Psi}}$ under the setting of $\mathbf{z}_1,\cdots,\mathbf{z}_N$ being uniformly distributed samples on $[-\pi,\pi)^d$. Numerical results are presented in Section 4 to support the theoretical finding, and we finally give some concluding remarks in Section 5.

2 Coherence analysis for Problem 1

This section is devoted to the coherence analysis of Problem 1.

2.1 Preliminary results

Recall that $f\in\Pi_s(\Gamma)$, i.e.,

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\Gamma} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}},$$

where $\|\mathbf{c}\|_0\le s$ with $\mathbf{c}=(c_{\mathbf{k}})_{\mathbf{k}\in\Gamma}$. We set $\mathbf{V}:=(\mathbf{v}_1,\cdots,\mathbf{v}_k)\in\mathbb{R}^{d\times k}$ where $\mathbf{v}_1,\cdots,\mathbf{v}_k\in\mathbb{R}^d$ are k directions. We first introduce the following definition.

Definition 2.1. For any $\mathbf{k},\mathbf{k}'\in\Gamma$, we define the distance between $\mathbf{k}$ and $\mathbf{k}'$ as follows:

$$\|\mathbf{k}-\mathbf{k}'\|_{\mathbf{V}} := \|\mathbf{V}^T\mathbf{k}-\mathbf{V}^T\mathbf{k}'\|_\infty.$$

The directions $\mathbf{v}_1,\cdots,\mathbf{v}_k$ are said to be *admissible* with respect to Γ if $\|\mathbf{k}-\mathbf{k}'\|_{\mathbf{V}}\neq 0$ for any $\mathbf{k},\mathbf{k}'\in\Gamma$ with $\mathbf{k}\neq\mathbf{k}'$.

Note that $\|\mathbf{k}-\mathbf{k}'\|_{\mathbf{V}}=0$ if and only if $\mathbf{V}^T\mathbf{k}=\mathbf{V}^T\mathbf{k}'$. Hence, if $\mathrm{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_k\}=\mathbb{R}^d$, then $\mathbf{V}=(\mathbf{v}_1,\cdots,\mathbf{v}_k)$ is admissible with respect to Γ. We next present an example to show that there exists $\mathbf{v}_1,\cdots,\mathbf{v}_k\in\mathbb{R}^d$ which are admissible with respect to Γ but $\mathrm{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_k\}\neq\mathbb{R}^d$. Suppose that $d=2$ and $\Gamma=[-q,q]^2\cap\mathbb{Z}^2$. Take $\mathbf{v}_1:=[1,\sqrt{2}]^T\in\mathbb{R}^2$. Notice that

$$\{\eta\in\mathbb{R}^2:\langle\eta,\mathbf{v}_1\rangle=0\} = \{\alpha\cdot(-\sqrt{2},1)^T\in\mathbb{R}^2:\alpha\in\mathbb{R}\}.$$

Hence, we know that the set

$$\{\eta\in\mathbb{R}^2:\langle\eta,\mathbf{v}_1\rangle=0\}\cap\mathbb{Z}^2$$

contains only zero point, and this implies that $\mathbf{v}_1$ is admissible with respect to Γ but $\mathrm{span}\{\mathbf{v}_1\}\neq\mathbb{R}^2$.

Recall that in Problem 1, we would like to recover $f\in\Pi_s(\Gamma)$ from

$$f(\mathbf{z}_j)=f_j, \quad D_{\mathbf{v}_t}f(\mathbf{z}_j)=f_{j,t}, \quad t=1,\cdots,k, \quad \mathbf{z}_j\in\Xi, \tag{2.1}$$

where $D_{\mathbf{v}_t}f(\mathbf{z}_j):=\langle\nabla f(\mathbf{x}),\mathbf{v}_k\rangle|_{\mathbf{x}=\mathbf{z}_j}$ and $\mathbf{v}_t\in\mathbb{R}^d$. We also recall the definition of the interpolation matrix corresponding to $\{e^{i\mathbf{k}\cdot\mathbf{x}}\}_{\mathbf{k}\in\Gamma}$ and Ξ:

$$\mathbf{\Phi}=\left[e^{i\mathbf{k}\cdot\mathbf{z}_j}\right]_{\mathbf{z}_j\in\Xi,\mathbf{k}\in\Gamma}, \quad \mathbf{\Phi}_t=\left[D_{\mathbf{v}_t}e^{i\mathbf{k}\cdot\mathbf{z}_j}\right]_{\mathbf{z}_j\in\Xi,\mathbf{k}\in\Gamma}, \quad 1\le t\le k,$$

and

$$\tilde{\mathbf{\Phi}}:=\begin{pmatrix}\mathbf{\Phi}\\ \mathbf{\Phi}_1\\ \vdots\\ \mathbf{\Phi}_k\end{pmatrix}. \tag{2.2}$$

We next study the coherence of $\tilde{\mathbf{\Phi}}$, i.e., $\mu(\tilde{\mathbf{\Phi}})$. To this end, we first introduce a refinement of Cauchy-Schwarz inequality.

Lemma 2.1. [34, P.289] *Suppose that* $\mathbf{x}\in\mathbb{R}^d$ *and* $\mathbf{y}\in\mathbb{R}^d$ *are not proportional. Suppose that* $\mathbf{u}\in\mathbb{R}^d$ *satisfy* $\langle\mathbf{u},\mathbf{x}\rangle=0$, $\langle\mathbf{u},\mathbf{y}\rangle=1$. *Then*

$$\frac{\langle\mathbf{x},\mathbf{y}\rangle^2}{\|\mathbf{x}\|^2\cdot\|\mathbf{y}\|^2}\leq 1-\frac{1}{\|\mathbf{y}\|^2\cdot\|\mathbf{u}\|^2}.$$

2.2 Coherence analysis

The main result of this section is summarized as the following theorem:

Theorem 2.1. *Suppose that* $\mathbf{V}=(\mathbf{v}_1,\cdots,\mathbf{v}_k)\in\mathbb{R}^{d\times k}$ *is admissible with respect to the index set* Γ. *Then we have*

$$\mu(\tilde{\mathbf{\Phi}})\leq\lambda\cdot\mu(\mathbf{\Phi}),$$

where

$$\lambda\leq\left(1-\frac{\Gamma_{\min}}{\max_{\mathbf{k}\in\Gamma}(1+\|\mathbf{V}^T\mathbf{k}\|_\infty^2)\max_{\mathbf{k}\in\Gamma}(1+\|\mathbf{V}^T\mathbf{k}\|_2^2)}\right)^{1/2},$$

and

$$\Gamma_{\min}:=\min_{\mathbf{k}\neq\mathbf{k}',\mathbf{k},\mathbf{k}'\in\Gamma}\|\mathbf{k}-\mathbf{k}'\|_{\mathbf{V}}^2.$$

Proof. To state conveniently, we denote by $\mathbf{a}_{\mathbf{k}}$ the column vector of $\mathbf{\Phi}$, i.e.,

$$\mathbf{a}_{\mathbf{k}}:=[e^{i\mathbf{k}\cdot\mathbf{z}_j}:\mathbf{z}_j\in\Xi].$$

