ARTICLE IN PRESS

Journal of Computational and Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

The polynomial Fourier transform with minimum mean square error for noisy data

Naoki Saito [a], Yi Wang [b,*]

[a] Department of Mathematics, University of California, Davis, CA 95616, USA

[b] Department of Mathematics, Auburn University Montgomery, P.O. Box 244023, Montgomery, AL 36124, USA

ARTICLE INFO

Article history:
Received 3 July 2009

Keywords:
Algebraic polynomials
Trigonometric polynomials
Least squares approximation
Mean square error
Noisy data

ABSTRACT

In 2006, Naoki Saito proposed a Polyharmonic Local Fourier Transform (PHLFT) to decompose a signal $f \in L^2(\Omega)$ into the sum of a *polyharmonic component* u and a *residual* v, where Ω is a bounded and open domain in $\mathbb{R}^d$. The solution presented in PHLFT in general does not have an error with minimal energy. In resolving this issue, we propose the least squares approximant to a given signal in $L^2([-1, 1])$ using the combination of a set of algebraic polynomials and a set of trigonometric polynomials. The maximum degree of the algebraic polynomials is chosen to be small and fixed. We show in this paper that the least squares approximant converges uniformly for a Hölder continuous function. Therefore Gibbs phenomenon will not occur around the boundary for such a function. We also show that the PHLFT converges uniformly and is a near least squares approximation in the sense that it is arbitrarily close to the least squares approximant in L^2 norm as the dimension of the approximation space increases. Our experiments show that the proposed method is robust in approximating a highly oscillating signal. Even when the signal is corrupted by noise, the method is still robust. The experiments also reveal that an optimum degree of the trigonometric polynomial is needed in order to attain the minimal l^2 error of the approximation when there is noise present in the data set. This optimum degree is shown to be determined by the intrinsic frequency of the signal. We also discuss the energy compaction of the solution vector and give an explanation to it.

1. Introduction

In 2006, Naoki Saito proposed a Polyharmonic Local Sine Transform (PHLST) [1] in an attempt to develop a local Fourier analysis and synthesis method without encountering the infamous Gibbs phenomenon. PHLST is also used to resolve several problems occurring in the Local Trigonometric Transforms (LTTs) of Coifman and Meyer [2] and Malvar [3,4], such as the overlapping windows and the slopes of the bell functions (see [1] for the details on how PHLST resolves these problems).

Let us pause for a moment to define some notations. To index infinitely countable sets, we adopt the following standard conventions: let $\mathbb{N}$ be the set of natural numbers, and the set $\mathbb{Z}_+ := \{0\} \cup \mathbb{N}$. To enumerate finite sets, we define $\mathbb{Z}_k := \{0, 1, \ldots, k-1\}$, and $\mathbb{N}_k := \{1, 2, \ldots, k\}$.

PHLST first segments a given function (or input data) $f(\boldsymbol{x})$, $\boldsymbol{x} \in \Omega \subset \mathbb{R}^d$ supported on an open and bounded domain Ω into a set of disjoint blocks $\{\Omega_j : j \in \mathbb{Z}_M\}$ for a positive integer M such that $\overline{\Omega} = \bigcup_{j\in\mathbb{Z}_M} \overline{\Omega}_j$. Denote by f_j the restriction of the function f to $\overline{\Omega}_j$, i.e., $f_j = \chi_{\overline{\Omega}_j} f$, where $\chi_{\overline{\Omega}_j}$ is the characteristic function on the set $\overline{\Omega}_j$, $j \in \mathbb{Z}_M$. Then PHLST decomposes

* Corresponding author.
E-mail address: ywang2@aum.edu (Y. Wang).

0377-0427/$ – see front matter
doi:10.1016/j.cam.2010.02.039

each f_j into two components as $f_j = u_j + v_j$. The components u_j and v_j are referred to as the *polyharmonic component* and the *residual*, respectively. The polyharmonic component is obtained by solving the following *polyharmonic equation*:

$$\Delta^m u_j = 0 \quad \text{in } \Omega_j,\ m \in \mathbb{N} \tag{1.1}$$

with a set of given boundary values and normal derivatives

$$\frac{\partial^{q_\ell} u_j}{\partial \nu^{q_\ell}} = \frac{\partial^{q_\ell} f}{\partial \nu^{q_\ell}} \quad \text{on } \partial\Omega_j,\ \ell \in \mathbb{Z}_m, \tag{1.2}$$

where $\Delta = \sum_{i=1}^{d} \partial^2/\partial x_i^2$ is the Laplace operator in $\mathbb{R}^d$. The natural number m is called the degree of polyharmonicity, and q_ℓ is the order of the normal derivative. These boundary conditions (1.2) enforce that the solution u_j interpolates the function values and the normal derivatives of orders $q_1, \ldots, q_{m-1}$ of the original signal f along the boundary $\partial\Omega_j$. The parameter q_0 is normally set to 0, which means that $u_j = f$ on the boundary $\partial\Omega_j$, i.e., the Dirichlet boundary condition. If the blocks $\Omega_j, j \in \mathbb{Z}_M$, are all rectangles (of possibly different sizes), PHLST sets $q_\ell = 2\ell$, i.e., only normal derivatives of *even orders* are interpolated. It is not necessary to match normal derivatives of odd orders when the blocks Ω_j's are rectangular domains. This is because the Fourier sine series of the residual v_j is equivalent to the complex Fourier series of the periodized v_j after odd reflection with respect to the boundary $\partial\Omega_j$, hence the continuity of the normal derivatives of odd orders (up to order $2m-1$) is automatically guaranteed. Thanks to these boundary conditions, the residual component can be expanded into a Fourier sine series without facing the Gibbs phenomenon, and the Fourier sine expansion coefficients of the residual v_j decay rapidly, i.e., in the order $O(\|\boldsymbol{k}\|^{-2m-1})$, provided that there is no other intrinsic singularity in the domain Ω_j, where $\boldsymbol{k}$ is the frequency index vector.

In our joint work [5], we implemented PHLST up to polyharmonicity of degree 5. The corresponding algorithm is called PHLST5. In that work, we derived a fast algorithm to compute a 5th degree polyharmonic function that satisfies certain boundary conditions. Although the Fourier sine coefficients of the residual of PHLST5 possess the same decaying rate as in LLST (Laplace Local Sine Transform, the simplest version of PHLST with polyharmonicity of degree 1), by using additional information of first order normal derivative from the boundary, the blocking artifacts are largely suppressed in PHLST5 and the residual component becomes much smaller than that of LLST. Therefore PHLST5 provides a better approximation result. Due to the difficulty of estimating higher order derivatives, we consider PHLST5 as the practical limitation of implementing PHLST with higher degree polyharmonicity.

Soon after developing PHLST, N. Saito and K. Yamatani extended it to the *Polyharmonic Local Cosine Transform* (PHLCT) [6]. The PHLCT allows the Fourier cosine coefficients of the residual decay in the order $O\left(\|\boldsymbol{k}\|^{-2m-2}\right)$ by setting $q_\ell = 2\ell + 1,\ \ell \in \mathbb{Z}_m$ in the boundary conditions (1.2) and by introducing an appropriate source term on the right-handed side of the polyharmonic equation (1.1). In that work, an efficient algorithm was developed to improve the quality of images already severely compressed by the popular JPEG standard, which is based on Discrete Cosine Transform (DCT).

Finally, N. Saito introduced the *Polyharmonic Local Fourier Transform* (PHLFT) [1] by setting $q_\ell = \ell,\ \ell \in \mathbb{Z}_m$ in Eq. (1.2) and by replacing the Fourier sine series with the complex Fourier series in expanding the v_j components. With some sacrifice of the decay rate of the expansion coefficients, i.e., of order $O\left(\|\boldsymbol{k}\|^{-m-1}\right)$ instead of order $O\left(\|\boldsymbol{k}\|^{-2m-1}\right)$ or of order $O\left(\|\boldsymbol{k}\|^{-2m-2}\right)$, PHLFT allows one to compute local Fourier magnitudes and phases without facing the Gibbs phenomenon. PHLFT also can capture the important information of orientation much better than PHLST and PHLCT. Moreover, it is fully invertible and should be useful for various filtering, analysis, and approximation purposes.

Although the Fourier coefficients of the residual v decay rapidly, it is virtually useless for the purpose of approximation. *Therefore, in practice we shall not only seek fast decaying rate of the Fourier coefficients of the residual v, but also a residual v of a small energy*. However, the residual v in PHLST (or in PHLCT or PHLFT) in general does not necessarily have minimal energy. In resolving this issue, we propose the least squares approximant to a given signal using the combination of a set of algebraic polynomials and a set of trigonometric polynomials. The maximum degree of the algebraic polynomials is chosen to be small and fixed. We show in this paper that the least squares approximant converges uniformly for a Hölder continuous function. Therefore the Gibbs phenomenon will not occur on the boundary for such functions. We also show that the PHLST (or PHLCT, PHLFT) converges uniformly and is a near least squares approximation in the sense that it is arbitrarily close to the least squares approximant in the L^2 norm as the dimension of the approximation space increases. Our experiments show the proposed method is robust in approximating a highly oscillating signal. Even when the signal is corrupted by noise, the method is still robust. The experiments also reveal that an optimum degree of trigonometric polynomial is needed in order to attain minimal l^2 error of the approximation when there is noise present in the data set. This optimum degree is shown to be determined by the intrinsic frequency of the signal. We also discuss the energy compaction of the solution vector and give an explanation to it.

2. Problem formulation and characterization of the solution

Let f be a noise corrupted and finite-energy signal on the interval $J := [-1, 1]$, that is $f \in L^2(J)$. The L^2 norm of a function $f \in L^2(J)$ is denoted by $\|f\|$, that is $\|f\|^2 := \int_J |f(x)|^2\, dx$. Other norms used in this paper will be indicated by the appropriate subscripts.

Let q and p be nonnegative integers. Denote by $\mathcal{T}_q := \operatorname{span}\left\{\frac{1}{\sqrt{2}}e^{(\mathrm{i}j\pi\cdot)}, |j| \le q\right\}$ the space of trigonometric polynomials of degree no greater than q, and by $\mathcal{P}_p$ the set of algebraic polynomials with degree no greater than p. Let us consider the minimization problem: *finding* $u_p \in \mathcal{P}_p$ *and* $v_q \in \mathcal{T}_q$ *such that*

$$\|f - u_p - v_q\| = \inf_{\substack{u\in\mathcal{P}_p,\\ v\in\mathcal{T}_q}} \|f - u - v\|. \tag{2.3}$$

Our intention is to keep p small and fixed, such as $p = 1, 2, 3$, *and q can be large as needed*. Equivalently, by defining $n := p+q+1$ and

$$V_n := \mathcal{P}_p \cup \mathcal{T}_q,$$

we can recast the minimization problem (2.3) as: finding $f_n := u_p + v_q \in V_n$, such that

$$\|f - f_n\| = \inf_{h\in V_n} \|f - h\|. \tag{2.4}$$

We refer to the approximant f_n obtained via (2.3) or (2.4) as *one-step least squares approximation*. To assert the existence and uniqueness of the solution to the minimization problems (2.3) or (2.4), we recall the following proposition.

Proposition 2.1 ([7, p. 114]). *Assume V is a convex and closed finite-dimensional subset of an inner product space H. There is a unique element* $\hat{f} \in V$ *such that*

$$\|f - \hat{f}\| = \inf_{h\in V} \|f - h\|, \quad f \in H.$$

The set V_n defined earlier is in fact a finite-dimensional subspace of $L^2(J)$, hence convex and closed. Therefore by Proposition 2.1, the following theorem holds.

Theorem 2.2. *There exists a unique best approximant* $f_n \in V_n$ *to the minimization problem* (2.3) *or* (2.4)*.*

The set $\{V_n, n \in \mathbb{N}\}$ forms a nested sequence of subspaces of $L^2(J)$, i.e.,

$$V_1 \subset V_2 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots \subset L^2(J).$$

In fact, since it is known that the set $\left\{\frac{1}{\sqrt{2}}e^{(\mathrm{i}j\pi\cdot)} : j \in \mathbb{Z}\right\}$ is an orthonormal basis for $L^2(J)$, we must have

$$\overline{\bigcup_{n\in\mathbb{N}} V_n} = L^2(J).$$

Therefore, in the L^2 norm, the least squares approximant f_n converges to f, i.e.,

$$\|f - f_n\| \to 0 \quad \text{as } n \to \infty. \tag{2.5}$$

Next lemma is crucial in characterizing the one-step least squares approximation. We shall use $\mathcal{F}_q(f)$ to denote the partial Fourier series of terms of degree $\le q$ of a function $f \in L^2(J)$.

Lemma 2.3. *Suppose* $f_n = u_p + v_q$ *is the one-step least squares approximation of* $f \in L^2(J)$, *where* $u_p \in \mathcal{P}_p$ *and* $v_q \in \mathcal{T}_q$. *Then* v_q *must be the partial Fourier series of* $f - u_p$ *for each* $q \in \mathbb{Z}_+$, *i.e.,*

$$v_q = \mathcal{F}_q(f - u_p) = \sum_{|j|\le q} \frac{1}{\sqrt{2}} c_j e^{(\mathrm{i}j\pi\cdot)}$$

and

$$f - u_p = \lim_{q\to\infty} v_q = \sum_{j\in\mathbb{Z}} \frac{1}{\sqrt{2}} c_j e^{(\mathrm{i}j\pi\cdot)},$$

where the sequence $(c_j : j \in \mathbb{Z}) \in l^2(\mathbb{Z})$ *and* $c_j = \langle f - u_p, \frac{1}{\sqrt{2}} e^{\mathrm{i}j\pi\cdot}\rangle$ *for* $j \in \mathbb{Z}$. *Here* $\langle \cdot, \cdot \rangle$ *denotes the inner product of two functions on the interval* $[-1, 1]$.

Proof. Since $u_p + v_q$ is the least squares approximation of f, we must have

$$\|f - (u_p + v_q)\| \le \|f - (u + v)\|, \quad \text{for any } u \in \mathcal{P}_p, v \in \mathcal{T}_q.$$

In particular, fix $u = u_p$,

$$\|f - (u_p + v_q)\| \le \|f - (u_p + v)\|, \quad \text{for any } v \in \mathcal{T}_q.$$

This is equivalent to

$$\|(f-u_p)-v_q\| \le \|(f-u_p)-v\|, \quad \text{for any } v \in \mathcal{T}_q,$$

which implies that v_q is the best L^2 approximation in $\mathcal{T}_q$ of $f-u_p$. But this is equivalent to saying that v_q is the partial Fourier series of $f-u_p$ of degree q. □

In fact as $n \to \infty$ (hence $q \to \infty$), there are two sequences of functions created: $(u_p^{(n)} : n \in \mathbb{N})$ and $(v_q^{(n)} : n \in \mathbb{N})$, and the corresponding coefficient sequence $(c_j) : j \in \mathbb{Z}$ also depends on n. To ease the burden of notations, we will suppress the superscripts. But keep in mind for the least squares approximant $f_n = u_p + v_q$, *both u_p and v_q vary as q increases while p is kept fixed*.

