Journal of Computational Physics 371 (2018) 363–381

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Sequential function approximation with noisy data ☆

Yeonjong Shin, Kailiang Wu, Dongbin Xiu *

Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 13 September 2017
Received in revised form 24 April 2018
Accepted 22 May 2018
Available online 29 May 2018

Keywords:
Approximation theory
Randomized algorithm
Noisy data

ABSTRACT

We present a sequential method for approximating an unknown function sequentially using random noisy samples. Unlike the traditional function approximation methods, the current method constructs the approximation using one sample at a time. This results in a simple numerical implementation using only vector operations and avoids the need to store the entire data set. The method is thus particularly suitable when data set is exceedingly large. Furthermore, we present a general theoretical framework to define and interpret the method. Both upper and lower bounds of the method are established for the expectation of the results. Numerical examples are provided to verify the theoretical findings.

1. Introduction

We are concerned with the problem of approximating an unknown function $f(\mathbf{x})$ using its sample data $f(\mathbf{x}_i), i = 1, \ldots,$ where $\mathbf{x}, \mathbf{x}_i \in \mathbb{R}^d$, $d \geq 1$. This is a central task of approximation theory, which has been developed for a long time. The classical setting is that, given a set of finite number of sample data $\{f(\mathbf{x}_1), \ldots f(\mathbf{x}_M)\}$, $M \geq 1$, find a function $p(\mathbf{x}) \approx f(\mathbf{x})$. Numerous methods exist, for example, least squares, interpolation, etc. The topic has been extensively discussed in a large amount of literature. Here we mention only a few early books for reference, cf., [5,6,11,12].

In this paper we discuss a different approach, which seeks to construct the approximation p iteratively using only one sample at a time. This results in a sequential approximation, which can become progressively more accurate as more data are used. The implementation of the method requires only vector operations. In contrast, most of the classical approximation methods require operations on a Vandermonde-like matrix, also known as model matrix, whose number of rows is the number of samples and number of columns is the number of basis functions used to construct p. When the data set is extraordinarily large, the classical methods can be challenging to implement, as they require declaring, storing, and operating on, the extremely large model matrix. On the contrary, the current method uses one sample at a time and is not affected by the size of the data set. The sequential approximation (SA) method is rooted on the recent work of randomized approximation method [13,16], which was motivated by randomized Kaczmarz methods for linear system of equations [14] that were further analyzed and extended in a series of work, cf., [10,17,4,7,9,3,15]. Convergence and error analysis of the SA method were established in [13,16] for noiseless data. It was shown that the method can be highly efficient for extremely large data set, especially for approximations in very high dimensional space. The major contributions of the current paper are in two fronts: (1) We establish a broader framework for the SA method. In particular, we define a cost function which measures the difference between the approximation and the current data sample and the difference between two consecutive approximations. We then prove that the SA scheme is the minimization of the cost function. (2) Utilizing

☆ This work was partially supported by AFOSR FA95501410022 and NSF DMS 1418771.
* Corresponding author.
E-mail addresses: shin.481@osu.edu (Y. Shin), wu.3423@osu.edu (K. Wu), xiu.16@osu.edu (D. Xiu).

https://doi.org/10.1016/j.jcp.2018.05.042
0021-9991/

this new framework, we extend the application of the SA method to noisy data, which was not considered in the earlier work [13,16]. We then further provide analysis of the numerical errors and establish convergence of the method. We prove that when the noisy data samples are collected at independent and identically distributed (i.i.d.) random locations, the expectation of the errors in the approximation shall have (tight) upper and lower bounds. Furthermore, when the sampling locations follow a certain "optimal" probability measure, we derive the numerical error in a form of equality.

The paper is organized as follows. After setting up notation and preliminaries in Section 2, we present the general framework for defining the sequential approximation methods in Section 3. Two special cases, which correspond to two practical implementation procedures, are discussed. Theoretical results are presented for the errors of the methods. We then present extensive numerical results in Section 4 to verify the theoretical results, before concluding the paper.

2. Setup

Consider the problem of approximating an unknown function $f : D \to \mathbb{R}$ using its samples, where $D \subseteq \mathbb{R}^d$, $d \geq 1$, and is equipped with a measure ω. Let $\mathbf{x} = (x_1, \ldots, x_d)$ be the coordinate and $f \in L^2_\omega(D)$, the standard Hilbert space with inner product

$$(g, h)_{L^2_\omega} := \int_D g(\mathbf{x}) h(\mathbf{x}) d\omega(\mathbf{x}),$$

and the corresponding induced norm $\|\cdot\|_{L^2_\omega}$. The measure ω is assumed to be absolute continuous, and satisfy $\int_D d\omega = 1$.

Let $\mathbf{x}_k \in D$, $k = 1, \cdots$, be a sequence of sample points and $f(\mathbf{x}_k)$ be the function values at the samples. Let

$$f_k = f(\mathbf{x}_k) + \epsilon_k, \qquad k = 1, \ldots, \tag{2.1}$$

be the sample data, where ϵ_k, $k = 1, \cdots$, are independent random noises with zero mean value and bounded variance, i.e.,

$$\mathbb{E}(\epsilon_k) = 0, \qquad \mathbb{E}(\epsilon_k^2) = \sigma^2(\mathbf{x}_k) \leq \sigma_B^2 < +\infty, \qquad k = 1, \ldots. \tag{2.2}$$

The case of $\sigma_B = 0$ corresponds to the noiseless case.

Let $V \subset L^2_\omega(D)$ be a finite dimensional linear subspace with $\dim V = N \geq 1$. We then seek $p \in V$ as an approximation to f. Let $\{\psi_j(\mathbf{x}), j = 1, \ldots, N\}$ be a basis of V. Without loss of generality, we assume the basis functions are orthogonal with respect to the measure ω, i.e.,

$$\left(\psi_i, \psi_j\right)_{L^2_\omega} = \delta_{ij}, \qquad 1 \leq i, j \leq N. \tag{2.3}$$

(Note that non-orthogonal basis can always be orthogonalized by the Gram–Schmidt procedure.) The approximation can then be expressed as

$$p(\mathbf{x}) = \sum_{j=1}^{N} c_j \psi_j(\mathbf{x}). \tag{2.4}$$

Upon using vector notation

$$\mathbf{\Psi}(\mathbf{x}) = \left(\psi_1(\mathbf{x}), \ldots, \psi_N(\mathbf{x})\right)^\top, \qquad \mathbf{c} = \left(c_1, \ldots, c_N\right)^\top, \tag{2.5}$$

the approximation can be written as

$$p(\mathbf{x}) = \langle \mathbf{c}, \mathbf{\Psi}(\mathbf{x}) \rangle, \tag{2.6}$$

where $\langle \cdot, \cdot \rangle$ is the standard vector inner product. The obvious goal is to compute the expansion coefficients $\mathbf{c}$.

The best L^2_ω approximation of f in V is its orthogonal projection onto V, i.e.,

$$\mathcal{P}_V f = \sum_{k=1}^{N} \hat{c}_k \psi_k(\mathbf{x}) = \langle \hat{\mathbf{c}}, \mathbf{\Psi}(\mathbf{x}) \rangle, \tag{2.7}$$

where

$$\hat{\mathbf{c}} = \left(\hat{c}_1, \ldots, \hat{c}_N\right)^\top, \qquad \hat{c}_k = (f, \psi_k)_{L^2_\omega} \quad 1 \leq k \leq N. \tag{2.8}$$

Throughout this paper we shall use $\mathcal{P}_V f$ as the reference solution to examine the quality of the numerical approximation p, in the form of $\|p - \mathcal{P}_V f\|_{L^2_\omega}$, or, equivalently, $\|\mathbf{c} - \hat{\mathbf{c}}\|$ in vector 2-norm.

3. Sequential approximation methods

In this section we present the sequential approximation methods. We first present a general mathematical framework to define the methods. We then discuss a few special cases for practical implementation. Finally, we provide theoretical analysis for the errors of the methods.

3.1. General method and its variations

The sequential approximation method is of iterative nature. From the linear subspace V, we choose an arbitrary initial approximation $p^{(0)}(\mathbf{x})$. Then, at the k-th step, $k = 1, \ldots$, when the sample $(\mathbf{x}_k, f_k)$ is available, we seek the k-th approximation as the solution of the following minimization problem.

$$p^{(k)}(\mathbf{x}; \gamma_k) = \underset{p \in V}{\operatorname{argmin}} \left(\left\| p(\mathbf{x}) - p^{(k-1)}(\mathbf{x}) \right\|_{L^2_\omega}^2 + \frac{1}{\gamma_k} |p(\mathbf{x}_k) - f_k|^2 \right), \tag{3.1}$$

where $\gamma_k \geq 0$, $k = 1, \cdots$, are parameters, and the limiting case of $\gamma_k = 0$ is defined as

$$p^{(k)}(\mathbf{x}; 0) := \underset{p \in V}{\operatorname{argmin}} \left(\left\| p(\mathbf{x}) - p^{(k-1)}(\mathbf{x}) \right\|_{L^2_\omega}^2 \right), \quad \text{subject to} \quad p(\mathbf{x}_k) = f_k. \tag{3.2}$$

Hereafter we will omit the explicit dependence on γ_k and simply write $p^{(k)}(\mathbf{x})$, unless confusion arises otherwise.

The solution of this optimization problem can be explicitly derived.

Theorem 3.1. *For $k = 1, \ldots$, let $p^{(k)} \in V$ be expressed as*

$$p^{(k)}(\mathbf{x}) = \sum_{j=1}^{N} c_j^{(k)} \psi_j(\mathbf{x}) = \langle \mathbf{c}^{(k)}, \Psi(\mathbf{x}) \rangle, \tag{3.3}$$

where the orthonormal basis (2.3) is used and the vector notations (2.5) are employed. Then the solution to (3.1), which includes the special case (3.2), is

$$p^{(k)}(\mathbf{x}) = p^{(k-1)}(\mathbf{x}) + \frac{f_k - p^{(k-1)}(\mathbf{x}_k)}{\|\Psi(\mathbf{x}_k)\|_2^2 + \gamma_k} \Psi(\mathbf{x}_k)^\top \Psi(\mathbf{x}), \quad k = 1, \ldots, \tag{3.4}$$

or, equivalently, in the form of the expansion coefficients

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \frac{f_k - \langle \mathbf{c}^{(k-1)}, \Psi(\mathbf{x}_k) \rangle}{\|\Psi(\mathbf{x}_k)\|_2^2 + \gamma_k} \Psi(\mathbf{x}_k), \quad k = 1, \ldots. \tag{3.5}$$

Proof. See Appendix A. □

Equation (3.5) thus defines a simple algorithm for the implementation of sequential approximation, as follows.

- Start from an arbitrary initial choice of the coefficient vector $\mathbf{c}^{(0)}$, which defines the arbitrary initial approximation $p^{(0)}(\mathbf{x})$.
- At k-th step, when the sample $(\mathbf{x}_k, f_k)$ is available, use the scheme (3.5) to update the coefficient vector, which defines the corresponding approximation $p^{(k)}(\mathbf{x})$.