A simple observation is that

$$\mu(\mathbf{\Phi})=\frac{1}{N}\max_{\mathbf{k}\neq\mathbf{k}',\mathbf{k},\mathbf{k}'\in\Gamma}|\langle\mathbf{a}_{\mathbf{k}},\mathbf{a}_{\mathbf{k}'}\rangle|.$$

Notice that we have

$$D_{\mathbf{v}_t}e^{i\mathbf{k}\cdot\mathbf{z}_j}=\langle\mathbf{v}_t,\mathbf{k}\rangle e^{i\mathbf{k}\cdot\mathbf{z}_j}.$$

Hence, the column vectors of $\mathbf{\Phi}_t$ are $\{\langle\mathbf{v}_t,\mathbf{k}\rangle\mathbf{a}_{\mathbf{k}}:\mathbf{k}\in\Gamma\}$. Then a simple observation yields

$$\mu(\tilde{\mathbf{\Phi}})=\max_{\mathbf{k}\neq\mathbf{k}',\mathbf{k},\mathbf{k}'\in\Gamma}|\langle\mathbf{a}_{\mathbf{k}},\mathbf{a}_{\mathbf{k}'}\rangle|\frac{|1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}\rangle\langle\mathbf{v}_t,\mathbf{k}'\rangle|}{\sqrt{1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}\rangle^2}\sqrt{1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}'\rangle^2}},$$

which implies

$$\mu(\tilde{\mathbf{\Phi}})\leq\lambda\cdot\mu(\mathbf{\Phi})$$

with

$$\lambda=\max_{\mathbf{k}\neq\mathbf{k}',\mathbf{k},\mathbf{k}'\in\Gamma}\frac{|1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}\rangle\langle\mathbf{v}_t,\mathbf{k}'\rangle|}{\sqrt{1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}\rangle^2}\sqrt{1+\sum_{t=1}^k\langle\mathbf{v}_t,\mathbf{k}'\rangle^2}}.\tag{2.3}$$

For fixed $\mathbf{k},\mathbf{k}'\in\Gamma$ with $\mathbf{k}\neq\mathbf{k}'$, we next study the quantity

$$\frac{|1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}\rangle\langle\mathbf{v}_t,\mathbf{k}'\rangle|}{\sqrt{1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}\rangle^2}\cdot\sqrt{1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}'\rangle^2}}.$$

We let $\mathbf{x}:=(1,\langle\mathbf{v}_1,\mathbf{k}\rangle,\cdots,\langle\mathbf{v}_k,\mathbf{k}\rangle)$, $\mathbf{y}:=(1,\langle\mathbf{v}_1,\mathbf{k}'\rangle,\cdots,\langle\mathbf{v}_k,\mathbf{k}'\rangle)$ and

$$t_0:=\mathrm{argmax}_{1\le t\le k}|\langle\mathbf{v}_t,\mathbf{k}-\mathbf{k}'\rangle|.$$

Suppose that $\mathbf{u}=(u_1,\cdots,u_d)\in\mathbb{R}^d$ where $u_1=\frac{-\langle\mathbf{v}_{t_0},\mathbf{k}\rangle}{\langle\mathbf{v}_{t_0},\mathbf{k}'-\mathbf{k}\rangle}$, $u_{t_0}=\frac{1}{\langle\mathbf{v}_{t_0},\mathbf{k}'-\mathbf{k}\rangle}$ and other entries are 0. It can be shown that $\mathbf{u}$ satisfies $\langle\mathbf{x},\mathbf{u}\rangle=0$ and $\langle\mathbf{y},\mathbf{u}\rangle=1$. Then, Lemma 2.1 implies that

$$\frac{|1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}\rangle\langle\mathbf{v}_t,\mathbf{k}'\rangle|}{\sqrt{1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}\rangle^2}\cdot\sqrt{1+\sum_{t=1}^{k}\langle\mathbf{v}_t,\mathbf{k}'\rangle^2}}\le\sqrt{1-\frac{1}{\|\mathbf{y}\|^2\cdot\|\mathbf{u}\|^2}}.$$

Notice that

$$\|\mathbf{u}\|^2\le\frac{1+\|\mathbf{V}^T\mathbf{k}\|_\infty^2}{\|\mathbf{k}'-\mathbf{k}\|_{\mathbf{V}}^2}\le\frac{1+\|\mathbf{V}^T\mathbf{k}\|_\infty^2}{\Gamma_{\min}},\quad \|\mathbf{y}\|^2=1+\|\mathbf{V}^T\mathbf{k}\|_2^2.$$

Then, we obtain

$$\lambda\le\left(1-\frac{\Gamma_{\min}}{\max_{\mathbf{k}\in\Gamma}(1+\|\mathbf{V}^T\mathbf{k}\|_\infty^2)\max_{\mathbf{k}\in\Gamma}(1+\|\mathbf{V}^T\mathbf{k}\|_2^2)}\right)^{1/2}. \tag{2.4}$$

This completes the proof. □

For arbitrary sampling points, the above Theorem shows that the inclusion of gradient information can indeed decrease the MIP of the interpolation matrix, under the condition that the $(\mathbf{v}_1,\cdots,\mathbf{v}_k)$ is admissible with respect to the index set Γ. For the special case where $\Gamma=[-q,q]^d\cap\mathbb{Z}^d$ and $\mathbf{v}_j=\mathbf{e}_j, j=1,\cdots,d$, we have the following corollary:

Corollary 2.1. *Suppose that* $\Gamma=[-q,q]^d\cap\mathbb{Z}^d, \mathbf{V}=(\mathbf{v}_1,\cdots,\mathbf{v}_d)\in\mathbb{R}^{d\times d}$ *and* $\mathbf{v}_j=\mathbf{e}_j$ *where* $\mathbf{e}_j$ *denotes the vector with jth being* 1 *and all other entries being* 0*. We have*

$$\mu(\tilde{\mathbf{\Phi}})\le\left(1-\frac{1}{(1+q^2)(1+dq^2)}\right)^{1/2}\mu(\mathbf{\Phi}).$$

Proof. For $\mathbf{V}=(\mathbf{e}_1,\cdots,\mathbf{e}_d)$ we can derive that

$$\Gamma_{\min}=\min_{\mathbf{k}\neq\mathbf{k}',\mathbf{k},\mathbf{k}'\in\Gamma}\|\mathbf{k}-\mathbf{k}'\|_{\mathbf{V}}^2=1.$$

Similarly, we have

$$\max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|_\infty^2=q^2,\qquad \max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|_2^2=dq^2.$$