We shall in the sequel also consider the least squares approximant on the interval $I := [0, 1]$ of a function $g \in L^2(I)$ using the combination of algebraic polynomials with sine polynomials only or cosine polynomials only in place of trigonometric polynomials. For simplicity, we still use $\mathcal{P}_p$ to denote the set of algebraic polynomials of degree no greater than p on I. Denote the set of sine polynomials of degree no greater than q by

$$\mathcal{S}_q := \left\{\sqrt{2}\sin(j\pi\,\cdot) : j \in \mathbb{N}_q\right\},$$

and the set of cosine polynomials of degree no greater than q by

$$\mathcal{C}_q := \left\{\sqrt{2}\cos(j\pi\,\cdot) : j \in \mathbb{N}_q\right\}.$$

Recall the Fourier sine series of g on I coincides with the (full) Fourier series of the periodized odd reflection of g, which we denote by g_{odd}. Likewise, the Fourier cosine series of g on I is the (full) Fourier series of the periodized even reflection of g, which we denote by g_{even}. Sometimes we will suppress the subscripts for even or odd extension if the meaning is clear from the context. We shall still use $\mathcal{F}_q(g)$ to denote the truncated sine series or cosine series of a function g on I. The following corollaries are immediate results of Lemma 2.3.

Corollary 2.4. *Suppose $f_n = u_p + v_q$ is the one-step least squares approximation of $f \in L^2(I)$, where $u_p \in \mathcal{P}_p$ and $v_q \in \mathcal{S}_q$. Then v_q must be the partial sine series of $f-u_p$ for each $q \in \mathbb{N}$, i.e.,*

$$v_q = \mathcal{F}_q(f-u_p) = \sum_{j\in\mathbb{N}_q} \sqrt{2}c_j \sin(j\pi\,\cdot)$$

and

$$f-u_p = \lim_{q\to\infty} v_q = \sum_{j\in\mathbb{N}} \sqrt{2}c_j \sin(j\pi\,\cdot),$$

where the sequence $(c_j : j \in \mathbb{N}) \in l^2(\mathbb{Z})$.

Corollary 2.5. *Suppose $f_n = u_p + v_q$ is the one-step least squares approximation of $f \in L^2(I)$, where $u_p \in \mathcal{P}_p$ and $v_q \in \mathcal{C}_q$. Then v_q must be the partial cosine series of $f-u_p$ for each $q \in \mathbb{N}$, i.e.,*

$$v_q = \mathcal{F}_q(f-u_p) = \sum_{j\in\mathbb{Z}_{q+1}} \sqrt{2}c_j \cos(j\pi\,\cdot)$$

and

$$f-u_p = \lim_{q\to\infty} v_q = \sum_{j\in\mathbb{Z}_+} \sqrt{2}c_j \cos(j\pi\,\cdot),$$

where the sequence $(c_j : j \in \mathbb{Z}_+) \in l^2(\mathbb{Z})$.

Lemma 2.3 and Corollaries 2.4 and 2.5 allow us to apply the classical Fourier analysis to the least squares approximant $u_p + v_q$.

Theorem 2.6. *Let $f \in L^2(J)$. Suppose that $\tilde{u}_p \in \mathcal{P}_p$ is an arbitrary polynomial of degree no greater than p, and $\tilde{v}_q := \mathcal{F}_q(f-\tilde{u}_p)$ is the partial Fourier series of $f-\tilde{u}_p$ of degree q. Define the function*

$$\tilde{f}_n = \tilde{u}_p + \tilde{v}_q.$$

Then there exist $N \in \mathbb{N}$, such that whenever $n \ge N$, $\tilde{f}_n$ is arbitrarily close to the one-step least squares approximant f_n in the L^2 norm. That is, for any given $\epsilon > 0$, there exits N, such that

$$\|f_n - \tilde{f}_n\| < \epsilon, \quad \text{for } n \ge N,$$

where the norm $\|\cdot\|$ is the L^2 norm.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

Proof. By Lemma 2.3, we know for the one-step least squares approximant $u_p + v_q$,

$$f - u_p = \lim_{q\to\infty} v_q.$$

By the assumption of the theorem, we also obtain

$$f - \tilde{u}_p = \lim_{q\to\infty} \tilde{v}_q.$$

The above two equalities are in the L^2 sense and they yield

$$u_p + \lim_{q\to\infty} v_q = \tilde{u}_p + \lim_{q\to\infty} \tilde{v}_q.$$

But this implies that

$$\lim_{q\to\infty} (u_p + v_q) = \lim_{q\to\infty} (\tilde{u}_p + \tilde{v}_q)$$

in the L^2 norm and the theorem follows. □

Although u_p varies for each q as q increases in finding the one-step least squares approximant $f_n = u_p + v_q$, according to Theorem 2.6, one can find an approximant $\tilde{f}_n$ arbitrarily close to the one-step least squares approximant f_n by fixing the polynomial component $\tilde{u} = \varphi$, and then computing the partial Fourier series $\tilde{v}_q$ of $f - \varphi$. We summarize this in the following corollary.

Corollary 2.7. *Given $f \in L^2(J)$, its one-step least squares approximant $f_n = u_p + v_q$ from the space $\mathcal{P}_p \cup \mathcal{T}_q$ can be approximated by choosing a fixed polynomial $\varphi \in \mathcal{P}_p$. That is,*

$$f_n \approx \varphi + \tilde{v}_q, \tag{2.6}$$

where $\tilde{v}_q$ is the partial Fourier series of degree q of $f - \varphi$. This approximation can have arbitrary accuracy in the sense that

$$\lim_{n\to\infty} f_n = \varphi + \lim_{q\to\infty} \tilde{v}_q,$$

in the L^2 norm.

The results of Theorem 2.6 and Corollary 2.7 can be easily adapted to the cases of using sine polynomials or cosine polynomials. We leave it to the readers. Since in Eq. (2.6), the choice of φ is arbitrary, we define a near least squares approximation for the class of continuous functions.

Definition 2.8. Let f be a continuous function on the interval $[-1, 1]$. Let φ be a polynomial of degree p passing through the two points $(-1, f(-1))$ and $(1, f(1))$, and let the truncated Fourier series of $f - \varphi$ be $\bar{v}_q$ of degree q. The *near least squares approximation* of f from the approximating space $\mathcal{P}_p \cup \mathcal{T}_q$ is

$$\bar{f}_n := \varphi + \bar{v}_q.$$

Remark. For the near least squares approximation on the interval [0, 1] formed by sine polynomials, we define the polynomial φ to be the one passing through $(0, f(0))$ and $(1, f(1))$. For the near least squares approximation on the interval [0, 1] formed by cosine polynomials, φ can be an arbitrary polynomial.

We are more concerned if the one-step least squares approximation converges uniformly on the interval $[-1, 1]$ (or [0, 1]), hence also strongly in the L^2 norm and pointwise. To that end, we recall the well-known Jackson's Theorem [8] on the uniform error bound for algebraic polynomial approximation or trigonometric polynomial approximation, and the fact that the *uniform norm* of $\mathcal{F}_q(f)$ is of order $O(\log q)$ [9, p. 67], we present the following proposition.

Proposition 2.9 ([7, p. 128]). *Let the 2-periodic function f possess continuous derivatives up to order $k \geq 0$. Further assume that its k-th derivative is Hölder continuous of order $\alpha \in (0, 1]$, then the uniform norm of the error of the approximation $\mathcal{F}_q(f)$*

$$\|f - \mathcal{F}_q(f)\|_\infty \leq \tau_k \frac{\log q}{q^{k+\alpha}}, \quad \text{for } q \geq 2,$$

where τ_k is a constant linearly dependent on the Hölder constant of the k-th derivative of f and otherwise independent of f. In particular, if f is continuous $(k = 0)$ for some Hölder constant $\alpha > 0$, then $\mathcal{F}_q(f)$ converges to f uniformly.

Next lemma shows that the near least squares approximant for a continuous function converges uniformly.

Lemma 2.10. *Assume f is Hölder continuous of order $\alpha \in (0, 1]$ on the interval $[-1, 1]$. Suppose its first derivatives is of bounded variation. The near least squares approximation $\bar{f}_n$ converges uniformly on the interval $[-1, 1]$ as $n \to \infty$, at least in the order of $O\left(\frac{\log q}{q^\alpha}\right)$, where q is the degree of the trigonometric polynomial. Moreover, the Fourier coefficient with frequency index j decays at least in the order of $O\left(\frac{1}{j^2}\right)$.*

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

Proof. Note in constructing the near least squares approximation, φ is chosen such that $(f-\varphi)(-1)=(f-\varphi)(1)=0$. Thus $(f-\varphi)$ can be continuously extended to a 2-periodic function $(f-\varphi)_{\text{per}}$ of Hölder constant α. By Proposition 2.9, $\bar{v}_q$, the truncated Fourier series $\mathcal{F}_q(f-\varphi)$ of the periodized $(f-\varphi)_{\text{per}}$, converges to $f-\varphi$ uniformly in the order indicated in the theorem. The decay rate of the Fourier coefficients follows from the classical Fourier analysis [9, p. 46]. □

Remark. For the near least squares approximant using cosine polynomials or sine polynomials, the result of Lemma 2.10 remains the same. In the case f has continuous derivatives of higher order $k>0$, such as at least $k\geq\left(\lfloor\frac{p+1}{2}\rfloor-1\right)$, one can find φ as the Hermite interpolating polynomial of f such that $f^{(i)}(-1)=\varphi^{(i)}(-1)$ and $f^{(i)}(1)=\varphi^{(i)}(1)$, $i\in\mathbb{Z}_{\lfloor\frac{p+1}{2}\rfloor}$, where $\lfloor x\rfloor$ is the largest integer $\leq x$. If we also assume that the $\left(\lfloor\frac{p+1}{2}\rfloor-1\right)$th order derivative of f is of α-Hölder continuous, consequently, the periodized extension of $f-\varphi$ is in $C^{\lfloor\frac{p+1}{2}\rfloor-1,\alpha}(\mathbb{R})$, where $\alpha\in(0,1]$. In this case, the near least squares approximant $\bar{f}_n$ converges to f in the possibly highest order $O\left(\frac{\log q}{q^{\lfloor\frac{p+1}{2}\rfloor-1+\alpha}}\right)$, and correspondingly the Fourier coefficients decay in the order of $O\left(\frac{1}{j^{\lfloor\frac{p+1}{2}\rfloor+1}}\right)$. This makes us quickly link the near least squares approximant in this case to a truncation of the PHLFT of a function f. We state it in the next theorem.

Theorem 2.11. *Let f be $2m$ $(m>0)$ times continuously differentiable on the interval $[-1,1]$. Assume its $(m-1)$ st derivative $f^{(m-1)}$ is Hölder continuous of order $\alpha\in(0,1]$. Then its PHLFT representation with the degree of polyharmonicity m is*

$$f=\phi+\frac{1}{\sqrt{2}}\sum_{j\in\mathbb{Z}}c_j\mathrm{e}^{(\mathrm{i}j\pi\cdot)},$$

where, ϕ is a Hermite interpolating polynomial of degree at most $2m-1$ satisfies the polyharmonic equation $\phi^{(2m)}=0$ and $f^{(\ell)}(-1)=\phi^{(\ell)}(-1), f^{(\ell)}(1)=\phi^{(\ell)}(1)$, $\ell\in\mathbb{Z}_m$. Hence the coefficients c_j, $j\in\mathbb{N}$, decay in the order of $O(j^{-m-1})$. Moreover,

$$\tilde{f}_n=\phi+\frac{1}{\sqrt{2}}\sum_{|j|\leq q}c_j\mathrm{e}^{(\mathrm{i}j\pi\cdot)}$$

is a near least squares approximant of f on $[-1,1]$ from the space $\mathcal{P}_{2m-1}\cup\mathcal{T}_q$, and $\tilde{f}_n$ converges to f uniformly in the order of $O\left(\frac{\log q}{q^{m-1+\alpha}}\right)$.

We note the periodized odd reflection of a function on $[0,1]$ automatically induces the continuity of the derivatives of odd order at the boundary, therefore to maximize the regularity of the periodized odd reflection of $f-\varphi$, one only needs to match the values of even order derivatives of φ with those of f on the boundary. This makes us link PHLST to the near least approximant. We state it in the next corollary.

Corollary 2.12. *Let f be $2m$ $(m>0)$ times continuously differentiable on the interval $[0,1]$. Assume its $(2m-1)$ st derivative $f^{(2m-1)}$ is Hölder continuous of order $\alpha\in(0,1]$. Then its PHLST representation with the degree of polyharmonicity m is*

$$f=\phi+\sqrt{2}\sum_{j\in\mathbb{N}}c_j\sin(j\pi\cdot),$$

where, ϕ is a Hermite interpolating polynomial of degree at most $2m-1$ satisfies the polyharmonic equation $\phi^{(2m)}=0$ and $f^{(2\ell)}(0)=\phi^{(2\ell)}(0), f^{(2\ell)}(1)=\phi^{(2\ell)}(1)$, $\ell\in\mathbb{Z}_m$. Hence the coefficients c_j, $j\in\mathbb{N}$, decay in the order of $O(j^{-2m-1})$. Moreover,

$$\tilde{f}_n=\phi+\sqrt{2}\sum_{j\in\mathbb{N}_q}c_j\sin(j\pi\cdot)$$

is a near least squares approximant of f on $[0,1]$ from the space $\mathcal{P}_{2m-1}\cup\mathcal{T}_q$, and $\tilde{f}_n$ converges to f uniformly in the order of $O\left(\frac{\log q}{q^{2m-1+\alpha}}\right)$.

Likewise, the periodized even reflection of a function on $[0,1]$ automatically induces the continuity of the even order derivatives on the boundary, therefore to maximize the regularity of the periodized even reflection of $f-\varphi$, one only needs to match the values of odd order derivatives of φ with those of f on the boundary. This makes us link PHLCT to the near least approximant. We state it in the next corollary.

Corollary 2.13. *Let f be $2m$ $(m>0)$ times continuously differentiable on the interval $[0,1]$. Assume its $2m$-th derivative $f^{(2m)}$ is Hölder continuous of order $\alpha\in(0,1]$. Then its PHLCT representation with the degree of polyharmonicity m is*

$$f=\phi+\sqrt{2}\sum_{j\in\mathbb{N}}c_j\cos(j\pi\cdot),$$

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

where, ϕ is a Hermite interpolating polynomial of degree at most $2m-1$ satisfies the polyharmonic equation $\phi^{(2m)} = 0$ and $f^{(2\ell+1)}(0) = \phi^{(2\ell+1)}(0), f^{(2\ell+1)}(1) = \phi^{(2\ell+1)}(1),\ \ell \in \mathbb{Z}_m$. Hence the coefficients $c_j, j \in \mathbb{N}$, decay in the order of $O(j^{-2m-2})$. Moreover,

$$\tilde{f}_n = \phi + \sqrt{2} \sum_{j\in\mathbb{Z}_{q+1}} c_j \cos(j\pi\,\cdot)$$

is a near least squares approximant of f on $[0, 1]$ from the space $\mathcal{P}_{2m-1} \cup \mathcal{T}_q$, and $\tilde{f}_n$ converges to f uniformly in the order of $O\left(\frac{\log q}{q^{2m+\alpha}}\right)$.

From the above discussion, we know a near least squares approximant is an approximation to the one-step least squares approximant that also converges uniformly. Now we are ready to show that the least squares approximant must converge uniformly for a continuous functions.