An alternative way to interpret the scheme (3.5) is to view it as stochastic gradient descent (SGD) ([2]) for minimizing the following objective function

$$\min_{\mathbf{c} \in \mathbb{R}^N} \int_D \Big(\langle \mathbf{c}, \Psi(\mathbf{x}) \rangle - f(\mathbf{x}) \Big)^2 d\omega(\mathbf{x}),$$

with randomly selected data points and a specific step size. However, we emphasize the following fact: in each iterative step of SA, the k-th solution $p^{(k)}(\mathbf{x})$ is the *exact* solution to the proposed optimization problem (3.1). The form of (3.1) has certain resemblance to Kalman filter, which was associated with the recursive least squares algorithm ([1]). Unlike any least squares methods, the SA method is matrix-free and involves only vectors.

The sequential approximation (3.1), or equivalently, its implementation scheme (3.5), depends on the choice of the parameter γ_k. This parameter is related to the noise level in the data and becomes zero for the noiseless case. Here we discuss two special cases.

3.1.1. Method I: $\gamma_k \equiv \tau$

A natural choice for the parameter γ_k is a constant $\tau \geq 0$. This results in the following scheme.

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \frac{f_k - \langle \mathbf{c}^{(k-1)}, \mathbf{\Psi}(\mathbf{x}_k)\rangle}{\|\mathbf{\Psi}(\mathbf{x}_k)\|_2^2 + \tau}\mathbf{\Psi}(\mathbf{x}_k), \quad k = 1, \ldots. \tag{3.6}$$

The constant τ can be chosen according to the noise level in the data, i.e. $\tau \propto \sigma_B$, where σ_B is defined in (2.2). For the noiseless case $\sigma_B = 0$, one may set $\tau = 0$, which corresponds to the limiting optimization case (3.2). These choices are intuitive from a practical point of view. On the other hand, our analysis in the following sections will show that the convergence of the method can be achieved by an arbitrary choice of $\tau \geq 0$. We also remark that the limiting case of $\tau = 0$ reproduces the randomized approximation method presented in [13,16], which considered only the noiseless case.

3.1.2. Method II: $\gamma_k = \tau \|\mathbf{\Psi}(\mathbf{x}_k)\|_2^2$

Again, let $\tau > 0$ be a constant. We set $\gamma_k = \tau \|\mathbf{\Psi}(\mathbf{x}_k)\|_2^2$. This gives us the following scheme.

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \frac{f_k - \langle \mathbf{c}^{(k-1)}, \mathbf{\Psi}(\mathbf{x}_k)\rangle}{(1+\tau)\|\mathbf{\Psi}(\mathbf{x}_k)\|_2^2}\mathbf{\Psi}(\mathbf{x}_k), \quad k = 1, \ldots. \tag{3.7}$$

We present this special case because it allows one to obtain certain "optimal" convergence, under certain conditions. Again, the value of τ can be set to be proportional to the noise level in the data, although the convergence of the method can be obtained for any positive value of τ.

3.2. Error analysis

In this section we present error analysis for the sequential approximation methods in the previous section. Our analysis is based on random sampling in the domain D and the errors are evaluated as expectations of the random sequences.

3.2.1. Notations

Since the parameter γ_k in the main method (3.5) depends on the iteration step, which in turn depends on the location of the sample $\mathbf{x}_k$, we write $\gamma_k = \gamma(\mathbf{x}_k)$. We then define the following measures

$$\begin{aligned} d\tilde{\mu} &= \frac{1}{\|\mathbf{\Psi}(\mathbf{x})\|_2^2 + \gamma(\mathbf{x})} d\mu, \\ d\tilde{\omega} &= \frac{\|\mathbf{\Psi}(\mathbf{x})\|_2^2}{\left(\|\mathbf{\Psi}(\mathbf{x})\|_2^2 + \gamma(\mathbf{x})\right)^2} d\mu, \\ d\tilde{\nu} &= d\tilde{\mu} - d\tilde{\omega}, \end{aligned} \tag{3.8}$$

and let $(\cdot,\cdot)_{L^2_{\tilde{\mu}}}$ and $(\cdot,\cdot)_{L^2_{\tilde{\nu}}}$ be their corresponding inner products. Let

$$\mathbf{\Sigma} = \left(\Sigma_{ij}\right)_{1\leq i,j\leq N}, \qquad \Sigma_{ij} = (\psi_i, \psi_j)_{L^2_{\tilde{\mu}}}, \quad 1 \leq i, j \leq N, \tag{3.9}$$

be a $N \times N$ covariance matrix. It is symmetric and positive definite, and has the following eigenvalue decomposition

$$\mathbf{\Sigma} = \mathbf{Q}^\top \mathbf{\Lambda} \mathbf{Q},$$

where $\mathbf{Q}$ is orthogonal and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ with

$$\lambda_{\max} := \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N =: \lambda_{\min} > 0. \tag{3.10}$$

Let

$$\mathbf{\Sigma}_* := \mathbf{\Sigma} + \left((\psi_i, \psi_j)_{L^2_{\tilde{\nu}}}\right)_{1\leq i,j\leq N} \tag{3.11}$$

be another covariance matrix, whose eigenvalues are

$$\lambda^*_{\max} := \lambda^*_1 \geq \lambda^*_2 \geq \cdots \geq \lambda^*_N =: \lambda^*_{\min} > 0. \tag{3.12}$$

Again, upon using vector notations, we denote

$$\mathbf{e} = \left(e_1, e_2, \ldots, e_N\right)^\top, \qquad \mathbf{e}_* = \left(e^*_1, e^*_2, \ldots, e^*_N\right)^\top, \tag{3.13}$$

where, for $j = 1, \ldots, N$,

$$e_j = \left(f - \mathcal{P}_V f, \psi_j\right)_{L^2_{\tilde{\mu}}}, \qquad e^*_j = \left(f - \mathcal{P}_V f, \psi_j\right)_{L^2_{\tilde{\nu}}}. \tag{3.14}$$

3.2.2. Errors for the general method (3.5)

For the general method (3.5), we first provide the following error bound on the solution coefficients.

Lemma 3.2. *Let $d\mu$ be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$, are drawn. Then the k-th iterative solution of (3.5) satisfies*

$$\left\|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\right\|_2^2 r_\ell^k + \frac{1-r_\ell^k}{\lambda^*_{\max}}\mathcal{E} + \mathcal{B}(r_\ell) \le \mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) \le \left\|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\right\|_2^2 r_u^k + \frac{1-r_u^k}{\lambda^*_{\min}}\mathcal{E} + \mathcal{B}(r_u),$$

where $\hat{\mathbf{c}}$ is the best L^2_ω projection coefficients (2.8),

$$r_\ell = 1 - \lambda^*_{\max}, \qquad r_u = 1 - \lambda^*_{\min}, \tag{3.15}$$

and

$$\mathcal{E} = \|f - \mathcal{P}_V f\|^2_{L^2_\omega} + \|\sigma\|^2_{L^2_\omega},$$

$$\mathcal{B}(r) := 2\mathbf{e}_*^\top \left(\mathbf{Q}^\top \mathbf{D}^{(k)}(r)\mathbf{Q}\left(\mathbf{\Sigma}(\hat{\mathbf{c}} - \mathbf{c}^{(0)}) + \mathbf{e}\right) + \frac{1-r^k}{1-r}\mathbf{\Sigma}^{-1}\mathbf{e}\right),$$

with

$$\mathbf{D}^{(k)}(r) := \operatorname{diag}\left\{d_1^{(k)}, \cdots, d_N^{(k)}\right\}, \qquad d_i^{(k)} = \frac{r^k - (1-\lambda_i)^k}{\lambda_i\left(1 - r - \lambda_i\right)}.$$

In the limit of $k \to \infty$,

$$\frac{2\mathbf{e}_*^\top \mathbf{\Sigma}^{-1}\mathbf{e} + \mathcal{E}}{\lambda^*_{\max}} \le \lim_{k\to\infty} \mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) \le \frac{2\mathbf{e}_*^\top \mathbf{\Sigma}^{-1}\mathbf{e} + \mathcal{E}}{\lambda^*_{\min}}.$$

Here, λ and λ^* are the eigenvalues defined in (3.10) and (3.12), respectively.

Proof. See Appendix B. □

We then immediately have the error bounds for the k-th iterative solution.

Theorem 3.3. *Let $d\mu$ be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$, are drawn. Then the k-th iterative solution of (3.4) satisfies*

$$\mathbb{E}\left(\left\|p^{(k)} - f\right\|^2_{L^2_\omega}\right) \le \left\|f - \mathcal{P}_V f\right\|^2_{L^2_\omega} + \left\|p^{(0)} - \mathcal{P}_V f\right\|^2_{L^2_\omega} r_u^k + \frac{1-r_\ell^k}{\lambda^*_{\max}}\mathcal{E} + \mathcal{B}(r_u),$$

and

$$\mathbb{E}\left(\left\|p^{(k)} - f\right\|^2_{L^2_\omega}\right) \ge \left\|f - \mathcal{P}_V f\right\|^2_{L^2_\omega} + \left\|p^{(0)} - \mathcal{P}_V f\right\|^2_{L^2_\omega} r_\ell^k + \frac{1-r_u^k}{\lambda^*_{\min}}\mathcal{E} + \mathcal{B}(r_\ell).$$

In the limit of $k \to \infty$,

$$\left\|f - \mathcal{P}_V f\right\|^2_{L^2_\omega} + \frac{2\mathbf{e}_*^\top \mathbf{\Sigma}^{-1}\mathbf{e} + \mathcal{E}}{\lambda^*_{\max}} \le \lim_{k\to\infty} \mathbb{E}\left(\left\|p^{(k)} - f\right\|^2_{L^2_\omega}\right) \le \left\|f - \mathcal{P}_V f\right\|^2_{L^2_\omega} + \frac{2\mathbf{e}_*^\top \mathbf{\Sigma}^{-1}\mathbf{e} + \mathcal{E}}{\lambda^*_{\min}}.$$

Proof. The proof is a direct consequence of Lemma 3.2, using the fact that $\left\|p^{(k)} - f\right\|^2_{L^2_\omega} = \left\|f - \mathcal{P}_V f\right\|^2_{L^2_\omega} + \left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2$. □

3.2.3. Errors for Method I (3.6) with $\gamma_k = \tau$

Since Method I (3.6) is a special case of the general method (3.5), its errors can be directly derived from the error bounds in the previous section. The following result is a different estimate of the upper bound. Although it is not as sharp as the result from Lemma 3.2, it has a simpler form.

Theorem 3.4. *Let $d\mu$ be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$, are drawn. Then the k-th iterative solution of (3.6) satisfies*

$$\mathbb{E}\left(\left\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\right\|_2^2\right) \le r_u^k \left\|\mathbf{c}^{(0)}-\hat{\mathbf{c}}\right\|_2^2 + \frac{1}{\lambda_{\min}}(1-r_u^k)\left(\|f-\mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \|\sigma\|_{L^2_{\tilde{\omega}}}^2\right), \tag{3.16}$$

where

$$r_u = 1-\lambda_{\min}. \tag{3.17}$$

Proof. See Appendix C. □

If we choose a specific sampling measure for the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, then we have the following sharp error bounds.