Then the desired results follows by using Theorem 2.1. □

Next, we shall present an example with showing that, without the admissible condition, it is possible $\mu(\Phi)=\mu(\tilde{\Phi})$. Consider the two dimensional case with $d=2$ and $\Gamma=[-q,q]^2\cap\mathbb{Z}^d$. Suppose that $p>2q+1$ is a prime number and let us consider the following samples

$$\Xi=\{2\pi(j,j^2)/p : j=0,\cdots,p-1\}.$$

We denote the interpolation matrix corresponding to Ξ as Ψ, i.e.,

$$\Psi := \Big[\exp(2i\pi(k_1 j+k_2 j^2)/p)\Big]_{j=0,\cdots,p-1,(k_1,k_2)\in[-q,q]^2} \in \mathbb{C}^{p\times(2q+1)^2}. \tag{2.5}$$

We also set the matrix with gradient information as

$$\Psi_1 := \Big[D_{\mathbf{e}_1}\exp(2i\pi(k_1 j+k_2 j^2))\Big]_{j=0,\cdots,p-1,(k_1,k_2)\in[-q,q]^2} \in \mathbb{C}^{p\times(2q+1)^2} \tag{2.6}$$

Proposition 2.1. Let $\widetilde{\Psi}=\begin{pmatrix}\Psi\\ \Psi_1\end{pmatrix}$ where Ψ and Ψ_1 are defined in (2.5) and (2.6), respectively. Then, we have $\mu(\tilde{\Psi})=\mu(\Psi)$.

Proof. Notice that the column vector of Ψ are

$$\mathbf{a}_{k_1,k_2}=\Big[\exp(2i\pi(k_1 j+k_2 j^2)/p)\Big]_{j=0,\cdots,p-1}.$$

The Gauss sum formula shows that $\mu(\Psi)=1/\sqrt{p}$. Particularly, we have

$$\frac{|\langle \mathbf{a}_{k_1,k_2},\mathbf{a}_{k_1',k_2'}\rangle|}{\|\mathbf{a}_{k_1,k_2}\|\cdot\|\mathbf{a}_{k_1',k_2'}\|}=\frac{1}{\sqrt{p}},$$

provided that $k_1=k_1'$ but $k_2\neq k_2'$. Notice that the column vectors of $\tilde{\Psi}$ are

$$\tilde{\mathbf{a}}_{k_1,k_2} := \begin{pmatrix}\mathbf{a}_{k_1,k_2}\\ k_1\mathbf{a}_{k_1,k_2}\end{pmatrix}. \tag{2.7}$$

Then for (k_1,k_2) and (k_1',k_2') with $k_1=k_1', k_2\neq k_2'$, we have

$$\mu(\tilde{\Psi})=\frac{|\langle\tilde{\mathbf{a}}_{k_1,k_2},\tilde{\mathbf{a}}_{k_1',k_2'}\rangle|}{\|\tilde{\mathbf{a}}_{k_1,k_2}\|\cdot\|\tilde{\mathbf{a}}_{k_1',k_2'}\|}=\frac{|\langle\mathbf{a}_{k_1,k_2},\mathbf{a}_{k_1',k_2'}\rangle+k_1^2\langle\mathbf{a}_{k_1,k_2},\mathbf{a}_{k_1',k_2'}\rangle|}{(1+k_1^2)\cdot\|\mathbf{a}_{k_1,k_2}\|\cdot\|\mathbf{a}_{k_1',k_2'}\|}=\frac{1}{\sqrt{p}}.$$

The proof is completed. $\square$

2.3 An improved estimation

In this section, we provide an improved estimation by *weighting* the gradient information. Notice that for any constant $\alpha \in \mathbb{R}$, we have

$$\alpha \cdot D_{\mathbf{v}_t} f(\mathbf{z}_j) = \alpha \cdot f_{j,t}.$$

After multiplying a constant α, is it possible to decrease the coherence of the interpolation matrix $\tilde{\mathbf{\Phi}}$? We answer this question in the following theorem.

Theorem 2.2. *Suppose that* $\mathbf{W} = \mathrm{diag}(W_1, \cdots, W_{kN}) \in \mathbb{R}^{kN \times kN}$ *is a diagonal matrix with*

$$W_j = \begin{cases} 1, & 1 \le j \le N, \\ \alpha, & else. \end{cases} \tag{2.8}$$

Then, under conditions in Theorem 2.1*, we have*

$$\mu(\mathbf{W}\tilde{\mathbf{\Phi}}) \le \lambda \mu(\mathbf{\Phi}),$$

where

$$\lambda \le \left(1 - \frac{\alpha^2 \Gamma_{\min}}{\max_{\mathbf{k}\in\Gamma}(1+\alpha^2\|\mathbf{V}^T\mathbf{k}\|_\infty^2) \cdot \max_{\mathbf{k}\in\Gamma}(1+\alpha^2\|\mathbf{V}^T\mathbf{k}\|_2^2)}\right)^{1/2}. \tag{2.9}$$

Proof. The proof is similar with the one of Theorem 2.1. We just replace the definition of $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x} := (1, \alpha\langle \mathbf{v}_1, \mathbf{k}\rangle, \cdots, \alpha\langle \mathbf{v}_k, \mathbf{k}\rangle), \qquad \mathbf{y} := (1, \alpha\langle \mathbf{v}_1, \mathbf{k}'\rangle, \cdots, \alpha\langle \mathbf{v}_k, \mathbf{k}'\rangle).$$

And then the desired follows by taking $\mathbf{u} = (u_1, \cdots, u_d) \in \mathbb{R}^d$ where $u_1 = \frac{-\langle \mathbf{v}_{t_0}, \mathbf{k}\rangle}{\langle \mathbf{v}_{t_0}, \mathbf{k}'-\mathbf{k}\rangle}, u_{t_0} = \frac{1}{\alpha\langle \mathbf{v}_{t_0}, \mathbf{k}'-\mathbf{k}\rangle}$ and other entries are 0. □

Notice that the above upper bound in (2.9)

$$\left(1 - \frac{\alpha^2 \Gamma_{\min}}{\max_{\mathbf{k}\in\Gamma}(1+\alpha^2\|\mathbf{V}^T\mathbf{k}\|_\infty^2) \cdot \max_{\mathbf{k}\in\Gamma}(1+\alpha^2\|\mathbf{V}^T\mathbf{k}\|_2^2)}\right)^{1/2}$$

reach its minimum when

$$\alpha = \frac{1}{\sqrt{\max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|_\infty \cdot \max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|}}.$$

Particular, using similar arguments the as in Corollary 2.1, we have

Corollary 2.2. *Suppose that* $\Gamma=[-q,q]^d$, $\mathbf{V}=(\mathbf{v}_1,\cdots,\mathbf{v}_d)\in\mathbb{R}^{d\times d}$ *and* $\mathbf{v}_j=\mathbf{e}_j$ *where* $\mathbf{e}_j$ *denotes the vector with jth being* 1 *and all other entries being* 0*. If we set*

$$\alpha=\frac{1}{\sqrt{\max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|_\infty\cdot\max_{\mathbf{k}\in\Gamma}\|\mathbf{V}^T\mathbf{k}\|}}=\frac{1}{\sqrt{dq}},$$

then, it holds

$$\mu(\mathbf{W}\tilde{\mathbf{\Phi}})\leq\left(1-\frac{1}{(1+\sqrt{d})^2q}\right)^{1/2}\mu(\tilde{\mathbf{\Phi}}),$$

where $\mathbf{W}$ *is defined in Theorem* 2.2*.*

Remark 2.1. Notice that

$$\left(1-\frac{1}{(1+\sqrt{d})^2q}\right)^{1/2}\leq\left(1-\frac{1}{(1+q^2)(1+dq^2)}\right)^{1/2}.$$

Hence, by multiplying the matrix $\mathbf{W}$, one can obtain a better estimation over the result in Corollary 2.1.