Theorem 2.14. *Assume f is Hölder continuous of order $\alpha \in (0, 1]$ on the interval $[-1, 1]$. The one-step least squares approximant $f_n = u_p + v_q$ converges uniformly on the interval $[-1, 1]$, as $n \to \infty$, at least in the order of $O\left(\frac{\log q}{q^\alpha}\right)$. Moreover, the Fourier coefficients decay at least in the order of $O\left(\frac{1}{j^2}\right)$, and u_p must converge to a polynomial whose values at the boundary agree with those of f, as q approaches ∞.*

Proof. We recall from Lemma 2.3 that v_q is the partial Fourier series of $f - u_p$. Let us denote the Fourier series of $f - u_p$ by $f - u_p = \sum_{j\in\mathbb{Z}} \frac{1}{\sqrt{2}} c_j e^{(ij\pi\cdot)}$. Similarly, denote the Fourier series of $f - \varphi$ by $f - \varphi = \sum_{j\in\mathbb{Z}} \frac{1}{\sqrt{2}} d_j e^{(ij\pi\cdot)}$. Since $\|f - u_p - v_q\| \le \|f - \varphi - \bar{v}_q\|$, we have by using Parseval's identity,

$$\sum_{|j|>q} |c_j|^2 \le \sum_{|j|>q} |d_j|^2.$$

By Lemma 2.10 we know

$$\sum_{|j|>q} |d_j|^2 \le C \sum_{|j|>q} \left(\frac{1}{j^2}\right)^2,$$

for some constant C. Therefore

$$\sum_{|j|>q} |c_j|^2 \le C \sum_{|j|>q} \left(\frac{1}{j^2}\right)^2.$$

This implies c_j decays at least in the order of $O\left(\frac{1}{j^2}\right)$. Hence $\sum_{j\in\mathbb{N}} |c_j| < \infty$. Consequently v_q must converge uniformly to $f - u_p$, which in turn implies $f_n = u_p + v_q$ converges to f uniformly. The decaying order of c_j also implies that the periodized extension of $f - u_p$ is in $C^{0,\alpha}(\mathbb{R})$ as $q \to \infty$. The rest of the theorem immediately follows from this fact. □

The above theorem indicates that for an α-Hölder continuous function f on $[-1, 1]$ with nonzero values at the two end points, our proposed one-step least squares approximant and the near least squares approximant can eliminate the Gibbs phenomenon that would occur if only trigonometric polynomials are used.

When f has continuous derivatives of higher order $k > 0$, such as at least $k \ge \left(\lfloor \frac{p+1}{2} \rfloor - 1\right)$, we expect that the one-step least squares approximant should converge uniformly faster. By the remark after Lemma 2.10 and the previous theorem we deduce the following result.

Theorem 2.15. *Assume that f has continuous derivatives of order $k \ge \left(\lfloor \frac{p+1}{2} \rfloor - 1\right)$ on the interval $[-1, 1]$, and that the $\left(\lfloor \frac{p+1}{2} \rfloor - 1\right)$ st order derivative of f is of $\alpha \in (0, 1]$ Hölder continuous. The one-step least squares approximant $f_n = u_p + v_q$ converges uniformly on the interval $[-1, 1]$, as $n \to \infty$, at least in the order of $O\left(\frac{\log q}{q^{\lfloor \frac{p+1}{2} \rfloor - 1 + \alpha}}\right)$. Moreover, the Fourier coefficients decay at least in the order of $O\left(\frac{1}{j^{\lfloor \frac{p+1}{2} \rfloor + 1}}\right)$, and u_p must converge to a polynomial whose values at the boundary agree with those of f, and moreover whose values of normal derivatives at the boundary up to order $\lfloor \frac{p+1}{2} \rfloor - 1$ agree with those of f, as q approaches ∞.*

We readily present the results for the one-step least squares approximant by using only sine polynomials or only cosine polynomials in the next two corollaries.

Corollary 2.16. *Let f be $2m$ $(m > 0)$ times continuously differentiable on the interval $[0, 1]$. Assume its $(2m-1)$ st derivative $f^{(2m-1)}$ is Hölder continuous of order $\alpha \in (0, 1]$. The least squares approximant $f_n = u_p + v_q$ by using sine polynomials v_q*

converges uniformly on the interval $[0, 1]$, *as* $n \to \infty$, *at least in the order of* $O\left(\frac{\log q}{q^{2m-1+\alpha}}\right)$. *Moreover, the Fourier sine coefficients decay at least in the order of* $O\left(\frac{1}{j^{2m+1}}\right)$, *and* u_p *must converge to a polynomial whose values at the boundary agree with those of* f, *and whose values of normal derivatives at the boundary up to order* $2m-1$ *agree with those of* f, *as* q *approaches* ∞.

Corollary 2.17. *Let* f *be* $2m$ $(m > 0)$ *times continuously differentiable on the interval* $[0, 1]$. *Assume its* $2m$ *st derivative* $f^{(2m)}$ *is Hölder continuous of order* $\alpha \in (0, 1]$. *The one-step least squares approximant* $f_n = u_p + v_q$ *by using cosine polynomials* v_q *converges uniformly on the interval* $[0, 1]$, *as* $n \to \infty$, *at least in the order of* $O\left(\frac{\log q}{q^{2m+\alpha}}\right)$. *Moreover, the Fourier sine coefficients decay at least in the order of* $O\left(\frac{1}{j^{2m+2}}\right)$, *and* u_p *must converge to a polynomial whose values at the boundary agree with those of* f, *and whose values of normal derivatives at the boundary up to order* $2m$ *agree with those of* f, *as* q *approaches* ∞.

According to Corollary 2.7, one may try to approximate the one-step least squares approximant by the following approach:
Step 1: find $\hat{u}_p \in \mathscr{P}_p$, such that

$$\|f - \hat{u}_p\| = \inf_{u \in \mathscr{P}_p,} \|f - u\|.$$

Step 2: Find $\hat{v}_q \in \mathscr{T}_q$, such that

$$\|(f - \hat{u}_p) - \hat{v}_q\| = \inf_{v \in \mathscr{T}_q,} \|(f - \hat{u}_p) - v\|.$$

Let us denote $\hat{f}_n := \hat{u}_p + \hat{v}_q$. We name $\hat{f}$ as the *two-step least squares approximant*. Although the two-step least squares approximant still provides an approximation to the one-step least squares approximant in L^2 norm, in general, the sequence $(\hat{f}_n)$ does not converges to f uniformly if trigonometric polynomials are used. To explain this, let us assume f is continuous on $[-1, 1]$. Then in general $(f - \hat{u}_p)(-1) \neq (f - \hat{u}_p)(1)$, hence the periodized extension of $(f - \hat{u}_p)$ has discontinuity at -1 and 1. In this case, the Gibbs phenomenon occurs near -1 and 1. If the function f is continuous on $[0, 1]$, and using the sine polynomials to construct the two-step least squares approximant, in general $(f - \hat{u}_p)(0) \neq 0 \neq (f - \hat{u}_p)(1)$, hence the periodized odd extension has discontinuities at 0 and 1 that leads to Gibbs phenomenon. In both cases the Fourier (sine) coefficient with frequency index j decays in the order of $O\left(\frac{1}{j}\right)$. However, for a continuous function f on $[0, 1]$, and using the cosine polynomials to construct the two-step least squares approximant, no Gibbs phenomenon will occur because the even reflection does not introduce an discontinuities into the periodized extension, and the Fourier cosine coefficient with frequency index j decays in the order of $O\left(\frac{1}{j^2}\right)$. We point out a further fact in the following corollary when using sine polynomials to build the two-step least squares approximant.

Corollary 2.18. *Let* f *be continuous on* $[0, 1]$. *If we use sine polynomials to construct the two-step least squares approximant, the* L^∞ *error of the two-step least squares approximation remains constant when* q *is greater than certain number while* p *remains constant.*

Proof. Note that the two-step least squares approximation only creates discontinuities at the boundary. Therefore by Dirichlet's Theorem [9, p. 57], the two-step least squares approximant converges to f pointwise except the two boundary points. Observe that the two-step least squares approximant completely decouples the subspace of algebraic polynomials and the subspace of sine polynomials. Since any function in the set $\mathscr{S}_q$ vanishes at the boundary point 0 and 1, the error of the two-step approximant on the boundary is completely determined by the chosen degree p of algebraic polynomials, regardless of the degree q of trigonometric sine polynomials used in the approximation. Thus as q increases with p fixed, the error in the interior of the interval $[0, 1]$ deceases, while the error at the boundary points 0 and 1 remains constant and gradually becomes dominant. Hence even if q becomes greater than some number, the L^∞ error of the two-step least squares approximation does not change. □

3. Implementation consideration

We plan to compute and compare the one-step least squares approximant f_n, the near least squares approximant $\bar{f}_n$, the PHLST (or PHLFT, PHLCT) approximation $\tilde{f}_n$, and the two-step least squares approximant $\hat{f}_n$ of a given signal f. For f_n, a system of linear equations is needed to be solved. For other three types of approximant, solving a linear system is not needed, only Fourier series expansion is involved, therefore, DFT, DST, or DCT, all of which are implemented by the FFT algorithm, can be employed.

We next briefly discuss implementations of finding the one-step least squares approximant f_n. Let

$$V_n = \text{span}\left\{\phi_i \in L^2(J), i \in \mathbb{Z}_n\right\}, \quad n \in \mathbb{Z}_+.$$

Let the least squares approximant $f_n = \sum_{j\in\mathbb{Z}_n} c_j\phi_j$. Then the unknown coefficient vector $\boldsymbol{c} := \{c_j, j \in \mathbb{Z}_n\}$ is found by solving the *normal matrix equation*:

$$\boldsymbol{G}\boldsymbol{c} = \boldsymbol{d}, \tag{3.7}$$

where $\boldsymbol{G}$ is a *symmetric and positive semi-definite* Gram matrix:

$$\boldsymbol{G} := \left[g_{i,j}\right]_{i,j\in\mathbb{Z}_n}, \quad \text{with } g_{i,j} = (\phi_i, \phi_j) = \int_J \phi_i(x)\phi_j(x)\,\mathrm{d}x$$

and

$$\boldsymbol{d} = [d_i]_{i\in\mathbb{Z}_n}, \quad \text{with } d_i = (f, \phi_i) = \int_J f(x)\phi_i(x)\,\mathrm{d}x.$$

The Gram matrix $\boldsymbol{G}$ only depends on the chosen spanning set $\{\phi_i \in L^2(I), i \in \mathbb{Z}_n\}$ of the approximating space. $\boldsymbol{G}$ is nonsingular if and only if the set $\{\phi_i \in L^2(I) : i \in \mathbb{Z}_n\}$ is linearly independent. Hence if the set $\{\phi_i \in L^2(I) : i \in \mathbb{Z}_n\}$ is linearly independent, it immediately follows the existence and uniqueness of the minimization problems (2.3) or (2.4), and also the *positive definiteness* of $\boldsymbol{G}$. When $\boldsymbol{G}$ is positive definite, the Cholesky factorization of it is available. Assume the Cholesky factorization of the Gram matrix $\boldsymbol{G}$ is

$$\boldsymbol{G} = \boldsymbol{R}^T\boldsymbol{R},$$

where $\boldsymbol{R}$ is an upper triangular matrix. Then we have

$$\boldsymbol{c} = \boldsymbol{R}^{-1}(\boldsymbol{R}^T)^{-1}\boldsymbol{d}. \tag{3.8}$$

However in the case when $\boldsymbol{G}$ becomes nearly singular, the solution by Eq. (3.8) is unreliable. We therefore shall instead consider the regularized SVD of the matrix $\boldsymbol{G}$ by setting to zeros those singular values smaller than a threshold, say, $\epsilon > 0$. Let the regularized SVD of $\boldsymbol{G}$ be

$$\boldsymbol{G}_\epsilon := \boldsymbol{U}_r\boldsymbol{D}_r\boldsymbol{V}_r^T,$$

where r is the number of singular values greater than the threshold ϵ, $\boldsymbol{U}_r$ is a matrix of size $n \times r$ with orthonormal columns, $\boldsymbol{D}_r$ is a $r \times r$ diagonal matrix with the singular values of $\boldsymbol{G}$ greater than the threshold ϵ, and $\boldsymbol{V}_r$ is a $n \times r$ matrix with orthonormal columns. Define the *pseudo-inverse* of $\boldsymbol{G}$ by

$$\boldsymbol{G}^+ := \boldsymbol{V}_r\boldsymbol{D}_r^{-1}\boldsymbol{U}_r^T.$$

Then the solution obtained by regularized SVD is denoted by

$$\boldsymbol{c}_\epsilon := \boldsymbol{G}^+\boldsymbol{d} = \boldsymbol{V}_r\boldsymbol{D}_r^{-1}\boldsymbol{U}_r^T\boldsymbol{d}. \tag{3.9}$$

We next specifically look at the Gram matrix $\boldsymbol{G}$ by using the example of normalized Legendre polynomials and sine polynomials on the interval [0, 1]. Let P_k be the normalized Legendre polynomial of degree k on [0, 1], $k \in \mathbb{Z}_+$, and $\psi_j(x) := \sqrt{2}\sin j\pi x, j \in \mathbb{N}$. It is easy to see that the set $\{P_k, k \in \mathbb{Z}_{p+1}\} \cup \{\psi_j, j \in \mathbb{N}_q\}$ is linearly independent for any finite $p \in \mathbb{Z}_+$ and $q \in \mathbb{N}$. Therefore the Gram matrix $\boldsymbol{G}$ is nonsingular. The choice of the basis functions allows us to specifically compute the Gram matrix $\boldsymbol{G}$ as follows.

$$\begin{aligned}
&g_{i,j} = (P_i, P_j) = \delta_{ij}, \quad i, j \in \mathbb{Z}_{p+1}\\
&g_{i+p,j+p} = (\psi_i, \psi_j) = \delta_{ij}, \quad i, j \in \mathbb{N}_q\\
&g_{i,j+p} = (P_i, \psi_j) = \sqrt{2}\int_0^1 P_i(x)\sin j\pi x\,\mathrm{d}x, \quad i \in \mathbb{Z}_{p+1},\ j \in \mathbb{N}_q.
\end{aligned}$$

The structure of the Gram matrix $\boldsymbol{G}$ is more easily seen by a block matrix, that is

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_{p+1} & \boldsymbol{B}_{p+1,q} \\ \boldsymbol{B}_{p+1,q}^T & \boldsymbol{I}_q \end{bmatrix} = \boldsymbol{I}_n + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{B}_{p+1,q} \\ \boldsymbol{B}_{p+1,q}^T & \boldsymbol{0} \end{bmatrix} \tag{3.10}$$

where, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $\boldsymbol{B}_{p+1,q}$ is a matrix of size $(p+1) \times q$ with

$$\boldsymbol{B}_{p+1,q} = [b_{ij}]_{i\in\mathbb{Z}_{p+1}, j\in\mathbb{N}_q} \quad \text{with } b_{ij} = g_{i,j+p} = (P_i, \psi_j).$$

Moreover we observe that the matrix $\boldsymbol{B}$ has a special "checkerboard" structure, namely it has alternating zeros in its rows. The following is an illustration of $\boldsymbol{B}$, where $*$ indicates a nonzero entry.

$$\boldsymbol{B} = \begin{bmatrix} * & 0 & * & 0 & \cdots \\ 0 & * & 0 & * & \cdots \\ * & 0 & * & 0 & \cdots \\ \vdots & & & & \vdots \end{bmatrix}.$$

That is,

$$b_{2i,2j} = b_{2i+1,2j-1} = 0, \quad i \in \mathbb{Z}_{p+1}, j \in \mathbb{N}_q,$$

or in terms of elements of $\boldsymbol{G}$,

$$g_{2i,2j+p} = g_{2i+1,2j-1+p} = 0, \quad i \in \mathbb{Z}_{p+1}, j \in \mathbb{N}_q.$$

The structure of $\boldsymbol{B}$ is confirmed by the following lemma.