Theorem 3.5. *Let*

$$d\mu = \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2+\tau}{N+\tau}d\omega,$$

be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, are drawn, then the k-th iterative solution of (3.6) satisfies

$$\begin{aligned}
&\mathbb{E}\big(\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\|_2^2\big) \le \|\mathbf{c}^{(0)}-\hat{\mathbf{c}}\|_2^2 r_u^k + \eta_u\left(1-r_u^k\right)\left(\|f-\mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right),\\
&\mathbb{E}\big(\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\|_2^2\big) \ge \|\mathbf{c}^{(0)}-\hat{\mathbf{c}}\|_2^2 r_\ell^k + \eta_\ell\left(1-r_\ell^k\right)\left(\|f-\mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right),
\end{aligned} \tag{3.18}$$

where

$$\begin{aligned}
&\Psi_{\sup} := \sup_{\mathbf{x}\in D}\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2, \qquad \Psi_{\inf} := \inf_{\mathbf{x}\in D}\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2,\\
&r_u = 1-\frac{\Psi_{\sup}+2\tau}{(N+\tau)(\Psi_{\sup}+\tau)}, \qquad \eta_u = \frac{\Psi_{\sup}}{\Psi_{\sup}+2\tau},\\
&r_\ell = 1-\frac{\Psi_{\inf}+2\tau}{(N+\tau)(\Psi_{\inf}+\tau)}, \qquad \eta_\ell = \frac{\Psi_{\inf}}{\Psi_{\inf}+2\tau}.
\end{aligned}$$

In the limit of $k\to\infty$, we have

$$\eta_\ell\left(\|f-\mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right) \le \lim_{k\to\infty}\mathbb{E}\big(\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\|_2^2\big) \le \eta_u\left(\|f-\mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right). \tag{3.19}$$

Proof. See Appendix D. □

It is worth mentioning that $\Psi_{\sup}$ in the theorem can be infinity. The conclusion still holds in that case, with $r_u = 1-\frac{1}{N+\tau}$ and $\eta_u = 1$.

3.2.4. Errors for Method II (3.7) with $\gamma_k = \tau\,\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2$

Again, we first present a different upper bound of the error, which, albeit not as sharp as that of Lemma 3.2, possesses a simpler form.

Theorem 3.6. *Let $d\mu$ be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, are drawn. Then the k-th iterative solution of (3.7) satisfies*

$$\mathbb{E}\big(\left\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\right\|_2^2\big) \le r_u^k\|\mathbf{c}^{(0)}-\hat{\mathbf{c}}\|_2^2 + \frac{1-r_u^k}{\lambda_{\min}}\left(\|f-\mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \frac{1}{1+\tau}\|\sigma\|_{L^2_{\tilde{\mu}}}^2\right),$$

where

$$r_u = 1-\lambda_{\min}.$$

Proof. The proof is similar to that of Theorem 3.4. □

We now present an error analysis based on a specific choice of sampling measure. Under this sampling measure, we establish the error in a form of *equality*, which implies that the upper and lower bounds meet.

Theorem 3.7. *Let*

$$d\mu = \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{N} d\omega,$$

be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ *are drawn. Then the k-th iterative solution of* (3.7) *satisfies*

$$\mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) = r^k \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 + \frac{1 - r^k}{1 + 2\tau}\left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right), \tag{3.20}$$

where

$$r = 1 - \frac{1 + 2\tau}{N(1+\tau)^2}.$$

In the limit of $k \to \infty$,

$$\lim_{k\to\infty} \mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) = \frac{1}{1 + 2\tau}\left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right). \tag{3.21}$$

Proof. See Appendix E. □

It is therefore obvious that one may choose a larger value for γ_k, resp. τ, to obtain smaller numerical error in the final converged solution, at the expense of slower rate of convergence r.

3.2.5. Errors for the special case of $\gamma_k \equiv 0$

One can always set the parameter γ_k in the main method (3.5) to be zero. This special case recovers the method presented in [13,16], which assumed the sample data are noiseless. Here we present an error estimate when data contain noises in the form of (2.1).

Theorem 3.8. *Assume* $\gamma_k \equiv 0$, *and let* $d\mu$ *be the sampling measure from which the i.i.d. samples of the sequence* $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ *are drawn. Then the k-th iterative solution of* (3.5) *satisfies*

$$\begin{aligned}
\mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) &\le r_u^k \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 + \frac{1 - r_u^k}{\lambda_{\min}}\left(\|f - \mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \|\sigma\|_{L^2_{\tilde{\mu}}}^2\right), \\
\mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) &\ge r_\ell^k \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 + \frac{1 - r_\ell^k}{\lambda_{\max}}\left(\|f - \mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \|\sigma\|_{L^2_{\tilde{\mu}}}^2\right),
\end{aligned} \tag{3.22}$$

where

$$r_u = 1 - \lambda_{\min}, \qquad r_\ell = 1 - \lambda_{\max}. \tag{3.23}$$

In the limit of $k \to \infty$,

$$\frac{1}{\lambda_{\max}}\left(\|f - \mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \|\sigma\|_{L^2_{\tilde{\mu}}}^2\right) \le \lim_{k\to\infty} \mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) \le \frac{1}{\lambda_{\min}}\left(\|f - \mathcal{P}_V f\|_{L^2_{\tilde{\mu}}}^2 + \|\sigma\|_{L^2_{\tilde{\mu}}}^2\right).$$

Proof. This is a direct corollary of Theorem 3.2 by noting $\mathbf{e}_* = \mathbf{0}$ and $d\tilde{\omega} = d\tilde{\mu}$ when $\gamma \to 0$. □

Using a specific sampling measure, the following result holds.

Theorem 3.9. *Assume* $\gamma_k = 0$, *and the*

$$d\mu = \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{N} d\omega, \tag{3.24}$$

be the sampling measure from which the i.i.d. samples of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ *are drawn. Then the k-th iterative solution of* (3.5) *satisfies*

$$\mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) = r^k \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 + (1 - r^k)\left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2\right), \tag{3.25}$$

where

$$r = 1 - 1/N. \tag{3.26}$$

Furthermore,

$$\lim_{k\to\infty} \mathbb{E}\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) = \|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2. \tag{3.27}$$

Proof. This is a direct corollary of Theorem 3.8. □

4. Numerical examples

In this section we present numerical examples to verify the properties of the proposed sequential approximation methods.

4.1. Setup

Without loss of generality, we consider polynomial approximation and let the subspace V be Π_n^d, the linear polynomial subspace of polynomials of degree up to $n \geq 1$. That is,

$$\Pi_n^d := \text{span}\{\mathbf{x}^{\mathbf{i}} = x_1^{i_1} \cdots x_d^{i_d}, |\mathbf{i}| \leq n\}, \tag{4.1}$$

where $\mathbf{i} = (i_1, \ldots, i_d)$ is multi-index with $|\mathbf{i}| = i_1 + \cdots + i_d$. Immediately, we have

$$N = \dim \Pi_n^d = \binom{n+d}{d} = \frac{(n+d)!}{n!d!}. \tag{4.2}$$

We choose the domain to be hypercube, i.e., $D = [-1, 1]^d$, and employ the normalized Legendre polynomials as the orthogonal basis. To demonstrate the properties in multiple dimensions $d > 1$, we present examples in $d = 2$ and $d = 10$.

For benchmarking purpose, we use the following four multivariate functions as target functions. These are from [8] and have been widely used for multi-dimensional function integration and approximation tests. They include

$$\begin{aligned}
f_1(\mathbf{x}) &= \exp\left(-\sum_{i=1}^d a_i^2 \left(\frac{x_i+1}{2} - \chi_i\right)^2\right); \quad \text{(GAUSSIAN)} \\
f_2(\mathbf{x}) &= \exp\left(-\sum_{i=1}^d a_i \left|\frac{x_i+1}{2} - \chi_i\right|\right); \quad \text{(CONTINUOUS)} \\
f_3(\mathbf{x}) &= \left(1 + \sum_{i=1}^d a_i \frac{(x_i+1)}{2}\right)^{-(d+1)}, \text{ where } a_i = \frac{1}{i^2}; \quad \text{(CORNER PEAK)} \\
f_4(\mathbf{x}) &= \prod_{i=1}^d \left(a_i^{-2} + \left(\frac{x_i+1}{2} - \chi_i\right)^2\right)^{-1}; \quad \text{(PRODUCT PEAK)},
\end{aligned} \tag{4.3}$$

where $\boldsymbol{a} = (a_1, \cdots, a_d)$ are parameters controlling the difficulty of the functions, and $\boldsymbol{\chi} = (\chi_1, \cdots, \chi_d)$ are shifting parameters.

The convergence results are examined by the errors in the approximation $p^{(k)}$ at k-th step iteration, $k = 1, \ldots$. To evaluate the errors we independently draw a set of dense samples, 20,000 for $d = 2$ and 100,000 for $d = 10$, uniformly in the domain and compute the difference between $p^{(k)}$ and the target function f at these sampling points. The vector 2-norm is then used to report the difference. This gives us a numerical estimation to the error $\left\|p^{(k)} - f\right\|_{L^2_\omega}$. We then use this error to verify the theoretical convergence results in Section 3.2.

To draw the random sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, we employ several different probability measures: uniform distribution, tensorized Chebyshev distribution, and the specific measures defined in the Theorems 3.5, 3.7, and 3.9. These specific measures are not standard probability measures, but can be effectively produced by using Markov chain Monte Carlo method.

4.2. Two-dimensional case $d = 2$

We first present extensive numerical results in $d = 2$. All the results are averaged over 50 independent sample sequences simulations. This is to verify our theoretical error estimates, which are expressed in term of expectation.

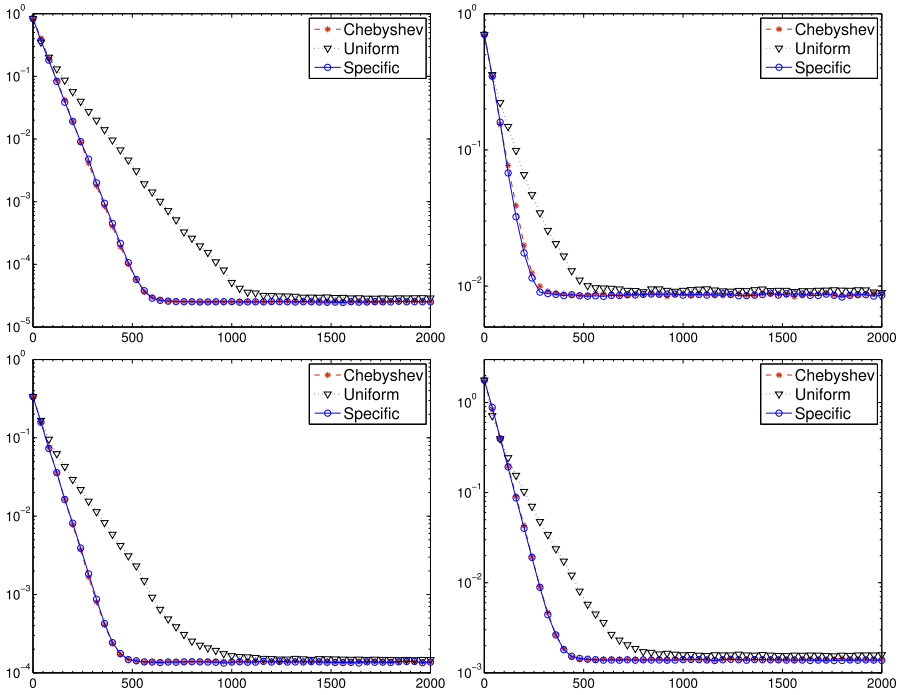

Fig. 4.1. Function approximation errors versus number of iterations for four test functions in (4.3) at $d = 2$ without noise. $\gamma_k = 0$ is used in our method. Top left: f_1 with $\boldsymbol{a} = (1, 1)$ and $\boldsymbol{\chi} = (0.375, 0.375)$; Top right: f_2 with $\boldsymbol{a} = (-2, 1)$ and $\boldsymbol{\chi} = (0.25, -0.75)$; Bottom left: f_3; Bottom right: f_4 with $\boldsymbol{a} = (-1.5, 1)$ and $\boldsymbol{\chi} = (0.375, 0.375)$.

