Discrete least-squares radial basis functions approximations

Siqing Li[b,a,*], Leevan Ling[a], Ka Chun Cheung[c]

[a] *Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong*
[b] *College of Mathematics, Taiyuan University of Technology, Shanxi, China*
[c] *NVIDIA AI Technology Center (NVAITC) | NVIDIA*

Abstract

We consider discrete least-squares methods using radial basis functions. A general ℓ^2-Tikhonov regularization with W_2^m-penalty is considered. We provide error estimates that are comparable to kernel-based interpolation in cases which the function it is approximating is within and is outside of the native space of the kernel. Our proven theories concern the denseness condition of collocation points and selection of regularization parameters. In particular, the unregularized least-squares method is shown to have $W_2^\mu(\Omega)$ convergence for $\mu > d/2$ on smooth domain $\Omega \subset \mathbb{R}^d$. For any properly regularized least-squares method, the same convergence estimates hold for a large range of $\mu \geq 0$. These results are extended to the case of noisy data. Numerical demonstrations are provided to verify the theoretical results. In terms of applications, we also apply the proposed method to solve a heat equation whose initial condition has huge oscillation in the domain.

Keywords: Error estimate, Meshfree approximation, Kernel methods, Tikhonov regularization, Noisy data
2010 MSC: 65D15, 65N35, 41A63.

1. Introduction

Given a set $X = \{x_1, \dots, x_{n_X}\}$ of data points in some bounded domain $\Omega \subset \mathbb{R}^d$, on each of which function value $f_i = f(x_i) \in \mathbb{R}$, $1 \leq i \leq n_X$, was specified via some unknown function f. One important application is to reconstruct f based on data; commonly used methods include interpolation or function approximation. The function reconstruction process by radial basis functions (RBF) seeks an interpolant or approximant from the trial space

$$\mathcal{U}_{Z,\Phi,\Omega} := \operatorname{span}\{\Phi(\cdot, z_j) \,:\, z_j \in Z\},$$

defined by some translation invariant symmetric positive definite kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$ centered at a set $Z = \{z_1, \dots, z_{n_Z}\} \subset \Omega$ of trial centers. In RBF interpolation, we pick

*Corresponding author
Email addresses: `lisq3@sustech.edu.cn` (Siqing Li), `lling@hkbu.edu.hk` (Leevan Ling), `chcheung@nvidia.com` (Ka Chun Cheung)

$Z = X$ and the reconstructed interpolant of the form

$$s = \sum_{j=1}^{n_Z} \alpha_j \Phi(\cdot - z_j), \tag{1}$$

which can be uniquely determined by the solution $\vec{\alpha} = [\alpha_1, \dots, \alpha_{n_Z}]^T$ of the (square) linear system

$$\Phi(X, X)\vec{\alpha} = \vec{f}_{|X},$$

where $[\Phi(X, X)]_{i,j} = \Phi(x_i, x_j)$, $1 \le i, j \le n_X$, is the interpolation matrix of Φ on X and $\vec{f}_{|X} = [f_1, \dots, f_{n_X}]^T$ is the vector of data values. Traditional interpolation error estimates were proven under the assumption that f lies in the reproducing kernel Hilbert space, a.k.a. the native space, $\mathcal{N}_{\Phi,\Omega}$ associated with the RBF kernel Φ. Convergence estimates were measured in terms of the *fill distance* of trial centers Z as

$$h_Z := \sup_{z\in\Omega} \min_{z_j \in Z} \|z - z_j\|_{\ell_2(\mathbb{R}^d)}. \tag{2}$$

Generally speaking, smoother kernels, which imply a smoother unknown function $f \in \mathcal{N}_{\Phi,\Omega}$, yield faster convergence rates. Readers are referred to the monographs [1, 2, 3, 4] for details and to [5] for f not in the native space.