Lemma 3.1. *Let P_k be the normalized Legendre polynomial of degree k on $[0, 1]$, then we have*

$$\int_0^1 P_{2k}(x) \sin 2j\pi x \, dx = \int_0^1 P_{2k+1}(x) \sin(2j-1)\pi x \, dx = 0, \quad k \in \mathbb{Z}_+, j \in \mathbb{N}. \tag{3.11}$$

In particular,

$$\int_0^1 \sqrt{2} \sin j\pi x P_0(x) \, dx = \frac{\sqrt{2}}{j\pi}(1 - (-1)^j) \tag{3.12}$$

$$\int_0^1 \sqrt{2} \sin j\pi x P_1(x) \, dx = -\frac{\sqrt{6}}{j\pi}(1 + (-1)^j) \tag{3.13}$$

$$\int_0^1 \sqrt{2} \sin j\pi x P_2(x) \, dx = \frac{\sqrt{10}(j^2\pi^2 - 12)}{j^3\pi^3}(1 - (-1)^j) \tag{3.14}$$

$$\int_0^1 \sqrt{2} \sin j\pi x P_3(x) \, dx = -\frac{\sqrt{14}(j^2\pi^2 - 60)}{j^3\pi^3}(1 + (-1)^j). \tag{3.15}$$

Proof. It is well known that Legendre polynomials are symmetric or antisymmetric about the midpoint of the defining interval, i.e.,

$$P_k(x) = (-1)^k P_k(1-x), \quad k \in \mathbb{Z}_+.$$

On the other hand, it is easy to check that

$$\sin(j\pi x) = (-1)^{j+1} \sin(j\pi(1-x)), \quad j \in \mathbb{N},$$

i.e., $\sin(j\pi x)$ is also symmetric or antisymmetric on the interval $[0, 1]$. Therefore, we see, when k and j are both even or both odd, P_k and $\sin(j\pi\,\cdot)$ have opposite symmetry, hence Eq. (3.11) holds.

While when k and j have opposite parity, P_k and $\sin(j\pi\,\cdot)$ are either both symmetric or both antisymmetric about $x = \frac{1}{2}$, hence the integral $\int_0^1 P_k(x) \sin j\pi x \, dx$ is nonzero.

Eqs. (3.12)–(3.15) are results of direct computation. □

The 'checkerboard' pattern of alternating zeros of $\boldsymbol{B}$ allows us to say something about the inverse of $\boldsymbol{G}$.

Theorem 3.2. *The inverse of the Gram matrix $\boldsymbol{G}$ as in (3.10) has the same pattern of alternating zeros as $\boldsymbol{G}$. That is, $\boldsymbol{G}$ and $\boldsymbol{G}^{-1}$ have the same 'checkerboard' pattern.*

Proof. Let $\boldsymbol{P}$ be a permutation matrix. Hence $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$ and the transformation $\boldsymbol{A} \to \boldsymbol{PAP}^T$ permutes the rows and columns of $\boldsymbol{A}$ in the same way. Because the pattern of alternating zeros of $\boldsymbol{G}$, there exists a permutation matrix $\boldsymbol{P}$ of the same order of $\boldsymbol{G}$ such that

$$\boldsymbol{PGP}^T = \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2 \end{bmatrix},$$

where if p is odd, then $\boldsymbol{G}_1$ consists of the entries in odd rows and odd columns of $\boldsymbol{G}$, and $\boldsymbol{G}_2$ consists of the entries in even rows and even columns of $\boldsymbol{G}$; If p is even, then $\boldsymbol{G}_1$ consists of the entries in rows and columns of $\boldsymbol{G}$ with indices $\{1, 3, \ldots, p+1, p+2, p+4, \ldots, p+q+\frac{1}{2}(1-(-1)^q)\}$, and $\boldsymbol{G}_2$ consists of the entries in rows and columns of $\boldsymbol{G}$ with indices $\{2, 4, \ldots, p, p+3, p+5, \ldots, p+q+\frac{1}{2}(1+(-1)^q)\}$. Taking the inverse of both sides in the above equation produces

$$\boldsymbol{PG}^{-1}\boldsymbol{P}^T = \begin{bmatrix} \boldsymbol{G}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{-1} \end{bmatrix}.$$

Therefore,

$$\boldsymbol{G}^{-1} = \boldsymbol{P}^T \begin{bmatrix} \boldsymbol{G}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{-1} \end{bmatrix} \boldsymbol{P}.$$

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

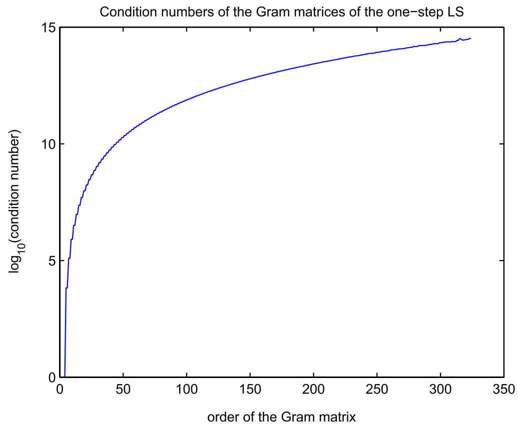

Fig. 3.1. Condition numbers of the Gram matrices of the one-step least squares approximation with $p = 3$.

Noting $\boldsymbol{P}$ and $\boldsymbol{P}^T$ are inverses of each other, the above equation precisely implies that the matrix $\boldsymbol{G}^{-1}$ is obtained by permuting reversely the rows and columns of the matrix

$$\begin{bmatrix} \boldsymbol{G}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2^{-1} \end{bmatrix}.$$

This immediately implies $\boldsymbol{G}^{-1}$ has the same pattern of alternating zeros as $\boldsymbol{G}$. □

From the above proof, we see that $\boldsymbol{G}$ after permuting the rows and columns, becomes a block diagonal matrix. Thus the linear system $\boldsymbol{Gc} = \boldsymbol{d}$ can be replaced by two smaller sub-systems of about half of the size of the original system, and the computing speed is improved.

Since $\boldsymbol{G}$ is symmetric, positive definite and the two matrices $\boldsymbol{G}$ and $\begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2 \end{bmatrix}$ have the same set of eigenvalues, we readily conclude the following two corollaries.

Corollary 3.3. *The matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are both symmetric and positive definite.*

Corollary 3.4. *The condition number of $\boldsymbol{G}$ equals to the condition number of*

$$\begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2 \end{bmatrix}.$$

In Fig. 3.1, the condition numbers for the Gram matrices by one-step least square approximation is drawn as a function of the orders of the Gram matrices. Recall the order of the Gram matrix is $n := p + 1 + q$. The value of p is chosen as 3, i.e., the highest degree of the Legendre polynomial is chosen to be 3.

From Fig. 3.1, we see that the condition number of the Gram matrix increases approximately exponentially. This indicates that the Gram matrix for the one-step least squares approximation is ill-conditioned.

4. Experiments

In this section, through experiments we compare the one-step least squares approximation, the two-step least squares approximation, the near least squares approximation and the PHLST approximation. For simplicity, we will only show computation results by using a combination of algebraic polynomials and sine polynomials on the interval [0, 1]. The results of using complex trigonometric polynomials and using cosine polynomials are similar. To check the validity of our proposed methods, we will first look at discrete signal without noise, then we will look at various discrete data corrupted by white Gaussian noise.

When the one-step least squares approximation is used, the linear system (3.7) is solved by the regularized SVD as in (3.9) with the truncation threshold $\epsilon = 10^{-4}$. For other approximation methods, the linear system in each step is just an identity matrix of appropriate order. Therefore the solution of the linear system in each step just equals to the vector on the right-hand side in the system at each step.

Let the uncorrupted discrete data be denoted by $\boldsymbol{f}_0[k]$, $k \in \mathbb{Z}_{N+1}$. We then corrupt the data $\boldsymbol{f}_0[k]$ by additive white Gaussian noise $\boldsymbol{w}[k]$ with a specific SNR, where the SNR is the power ratio of signal to noise per sample in decibel. Recall for white Gaussian noise, its power equals to its variance due to its zero mean. Therefore for an un-noisy discrete signal $(s_i)_{i\in\mathbb{Z}_K}$ of length K, and if the variance of the Gaussian noise is σ_w^2, then the SNR is defined as

$$\mathrm{SNR} := 10\log_{10}\frac{\frac{1}{K}\sum_{i\in\mathbb{Z}_K} s_i^2}{\sigma_w^2}.$$

The raw noisy data is

$$\boldsymbol{f}[k] = \boldsymbol{f}_0[k] + \boldsymbol{w}[k].$$

The error is calculated by comparing the approximant $\hat{\boldsymbol{f}}[k]$ to the uncorrupted data $\boldsymbol{f}_0[k]$. Specifically, we define the error vector by

$$\boldsymbol{e}[k] := \boldsymbol{f}_0[k] - \hat{\boldsymbol{f}}[k].$$

The l^2 error and l^∞ error of the approximation is calculated by $\|\boldsymbol{e}\|_{l^2}$ and $\|\boldsymbol{e}\|_{l^\infty}$ respectively. The relative error is then defined as $\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{f}_0\|}$ for each norm.

For the near least squares approximation, we simply chose the following cubic polynomial

$$\varphi(x) = (\boldsymbol{f}[N] - \boldsymbol{f}[0])x + \boldsymbol{f}[0] + x(1-x)^2.$$

Namely φ is obtained by the line connecting the two endpoints $(0, \boldsymbol{f}[0])$ and $(1, \boldsymbol{f}[N])$ plus a cubic polynomial with a simple zero at $x = 0$ and a double zero at $x = 1$. We remark that any polynomial of degree 3 that interpolates the signal at the end points serves the purpose here.

For the PHLST approximation, theoretically, in addition to interpolating the function values at the boundary, only values of second derivative are needed to be matched at the boundary. However, for simplicity, as well as the fact as pointed out in [5] that, the values of boundary derivatives of higher order tend to be chaotically huge if they are estimated by higher order polynomial fit, we instead find the Hermite interpolating polynomial interpolating the function values and the first derivative on the boundary. The values of the first derivative at 0 and 1 are approximated by the *first order difference quotient* using the first two data points and the last two data points respectively.

We shall choose the shifted and normalized Legendre polynomials of degrees at most 3 on the interval [0, 1] as the basis for the subspace of the polynomials, namely, $p = 3$. Specifically

$$\begin{aligned}
P_0(x) &= 1;\\
P_1(x) &= \sqrt{3}(2x-1);\\
P_2(x) &= \sqrt{5}(6x^2-6x+1);\\
P_3(x) &= \sqrt{7}(20x^3-30x^2+12x-1).
\end{aligned}$$

The nonzero entries of the submatrix $\boldsymbol{B}$ of the Gram matrix $\boldsymbol{G}$ can be obtained directly by Eqs. (3.12)–(3.15). All the inner products to calculate the vector $\boldsymbol{d}$ in (3.7) are carried out through the standard quadrature rule: if the number of data points is odd, using Simpson's rule; while the number of data points is even, using the trapezoidal rule. Although the trapezoidal rule is generally only of second order, it is highly accurate for periodic functions [10,11]. Moreover, using the trapezoidal rule to estimate the Fourier series coefficients leads to DFT, DST, and DCT [12,11]. Therefore when the signal is given in the form of equally spaced discrete data, the solution vector for the sine polynomials obtained for the two-step least squares approximant, the near least squares approximant or the PHLST approximant is equivalent to the DST of the given data set.

Example 4.1 (*Uncorrupted Data*). In this experiment, to check the validity of our proposed methods, we consider the uncorrupted data $\boldsymbol{f}_0[k]$ generated by the signal

$$f(x) = \sqrt{x^5} + \sin(100x^2) \tag{4.16}$$

on the interval [0, 1] with length 1025.

In Fig. 4.2, the left curves depict the relative l^2 error in log scale vs. the order of the Gram matrix for the various approximations. The order of the Gram matrix is incremented by 1, and ranges from 4 to 320 (The same choice of orders of the Gram matrix is used in the subsequent computations). We can see that in terms of the l^2 error, as expected, the two-step least squares approximation is much worse than the other three methods due to the much slower decay rate of the Fourier coefficients. The one-step least squares approximation offers the least l^2 error within the accuracy of computation. However, both the near least squares approximation and the PHLST are almost as good as the one-step least squares approximation in terms of l^2 error. In fact, in this experiment, the PHLST approximation offers smaller l^2 error when $n \geq 235$ than that of one-step least squares approximation; while the near least squares approximation offers smaller l^2 error than that of the one-step least squares approximation when $n \geq 253$. We think this is because the one-step least squares approximation has an ill-conditioned Gram matrix, hence the computation is not reliable when the order of the Gram matrix is relatively

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

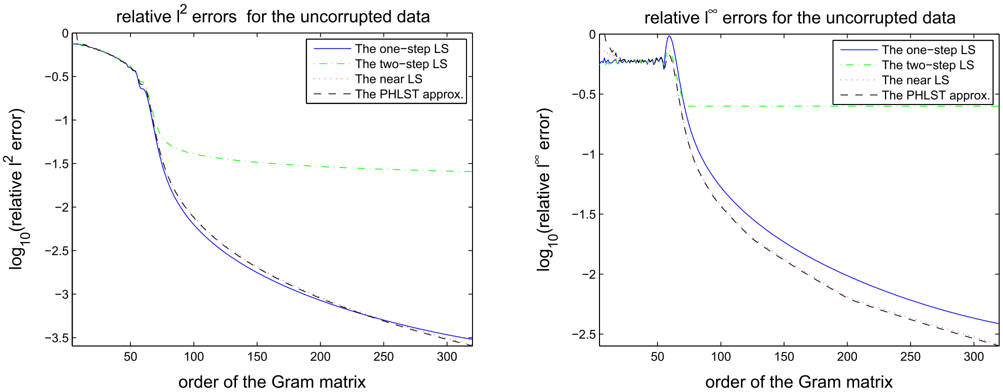

Fig. 4.2. $\log_{10}\frac{\|\boldsymbol{e}\|_{l^2}}{\|\boldsymbol{f}_0\|_{l^2}}$ and $\log_{10}\frac{\|\boldsymbol{e}\|_{l^\infty}}{\|\boldsymbol{f}_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the uncorrupted data.

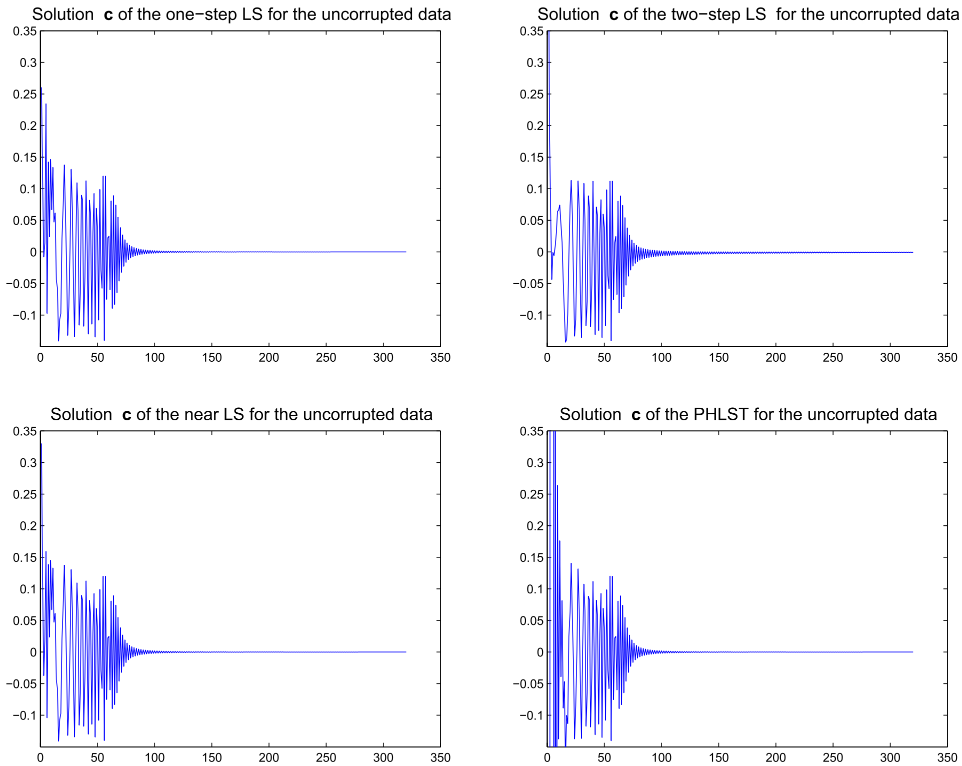

Fig. 4.3. Solution vector $\boldsymbol{c}$ when $n = 320$ for the uncorrupted data.

large. Moreover, the PHLST, although nearly performs the same as the near least squares approximation, consistently offers slightly smaller l^2 error than that of the latter. This is because its Fourier coefficients decay faster.