3 RIP analysis for Problem 2

This section is devoted to the RIP analysis of Problem 2. The aim of Problem 2 is to recover $f\in\Pi_s(\Gamma)$ with using

$$y_j=D_{\mathbf{v}_j}^{\tau_j}f(\mathbf{z}_j),\qquad j=1,\cdots,N, \tag{3.1}$$

where $\mathbf{v}_j\in\mathbb{R}^d,\tau_j\in\mathbb{Z}_{\geq0}$. When $\tau_j=0$, the above condition reduces to the case where we know only the function values at $\mathbf{z}_j$, i.e., $y_j=f(\mathbf{z}_j)$. We use the following ℓ_1 minimization to recover f:

$$\operatorname{argmin}\|\tilde{\mathbf{c}}\|_1\quad\text{subject to}\quad\tilde{\mathbf{\Psi}}\tilde{\mathbf{c}}=\mathbf{f}, \tag{3.2}$$

where $\tilde{\mathbf{\Psi}}=\mathbf{\Psi}\tilde{\mathbf{W}}$ with $\mathbf{\Psi}:=\left[D_{\mathbf{v}_j}^{\tau_j}\phi_t(\mathbf{z}_j)\right]_{t=1,\cdots,M,j=1,\cdots,N}\in\mathbb{C}^{N\times M}$, and

$$\tilde{\mathbf{W}}:=\operatorname{diag}\left(\frac{1}{\sqrt{\sum_{j=1}^N|\langle\mathbf{v}_j,\mathbf{k}\rangle|^{2\tau_j}}}:\mathbf{k}\in\Gamma\right)\in\mathbb{C}^{M\times M}.$$

Notice that the above normalizing matrix $\tilde{\mathbf{W}}$ is chosen such that the column norm of $\tilde{\mathbf{\Psi}}$ is normalized. After solving problem (3.2), we should re-normalize the solution, i.e., the desired solution is $\mathbf{c}^\#=\tilde{\mathbf{W}}\tilde{\mathbf{c}}^\#$ where $\tilde{\mathbf{c}}^\#$ denotes the solution to (3.2).

We set

$$Z_{\mathbf{k}}:=\{j:\langle\mathbf{v}_j,\mathbf{k}\rangle^{\tau_j}=0,1\leq j\leq N\},\quad\kappa:=\max_{\mathbf{k}\in\Gamma}\#Z_{\mathbf{k}}.$$

Here, we set $\langle \mathbf{v}_j,\mathbf{k}\rangle^{\tau_j}=1$ if $\tau_j=0$. We also set

$$R_0:=\frac{\max_{j\in Z^c_{\mathbf{k}}}|\langle \mathbf{v}_j,\mathbf{k}\rangle|^{\tau_j}}{\min_{j\in Z^c_{\mathbf{k}}}|\langle \mathbf{v}_j,\mathbf{k}\rangle|^{\tau_j}},$$

where $Z^c_{\mathbf{k}}:=\{1,\cdots,N\}\backslash Z_{\mathbf{k}}$. We are now ready to give the main result of this section which presents the RIP of $\tilde{\mathbf{\Psi}}$. Particularly, motivated by the results in [43] and [37], we have

Theorem 3.1. *Suppose that the random sampling points $\mathbf{z}_1,\cdots,\mathbf{z}_N$ are uniformly distributed on $[-\pi,\pi)^d$. Assume that $\mathbf{v}_j\in\mathbb{R}^d,\tau_j\in\mathbb{Z}_{\geq 0},j=1,\cdots,N$ satisfy $\sum_{j=1}^N\langle \mathbf{v}_j,\mathbf{k}\rangle^{2\tau_j}\neq 0$ for any $\mathbf{k}\in\Gamma$. For $0<\delta\leq 1/2$ and $0<\epsilon<1$, if*

$$N\geq 2\frac{(C_0R_0)^2}{\delta^2}s(\ln 100s)^2\ln(4M)\ln(10N)\ln\frac{\beta}{\epsilon}+\kappa, \tag{3.3}$$

then we have

$$\mathbb{P}(\delta_s\leq\delta)\geq 1-\epsilon,$$

where δ_s is the s-order RIP constant of $\tilde{\mathbf{\Psi}}$, $M:=\#\Gamma$ and C_0,β are universal constants.

We remark that improved RIP bounds are now available, see e.g. [2,14]. According to Theorem 1.1, the above theorem shows that the ℓ_1 minimization can produce a solution to Problem 2 with probability at least $1-\epsilon$ provided the number of samplings N satisfies (3.3) with $\delta=1/3$. Inspired by Theorem 3.1, we conclude that the gradient information can play a similar role as that of the function evaluations, up to a constant R_0 (that might be large), when the samples are chosen uniformly according to the uniform measure. The proof of Theorem 3.1 follows the ideas in [43] (see also [37]), and is postponed to the end of this section. To prove Theorem 3.1, we first introduce some useful lemmas.

Lemma 3.1. [27] *Assume that ¸ $=(\xi_j)_{j=1}^N$ is a sequence of independent random vectors in $\mathbb{C}^N$ equipped with a (semi-)norm $\|\cdot\|$ having expectations $\mathbf{x}_j=\mathbb{E}\xi_j$. Then for $1\leq p<\infty$*

$$\left(\mathbb{E}\|\sum_{j=1}^N(\xi_j-\mathbf{x}_j)\|^p\right)^{1/p}\leq 2\left(\mathbb{E}\|\sum_{j=1}^N\epsilon_j\xi_j\|^p\right)^{1/p},$$

where **ffl** *$=(\epsilon)_{j=1}^N\in\{-1,1\}^N$ is a Rademacher sequence independent of ¸.*

For $B\in\mathbb{C}^{M\times M}$, the semi-norm $|||\cdot|||_s$ is defined as following

$$|||B|||_s:=\max_{\mathbf{x}\in\Sigma_s}|\langle B\mathbf{x},\mathbf{x}\rangle|,\quad \Sigma_s:=\{\mathbf{x}\in\mathbb{C}^M:\|\mathbf{x}\|_2\leq 1,\|\mathbf{x}\|_0\leq s\}.$$

Then we have

Lemma 3.2. [37,43] *Assume that* $\mathbf{x}_1,\cdots,\mathbf{x}_N\in\mathbb{C}^M$ *satisfy* $\max_j\|\mathbf{x}_j\|_\infty\le K$. *Assume that* $s\le N$. *Then*

$$\left(\mathbb{E}\left|\left|\left|\sum_{j=1}^N\epsilon_j\mathbf{x}_j^*\mathbf{x}_j\right|\right|\right|_s^p\right)^{1/p}\le D_{p,s,N,M}\sqrt{\left|\left|\left|\sum_{j=1}^N\mathbf{x}_j^*\mathbf{x}_j\right|\right|\right|_s},$$

where $D_{p,s,N,M}=C_pK\sqrt{s}\ln(100s)\sqrt{\ln(4M)\cdot\ln(10N)}$,

$$C_p=\begin{cases}C_1', & p=1,\\ C_2'\beta^{1/p}\sqrt{p}, & p\ge 2,\end{cases}$$

and $C_1'=94.81$, $C_2'\approx 82.56$, $\beta=6.028$.