Still assuming that the data points in X and values $f_{|X}$ were given. A more general setting removes the restriction that the sets Z of trial centers and X were identical, but we do insist on having $n_Z \le n_X$ to yield different minimization problems. In comparison, there are far fewer theories for least-squares function approximation by RBF. A *continuous least-squares problem* takes the form

$$\arg\inf \left\{ \|f - s\|_{L_2(\Omega)} \, : \, s \in \mathcal{U}_{Z,\Phi,\Omega} \right\}. \tag{3}$$

Suppose $\Omega \subset \mathbb{R}^d$ is a bounded domain that satisfies a cone condition and has a Lipschitz boundary. Suppose further that the Fourier transform of the kernel Φ on $\mathbb{R}^d$ decays like

$$c_\Phi (1 + \|\omega\|_2^2)^{-m} \le \widehat{\Phi}(\omega) \le C_\Phi (1 + \|\omega\|_2^2)^{-m} \quad \text{for all } \omega \in \mathbb{R}^d, \tag{4}$$

for some $\lfloor m \rfloor > d/2$ with two positive constants $0 < c_\Phi \le C_\Phi$. Under these assumptions, the native space $\mathcal{N}_{\Phi,\Omega}$ is norm equivalent to the Sobolev space $W_2^m(\Omega)$. If $f \in W_2^m(\Omega)$, then it was shown in [6] and [7, Proposition 3.2] that

$$\min_{s \in \mathcal{U}_{Z,\Phi,\Omega}} \|f - s\|_{L_2(\Omega)} \le C h_Z^m \|f\|_{W_2^m(\Omega)}$$

for some $C > 0$ independent of Z and f.

One may also consider the *discrete least-squares problem*

$$\arg\inf \left\{ \|f - s\|_{\ell_2(X)}^2 := \sum_{i=1}^{n_X} [f(x_i) - s(x_i)]^2 \, : \, s \in \mathcal{U}_{Z,\Phi,\Omega} \right\}, \tag{5}$$

which can be solved by the solution of the overdetermined linear system

$$\Phi(X,Z)\vec{\alpha} = \vec{f}_{|X},$$

where $[\Phi(X,Z)]_{i,j} = \Phi(x_i, z_j)$, $1 \le i \le n_X$ and $1 \le j \le n_Z$ comprise the collocation matrix of Φ on Z at X. Under the same assumptions, [8, Theorem 2.3] implies an error estimate for the discrete error norm

$$\min_{s\in\mathcal{U}_{Z,\Phi,\Omega}} \|f-s\|_{\ell_2(X)} \le C n_X^{1/2} \rho_X^{d/2} h_Z^m \|f\|_{W_2^m(\Omega)}$$

for some constant $C > 0$ independent of X, Z, and f. Here, $\rho_X = h_X/q_X$ is the *mesh ratio* of X defined by its fill distance as in (2) and the *separation distance*

$$q_X := \frac{1}{2}\min_{i\neq j} \|x_i - x_j\|_{\ell_2(\mathbb{R}^d)}.$$

A trivial application of discrete least-squares approximation is on parabolic PDEs. If one consider the regularity estimates [9] for parabolic problems

$$u_t + \mathcal{L}u = f \in L^2(0,T;L^2(\Omega))$$

with homogenous boundary condition and initial condition $u = g \in H_0^1(\Omega)$, i.e.,

$$\begin{aligned}\operatorname*{ess\,sup}_{0\le t\le T} \|u(t)\|_{H_0^1(\Omega)} + \|u\|_{L^2(0,T;H^2(\Omega))} + \|u_t\|_{L^2(0,T;L^2(\Omega))}\\ \le C\big(\|f\|_{L^2(0,T;L^2(\Omega))} + \|g\|_{H_0^1(\Omega)}\big),\end{aligned}$$

it is desirable to have a H^1-approximation for the initial conditions g in the form of (1). Then, the evolution of the time dependent coefficients $\alpha(t)$ can be obtained by solving ODE systems. In this paper, we derive some error estimates in continuous norms for the discrete least-squares function approximation by radial basis functions.