In Fig. 4.2, the plots on the right depict the relative l^∞ error in log scale vs. the order of the Gram matrix. We see the l^∞ error does not behave quite consistently with the l^2 error. Again the two-step least squares approximant performs the worst owing to its non-uniform convergence. Note the l^∞ error using the two-step least squares approximation keeps the same level for $n \gtrsim 70$, confirming Corollary 2.18, while the l^∞ error of using the other three approximation methods steadily decreases as n increases owing to their uniform convergence. Again, like the l^2 error, the near least squares approximation behaves almost as good as the PHLST, although the latter offers slightly smaller l^∞ error.

We next plot the solution vector $\boldsymbol{c}$ with $n = 320$ for the same uncorrupted data set by using each of the four approximation methods.

From Fig. 4.3, we see the absolute values of c_i, $i > 80$ are much smaller compared to the values of c_i, $1 \le i \le 80$. For example, the ratio of the mean of the absolute values of c_i, $1 \le i \le 80$ to that of c_i, $81 \le i \le 320$ is about 41 for the two-step least squares approximation. This ratio is about 127, 211, 1187 for the one-step least squares approximation, the near least squares approximation, and the PHLST approximation respectively. The values of the ratios confirm our theorems

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

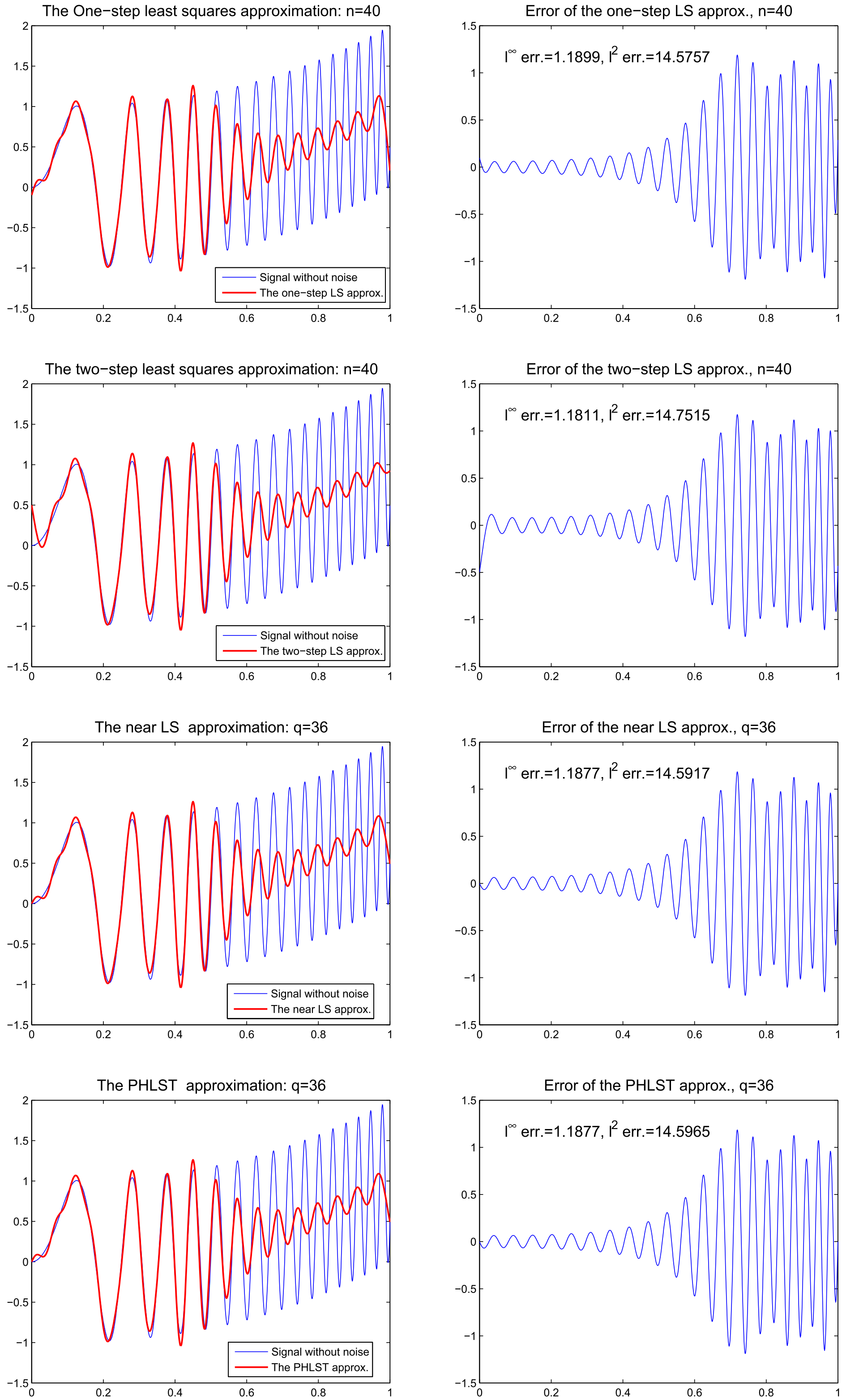

Fig. 4.4. Approximation results, $n = 40$ ($q = 36$) for uncorrupted data.

about the decay rate of the Fourier coefficients. However, the ratio for the one-step least squares approximant should be smaller than or compatible with that of the PHLST. We think this is due to the ill-conditioned linear system which produces

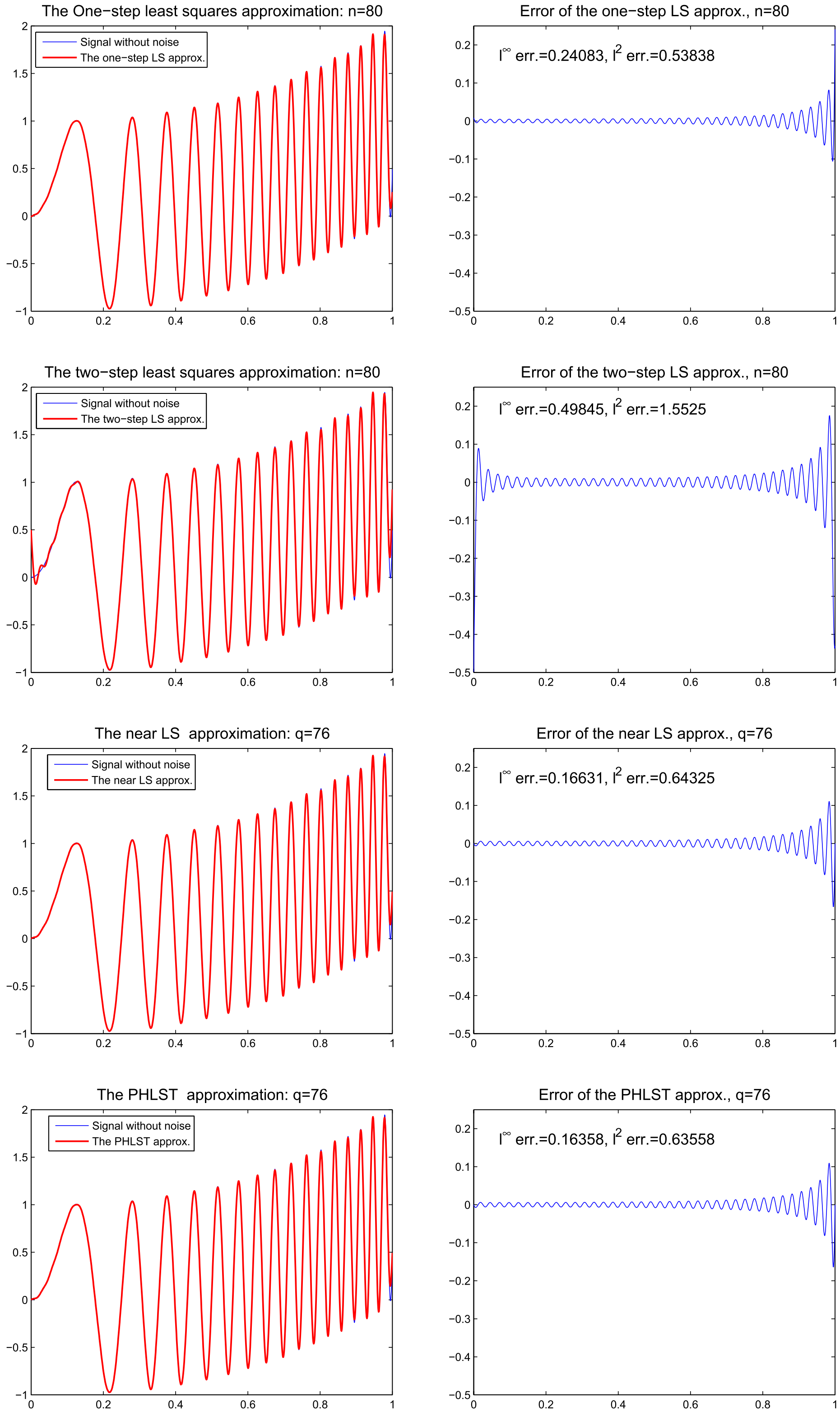

Fig. 4.5. The approximation results, $n = 80$ ($q = 76$) for uncorrupted data.

the solution vector with relatively larger numerical error. The ratio is much larger for the PHLST approximation because the first four coefficients for the corresponding Legendre polynomial basis functions are relatively large due to estimating the

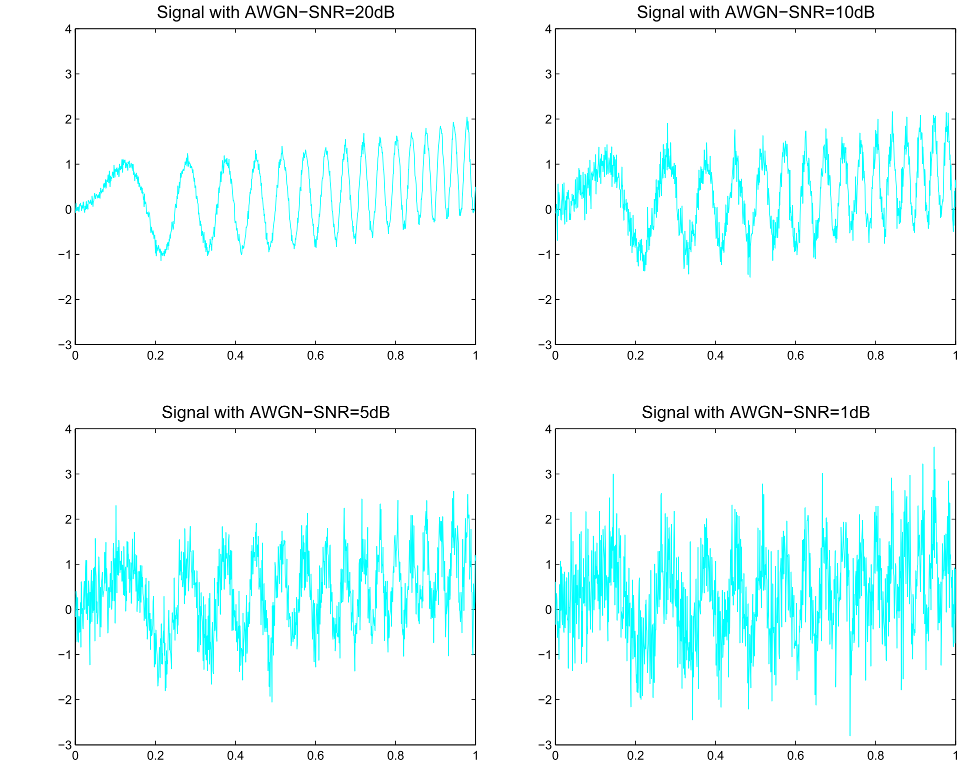

Fig. 4.6. The noisy data with various SNRs.

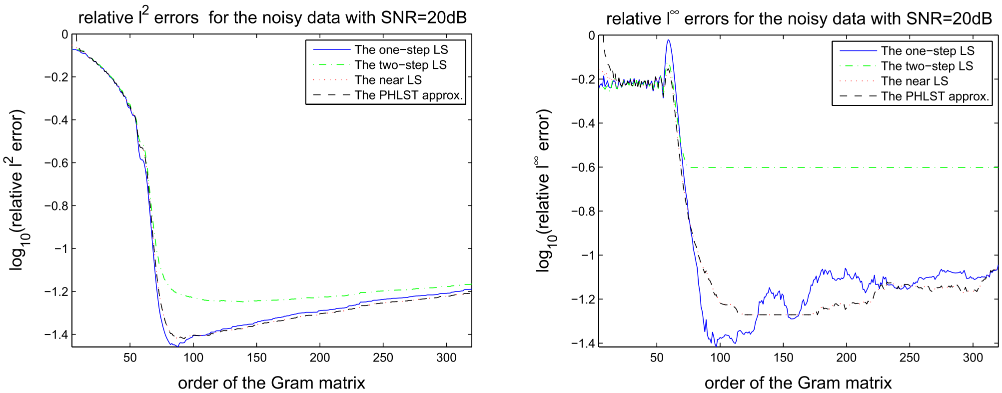

Fig. 4.7. $\log_{10} \frac{\|\boldsymbol{e}\|_{l^2}}{\|\boldsymbol{f}_0\|_{l^2}}$ and $\log_{10} \frac{\|\boldsymbol{e}\|_{l^\infty}}{\|\boldsymbol{f}_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the noisy data with SNR $= 20$ dB.

values of the first derivative of the function f from the given data set. But even the ratio of the mean of the absolute values of c_i, $5 \le i \le 80$ to that of c_i, $81 \le i \le 320$ is about 626. This suggests when we form the approximation, choosing the value of q around 75 (so n is about 80) should be sufficient. This phenomenon is named as *energy compaction* in [13, pp. 596–598] when they discussed DCT-2 (type 2). However, they did not explain why this phenomenon happens. We give an explanation through Proposition 5.1, Corollaries 5.2–5.3 in the next section.

Next we specifically look at the approximating results with two different values of q. The values of q are chosen as $q = 36$ and $q = 76$, hence the orders of the matrices are $n = 4 + 36 = 40$ and $n = 4 + 76 = 80$ respectively.

From Fig. 4.5, it is clear that the two-step least squares approximant is much worse than the other approximants in terms of both the l^∞ norm and in the l^2 norm. The one-step least squares approximation offers the least l^2 error, while both the near least squares approximation and the PHLST approximation offer the smaller l^∞ error. The PHLST approximation offers slightly smaller l^2 error and l^∞ error than those of the near least squares approximation.