Lemma 3.3. [45] *Suppose that* Z *is a random variable satisfying*

$$(\mathbb{E}|Z|^p)^{1/p}\le\alpha\beta^{1/p}p^{1/\gamma}\quad\textit{for all}\quad p\ge p_0$$

for some constants $\alpha,\beta,\gamma,p_0>0$. *Then for all* $u\ge p_0^{1/\gamma}$ *we have*

$$\mathbb{P}(|Z|\ge e^{1/\gamma}\alpha u)\le\beta e^{-u^\gamma/\gamma}.$$

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. We denote by $\{\tilde{\mathbf{X}}_j,j=1,\cdots,N\}$ the column vectors of $\tilde{\mathbf{\Psi}}$. Note that $\tilde{\mathbf{x}}_j$ only depends on $\mathbf{z}_j$ and hence they are independent vectors. Then a simple observation is that

$$\tilde{\mathbf{\Psi}}^*\tilde{\mathbf{\Psi}}=\sum_{j=1}^N\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j,\qquad \mathbb{E}\tilde{\mathbf{\Psi}}^*\tilde{\mathbf{\Psi}}=\mathbb{E}\sum_{j=1}^N\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j=\mathrm{Id}.$$

We now consider

$$\delta_s=|||\tilde{\mathbf{\Psi}}-\mathrm{Id}|||_s=\left|\left|\left|\sum_{j=1}^N\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j-\mathrm{Id}\right|\right|\right|_s=\left|\left|\left|\sum_{j=1}^N(\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j-\mathbb{E}\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j)\right|\right|\right|_s.$$

By Lemma 3.1, we have

$$E_p:=\left(\mathbb{E}\left|\left|\left|\sum_{j=1}^N\left(\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j-\mathbb{E}\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j\right)\right|\right|\right|_s^p\right)^{1/p}\le 2\left(\mathbb{E}\left|\left|\left|\sum_{j=1}^N\epsilon_j\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j\right|\right|\right|_s^p\right)^{1/p},\tag{3.4}$$

where $(\epsilon_1,\cdots,\epsilon_N)\in\{-1,1\}^N$ is a Rademacher sequence. By combining (3.4) and Lemma 3.2, we obtain

$$\begin{aligned}E_p^p&\le(2D_{p,s,N,M})^p\cdot\left|\left|\left|\sum_{j=1}^N\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j\right|\right|\right|_s^{p/2}\\&\le(2D_{p,s,N,M})^p\cdot\left(\left|\left|\left|\sum_{j=1}^N\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j-\mathrm{Id}\right|\right|\right|_s+1\right)^{p/2},\end{aligned}\tag{3.5}$$

where $D_{p,s,N,M}$ is defined in Lemma 3.2. The above equation (3.5) implies that

$$E_p \leq 2D_{p,s,N,M}\sqrt{E_p+1}.$$

Hence, we have

$$E_p \leq 2D_{p,s,N,M}^2+2D_{p,s,N,M}\sqrt{D_{p,s,N,M}^2+1}.$$

If $D_{p,s,N,M}\leq 1/4$ then

$$E_p \leq \frac{1+\sqrt{17}}{2}\cdot D_{p,s,N,M}. \tag{3.6}$$

Now, we set

$$\theta_s := \min\{\frac{1}{2},\delta_s\} = \min\left\{\frac{1}{2},\left\|\left|\sum_{j=1}^{N}\left(\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j-\mathbb{E}\tilde{\mathbf{X}}_j^*\tilde{\mathbf{X}}_j\right)\right\|\right|_s\right\}.$$

We claim that

$$(\mathbb{E}\theta_s^p)^{1/p} \leq \min\left\{\frac{1}{2},E_p\right\} \leq \frac{1+\sqrt{17}}{2}\cdot D_{p,s,N,M}. \tag{3.7}$$

In fact, when $D_{p,s,N,M}\leq \frac{1}{4}$, (3.6) implies that $E_p\leq \frac{1+\sqrt{17}}{2}\cdot D_{p,s,N,M}$. For the case where $D_{p,s,N,M}\geq \frac{1}{4}$, we have

$$\frac{1}{2} \leq \frac{1+\sqrt{17}}{2}\cdot D_{p,s,N,M}.$$

Hence, (3.7) follows.

Then by Lemma 3.3, we obtain that for all $u\geq\sqrt{2}$ we have

$$\mathbb{P}\left(|\theta_s| \geq C_0K\cdot\sqrt{s}\cdot\ln(100s)\cdot\sqrt{\ln(4M)\ln(10N)}\cdot u\right) \leq \beta\exp(-u^2/2), \tag{3.8}$$

where

$$C_0 = \frac{1+\sqrt{17}}{2}\cdot\sqrt{e}\cdot C_2',$$

and C_2' is defined in Lemma 3.2. Notice that

$$K = \|\tilde{\mathbf{X}}_j\|_\infty \leq \frac{R_0}{\sqrt{N-\kappa}}$$

and $0<\delta\leq 1/2$. By taking $u=\sqrt{2}\sqrt{\ln\frac{\beta}{\epsilon}}$ in (3.8), we obtain that

$$\mathbb{P}(\delta_s\leq\delta) = \mathbb{P}(\theta_s\leq\delta) \geq 1-\epsilon,$$

provided that

$$N \geq 2\frac{(C_0R_0)^2}{\delta^2}s(\ln 100s)^2\ln(4M)\ln(10N)\ln\frac{\beta}{\epsilon}+\kappa.$$

This completes the proof. □

Note that the lower bound in (3.3) depends on the κ and R_0. To obtain a good bound, we should require the index set Γ and the directions $\mathbf{V}$ have some special structure so that κ and R_0 are small. Particularly, we have the following corollary:

Corollary 3.1. *Suppose that $\Gamma=[1,q]^d\cap\mathbb{Z}^d$ with $q\geq 2$, $\tau_j\in\{0,1\}$ and $\mathbf{v}_j\in\{\mathbf{e}_1,\cdots,\mathbf{e}_d\}$, $j=1,\cdots,N$. Assume that the random sampling points $\mathbf{z}_1,\cdots,\mathbf{z}_N$ are uniform distribution on $[-\pi,\pi)^d$. For $0<\delta\leq 1/2$ and $0<\epsilon<1$, if*

$$N\geq 2\frac{(C_0q)^2}{\delta^2}s(\ln 100s)^2\ln(4M)\ln(10N)\ln\frac{\beta}{\epsilon}$$

then we have

$$\mathbb{P}(\delta_s\leq\delta)\geq 1-\epsilon,$$

where δ_s is the s-order RIP constant of $\tilde{\mathbf{\Psi}}$, $M=(q-1)^d$ and C_0,β are universal constants.