2. Stability of discrete least-squares problems

We consider a more general setting of the discrete least-squares problem with a smoothness penalty given by the native space norm of some kernel Φ satisfying (4). For any $s \in \mathcal{U}_{Z,\Phi,\Omega}$ in the form of (1), its native space norm is given by

$$\|s\|^2_{\mathcal{N}_{\Phi,\Omega}} = \vec{\alpha}^T \Phi(Z,Z)\vec{\alpha}.$$

We still assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain satisfying a cone condition. Now, for any regularization parameter $\lambda \ge 0$, we define the corresponding regularized least-squares approximant to data $(X, \vec{f}_{|X})$ by

$$s_\lambda := \operatorname*{arg\,inf}_{s\in\mathcal{U}_{Z,\Phi,\Omega}} \Big(\|s-f\|^2_{\ell_2(X)} + \lambda^2\|s\|^2_{\mathcal{N}_{\Phi,\Omega}}\Big). \tag{6}$$

In this paper, we apply the native space norm in the regularization term, which is related to kriging [10] in statistics. In statistics, other regularization techniques also play important roles in applications. For example, the ℓ_1 and ℓ_2 penalty are used in Lasso regularized method [11] and ridge regression [12], respectively. Combination of ℓ_1 and ℓ_2 yields the Elastic Net regularization method [13]. Recently, a new regularized method called principal components lasso (pcLasso) was proposed [14].

In a recently published monograph [15, Section 8.6], it was proven that the solution s_λ of (6) is stable with respect to the approximand:

$$\|s_\lambda - f\|_{\ell_2(X)}^2 + \lambda^2 \|s_\lambda\|_{\mathcal{N}_{\Phi,\Omega}}^2 \le (1 + \lambda^2/n_X)\|f\|_{\mathcal{N}_{\Phi,\Omega}}^2 \quad \text{for all } f \in \mathcal{N}_{\Phi,\Omega},$$

and to the regularization parameter:

$$\|s_\lambda\|_{\mathcal{N}_{\Phi,\Omega}} \le \|s_0\|_{\mathcal{N}_{\Phi,\Omega}} \quad \text{for all } \lambda \ge 0,$$

where s_0 denotes the unregularized solution with $\lambda = 0$. Moreover, the regularized solution s_λ converges to s_0 in the sense that

$$\|s_\lambda - s_0\|_{\mathcal{N}_{\Phi,\Omega}}^2 = \mathcal{O}(\lambda^2) \quad \text{and} \quad \|s_\lambda - s_0\|_{\ell_2(X)}^2 = o(n_X \lambda^2) \quad \text{for all } \lambda \searrow 0.$$

To obtain convergent estimates to f, our first goal is to derive a stability estimate for (6) within the trial space $\mathcal{U}_{Z,\Phi,\Omega}$. By the sampling inequality in [16, Theorem 3.3], there exists some constant that depends only on Ω, m and μ such that the followings hold:

$$\|s\|_{W_2^\mu(\Omega)} \le C_{\Omega,m,\mu} \left(h_X^{d/2-\mu} \|s\|_{\ell_2(X)} + h_X^{m-\mu} \|s\|_{W_2^m(\Omega)} \right) \quad \text{for } 0 \le \mu \le m, \tag{7}$$

for any $s \in W_2^m(\Omega)$ with $m > d/2$ and any discrete sets $X \subset \Omega$ with sufficiently small mesh norm h_X. By the inequality $(a+b)^2 \le 2(a^2+b^2)$ for any $a, b \ge 0$, we have

$$\begin{aligned} \|s\|_{W_2^\mu(\Omega)}^2 &\le Ch_X^{d-2\mu} \left(\|s\|_{\ell_2(X)}^2 + h_X^{2m-d} \|s\|_{W_2^m(\Omega)}^2 \right) \\ &\le Ch_X^{d-2\mu} \left(\|s\|_{\ell_2(X)}^2 + \lambda^2 \|s\|_{W_2^m(\Omega)}^2 + (h_X^{2m-d} - \lambda^2)_+ \|s\|_{W_2^m(\Omega)}^2 \right) \end{aligned}$$

for $0 \le \mu \le m$ with $(x)_+ = \max(x, 0)$. Denote the critical regularization parameter

$$\lambda_* := h_X^{m-d/2}.$$

For $\lambda \ge \lambda_*$, we immediately see that there is a constant depending on Ω, m, and μ such that

$$\|s\|_{W_2^\mu(\Omega)}^2 \le Ch_X^{d-2\mu} \left(\|s\|_{\ell_2(X)}^2 + \lambda^2 \|s\|_{W_2^m(\Omega)}^2 \right) \quad \text{for } 0 \le \mu \le m, \tag{8}$$

holds for any $s \in W_2^m(\Omega)$.