From Figs. 4.4 and 4.5, we can see clearly that as q increases, namely, as more and more sine polynomials are used, the approximation can *progressively* capture more and more oscillations in the data.

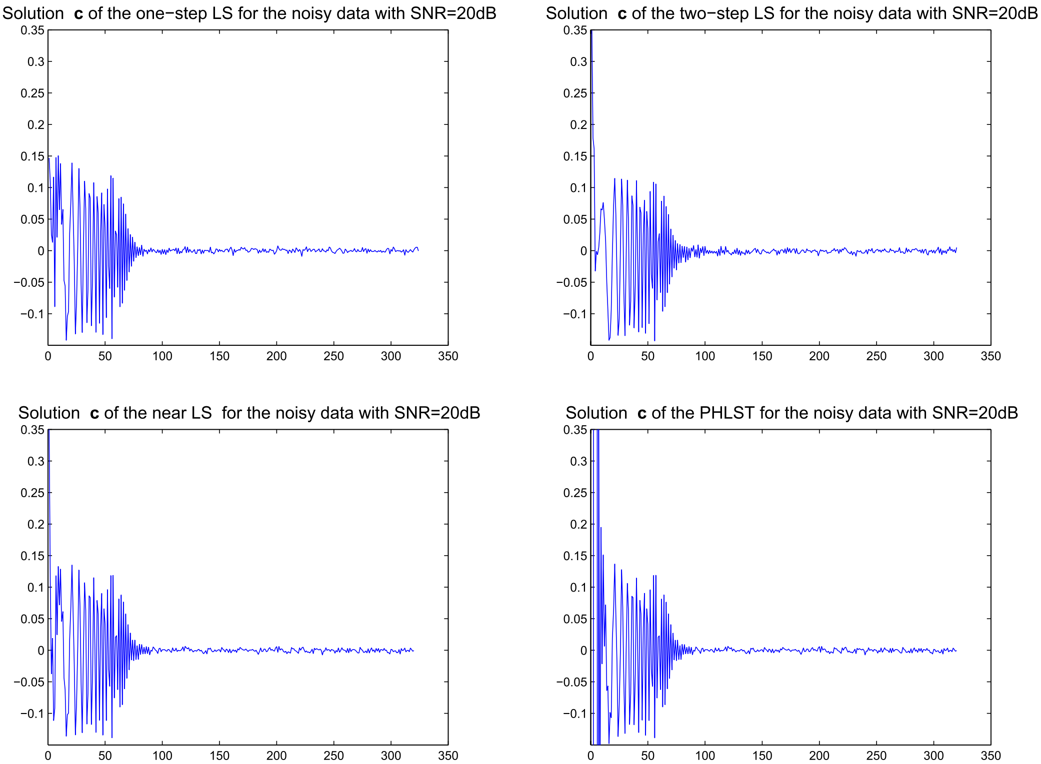

Fig. 4.8. Solution vector $\boldsymbol{c}$ when $n = 320$ for the noisy data with SNR = 20 dB.

From Figs. 4.4 and 4.5, we also can see the two-step least squares approximation has Gibbs phenomenon occurring at the two end points, while the other three approximations offer no error at the left boundary, and slightly larger error near the right boundary (The near least squares approximation and the PHLST approximation actually offer no error at the right boundary). This is because our chosen signal has larger instantaneous frequency (in the sense of Cohen [14, p. 15]) near the right boundary, hence needing more pure harmonics of higher frequencies to approximate the signal near the right boundary.

Example 4.2 (*Corrupted Data*). In this experiment, we shall consider discrete data corrupted by additive Gaussian noise. For convenience, we still use Eq. (4.16) to generate 1025 data points on the interval [0, 1]. We then corrupted the data by white Gaussian noise with SNR = 20 dB. The noisy signals with various SNRs are shown in Fig. 4.6.

In Fig. 4.7, the left curves depict the relative l^2 error vs. the order of the Gram matrix for all approximation methods, and the right curves depict the relative l^∞ error vs. the order of the Gram matrix for all approximation methods. Again similar behaviors of l^2 error and l^∞ error are observed for the four methods as for uncorrupted data.

Compare Figs. 4.2 and 4.7: the l^2 error of all methods for uncorrupted data does decrease as the degree of sine polynomials increases, as can be seen in Fig. 4.2, but the l^2 error of all methods for corrupted data attains a minimum about at $n = 80$. As we recall that, except the one-step least squares approximation, all other three approximations are well-conditioned. This implies *the fact that an optimum number of sine polynomials is needed for noisy data has nothing to do with the condition number of the approximation method*.

Next, we plot the solution vector $\boldsymbol{c}$ when $n = 320$ for the same noisy data corrupted by white Gaussian with SNR = 20 dB in Fig. 4.8. From this figure, we see the absolute values of c_i, $i > 80$ are much smaller compared to the values of c_i, $1 \le i \le 80$. As before, we computed the ratio of the mean of the absolute values of c_i, $1 \le i \le 80$ to that of c_i, $81 \le i \le 320$. This ratio for the current corrupted data with SNR = 20 dB is about 26, 34, 36, 934 for the two-step least squares approximation, the one-step least squares approximation, the near least squares approximation, and the PHLST approximation respectively.

Note the behavior of the solution vector for noisy data essentially remains the same as that for uncorrupted data. Namely in either case, about the first $n = 80$ entries in the solution vector $\boldsymbol{c}$ are significantly greater than the remaining entries. Moreover, this is the same optimum n for which the l^2 error attains minimum when noise is present in the data. This make us conjecture that *the optimum number of sine polynomials is determined by the intrinsic frequency of the data rather than the regularity of the data*. This conjecture is further corroborated by Examples 4.3 and 5.4.

We next specifically look at the approximating results with two different values of q. The values of q are chosen as $q = 36$ and $q = 76$, hence the orders of the matrix are $n = 4 + 36 = 40$ and $n = 4 + 76 = 80$ respectively.

From Figs. 4.9 and 4.10, we can see clearly that as q increases, namely, as the degree of the sine polynomial gets higher, the approximation can capture more and more oscillations in the data.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

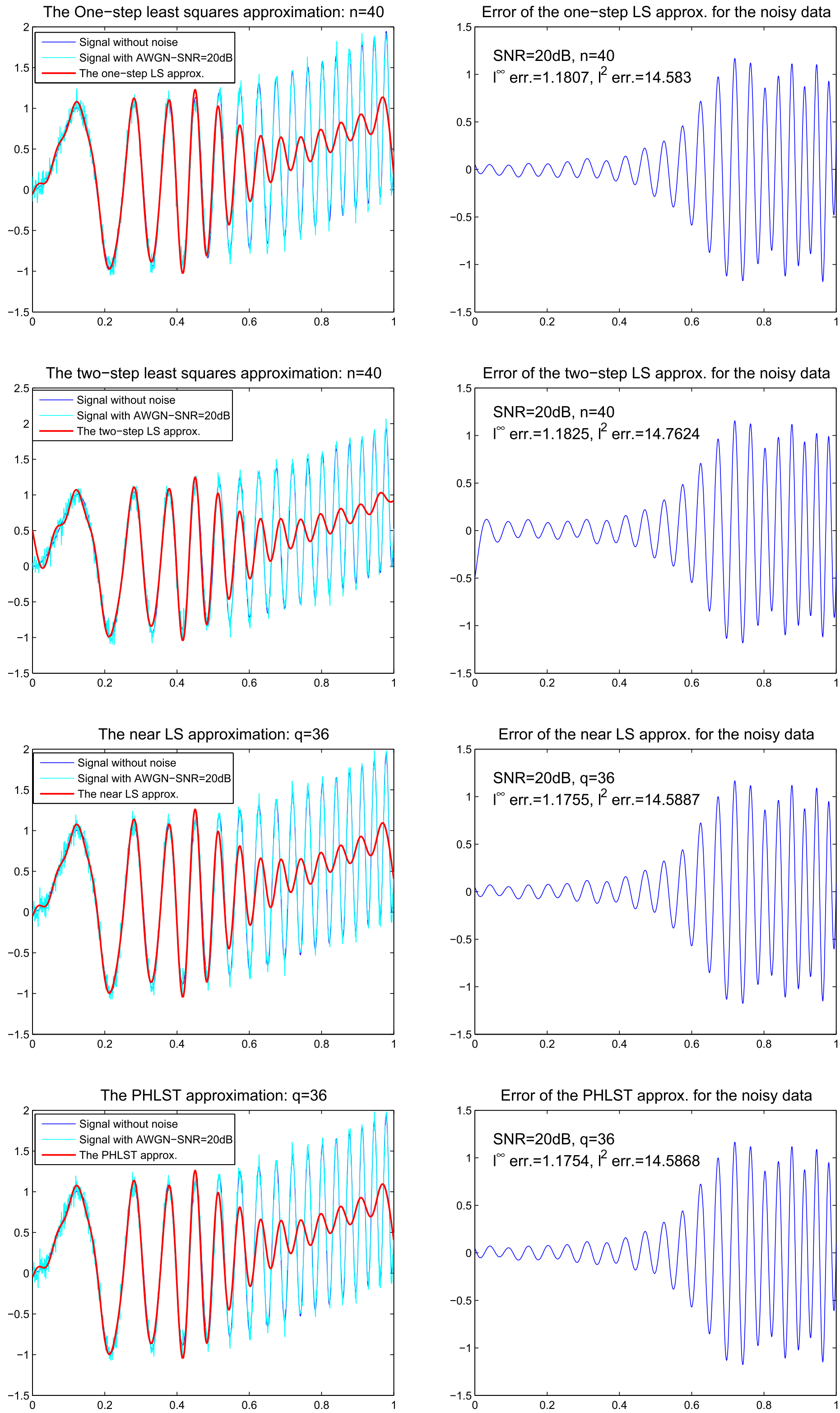

Fig. 4.9. Approximation results, $n = 40$ ($q = 36$) for the noisy data with SNR $= 20$ dB.

From Fig. 4.10, it is clear that the two-step least squares approximant is worse than the other three approximants in terms of both the l^∞ norm and the l^2 norm. Again, the near least squares approximant and the PHLST approximant are nearly as good as the one-step least squares approximant.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

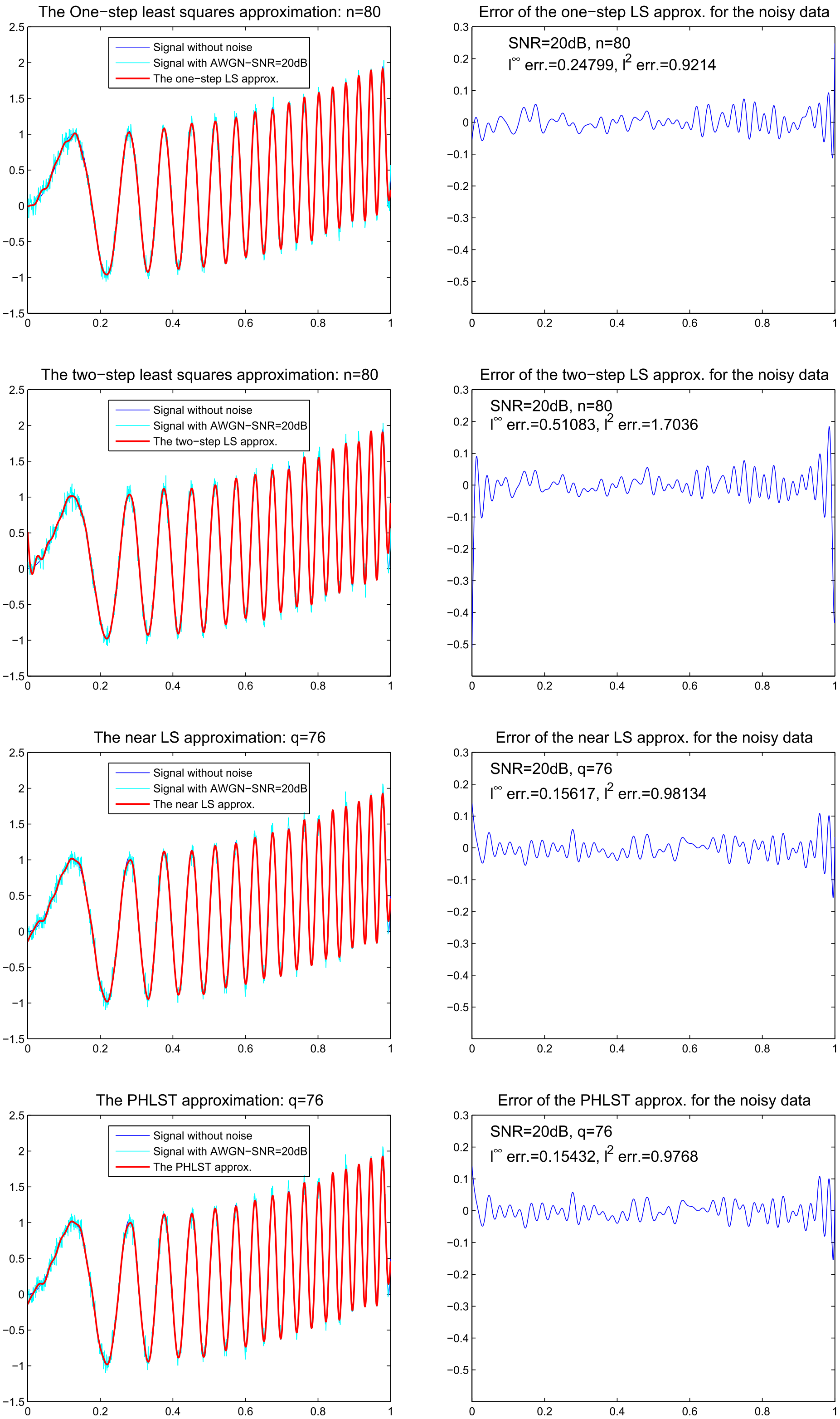

Fig. 4.10. Approximation results, $n = 80$ ($q = 76$) for the noisy data with SNR = 20 dB.

Example 4.3 (*A Different Data Set*). In this experiment, we perform our algorithms on a different data set. We use a different function

$$f(x) = \sqrt{x^5} + \sin(200x^2) \tag{4.17}$$

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

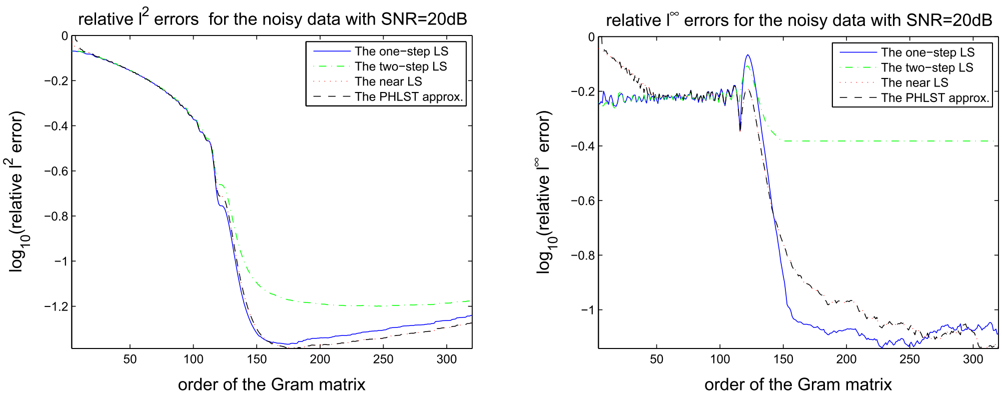

Fig. 4.11. $\log_{10} \frac{\|\boldsymbol{e}\|_{l^2}}{\|\boldsymbol{f}_0\|_{l^2}}$ and $\log_{10} \frac{\|\boldsymbol{e}\|_{l^\infty}}{\|\boldsymbol{f}_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the 2nd noisy data set.