4.2.1. Noiseless data

We first consider the case where the data contain no noise. Consequently, we set the parameter γ_k in the methods to be 0. The errors in $p^{(k)}$ with respect to the iteration count k are plotted in Fig. 4.1, for all four test functions using Legendre polynomials of degree $n = 6$. The sampling probability measures include the Chebyshev measure, uniform measure and the specific (optimal) measure (3.24) in Theorem 3.9. Exponential convergence rate can be observed, where the Chebyshev measure and the optimal measure exhibit similarly faster convergence.

We then examine the errors in the converged numerical results (at $k \to \infty$), for polynomial degrees $n = 1$ to $n = 8$. The results are shown in Fig. 4.2. For comparison, the errors of the orthogonal projection $\mathcal{P}_\Pi f$ are also shown, which are approximated by high-order numerical quadrature. One can see that the numerical errors of the proposed method are very close to the errors of the orthogonal projection, as suggested by Theorems 3.8 and 3.9.

To verify the theoretical error bounds in Theorems 3.8 and 3.9, we accurately compute all the terms in the theoretical estimates. (This is possible because we know the form of the target functions.) The comparison between the theoretical bounds (in expectation) and numerical curves are shown in Fig. 4.3 for f_1 and $n = 6$. With the specific (optimal) measure, the upper and lower bounds become the same, as the error analysis Theorem 3.9 is in the form of equality. The Chebyshev measure produces very tight upper and lower bounds. This suggests that the Chebyshev measure, which be easily sampled, is a good alternative to the optimal measure. The uniform measure produces wider bounds.

4.2.2. Noisy data

We now consider the case when data contain noises. We set the noises in the functions f_1 and f_2 to be Gaussian distribution $\mathcal{N}(0, 0.01^2)$ and the noises in the functions f_3 and f_4 to follow uniform distribution $\mathcal{U}(-0.01, 0.01)$. (The results behave very similarly for other choices of noises.)

We first employ the Method 1 with $\tau = \sigma^2$. Fig. 4.4 gives the convergence of the errors with respect to the iteration count for all the test functions in (4.3), for polynomial degree up to $n = 6$. Again, we observe the exponential type convergence of the errors, and the results by Chebyshev and the specific measure in Theorem 3.5 are very similar. Due to the presence of the noise, the converged errors become larger than those from the noiseless case shown in Fig. 4.1.

It is discovered that increasing τ can reduce the approximation errors in the converged solutions, see Fig. 4.5 for the results obtained by Method I with $\tau = 500$. The convergence rate, however, becomes slower, and the behavior of uniform sampling measure becomes closer to the specific one. Fig. 4.6 reports the approximation errors in the converged solutions for different polynomial degrees from 1 to 8. Those plots further validate the above observation. Such impact of the parameter choice τ is already visible from the analysis, see, for example, Theorem 3.7. More detailed analysis of the parameter τ shall be pursued in a separate study.

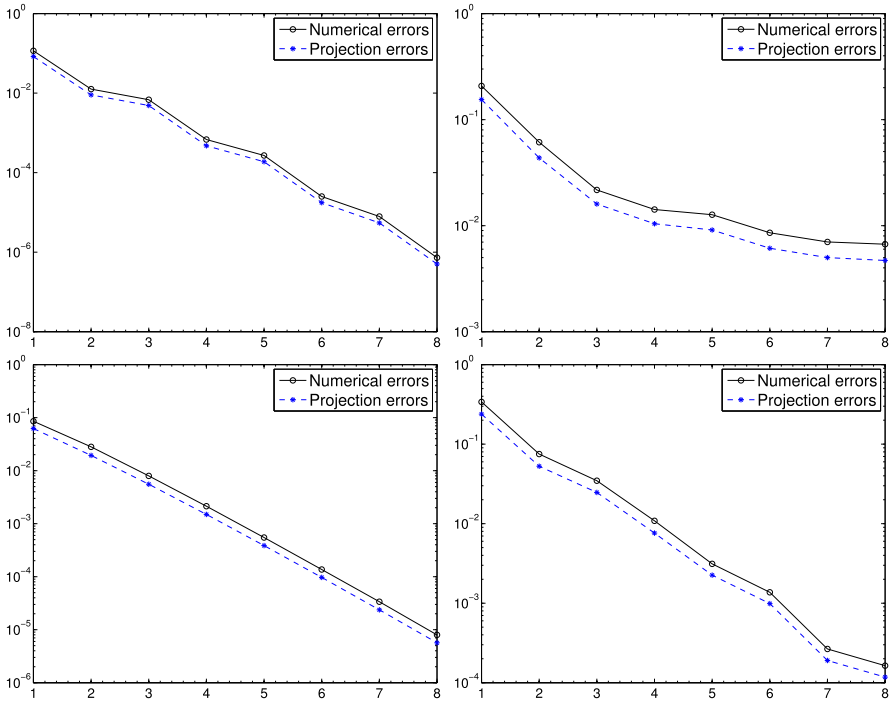

Fig. 4.2. Approximation errors in the converged solutions with respect to the polynomial degrees at $d = 2$ for no noise data. $\gamma_k = 0$ is used in our method. Top left: f_1 with $\boldsymbol{a} = (1, 1)$ and $\boldsymbol{\chi} = (0.375, 0.375)$; Top right: f_2 with $\boldsymbol{a} = (-2, 1)$ and $\boldsymbol{\chi} = (0.25, -0.75)$; Bottom left: f_3; Bottom right: f_4 with $\boldsymbol{a} = (-1.5, 1)$ and $\boldsymbol{\chi} = (0.375, 0.375)$.

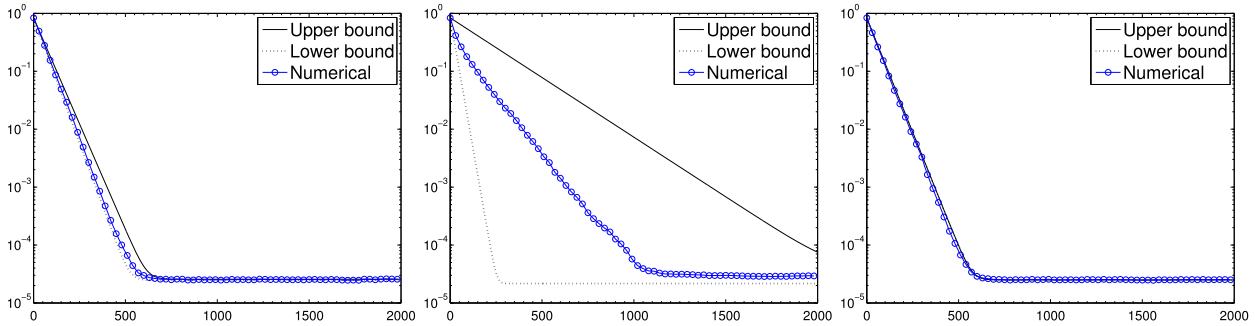

Fig. 4.3. Numerical approximation errors and theoretical error bounds from Theorem 3.9 at $d = 2$ for $\gamma_k = 0$ and no noise data: Chebyshev (left), uniform (middle) and specific (right) measure.

We now verify the theoretical error bounds in Theorems 3.2, 3.3, 3.4 and 3.5, by numerically computing all the terms in the theoretical analysis. The comparison between the theoretical bounds (in expectation) and numerical curves for f_1 and $n = 6, \tau = 500$ is displayed in Fig. 4.7, where "simple bounds" denotes those given by Theorems 3.4 and 3.5. The results are consistent with the theoretical findings, as the "simple bounds" from Theorems 3.4 and 3.5 are not as sharp as the bounds from Theorem 3.2. (Again, the "simple bounds" possess simpler form and are easier to evaluate.)

We then test the proposed Method II with $\tau = \sigma^2$. Fig. 4.8 shows the error convergence for all the test functions in (4.3) by polynomials of degree up to $n = 6$. It is seen that Method II also exhibits the exponential type convergence with respect to iteration count. The convergence behaviors are very similar to those of Method I in Fig. 4.4.

Again, we find that increasing τ can reduce the approximation errors in the converged solutions, albeit at slower rate of convergence. See Fig. 4.9 for the results obtained by Method II with $n = 6$ and $\tau = 20$. This is further validated by the results given in Fig. 4.10, where the approximation errors in the converged solutions by using different polynomial degrees are displayed.

We now verify the theoretical error bounds in Theorems 3.2, 3.3, 3.6 and 3.7, by numerically computing the terms in the theory. The comparison between the theoretical bounds (in expectation) and the actual numerical convergence for the function f_1 with $n = 6, \tau = 20$ is shown in Fig. 4.11, where the simple upper bound is that given in Theorem 3.6. The numerical bounds are consistent with the theoretical bounds.

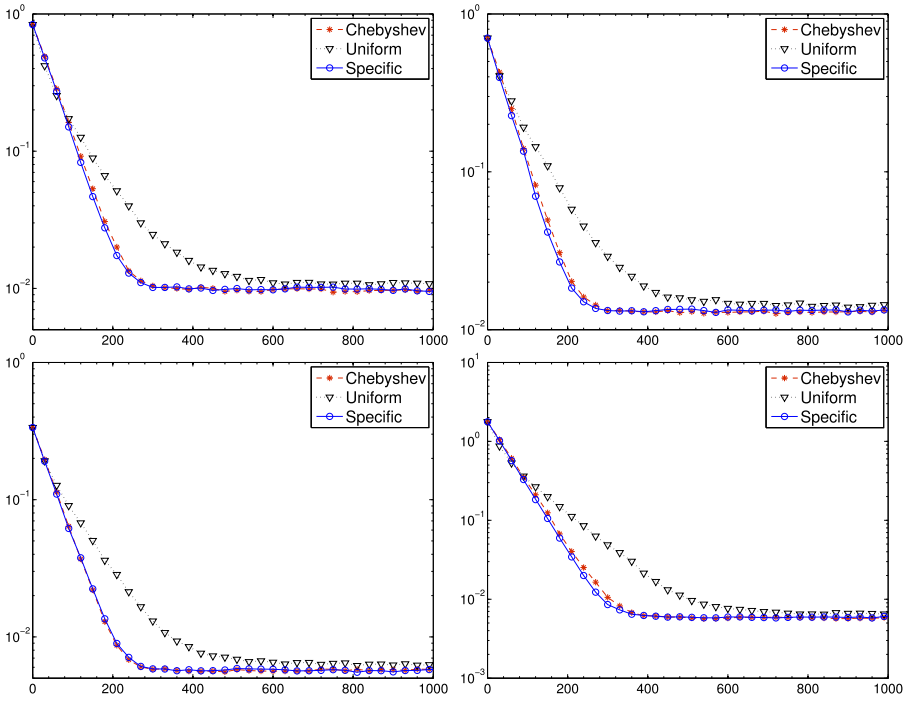

Fig. 4.4. Same as Fig. 4.1 except for the data with noise and Method I with $\tau = \sigma^2$.