Proof. Since $\Gamma=[1,q]^d$ and $\mathbf{v}_j\in\{\mathbf{e}_1,\cdots,\mathbf{e}_d\}$, we have $\kappa=0$ and $R_0=q$. The desired result follows by using Theorem 3.1. $\square$

4 Numerical examples

In this section we provide with numerical experiments to show the performance of the gradient-enhanced ℓ_1 recovery of sparse Fourier expansions $f\in\Pi_s(\Gamma)$. The main purpose is to show that the inclusion of gradient information can indeed enhance the recovery ability of the ℓ_1 minimization. To this end, we shall compare the gradient-enhanced ℓ_1 approach with the standard ℓ_1 recovery. Throughout this section, we take $\Gamma=[-q,q]^d\cap\mathbb{Z}^d$. In all our figures, we shall denote by "*gradient-enhanced*" the numerical results obtained by using gradient-enhanced ℓ_1 approach, while we shall use "*standard*" to denote the numerical results by using the standard ℓ_1 approach. To solve the ℓ_1 minimization, we shall employ the available tools SPGL1 from [47] that was implemented in MATLAB.

Example 4.1. We first present a numerical example under the setting of Problem 1. We take $d=2$, $q=10$ and $\mathbf{v}_1=\mathbf{e}_1$, $\mathbf{v}_2=\mathbf{e}_2$. The N sampling points $\{\mathbf{z}_j\}_{j=1}^N$ are produced according to the uniform distribution on $[-\pi,\pi)^d$. Given the sparsity s, the support set of $f\in\Pi_s(\Gamma)$ is drawn from the uniform distribution over the set of all subsets of $[-q,q]^d$ with the size s and the nonzero coefficients of f have the Gaussian distribution with mean zero and standard deviation one. Then we have N function values $f(\mathbf{z}_j)$, $j=1,\cdots,N$ and we also have $2N$ derivative values $D_{\mathbf{v}_1}f(\mathbf{z}_j),D_{\mathbf{v}_2}f(\mathbf{z}_j)$, $j=1,\cdots,N$. The full gradient-enhanced approach uses the $3N$ values (100% information, i.e., the N function values and the $2N$ derivative values). A 50% gradient-enhanced approach uses N function values and N derivative values (one randomly chosen derivative from $D_{\mathbf{v}_1}f(\mathbf{z}_j),D_{\mathbf{v}_2}f(\mathbf{z}_j)$ for each $j\in[1,N]$). In Fig. 1 (Left), we show the recovery success rate against the number of sampling points N with a fixed sparsity $s=5$. For each number N, we repeat the experiment 500 times and calculate the success rate. The results show that the use of gradient information

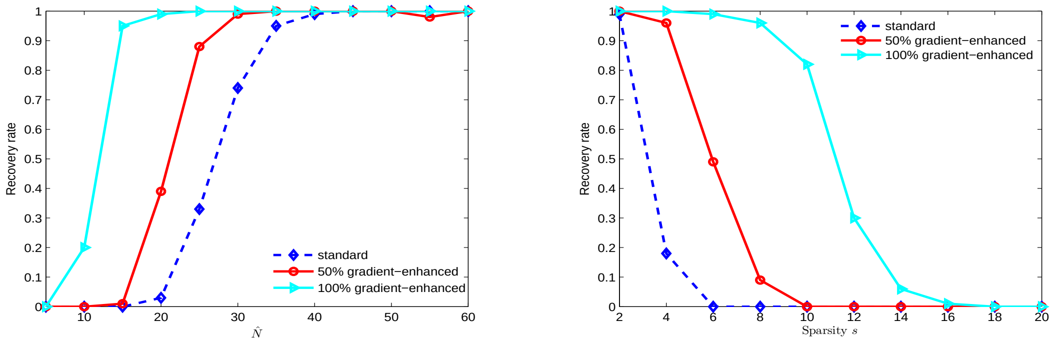


Figure 1: Left: Recovery rate against number of samples, $s=5$; Right: Recovery rate against sparsity, $m=20$. Two dimensional tests with $d=2$, $q=10$.

can indeed improve the recovery rate, and furthermore, the more gradient information is included, the better recovery rate is obtained. Fig. 1 (Right) depicts the recovery rate against the sparsity s with a fixed number of random sampling points $N=20$. Again, we repeat the experiment 500 times for each fixed sparsity s and calculate the recovery rate. Similar conclusion can be made as the one obtained by the left plot.

Example 4.2. The setting of this numerical experiment is similar with the one in Example 4.1 with $d=5$ and $q=2$. In Fig. 2 (Left), we show the recovery rate against the number of sampling points N with a fixed sparsity $s=6$. In this example, we test the 20% and 40% gradient-enhanced approach which means that partial derivatives with respect to one and two variables are involved in the ℓ_1 approach, respectively. Again, the better performance can be observed when the more gradient information is included. Similarly, Fig. 2 (right) depicts the recovery rate against the sparsity with a fixed number of random samples $N=30$.

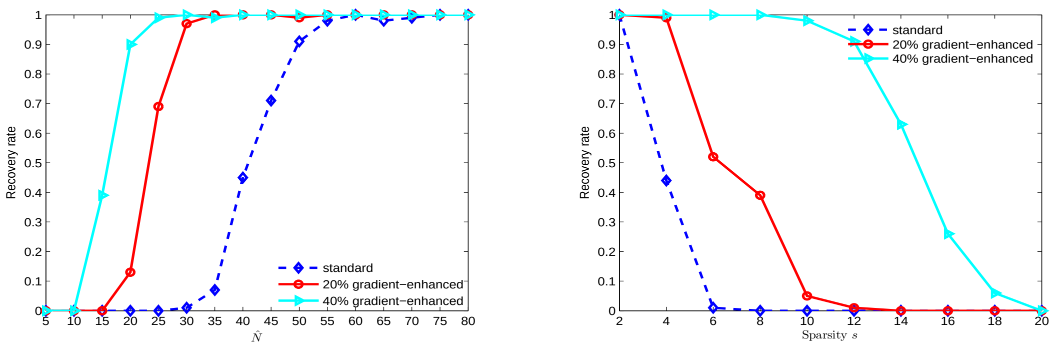


Figure 2: Left: Recovery rate against number of samples, $s=6$; Right: Recovery rate against sparsity, $m=30$. Five dimensional tests with $d=5$, $q=2$.