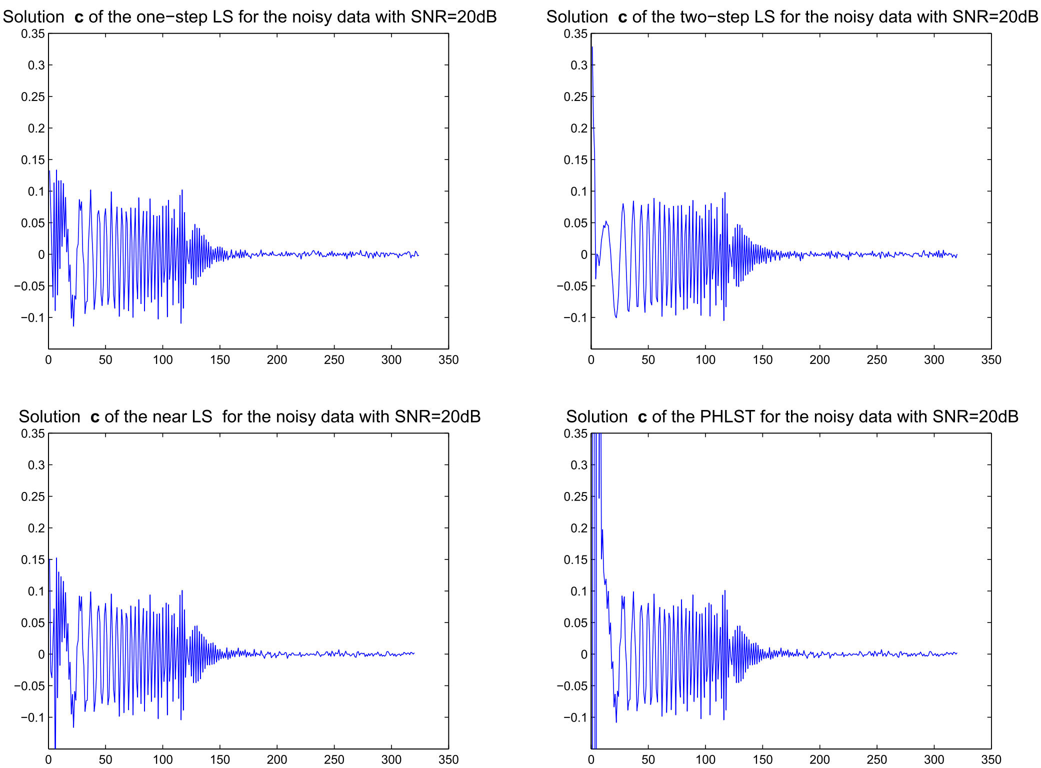

Fig. 4.12. Solution vector $\boldsymbol{c}$ when $n = 320$ and SNR $= 20$ for the 2nd noisy data set.

to generate 1025 data points on the interval [0, 1]. We then corrupted the data by white Gaussian noise with SNR = 20 dB.

We record the results in Figs. 4.11 and 4.12. Compared to the results in Figs. 4.7 and 4.8, one can immediately see that for the new data set in this experiment, it is seen that *the optimum degree of the sine polynomial is about* 150, *as opposed to the first data set in the previous experiments, the optimum degree of the sine polynomial is about* 75. We notice for the 2nd data set, its *instantaneous frequency* (in the sense of Cohen [14, p. 15]) is doubled.

5. Effects of noise on the approximation

From previous section, we observed that an optimum degree of sine polynomials needs to be estimated for noisy data in order to obtain the best approximation result in terms of the l^2 error. This optimum degree has nothing to do with the condition number of the approximation method, nor to do with the regularity of the signal. After some reflection, we now answer this question. As we conjectured, this is due to the intrinsic (instantaneous) frequency contents of the signal. We give an affirmative answer in the following.

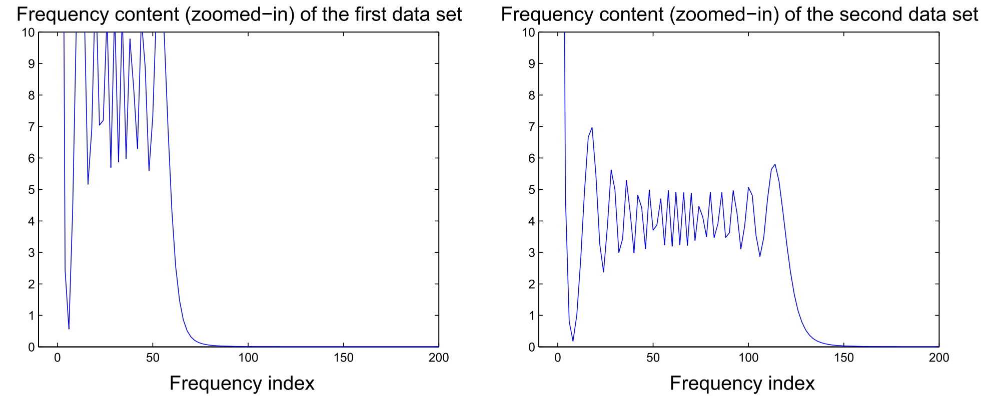

Fig. 5.13. DFT of the two uncorrupted data sets. Left: 1st data set; Right: 2nd data set.

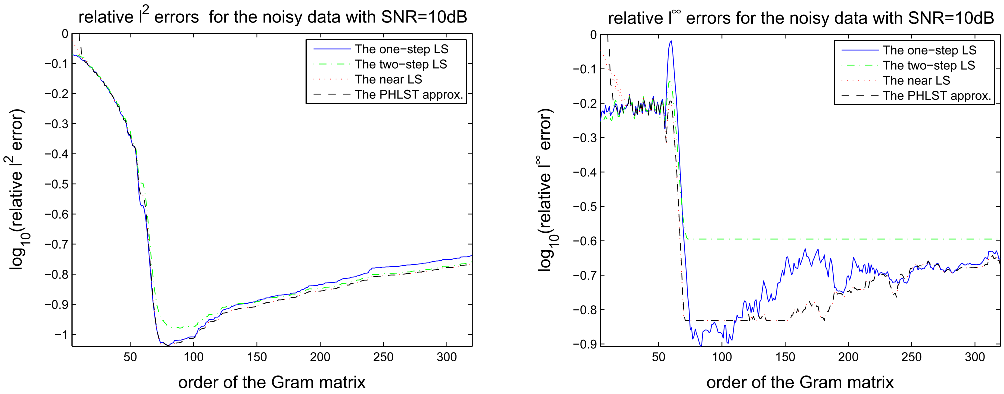

Fig. 5.14. $\log_{10}\frac{\|e\|_{l^2}}{\|f_0\|_{l^2}}$ and $\log_{10}\frac{\|e\|_{l^\infty}}{\|f_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the noisy data with SNR = 10 dB.

Proposition 5.1. *Given a signal f in $L^2([-1,1])$ and assuming that its periodized extension has only finite frequency components, i.e., $f = \sum_{-N\le j\le N} c_j e^{(ij\pi\cdot)}$. Let $\tilde{f}$ be the corrupted signal with white Gaussian noise w, i.e., $\tilde{f} = f + w$. Assume that $\tilde{f} = \sum_{j\in\mathbb{Z}} \tilde{c}_j e^{(ij\pi\cdot)}$, then the L^2 error $\|\tilde{f} - f\|$ is minimized when $\tilde{f}$ is truncated as a trigonometric polynomial of degree N.*

Proof. The proof is done by using the Parseval's identity. Recall the Parseval's identity for the Fourier series of a signal $s \in L^2([-1,1])$ reads as:

$$\frac{1}{2}\|s\|^2 = \frac{1}{2}\int_{-1}^{1} |s(t)|^2\,dt = \sum_{j\in\mathbb{Z}} |d_j|^2,$$

where the sequence $(d_j)_{j\in\mathbb{Z}}$ is the coefficient sequence of the Fourier series of s. Noting also the linearity of Fourier series, we have

$$\frac{1}{2}\|\tilde{f} - f\|^2 = \sum_{|j|\le N} |\tilde{c}_j - c_j|^2 + \sum_{|j|>N} |\tilde{c}_j|^2.$$

Clearly, if the second sum on the right-handed side of the above equation vanishes, $\|\tilde{f} - f\|$ attains its minimum. That is $\|\tilde{f} - f\|$ attains its minimum when $\tilde{f}$ is truncated as a trigonometric polynomial with the same degree N of the signal f. □

Above proposition indicates that when a time-limited signal is '*bandlimited*' in the sense its Fourier series contains only finite frequency components, with the presence of noise, the optimum degree of the approximating trigonometric polynomial equals to the highest degree of the frequency component in the signal. Additional harmonic waves with higher frequency only adds information of the noise.

Since the sine series of a function on [0, 1] is the Fourier series of its odd extension to $[-1, 1]$, and the cosine series of a function on [0, 1] is the Fourier series of its even extension to $[-1, 1]$, we immediately obtain the following two corollaries.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

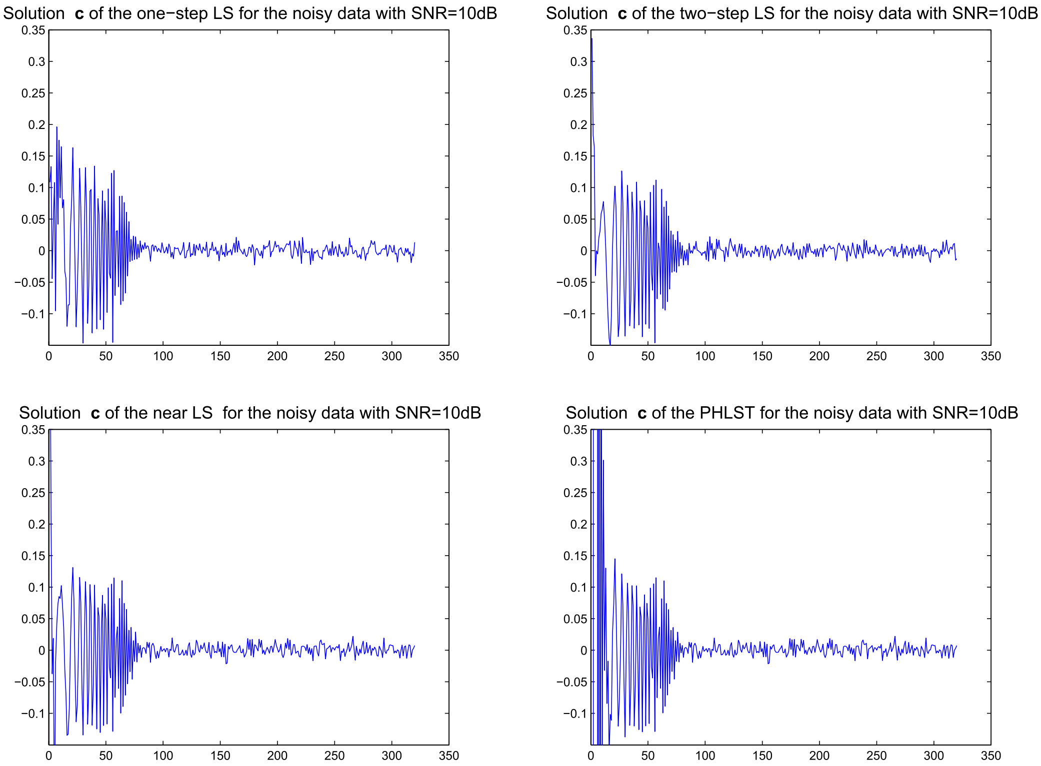

Fig. 5.15. Solution vector $\boldsymbol{c}$ when $n = 320$ for the noisy data with SNR $= 10$ dB.

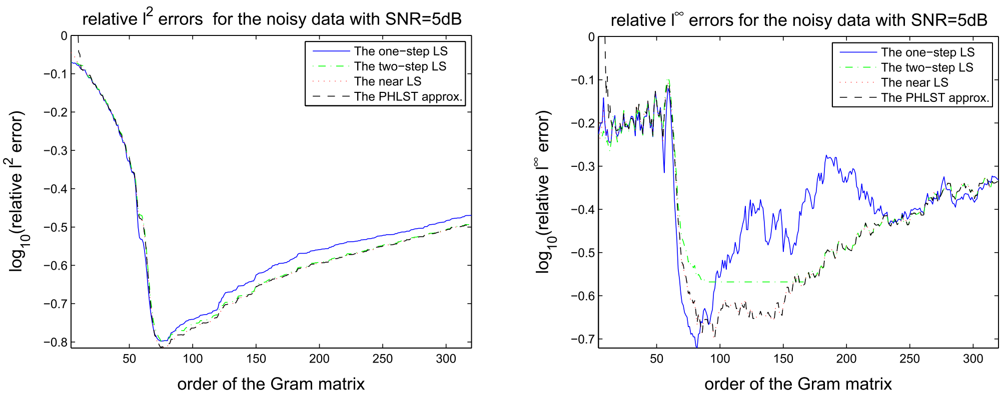

Fig. 5.16. $\log_{10} \frac{\|\boldsymbol{e}\|_{l^2}}{\|\boldsymbol{f}_0\|_{l^2}}$ and $\log_{10} \frac{\|\boldsymbol{e}\|_{l^\infty}}{\|\boldsymbol{f}_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the noisy data with SNR $= 5$ dB.