For $0 \le \lambda < \lambda^*$, we focus only on trial function $s \in \mathcal{U}_{Z,\Phi,\Omega} \subseteq \mathcal{N}_{\Phi,\Omega} = W_2^m(\Omega)$. Within the trial space $\mathcal{U}_{Z,\Phi,\Omega}$, we have a Bernstein inequality [17, Lemma 3.2], which states that there is a constant depending only on Ω, Φ, m, and μ such that

$$\|s\|_{W_2^m(\Omega)} \le C_{\Omega,\Phi,m,\mu}\, q_Z^{-m+\mu} \|s\|_{W_2^\mu(\Omega)} \quad \text{for } d/2 < \lfloor \mu \rfloor,\ \mu \le m \tag{9}$$

holds for all finite sets $Z \subset \Omega$ with separation distance q_Z. Although the original lemma there requires the integer index to satisfy $d/2 < \mu \le m$, a closer inspection of the proof shows that the updated condition in (9) on μ yields the appropriate extension to fractional orders. Thus, we have

$$\|s\|^2_{W_2^\mu(\Omega)} \leq Ch_X^{d-2\mu}\left(\|s\|^2_{\ell_2(X)} + \lambda^2\|s\|^2_{W_2^m(\Omega)} + (h_X^{2m-d} - \lambda^2)_+ q_Z^{-2m+2\mu}\|s\|^2_{W_2^\mu(\Omega)}\right).$$

For sufficiently dense X, in the sense that

$$C_{\Omega,\Phi,m,\mu} h_X^{d-2\mu}(h_X^{2m-d} - \lambda^2)_+ q_Z^{-2m+2\mu} \leq \frac{1}{2}, \tag{10}$$

we obtain the same stability estimate for $0 \leq \lambda < \lambda^*$ in the same form of (8) that holds for all $s \in \mathcal{U}_{Z,\Phi,\Omega}$. We summarize the result as follows:

Lemma 1. *Let a kernel $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ satisfying* (4) *with smoothness $\lfloor m \rfloor > d/2$ be given. Suppose $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain satisfying an interior cone condition. Let $Z \subset \Omega$ be a discrete set of trial centers with separation distance q_Z. Let $X \subset \Omega$ be another discrete set of collocation points with fill distance h_X. Then, there exists a constant depending on Ω, Φ, m, and μ such that*

$$\|s\|^2_{W_2^\mu(\Omega)} \leq Ch_X^{d-2\mu}\left(\|s\|^2_{\ell_2(X)} + \lambda^2\|s\|^2_{W_2^m(\Omega)}\right) \quad \textit{for any } \lambda \geq 0$$

holds for all $s \in \mathcal{U}_{Z,\Phi,\Omega}$ in two circumstances:

- *for $\lambda \geq h_X^{m-d/2}$, under the conditions*

$$0 \leq \mu \leq m,$$

  and the set X being sufficiently dense for (7) *to hold;*

- *or, for $0 \leq \lambda < h_X^{m-d/2}$, under the conditions*

$$d/2 < \lfloor \mu \rfloor, \ \mu \leq m,$$

  and the set X being dense enough with respect to Z and λ for (10) *to also hold.*

3. Error estimates

We consider cases $f \in W_2^\beta(\Omega)$ with $\lfloor \beta \rfloor > d/2$ and $\beta \leq m$; i.e., f is not in the native space of Φ if $\beta < m$.