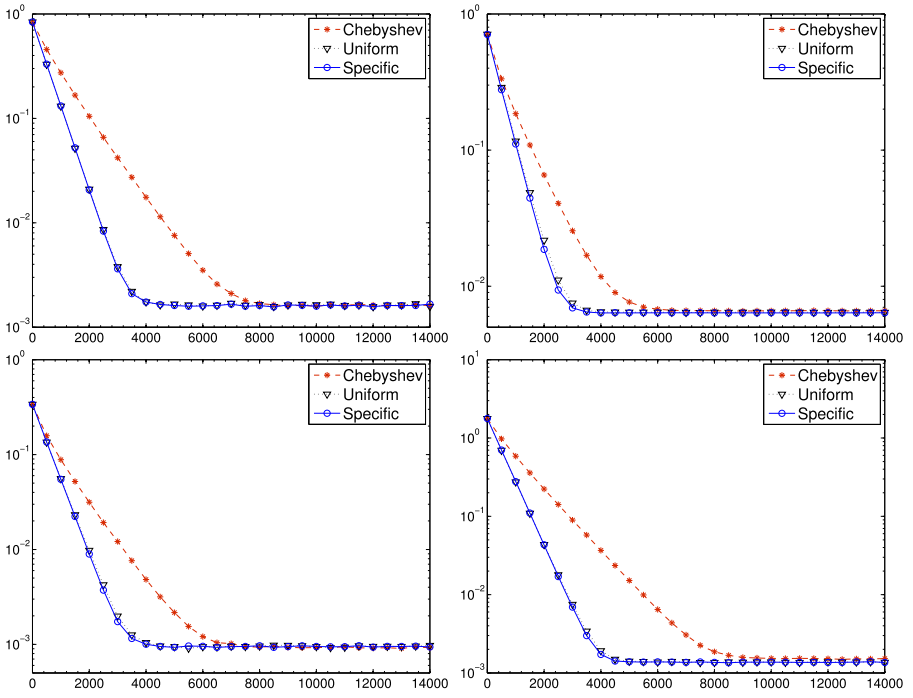

Fig. 4.5. Same as Fig. 4.1 except for the data with noise and Method I with $\tau = 500$.

4.3. *Ten-dimensional case* $d = 10$

We now focus on a higher dimension $d = 10$. Here all the numerical results are reported as those of single simulation. We only show the results of approximating f_1 in (4.3) and assume that the data contain Gaussian noises of $N(0, 0.01^2)$, i.e. with standard deviation $\sigma = 0.01$. The results for the other functions in (4.3) behave very similarly and will not be shown.

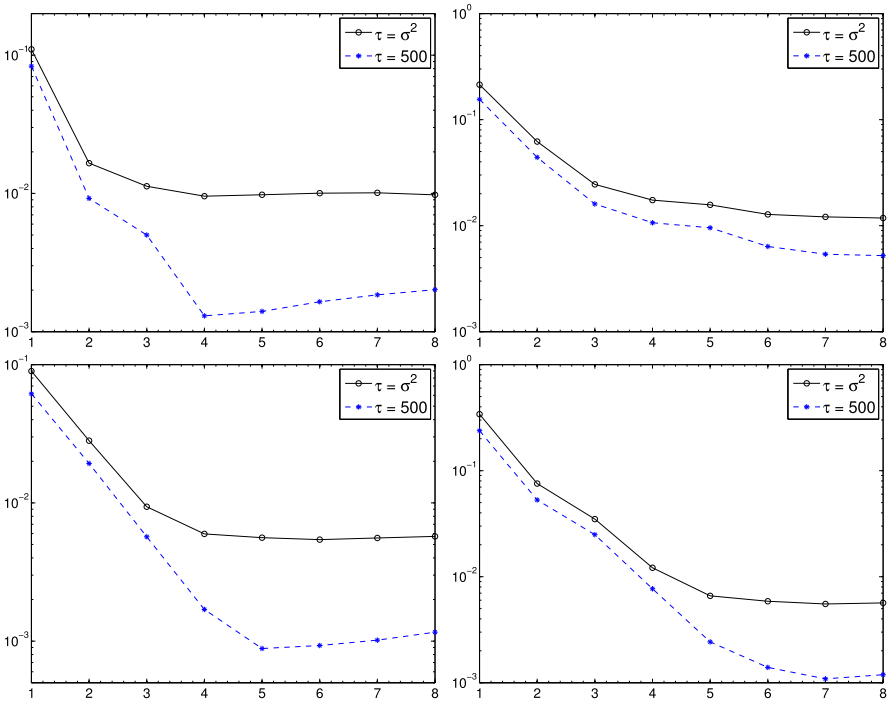

Fig. 4.6. Same as Fig. 4.2 except for the data with noise and using Method I.

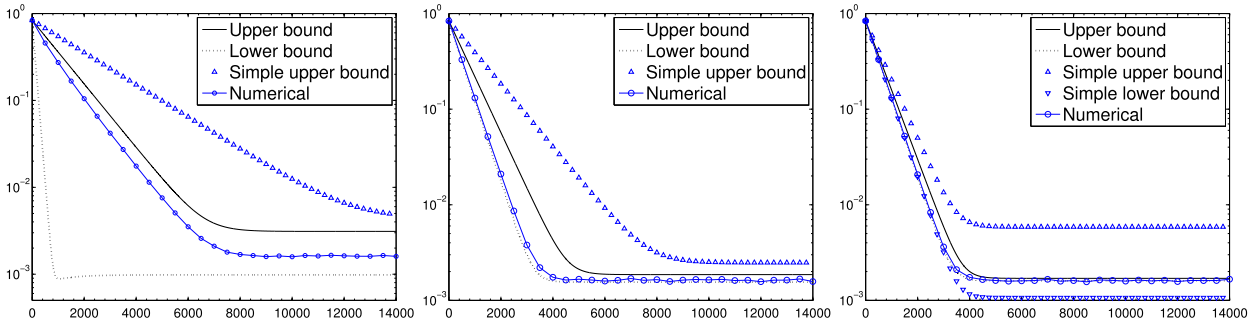

Fig. 4.7. Same as Fig. 4.3 except for the data with noise and Method I with $\tau = 500$.

We employ Method I with $\tau = \sigma^2$ and $\tau = 10^4$ and Method II with $\tau = \sigma^2$ and $\tau = 10$. Fig. 4.12 shows the error decay at polynomial degree $n = 6$ with respect to the number of iterations. We again observe the expected exponential type convergence of the errors, for both methods using different sampling measures. Similar to the two-dimensional case, with larger τ we obtain smaller errors in the converged solutions, at the expense of slower rate of convergence. To further demonstrate this phenomenon, we repeat the tests by using different degree of polynomials, and plot the approximation errors in the converged solutions with respect to n in Fig. 4.13. The errors decay at higher polynomial degree, as expected, and saturate at modest polynomial orders because the data noises become dominant. It is obvious that Increasing the parameter τ can reduce the errors in converged solutions for all polynomial degrees.

5. Conclusion

In this paper we presented a framework of sequential function approximation. The sequential approximation methods are of iterative nature and utilize one data sample at each step. They render simple numerical implementations and can handle practical problems with exceedingly large data set. This paper extends the earlier work of [13,16] to handle the case of noisy data and also presents a more broader theoretical framework. The convergence and error estimates of the methods were presented, along with numerous examples to demonstrate the effectiveness of the methods.

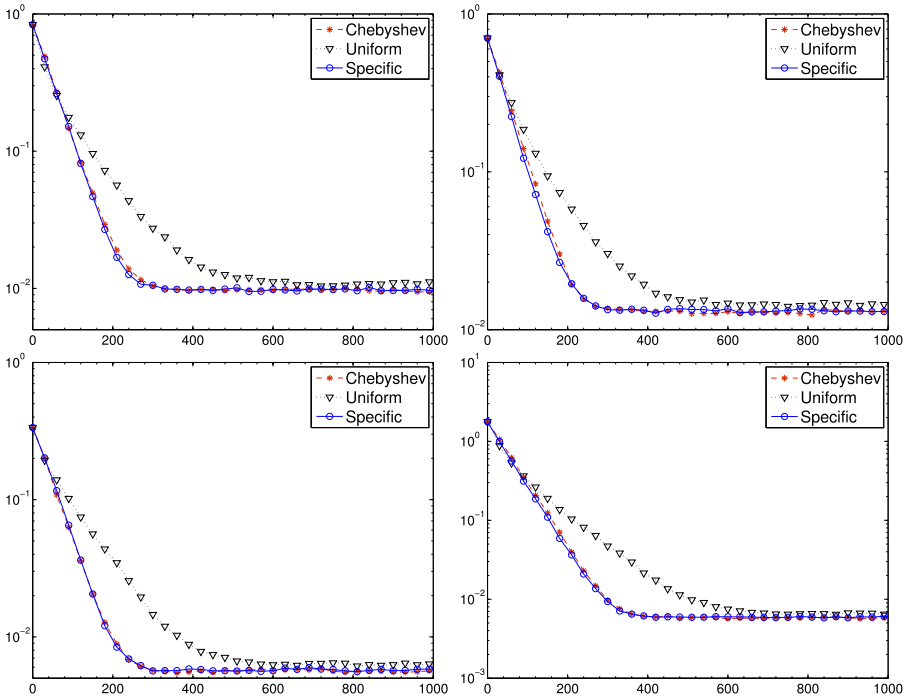

Fig. 4.8. Same as Fig. 4.1 except for the data with noise and Method II with $\tau = \sigma^2$.