Corollary 5.2. *Given a signal f in $L^2([0, 1])$ and assuming that its periodized odd extension has only finite frequency components, i.e., $f = \sum_{j \in \mathbb{N}_N} c_j \sin(j\pi \cdot)$. Let $\tilde{f}$ be the corrupted signal with white Gaussian noise w, i.e., $\tilde{f} = f + w$. Assume that $\tilde{f} = \sum_{j \in \mathbb{N}} \tilde{c}_j \sin(j\pi \cdot)$, then the L^2 error $\|\tilde{f} - f\|$ is minimized when $\tilde{f}$ is truncated as a sine polynomial of degree N.*

Corollary 5.3. *Given a signal f in $L^2([0, 1])$ and assuming that its periodized even extension has only finite frequency components, i.e., $f = \sum_{j \in \mathbb{N}_N} c_j \cos(j\pi \cdot)$. Let $\tilde{f}$ be the corrupted signal with white Gaussian noise w, i.e., $\tilde{f} = f + w$. Assume that $\tilde{f} = \sum_{j \in \mathbb{N}} \tilde{c}_j \cos(j\pi \cdot)$, then the L^2 error $\|\tilde{f} - f\|$ is minimized when $\tilde{f}$ is truncated as a cosine polynomial of degree N.*

Now we are ready to explain an optimum degree of sine polynomials exists in our previous experiments. Perform DFT on the two data sets generated by Eqs. (4.16) and (4.17) respectively. Note one should choose at least $2N + 1$ points to conduct DFT for a trigonometric polynomial of degree N to minimize the error of DFT [12]. The magnitude of DFT are shown in Fig. 5.13. It clearly shows that in the first data set generated by Eq. (4.16), the frequency component with index greater than 75 is almost negligible. This is the value of q of the optimal degree of sine polynomials used in the approximation (hence

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

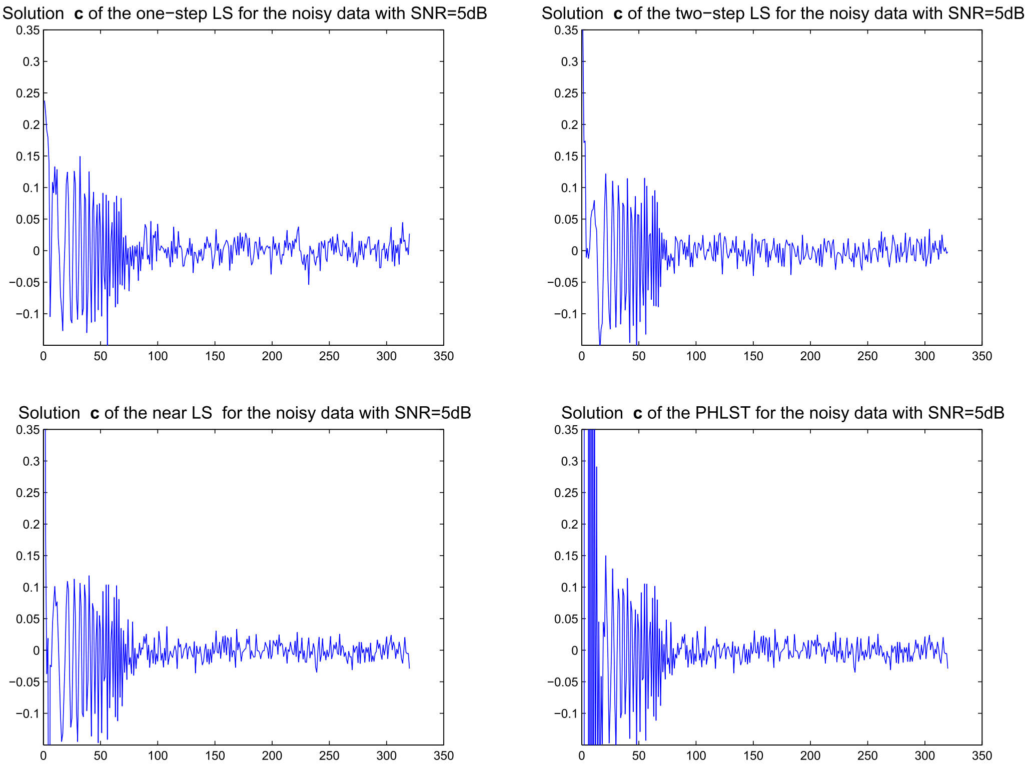

Fig. 5.17. Solution vector $\boldsymbol{c}$ when $n = 320$ for the noisy data with SNR $= 5$ dB.

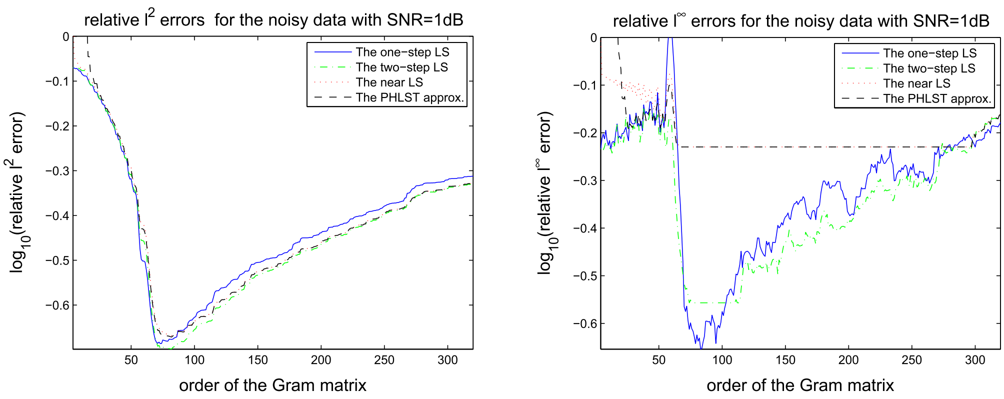

Fig. 5.18. $\log_{10} \frac{\|\boldsymbol{e}\|_{l^2}}{\|f_0\|_{l^2}}$ and $\log_{10} \frac{\|\boldsymbol{e}\|_{l^\infty}}{\|f_0\|_{l^\infty}}$ vs. the order of the Gram matrix for the noisy data with SNR $= 1$ dB.

$n = 4 + 75 = 79$). Likewise, Fig. 5.13 shows the second data set generated by Eq. (4.17) contains frequency components with frequency index approximately up to 150. This is the value of q of the optimal degree of sine polynomials used in the approximation (hence $n = 4 + 150 = 154$). Those values of q for either data set agree with our earlier observations in Examples 4.2 and 4.3.

We next see if the level of SNR of noise present in the data affects the optimal degree of sine polynomials in the approximation. The answer is clearly no.

Example 5.4 (*Corrupted Data with various SNR*). In this experiment, we perform the four approximation methods on the same data set generated by Eq. (4.16) with white Gaussian noises with various SNR (see Fig. 4.6). One can see, that similar results are obtained. Figs. 5.14–5.19 show our results for SNR $= 10$ dB, 5 dB, and 1 dB.

Note in all cases, *no matter what the level of SNR of the noise is, the optimum degree n of sine polynomials remains to be the same* $n = 80$, *which is the same for uncorrupted data*. Of course, we know from Corollary 5.2 that the optimum degree of sine polynomials is determined by the signal itself, not the noise.

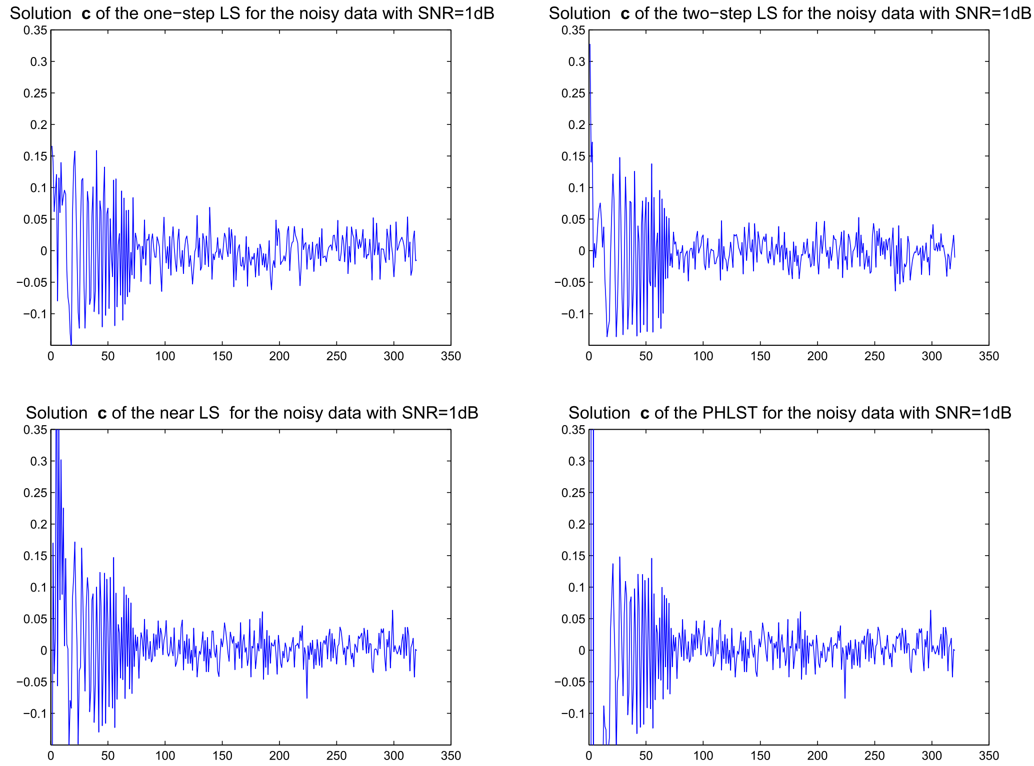

Fig. 5.19. Solution vector $\boldsymbol{c}$ when $n = 320$ for the noisy data with SNR $= 1$ dB.

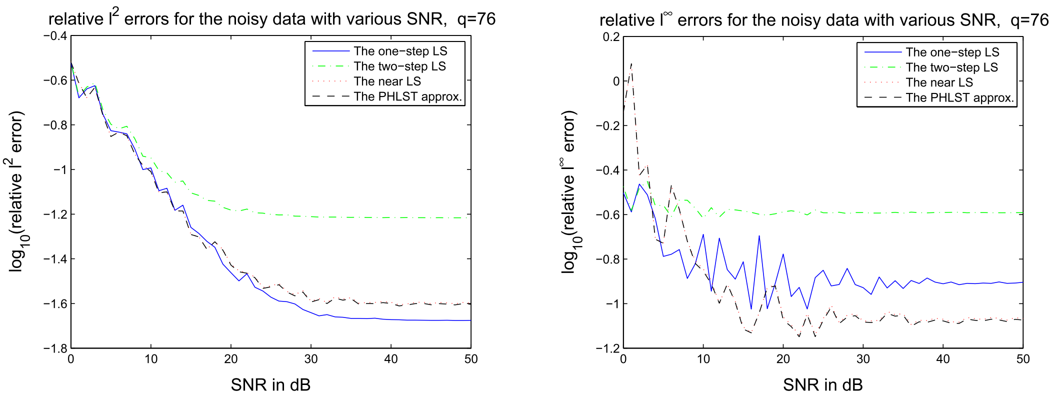

Fig. 5.20. $\log_{10} \frac{\|\boldsymbol{e}\|_{l^2}}{\|\boldsymbol{f}_0\|_{l^2}}$ and $\log_{10} \frac{\|\boldsymbol{e}\|_{l^\infty}}{\|\boldsymbol{f}_0\|_{l^\infty}}$ vs. various SNRs in dB.

Example 5.5 (*Approximations with Sine Polynomials of Fixed Degree for Noisy Data with Various SNR*). We now do experiments for a noisy data set with various SNR values. The degree of the sine polynomial is fixed to be $q = 76$, so the order of the Gram matrix is $n = 4 + 76 = 80$. We still use Eq. (4.16) to generate the uncorrupted data. We compute the corresponding relative l^2 error and relative l^∞ error for the data set corrupted with noise with SNR ranging from 0 to 50, incremented by 1. Note when SNR $= 0$ dB, it corresponds to the ratio of the power of signal to noise is 1 per sample.

From Fig. 5.20, as the SNR level becomes higher, the approximating errors in terms of both l^2 and in l^∞ become smaller, as expected.

On the other hand, by comparing the results of the four approximation methods, we see when the value of SNR of noise is relatively low (corresponding to higher level of noise), such as when SNR $\le$ 10 dB, the results (especially in terms of the l^2 error) of the two-step least squares approximation is nearly as good as other three approximation methods (see Figs. 5.14, 5.16, 5.18 and 5.20). When the value of SNR is very low, such as when SNR $\le$ 5 dB, both the two-step least squares approximation and the one-step least squares approximation also offer smaller l^∞ error than the other two methods, which normally give smaller l^∞ error for uncorrupted data or data with low level of noise. Recall the near least squares approximation and the PHLST approximation all depends on a polynomial passing through the two boundary data points.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039

This suggests that when the noise level present in the data is relatively high, the methods that depend on the point evaluation of the data would be vulnerable. On the other hand, one should rely on the methods that depend on energy (average) of the data for such low SNR cases.

6. Conclusions

From our analysis and experiments, we can conclude that

1. Among the four approximation methods: the one-step least squares approximation, two-step least squares approximation, the near least squares approximation, and the PHLST approximation, the one-step least squares approximation offers the least l^2 error. The near least squares approximation and the PHLST approximation are nearly as good as the one-step least squares approximation. They give the smaller l^∞ error for uncorrupted data and data with low level of noise.
2. An optimum degree of sine (cosine, or trigonometric) polynomials is needed for noisy data in order to obtain the best approximation result in terms of l^2 error. This optimum degree has nothing to do with the condition number of the approximation method, nor to do with the regularity of the signal. It is determined by the intrinsic (instantaneous) frequency of the signal. In practice, we need to estimate the optimum degree of the sine (cosine or trigonometric) polynomial from a given noisy data set. Potentially, we could use some information theoretic criterion such as the Minimum Description Length criterion (MDL) [15,16]. For a preliminary study of such estimation procedure, see [17].
3. When the noise level is relatively high in the data set, all the four approximation methods behave about the same in terms of l^2 error. When the noise level is very high, the two-step least squares approximant and the one-step least squares approximant not only offer small l^2 error, they also offer smaller l^∞ error than the near least squares approximation and the PHLST approximation.

Acknowledgements

This work was partially supported by the ONR Grants N00014-06-1-0615, N00014-07-1-0166, N00014-09-1-0041, N00014-09-1-0318, and the NSF Grant DMS-0410406.

References

[1] N. Saito, J.-F. Remy, The polyharmonic local sine transform: a new tool for local image analysis and synthesis without edge effect, Appl. Comput. Harmon. Anal. 20 (1) (2006) 41–73.
[2] R.R. Coifman, Y. Meyer, Remarques sur l'analyse de Fourier à fenêtre, C. R. Acad. Sci., Paris Sér. I 312 (1991) 259–261.
[3] H.S. Malvar, Lapped transforms for efficient transform/subband coding, IEEE Trans. Acoust. Speech Signal Process. 38 (1990) 969–978.
[4] H.S. Malvar, D.H. Staelin, The LOT: transform coding without blocking effects, IEEE Trans. Acoust. Speech Signal Process. 37 (1989) 553–559.
[5] J. Zhao, N. Saito, Y. Wang, PHLST5: a practical and improved version of polyharmonic local sine transform, J. Math. Imaging Vision 30 (2008) 23–41.
[6] K. Yamatani, N. Saito, Improvement of DCT-based compression algorithms using Poisson's equation, IEEE Trans. Image Process. 15 (12) (2006) 3672–3689.
[7] K. Atkinson, W. Han, Theoretical Numerical Analysis. A Functional Analysis Framework, in: Texts in Applied Mathematics, vol. 39, Springer, 2001.
[8] E.W. Cheney, Introduction to Approximation Theory, in: International Series in Pure and Applied Mathematics, McGraw-Hill, 1966.
[9] A. Zygmund, Trigonometric Series, Volume I and II, 2nd ed., Cambridge University Press, Cambridge, 1959.
[10] K.E. Atkinson, An Introduction to Numerical Analysis, 2nd ed., John Wiley & Sons, New York, 1989.
[11] M.S. Gockenbach, Partial Differential Equations: Analytical and Numerical Methods, SIAM, 2002.
[12] C. Gasquet, P. Witomski, Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets (R. Ryan, Trans.), Springer, 1998.
[13] A.V. Oppenheim, R.W. Schafer, Discrete-Time Signal Processing, 2nd ed., Prentice Hall, 1998.
[14] L. Cohen, Time-Frequency Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[15] J. Rissanen, Stochastic Complexity in Statistical Inquiry, World Scientific, Singapore, 1989.
[16] P.D. Grünwald, The Minimum Description Length Principle, The MIT Press, Cambridge, MA, 2007.
[17] N. Saito, E.R. Woei, Simultaneous segmentation, compression, and denoising of signals using polyharmonic local sine transform and minimum description length criterion, in: Proc. 13th IEEE Workshop on Statistical Signal Processing, IEEE, 2005, pp. 315–320.

Please cite this article in press as: N. Saito, Y. Wang, The polynomial Fourier transform with minimum mean square error for noisy data, Journal of Computational and Applied Mathematics (2010), doi:10.1016/j.cam.2010.02.039