Let $s_f \in \mathcal{U}_{Z,\Phi,\Omega}$ be the interpolant of f on Z in the trial space $\mathcal{U}_Z$. For any appropriate value of μ as specified in Lemma 1, we can manipulate the $W_2^\mu(\Omega)$ approximation error of $s_\lambda \in \mathcal{U}_{Z,\Phi,\Omega}$ in (6) by a chain of comparisons:

$$\begin{aligned}
\|s_\lambda - f\|^2_{W_2^\mu(\Omega)} &\le \|s_\lambda - s_f\|^2_{W_2^\mu(\Omega)} + \|s_f - f\|^2_{W_2^\mu(\Omega)} \\
&\le C h_X^{d-2\mu}\Big(\|s_\lambda - s_f\|^2_{\ell_2(X)} + \lambda^2 \|s_\lambda - s_f\|^2_{W_2^m(\Omega)}\Big) + \|s_f - f\|^2_{W_2^\mu(\Omega)} \\
&\le C h_X^{d-2\mu}\Big(\|s_\lambda - f\|^2_{\ell_2(X)} + \lambda^2 \|s_\lambda\|^2_{W_2^m(\Omega)} \\
&\qquad + \|s_f - f\|^2_{\ell_2(X)} + \lambda^2 \|s_f\|^2_{W_2^m(\Omega)}\Big) + \|s_f - f\|^2_{W_2^\mu(\Omega)}.
\end{aligned}$$

By the minimization property of s_λ, we obtain an upper bound in terms of the interpolant

$$\|s_\lambda - f\|^2_{W_2^\mu(\Omega)} \le 2C h_X^{d-2\mu}\Big(\|s_f - f\|^2_{\ell_2(X)} + \lambda^2 \|s_f\|^2_{W_2^m(\Omega)}\Big) + \|s_f - f\|^2_{W_2^\mu(\Omega)} \qquad (11)$$

for some positive constant $C = C_{\Omega,\Phi,m,\mu}$. Error estimates can now be derived based on the approximation power of RBF interpolants. For this, we rely on the results in [5] that some improved error bounds found in [7].

To begin, we bound the two continuous norms in (11). For any $0 \le \mu \le \beta \le m$, [5, Theorem 4.2] states that

$$\|s_f - f\|_{W_2^\mu(\Omega)} \le C h_Z^{\beta-\mu} \rho_Z^{m-\mu} \|f\|_{W_2^\beta(\Omega)},$$

and using [17, Lemma 3.2] followed by [5, Corollary 4.3] yields that

$$\begin{aligned}
\|s_f\|_{W_2^m(\Omega)} &\le C q_Z^{-m+\beta} \|s_f\|_{W_2^\beta(\Omega)} \qquad \text{for } \lfloor \beta \rfloor > d/2 \\
&\le C q_Z^{-m+\beta}(1 + C' \rho_Z^{m-\beta}) \|f\|_{W_2^\beta(\Omega)}.
\end{aligned}$$

It remains to handle the discrete norm in (11). If $\beta = m$ and $f \in W_2^m(\Omega)$, the zero lemma in [8, Theorem 2.3] ensures that

$$\begin{aligned}
\|s_f - f\|_{\ell_2(X)} &\le C n_X^{1/2} \rho_X^{d/2} h_Z^m \|s_f - f\|_{W_2^m(\Omega)} \\
&\le C n_X^{1/2} \rho_X^{d/2} h_Z^m \|f\|_{W_2^m(\Omega)}.
\end{aligned}$$

The last inequality follows from the optimality of s_f and norm equivalence between $\mathcal{N}_{\Phi,\Omega}$ and $W_2^m(\Omega)$.

For $f \in W_2^\beta(\Omega)$, with $\beta \in \mathbb{N}$ and $d/2 < \beta \le m$, not in the native space, we have

$$\|s_f - f\|_{\ell_2(X)} \le C n_X^{1/2} \rho_X^{d/2} \rho_Z^{m-\beta} h_Z^\beta \|f\|_{C^\beta(\bar{\Omega})}$$

by [7, Corollary 3.11]. Using the standard estimates $n_X \le C q_X^{-d} = C \rho_X^d h_X^{-d}$ and $\|f\|_{W_2^\beta(\Omega)} \le \|f\|_{C^\beta(\bar{\Omega})}$, we can complete the proof of the following theorem.