Example 4.3. The numerical experiment in this example is made under the setting of Problem 2. We take $d=2$, $q=5$ and $s=8$. The N sampling points are produced according to the unform distribution on $[-\pi,\pi)^d$. We use $N/4$ function values $f(\mathbf{z}_j)$, $j=1,\cdots,N/4$ and $3N/4$ derivative values $D_{\mathbf{e}_1}f(\mathbf{z}_j)$, $j=N/4+1,\cdots,N$ to recover $f\in\Pi_s(\Gamma)$. Fig. 3 compares the recovery rate of the gradient-enhance ℓ_1-approach with the one of the standard ℓ_1 approach which uses N function values. The numerical results show the derivative values play a similar role with the function values.

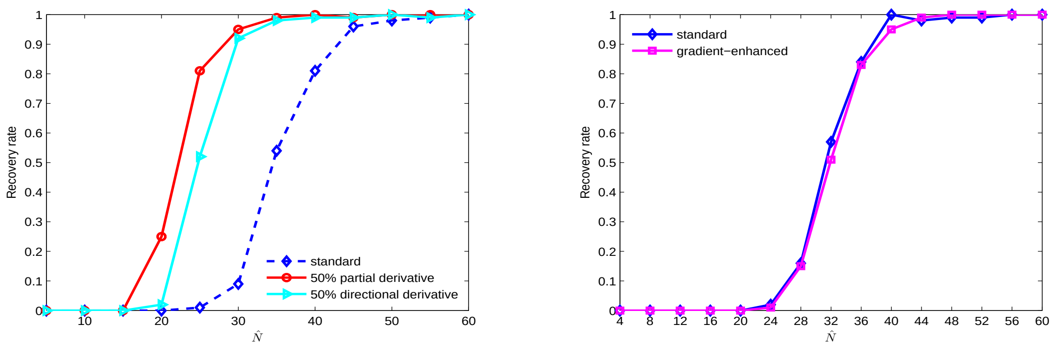
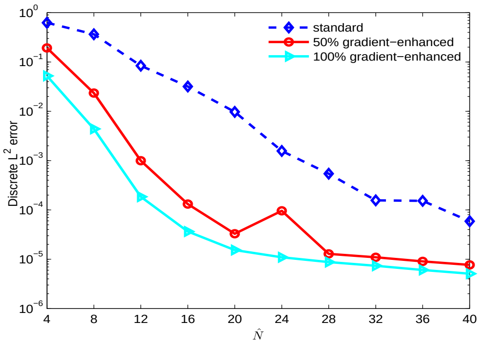


Figure 3: Left: Recovery rate against number of samples, $d=5$, $q=2$, $s=5$; Right: Recovery with 1/4 function values and 3/4 partial derivative information for $s=8$, $d=2$, $q=5$, $M=40$. Comparisons with standard ℓ^1 approach (with only function values, the blue line) are also provided.

Example 4.4. In this numerical experiment, we test the gradient-enhanced ℓ_1 minimization for function approximations. In this example, we take $d=2$. The aim is to use ℓ_1 minimization to find a sparse Fourier approximation of a given function. Notice that in this case the target is a function but not exact Fourier expansions. In Fig. 4 (Left), we consider to approximate the function $g(\mathbf{x})=\sin\left(\cos\left(\sum_{j=1}^d x_j\right)\right)$ by sparse Fourier expansions $f\in\Pi(\Gamma)$ where $\Gamma=[-q,q]^2$ and $q=10$ with random evaluations using the ℓ_1 approach. The right plot shows the approximation results for $g(\mathbf{x})=\cos\left(\sum_{j=1}^d x_j\right)\exp\left(\sin(\sum_{j=1}^d x_j)\right)$. In fact, we use the following programming to find the sparse Fourier approximation of g:

$$\min\|\mathbf{c}\|_1 \quad \text{s.t.} \quad f(\mathbf{z}_j)=g(\mathbf{z}_j), \quad D_{\mathbf{e}_t}f(\mathbf{z}_j)=D_{\mathbf{e}_t}g(\mathbf{z}_j), \quad t=1,\cdots,k, \ j=1,\cdots,N,$$

where $\mathbf{c}$ is the coefficient vectors of f. In Fig. 4, we use "50% gradient-enhance " to denote the case with $k=1$ meaning that we just use the partial derivative along the direction $\mathbf{e}_1$. Similarly, "100% gradient-enhance " denotes the case with $k=2$. In both cases, it is clear that the use of gradient information can dramatically enhance the approximation accuracy.



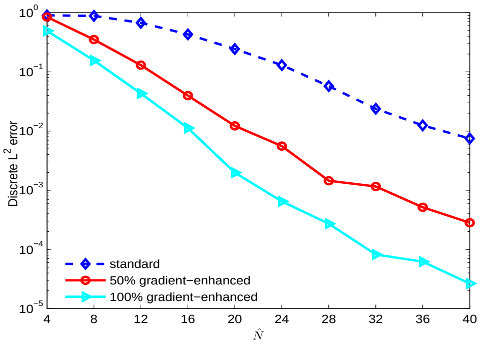


Figure 4: Discrete L^2 error against number of samples with random points, $d=2$, $q=10$.

5 Conclusion

In this work, we analyze a gradient-enhanced ℓ_1 recovery approach for the recover of sparse Fourier expansions. We have provided with conditions under which the inclusion of gradient information can indeed improve the recovery performance of the ℓ_1 minimization. Our analysis also shows that in some cases the gradient information can play similar roles as that of the function values. Several numerical examples are presented to support the theoretical statements. Potential applications to function interpolations are also discussed. There are, however, several issues to be addressed:

- According to the result for Problem 1, i.e. Theorem 2.1, we know the derivative information is helpful to decrease the MIC. We are interested in the following question: is the derivative information helpful to decrease the RIP constant? Motivated by the numerical results in Section 4, we conjecture the answer for the question is positive provided the sampling points $\mathbf{z}_1,\cdots,\mathbf{z}_N$ have the uniform distribution on $[-\pi,\pi)^d$.

- How to find the best way to include the derivative information? In view of Theorem 2.1, the choice of the directions $\{\mathbf{v}_\mathbf{j}\}$ and the constant α in equation (2.8) can affect the performance of the ℓ_1 minimization. To choose the optimal directions and the α is the subject of the future work. Furthermore, it is also interesting to extend the results in this paper to the case where the basis are orthogonal polynomials instead of Fourier basis.

Acknowledgments

Zhiqiang Xu was supported by NSFC grant (91630203, 11422113, 11331012, 11688101) and by National Basic Research Program of China (973 Program 2015CB856000). Tao Zhou was supported by the NSF of China (under grant numbers 11688101, 91630312, 91630203, 11571351, and 11731006), the science challenge project (No. TZ2018001), NCMIS, and the youth innovation promotion association (CAS).

References

[1] B. Adcock. Infinite-dimensional ℓ_1 minimization and function approximation from pointwise data. *Found Comput Math*, DOI: 10.1007/s10208-017-9350-3, 2017.