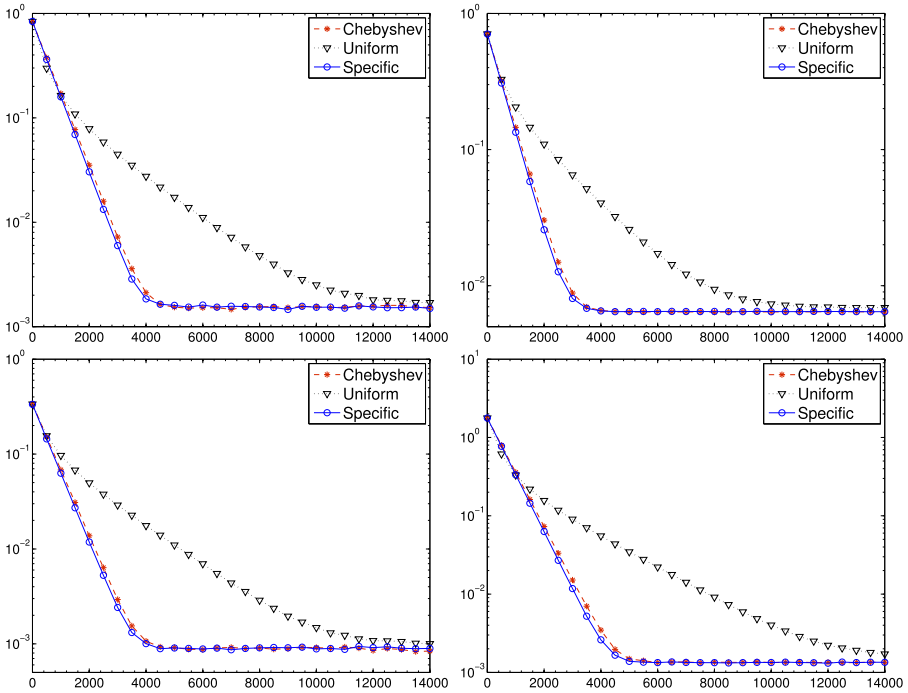

Fig. 4.9. Same as Fig. 4.1 except for the data with noise and Method II with $\tau = 20$.

Appendix A. Proof of Theorem 3.1

We start from the derivation of (3.5). By using the orthogonality of the basis ψ_i, the optimization problem (3.1) can be rewritten into the following equivalent problem

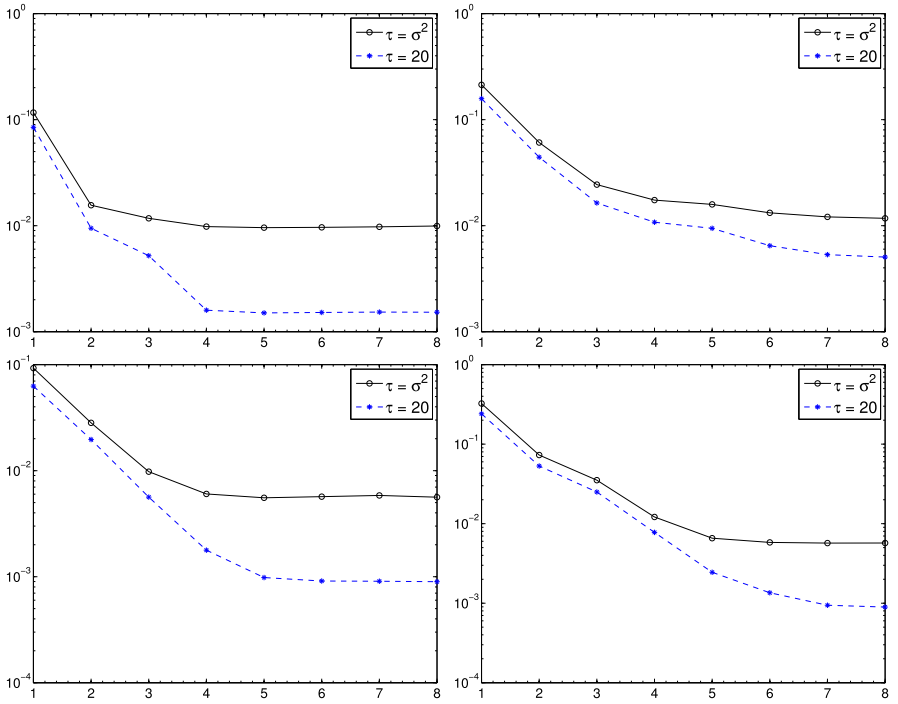

Fig. 4.10. Same as Fig. 4.2 except for the data with noise and using Method II.

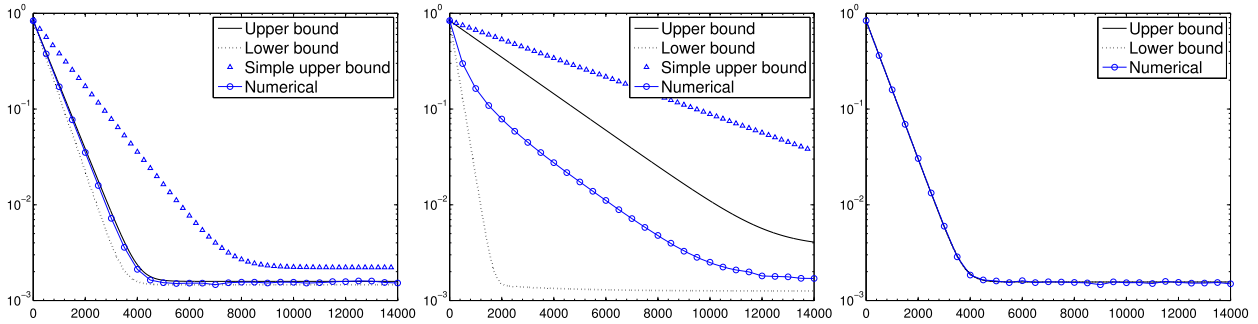

Fig. 4.11. Same as Fig. 4.3 except for the data with noise and Method II with $\tau = 20$.

$$\mathbf{c}^{(k)} = \underset{\mathbf{c}\in\mathbb{R}^N}{\operatorname{argmin}}\, \mathcal{J}_k(\mathbf{c}), \tag{A.1}$$

where

$$\mathcal{J}_k(\mathbf{c}) := \left\|\mathbf{c} - \mathbf{c}^{(k-1)}\right\|_2^2 + \frac{1}{\gamma_k}\left|\langle \mathbf{c}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle - f_k\right|^2.$$

The function $\mathcal{J}_k(\mathbf{c})$ is a convex, because its Hessian matrix is $2\left(\mathbf{I} + \frac{1}{\gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\boldsymbol{\Psi}(\mathbf{x}_k)^\top\right)$ which is positive definite. Thus the problem (A.1) is a convex quadratic programming, whose solution is unique and solves

$$\frac{\partial \mathcal{J}_k(\mathbf{c})}{\partial c_j} = 2\left(c_j - c_j^{(k-1)} + \frac{1}{\gamma_k}\left(\langle \mathbf{c}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle - f_k\right)\psi_j(\mathbf{x}_k)\right) = 0, \quad j = 1, 2, \cdots, N.$$

This can be written into vector form

$$\mathbf{c} - \mathbf{c}^{(k-1)} + \frac{1}{\gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\left(\boldsymbol{\Psi}(\mathbf{x}_k)^\top\mathbf{c} - f_k\right) = 0,$$

or,

$$\left(\mathbf{I} + \frac{1}{\gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\boldsymbol{\Psi}(\mathbf{x}_k)^\top\right)\mathbf{c} = \mathbf{c}^{(k-1)} + \frac{1}{\gamma_k}f_k\boldsymbol{\Psi}(\mathbf{x}_k). \tag{A.2}$$

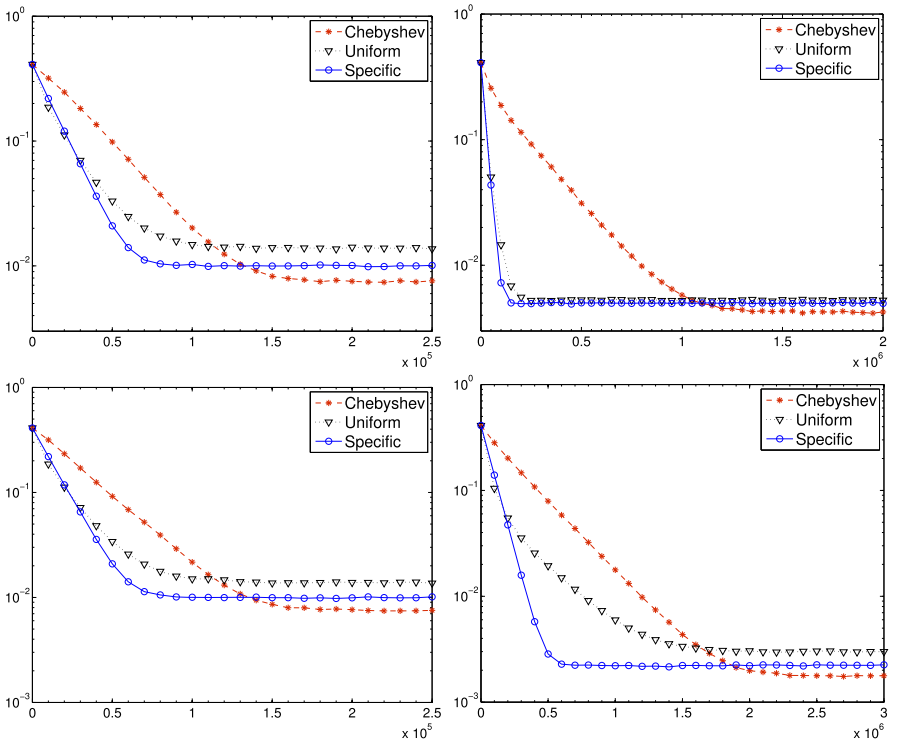

Fig. 4.12. Function approximation errors versus number of iterations for f_1 (with $a_i = 1$ and $\chi = 0.375$) in (4.3) at $d = 10$ with data noise obeying $N(0, 0.01^2)$. Top left: Method I with $\tau = \sigma^2$; Top right: Method I with $\tau = 10^4$; Bottom left: Method II with $\tau = \sigma^2$; Bottom right: Method II with $\tau = 10$.

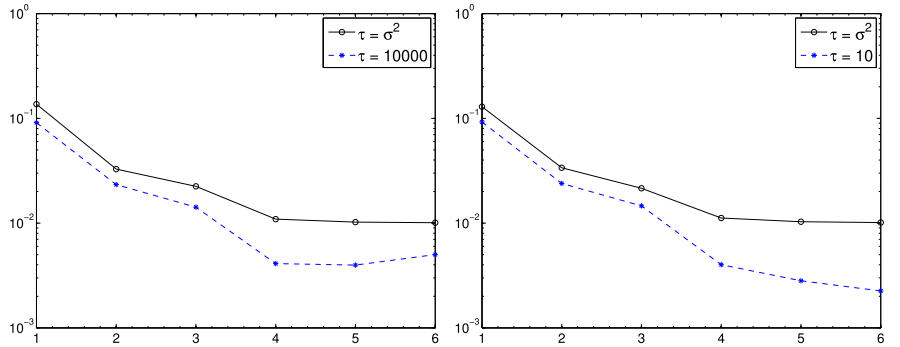

Fig. 4.13. Approximation errors given by Method I (left) and II (right) in the converged solutions with respect to the polynomial degrees.

Note that the inverse of $\mathbf{I} + \frac{1}{\gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\boldsymbol{\Psi}(\mathbf{x}_k)^\top$ is

$$\mathbf{I} - \frac{1}{\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2 + \gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\boldsymbol{\Psi}(\mathbf{x}_k)^\top.$$

Therefore, we obtain from (A.2) the solution to (A.1),

$$\mathbf{c}^{(k)} = \left(\mathbf{I} - \frac{1}{\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2 + \gamma_k}\boldsymbol{\Psi}(\mathbf{x}_k)\boldsymbol{\Psi}(\mathbf{x}_k)^\top\right)\left(\mathbf{c}^{(k-1)} + \frac{1}{\gamma_k} f_k \boldsymbol{\Psi}(\mathbf{x}_k)\right),$$

which delivers (3.5) upon further simplification. The result of (3.5) then follows directly from (3.4).

Next, we consider the limiting case of $\gamma_k = 0$ for problem (3.2). Its equivalent form is

$$\mathbf{c}^{(k)} = \underset{\mathbf{c}\in\mathbb{R}^N}{\operatorname{argmin}} \left\|\mathbf{c} - \mathbf{c}^{(k-1)}\right\|_2^2, \quad \text{subject to } \langle \mathbf{c}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle = f_k. \tag{A.3}$$

Let

$$\mathbf{c}^* = \mathbf{c}^{(k-1)} + \frac{f_k - \langle \mathbf{c}^{(k-1)}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle}{\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2}\boldsymbol{\Psi}(\mathbf{x}_k),$$

then $\langle \mathbf{c}^*, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle = f_k$, and

$$\left\|\mathbf{c} - \mathbf{c}^{(k-1)}\right\|_2^2 = \left\|\mathbf{c} - \mathbf{c}^*\right\|_2^2 + \left\|\mathbf{c}^* - \mathbf{c}^{(k-1)}\right\|_2^2, \tag{A.4}$$

where the relation

$$\begin{aligned}\langle \mathbf{c} - \mathbf{c}^*, \mathbf{c}^* - \mathbf{c}^{(k-1)}\rangle &= \frac{f_k - \langle \mathbf{c}^{(k-1)}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle}{\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2}\langle \mathbf{c} - \mathbf{c}^*, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle \\ &= \frac{f_k - \langle \mathbf{c}^{(k-1)}, \boldsymbol{\Psi}(\mathbf{x}_k)\rangle}{\|\boldsymbol{\Psi}(\mathbf{x}_k)\|_2^2}\left(f_k - f_k\right) = 0,\end{aligned}$$

has been used. It is then clear from Eq. (A.4) that $\mathbf{c}^*$ is the solution to problem (A.3).

Appendix B. Proof of Lemma 3.2

To prove Theorem 3.2, we first introduce a preliminary result. For notational convenience and without loss of generality, let us assume $\mathbf{c}^{(0)} = \mathbf{0}$. For $\mathbf{c}^{(0)} \neq \mathbf{0}$, we only need to consider the problem for approximating $f(\mathbf{x}) - \langle \mathbf{c}^{(0)}, \boldsymbol{\Psi}(\mathbf{x})\rangle$, and replace $\hat{\mathbf{c}}$ by $(\hat{\mathbf{c}} - \mathbf{c}^{(0)})$ and $\mathbf{c}^{(k)}$ by $(\mathbf{c}^{(k)} - \mathbf{c}^{(0)})$.

B.1. Preliminary

At the k-th iteration, we introduce the notation $\mathbb{E}_k$ to denote the expectation of the random sample $\mathbf{x}_k$ and the noise ϵ_k, conditioned upon the previous $(k-1)$ random variables $\{\mathbf{x}_j\}_{1\le j\le k-1}, \{\epsilon_k\}_{1\le j\le k-1}$, i.e.

$$\mathbb{E}_k(\cdot) = \mathbb{E}_{\mathbf{x}_k,\epsilon_k}(\cdot|\mathbf{x}_1, \cdots, \mathbf{x}_{k-1}, \epsilon_1, \cdots, \epsilon_{k-1}).$$

Lemma B.1. *Assume that $\mathbf{c}^{(0)} = \mathbf{0}$. The k-th iterative solution of the algorithm (3.5) satisfies*

$$\mathbb{E}(\mathbf{c}^{(k)}) = \mathbf{Q}^\top \hat{\mathbf{D}}^{(k)} \mathbf{Q}\tilde{\mathbf{c}}, \tag{B.1}$$

where $\tilde{\mathbf{c}} = \left(\tilde{c}_1, \tilde{c}_2, \cdots, \tilde{c}_N\right)^\top$ with $\tilde{c}_j = (f, \psi_j)_{L^2_{\tilde{\mu}}}$, and

$$\hat{\mathbf{D}}^{(k)} = \mathrm{diag}\{\hat{d}_1^{(k)}, \cdots, \hat{d}_N^{(k)}\}, \quad \hat{d}_i^{(k)} = \frac{1 - (1-\lambda_i)^k}{\lambda_i}.$$

Proof. By taking expectation on (3.5) we obtain

$$\mathbb{E}_k(\mathbf{c}^{(k)}) = \mathbf{c}^{(k-1)} + \tilde{\mathbf{c}} - \boldsymbol{\Sigma}\mathbf{c}^{(k-1)} = (\mathbf{I} - \boldsymbol{\Sigma})\,\mathbf{c}^{(k-1)} + \tilde{\mathbf{c}},$$

which implies

$$\mathbb{E}(\mathbf{c}^{(k)}) = (\mathbf{I} - \boldsymbol{\Sigma})\,\mathbb{E}(\mathbf{c}^{(k-1)}) + \tilde{\mathbf{c}}.$$

Upon iteratively using this recursive relation, we then have

$$\mathbb{E}(\mathbf{c}^{(k)}) = \sum_{j=1}^{k} (\mathbf{I} - \boldsymbol{\Sigma})^{j-1}\,\tilde{\mathbf{c}},$$

which further yields (B.1) via the eigenvalue decomposition of $\boldsymbol{\Sigma}$. □

B.2. Proof of Lemma 3.2

We now complete the proof of Lemma 3.2.

Proof. The proof starts from the following equality

$$\begin{aligned}\mathbb{E}_k\left(\left\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\right\|_2^2\right) &= \mathbb{E}_k\left(\left\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}} + \mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\right\|_2^2\right) \\ &= \left\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\right\|_2^2 + 2\left(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\right)^\top \mathbb{E}_k\left(\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\right) + \mathbb{E}_k\left(\left\|\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\right\|_2^2\right).\end{aligned} \tag{B.2}$$

The $\mathbb{E}_k\big(\mathbf{c}^{(k)}-\mathbf{c}^{(k-1)}\big)$ in the second term can be computed as

$$\begin{aligned}\mathbb{E}_k\big(\mathbf{c}^{(k)}-\mathbf{c}^{(k-1)}\big) &= \int_D \Big(f(\mathbf{x})-\mathcal{P}_V f(\mathbf{x})-\langle \mathbf{c}^{(k-1)}-\hat{\mathbf{c}},\boldsymbol{\Psi}(\mathbf{x})\rangle\Big)\boldsymbol{\Psi}(\mathbf{x})d\tilde{\mu}\\ &= \mathbf{e}-\boldsymbol{\Sigma}\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big),\end{aligned} \tag{B.3}$$

and the last term can be expressed as

$$\begin{aligned}\mathbb{E}_k\Big(\big\|\mathbf{c}^{(k)}-\mathbf{c}^{(k-1)}\big\|_2^2\Big) &= \int_D \mathbb{E}_{\epsilon_k}\Big(f(\mathbf{x})+\epsilon_k-\langle \mathbf{c}^{(k-1)},\boldsymbol{\Psi}(\mathbf{x})\rangle\Big)^2 d\tilde{\omega}\\ &= \int_D \Big(f(\mathbf{x})-\mathcal{P}_V f(\mathbf{x})-\langle \mathbf{c}^{(k-1)}-\hat{\mathbf{c}},\boldsymbol{\Psi}(\mathbf{x})\rangle\Big)^2 d\tilde{\omega}+\|\sigma\|_{L^2_{\tilde{\omega}}}^2\\ &= \mathcal{E}+\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)^\top\widetilde{\boldsymbol{\Sigma}}\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)-2\tilde{\mathbf{e}}^\top\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big),\end{aligned} \tag{B.4}$$

where $\widetilde{\boldsymbol{\Sigma}}:=\big((\psi_i,\psi_j)_{L^2_{\tilde{\omega}}}\big)_{N\times N}=2\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_*$, and $\tilde{\mathbf{e}}:=\big(\tilde{e}_1,\tilde{e}_2,\cdots,\tilde{e}_N\big)^\top=\mathbf{e}-\mathbf{e}_*$ with $\tilde{e}_j=(f-\mathcal{P}_V f,\psi_j)_{L^2_{\tilde{\omega}}}$.

By substituting (B.3) and (B.4) into (B.2), we then obtain

$$\begin{aligned}\mathbb{E}_k\big(\big\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\big\|_2^2\big) &= \big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)^\top\big(\mathbf{I}-\boldsymbol{\Sigma}_*\big)\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)+2\mathbf{e}_*^\top\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)+\mathcal{E}\\ &\le \big(1-\lambda^*_{\min}\big)\|\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\|_2^2+2\mathbf{e}_*^\top\big(\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big)+\mathcal{E}.\end{aligned} \tag{B.5}$$

It follows that

$$\begin{aligned}\mathbb{E}\big(\big\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\big\|_2^2\big) &\le \big(1-\lambda^*_{\min}\big)\mathbb{E}\big(\big\|\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big\|_2^2\big)+2\mathbf{e}_*^\top\big(\mathbb{E}(\mathbf{c}^{(k-1)})-\hat{\mathbf{c}}\big)+\mathcal{E}\\ &=: r_u\mathbb{E}\big(\big\|\mathbf{c}^{(k-1)}-\hat{\mathbf{c}}\big\|_2^2\big)+\mathcal{G}_{k-1},\end{aligned}$$

where

$$\mathcal{G}_{k-1}=2\mathbf{e}_*^\top\big(\mathbf{Q}^\top\hat{\mathbf{D}}^{(k-1)}\mathbf{Q}\tilde{\mathbf{c}}-\hat{\mathbf{c}}\big)+\mathcal{E},$$

and (B.1) has been used. Repeating the same operation $(k-1)$ times gives

$$\mathbb{E}\big(\big\|\mathbf{c}^{(k)}-\hat{\mathbf{c}}\big\|_2^2\big)\le\big\|\hat{\mathbf{c}}\big\|_2^2 r_u^k+\sum_{j=0}^{k-1}\mathcal{G}_j r_u^{k-j-1},$$

where the last term can be computed as

$$\begin{aligned}\sum_{j=0}^{k-1}\mathcal{G}_j r_u^{k-j-1} &= 2\mathbf{e}_*^\top\mathbf{Q}^\top\Big(\sum_{j=0}^{k-1}\hat{\mathbf{D}}^{(j)}r_u^{k-j-1}\Big)\mathbf{Q}\tilde{\mathbf{c}}+\big(\mathcal{E}-2\mathbf{e}_*^\top\hat{\mathbf{c}}\big)\sum_{j=0}^{k-1}r_u^{k-j-1}\\ &= 2\mathbf{e}_*^\top\Big(\mathbf{Q}^\top\mathbf{D}^{(k)}(r_u)\mathbf{Q}\tilde{\mathbf{c}}+\frac{1-r_u^k}{1-r_u}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{c}}\Big)+\big(\mathcal{E}-2\mathbf{e}_*^\top\hat{\mathbf{c}}\big)\frac{1-r_u^k}{1-r_u}\\ &= 2\mathbf{e}_*^\top\Big(\mathbf{Q}^\top\mathbf{D}^{(k)}(r_u)\mathbf{Q}\tilde{\mathbf{c}}+\frac{1-r_u^k}{1-r_u}\big(\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{c}}-\hat{\mathbf{c}}\big)\Big)+\frac{1-r_u^k}{\lambda^*_{\min}}\mathcal{E}\\ &= \mathcal{B}(r_u)+\frac{1-r_u^k}{\lambda^*_{\min}}\mathcal{E}.\end{aligned}$$