[2] B. Adcock, S. Brugiapaglia and C. Webster, Polynomial approximation of high-dimensional functions via compressed sensing, *arXiv:1703.06987*, 2017

[3] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: a new theory for compressed sensing. *Forum Math., Sigma*, 5. doi: 10.1017/fms.2016.32.

[4] A. K. Alekseev, I. M. Navon, and M. E. Zelentsov. The estimation of functional uncertainty using polynomial chaos and adjoint equations. *Int. J. Numer. Methods Fluids*, 67(3):328–341, 2011.

[5] Richard L. Burden and J. Douglas Faires. Numerical analysis. *Brooks Cole; 8 edition*, 7, 2004.

[6] R. Caflisch, S. Osher, H. Schaeffer, and G. Tran. Pdes with compressed solutions. *Commun. Math. Sci.*, 13(8):2155–2176, 2015.

[7] T. Cai, L. Wang, and G. Xu. Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inf. Theory*, 56:3516–3522, 2010.

[8] T. Cai and A. Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Appl. Comput. Harmon. Anal.*, 35:74–93, 2013.

[9] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 8(59):1207–1223, 2006.

[10] E. Candès and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *IEEE Trans. Inform. Theory*, 12(51):4203–4215, 2005.

[11] E. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 12(52):5406–5425, 2006.

[12] E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.

[13] Y. Cao, S. Li, L. Petzold, and R. Serban. Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. *SIAM J. Sci. Comput.*, 24(3):1076–1089, 2003.

[14] A. Chkifa, N. Dexter, H. Tran, and C. G. Webster. Polynomi al approximation via compressed sensing of high-dimensional functions on lower sets, Math. Comput., 2016. To appear (arXiv:1602.05823).

[15] I. Y. Chun, B. Adcock, and T. Talavage. Efficient compressed sensing sense pmri reconstruction with joint sparsity promotion. *IEEE Trans. Med. Imag.*, 31(1):354–368, 2015.

[16] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory*, 47:2845–2862, 2001.

[17] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc.Natl Acad. Sci. USA*, 102(27):9452–9457, 2005.

[18] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, 2009.

[19] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. *Birkhauser*, 2013.

[20] J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory*, 50:1341–1344, 2004.

[21] M. Geles and N. Pierce. An introduction to the adjoint approach to design. *Flow, turbulence and combustion*, 65(3-4):393–415, 2000.

[22] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inf. Theory*, 49:3320–3325, 2003.

[23] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies *J. Comput. Phys.*, 280(1) pp. 363–386, 2015.

[24] M. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process*, 57:2275–2284, 2009.

[25] J. D. Jakeman, M. S. Eldred, and K. Sargsyan. Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection. *J. Comput. Phys.*, 289:18–34, 2015.

[26] J. D. Jakeman, A. Narayan, and T. Zhou. A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions. *SIAM J. Sci. Comput.*, 39(3):A1114-A1144, 2017.

[27] M. Ledoux and M. Talagrand, Probability in Banach Spaces., Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.

[28] Y. Li, M. Anitescu, O. Roderick, and F. Hickernell. Orthogonal bases for polynomial regression with derivative information in uncertainty quantification. *Int. J. Uncertainty Quantification*, 4(1), 2011.

[29] W. Liu, D. Gong, and Z. Xu. One-bit compressed sensing by greedy algorithms, *Numer. Math. Theor. Meth. Appl.,* Vol. 9, No. 2, pp. 169-184, 2016.

[30] A. Mackey, H. Schaeffer, and S. Osher. On the compressive spectral method. *Multiscale Model. Simul.*, 12(4):1800–1827, 2014.

[31] S. Mukherjee and D. X. Zhou. Learning coordinate covariances via gradients. *J. Mach. Learn. Res.*, 7:519–549, 2006.

[32] A. Narayan and T. Zhou. Stochastic collocation methods on unstructured meshes, *Commun. Comput. Phys.*, 18(2015), pp. 1–36.

[33] V. Ozolins, R. Lai, R. Caflisch, and S. Osher. Compressed plane waves yield a compactly supported multiresolution basis for the laplace operator. *Proc. Natl. Acad. Sci.*, 111(5):1691–1696, 2014.

[34] A. Ostrowski, Vorlesungen über Differential-und Integralrechnung, Vol. 2, Birkhäuser, Basel. 1951.

[35] J. Peng, J. Hampton, and A. Doostan. On polynomial chaos expansion via gradient-enhanced ℓ_1-minimization. *J. Comput. Phys.*, 310(1):440–458, 2016.

[36] G. Pfander and H. Rauhut. Sparsity in time-frequency representations. *J. Fourier Anal.Appl.*, 16(1):233–260, 2010.

[37] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical Foundations and Numerical Methods for Sparse Recovery, Fornasier, M. (Ed.) Berlin,New York (DE GRUYTER)*, pages 1–92, 2010.

[38] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials II- Orthogonal matching pursuit versus basis pursuit, *Found. Comput. Math.*, 8(6) (2008)1615-3375.

[39] H. Rauhut. Random sampling of sparse trigonometric polynomials, *Appl. Comput. Harmon. Anal.*, 22(2007) 16-42.

[40] H. Rauhut and R. Ward. Interpolation via weighted ℓ_1 minimization. *arXiv:1308.0759*, 2013.

[41] O. Roderick, M. Anitescu, and P. Fischer. Polynomial regression approaches using derivative information for uncertainty quantification. *Nucl. Sci. Eng.*, 162(2):122–139, 2010.

[42] J. Romberg. Imaging via compressive sampling. *IEEE Signal Process. Magazine*, 25:14–20, 2008.

[43] M. Rudelson, R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.*, 61 (2008), 1025-1045.

[44] A. Spitzbart. A generalization of hermite's interpolation formula. *American Mathematical Monthly*, 67(1):42–46, 1960.

[45] J.A. Tropp, On the conditioning of random subdictionaries, *Appl. Comput. Harmon. Anal.*, 25 (2008), 1–24.

[46] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Trans. Inform. Theory*, 56:520–544, 2010.

[47] E. van den Berg and M. Friedlander. Spgl1: A solver for large-scale sparse reconstruction. *http://www.cs.ubc.ca/∼mpf/spgl1/*, 2007.

[48] Z. Wu. Hermite-birkhoff interpolation of scattered data by radial basis functions. *Approx. Theory Appl.*, 8:1–10, 1992.

[49] G. Xu and Z. Xu. Compressed Sensing Matrices from Fourier Matrices, *IEEE Trans. Inf. Theory*, 61(2015), 469-478.

[50] G. Xu and Z. Xu. On the ℓ_1-Norm Invariant Convex k-Sparse Decomposition of Signals, *Journal of the Operations Research Society of China,* 1 (2013), 537-541.

[51] Z. Xu. Deterministic sampling of sparse trigonometric polynomials. *Journal of Complexity*, 27(2):133–140, April 2011.

[52] Z. Xu and T. Zhou. On sparse interpolation and the design of deterministic interpolation points. *SIAM J. Sci. Comput.*, 36:A1752–A1769, 2014.