Here $\tilde{\mathbf{c}}=\boldsymbol{\Sigma}\hat{\mathbf{c}}+\mathbf{e}$ has been used in the last equality. This completes the proof of the upper bound. Similarly, one can obtain the lower bound. □

Appendix C. Proof of Theorem 3.4

Proof. The proof also starts from (B.2), where the last term is estimated as

$$\mathbb{E}_k\Big(\big\|\mathbf{c}^{(k)}-\mathbf{c}^{(k-1)}\big\|_2^2\Big)=\int_D\mathbb{E}_{\epsilon_k}\Big(f(\mathbf{x})+\epsilon_k-\langle\mathbf{c}^{(k-1)},\boldsymbol{\Psi}(\mathbf{x})\rangle\Big)^2\frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2+\tau}d\tilde{\mu}$$

$$\leq \int_D \Big(f(\mathbf{x}) - \mathcal{P}_V f(\mathbf{x}) - \langle \mathbf{c}^{(k-1)} - \hat{\mathbf{c}}, \boldsymbol{\Psi}(\mathbf{x}) \rangle \Big)^2 d\tilde{\mu} + \|\sigma\|^2_{L^2_{\tilde{\omega}}}$$
$$= \mathcal{E}_0 + \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big)^\top \boldsymbol{\Sigma} \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big) - 2\mathbf{e}^\top \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big),$$

with $\mathcal{E}_0 := \|f - \mathcal{P}_V f\|^2_{L^2_{\tilde{\mu}}} + \|\sigma\|^2_{L^2_{\tilde{\omega}}}$.

Hence, we have

$$\mathbb{E}_k\big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\big) = \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big)^\top (\mathbf{I} - \boldsymbol{\Sigma}) \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big) + \mathcal{E}_0$$
$$\leq (1 - \lambda_{\min}) \big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2 + \mathcal{E}_0.$$

It follows that

$$\mathbb{E}\big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\big) \leq (1 - \lambda_{\min})\, \mathbb{E}\big(\big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2\big) + \mathcal{E}_0.$$

The proof is completed after repeating the same operation $(k-1)$ times. □

Appendix D. Proof of Theorem 3.5

Proof. Under the assumption, $d\tilde{\mu} = \frac{d\omega}{N+\tau}$, and $\mathbf{e} = \mathbf{0}$, $\boldsymbol{\Sigma} = \frac{1}{N+\tau}\mathbf{I}$, the proof also starts from (B.2), with the last term estimated as

$$\mathbb{E}_k\Big(\big\|\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\big\|_2^2\Big) = \frac{1}{N+\tau} \int_D \mathbb{E}_\epsilon \Big(f(\mathbf{x}) + \epsilon_k - \langle \mathbf{c}^{(k-1)}, \boldsymbol{\Psi}(\mathbf{x}) \rangle \Big)^2 \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2 + \tau} d\omega$$
$$\leq \frac{\Psi_{\sup}}{(N+\tau)(\Psi_{\sup}+\tau)} \Big(\|f - \mathcal{P}_V f\|^2_{L^2_\omega} + \|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\|_2^2 + \|\sigma\|^2_{L^2_\omega} \Big).$$

Hence, we have

$$\mathbb{E}_k\Big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\Big) \leq r_u \big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2 + \frac{\Psi_{\sup}}{(N+\tau)(\Psi_{\sup}+\tau)} \Big(\|f - \mathcal{P}_V f\|^2_{L^2_\omega} + \|\sigma\|^2_{L^2_\omega} \Big)$$
$$=: r_u \big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2 + \frac{\Psi_{\sup} \mathcal{E}_1}{(N+\tau)(\Psi_{\sup}+\tau)}.$$

It follows that

$$\mathbb{E}\Big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\Big) \leq r_u \mathbb{E}\Big(\big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2\Big) + \frac{\Psi_{\sup} \mathcal{E}_1}{(N+\tau)(\Psi_{\sup}+\tau)}$$
$$\leq \cdots \leq r_u^k \big\|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\big\|_2^2 + \frac{\Psi_{\sup} \mathcal{E}_1}{(N+\tau)(\Psi_{\sup}+\tau)} \sum_{j=0}^{k-1} r_u^j$$
$$= r_u^k \big\|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\big\|_2^2 + \eta_u \mathcal{E}_1 (1 - r_u^k).$$

Similarly, one can prove the lower bound. □

Appendix E. Proof of Theorem 3.7

Proof. Under the assumption $\mathbf{e}_* = \mathbf{0}$, we have $\boldsymbol{\Sigma}_* = \frac{1+2\tau}{N(1+\tau)^2}\mathbf{I}$ because of $(1 + 1/\tau)d\tilde{\nu} = (1+\tau)d\tilde{\omega} = d\tilde{\mu} = \frac{d\omega}{N(1+\tau)}$. By using (B.5) we have

$$\mathbb{E}_k\big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\big) = \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big)^\top \big(\mathbf{I} - \boldsymbol{\Sigma}_*\big) \big(\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big) + \mathcal{E}$$
$$= r \big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2 + \frac{1}{N(1+\tau)^2} \Big(\|f - \mathcal{P}_V f\|^2_{L^2_\omega} + \|\sigma\|^2_{L^2_\omega} \Big).$$

It follows that

$$\mathbb{E}\big(\big\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\big\|_2^2\big) = r \mathbb{E}\big(\big\|\mathbf{c}^{(k-1)} - \hat{\mathbf{c}}\big\|_2^2\big) + \frac{1}{N(1+\tau)^2} \Big(\|f - \mathcal{P}_V f\|^2_{L^2_\omega} + \|\sigma\|^2_{L^2_\omega} \Big).$$

The proof is then completed after iteratively using the recursive relation. □

Appendix F. A general result

Theorems 3.5, 3.7 and 3.9 can be regarded as special cases of the following general result.

Theorem F.1. *Assume $\gamma_k = \gamma(\mathbf{x}_k)$, and the sampling measure is*

$$d\mu = \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2 + \gamma(\mathbf{x})}{N_*} d\omega,$$

with $N_* = N + \int_D \gamma(\mathbf{x}) d\omega$, then the k-th iterative solution of (3.5) satisfies

$$\mathbb{E}\left(\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\|_2^2\right) \le \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 r_u^k + \eta_u \left(1 - r_u^k\right) \left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2 \right),$$

and

$$\mathbb{E}\left(\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\|_2^2\right) \ge \|\mathbf{c}^{(0)} - \hat{\mathbf{c}}\|_2^2 r_\ell^k + \eta_\ell \left(1 - r_\ell^k\right) \left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2 \right),$$

where

$$r_u = 1 - \frac{2 - \Psi_u}{N_*}, \qquad \eta_u = \frac{\Psi_u}{2 - \Psi_u},$$
$$r_\ell = 1 - \frac{2 - \Psi_\ell}{N_*}, \qquad \eta_\ell = \frac{\Psi_\ell}{2 - \Psi_\ell},$$

with

$$\Psi_u := \sup_{\mathbf{x} \in D} \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2 + \gamma(\mathbf{x})}, \qquad \Psi_\ell := \inf_{\mathbf{x} \in D} \frac{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2}{\|\boldsymbol{\Psi}(\mathbf{x})\|_2^2 + \gamma(\mathbf{x})}.$$

Therefore

$$\eta_\ell \left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2 \right) \le \lim_{k \to \infty} \mathbb{E}\left(\|\mathbf{c}^{(k)} - \hat{\mathbf{c}}\|_2^2\right) \le \eta_u \left(\|f - \mathcal{P}_V f\|_{L^2_\omega}^2 + \|\sigma\|_{L^2_\omega}^2 \right).$$

Proof. The proof is similar to that of Theorem 3.5. □

References

[1] Björck Åke, Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, 1996.
[2] Léon Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010, Springer, 2010, pp. 177–186.
[3] C. Brezinski, M. Redivo-Zaglia, Convergence acceleration of Kaczmarz's method, J. Eng. Math. 93 (2015) 3–19.
[4] X. Chen, A.M. Powell, Almost sure convergence of the Kaczmarz algorithm with random measurements, J. Fourier Anal. Appl. 18 (2012) 1195–1214.
[5] E.W. Cheney, Introduction to Approximation Theory, McGraw–Hill, New York, 1966.
[6] P.J. Davis, Interpolation and Approximation, Dover, 1975.
[7] Y.C. Eldar, D. Needell, Acceleration of randomized Kaczmarz method via the Johnson–Lindenstrauss lemma, Numer. Algorithms 58 (2) (2011) 163–177.
[8] A. Genz, Testing multidimensional integration routines, in: Proc. of International Conference on Tools, Methods and Languages for Scientific and Engineering Computation, Elsevier North-Holland, Inc., 1984, pp. 81–94.
[9] J. Liu, S.J. Wright, An accelerated randomized Kaczmarz algorithm, Math. Comput. 85 (297) (2016) 153–178.
[10] D. Needell, Randomized Kaczmarz solver for noisy linear systems, BIT Numer. Math. 50 (2) (2010) 395–403.
[11] M.J.D. Powell, Approximation Theory and Methods, Cambridge University Press, 1981.
[12] T.J. Rivlin, An Introduction to the Approximation of Functions, Dover Publication Inc, 1969.
[13] Y. Shin, D. Xiu, A randomized algorithm for multivariate function approximation, SIAM J. Sci. Comput. 39 (2017) A983–A1002.
[14] T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, J. Fourier Anal. Appl. 15 (2) (2009) 262–278.
[15] T. Wallace, A. Sekmen, Acceleration of Kaczmarz using orthogonal subspace projections, in: Biomedical Sciences and Engineering Conference, BSEC, Oak Ridge, Tennessee, USA, May 21–23, 2013, IEEE, 2013, pp. 1–4.
[16] K. Wu, Y. Shin, D. Xiu, A randomized tensor quadrature method for high dimensional polynomial approximation, SIAM J. Sci. Comput. 39 (2017) A1811–A1833.
[17] A. Zouzias, N.M. Freris, Randomized extended Kaczmarz for solving least squares, SIAM J. Matrix Anal. Appl. 34 (2) (2013) 773–793.