Theorem 1. *Suppose all of the assumptions in Lemma 1 hold. Let s_λ be the discrete least-squares solution of* (6)*, then*

- *if* $f \in W_2^m(\Omega)$,

$$\|s_\lambda - f\|_{W_2^\mu(\Omega)} \leq C\Big(\rho_X^d h_X^{-\mu} h_Z^m + h_X^{d/2-\mu}\lambda + \rho_Z^{m-\mu} h_Z^{m-\mu}\Big)\|f\|_{W_2^m(\Omega)},$$

- *or, if* $f \in W_2^\beta(\Omega)$ *with* $\beta \in \mathbb{N}$, $d/2 < \lfloor \beta \rfloor$ *and* $\beta \leq m$,

$$\begin{aligned}\|s_\lambda - f\|_{W_2^\mu(\Omega)} \leq\ & C\Big(\rho_X^d \rho_Z^{m-\beta} h_X^{-\mu} h_Z^\beta \\ & + \rho_Z^{2m-2\beta} h_X^{d/2-\mu} h_Z^{\beta-m}\lambda + \rho_Z^{m-\mu} h_Z^{\beta-\mu}\Big)\|f\|_{C^\beta(\bar{\Omega})},\end{aligned}$$

holds for some constants depending on Ω, Φ, m, μ, *and* β.

The least amount of regularization used in (6) that does not impose any denseness requirement on the collocation set X is λ_*. The resulting approximation power is comparable to RBF interpolation.

Corollary 1. *Let* s_{λ_*} *be the regularized least-squares solution of* (6) *with regularization parameter* $\lambda_* = h_X^{m-d/2}$. *Under the assumption of Lemma 1, there exists some constant depending on* Ω, Φ, *and* m *such that*

$$\|s_{\lambda_*} - f\|_{W_2^\mu(\Omega)} \leq C\Big(\rho_X^d h_X^{-\mu} h_Z^m + h_X^{m-\mu} + \rho_Z^{m-\mu} h_Z^{m-\mu}\Big)\|f\|_{W_2^m(\Omega)} \quad \text{for } 0 \leq \mu \leq m,$$

holds for all $f \in W_2^m(\Omega)$. *In particular, we have*

$$\|s_{\lambda_*} - f\|_{L_2(\Omega)} \leq C\Big(\big(\rho_X^d + \rho_Z^m\big)h_Z^m + h_X^m\Big)\|f\|_{W_2^m(\Omega)}.$$

3.1. Noisy data

When data are contaminated by measurement or some other sort of error, we only have some noisy data $\vec{f}_{\delta|X}$ to work with instead of using the exact data value $\vec{f}_{|X}$ in (6), the regularized solution with noisy data takes the form

$$s_{\delta,\lambda} := \underset{s \in \mathcal{U}_{Z,\Phi,\Omega}}{\arg\inf} \Big(\|s - f_\delta\|^2_{\ell_2(X)} + \lambda^2 \|s\|^2_{\mathcal{N}_{\Phi,\Omega}}\Big). \tag{12}$$

Here, we do not require that f_δ be a function and only its values at X are required. Following the same line of logic in deriving (11), we have

$$\begin{aligned}\|s_{\delta,\lambda} - f\|^2_{W_2^\mu(\Omega)} \leq\ & C h_X^{d-2\mu}\Big(\|s_{\delta,\lambda} - f\|^2_{\ell_2(X)} + \lambda^2\|s_{\delta,\lambda}\|^2_{W_2^m(\Omega)} \\ & + \|s_f - f\|^2_{\ell_2(X)} + \lambda^2\|s_f\|^2_{W_2^m(\Omega)}\Big) + \|s_f - f\|^2_{W_2^\mu(\Omega)} \\ \leq\ & 2C h_X^{d-2\mu}\Big(\|s_f - f\|^2_{\ell_2(X)} + \lambda^2\|s_f\|^2_{W_2^m(\Omega)}\Big) + \|s_f - f\|^2_{W_2^\mu(\Omega)} \\ & + 2C h_X^{d-2\mu}\|f - f_\delta\|^2_{\ell_2(X)},\end{aligned}$$

that differs from (11) only in the last term.

Corollary 2. *Suppose all of the assumptions in Lemma 1 hold. Let $s_{\delta,\lambda}$ be the discrete least-squares solution of* (12) *with noisy data $\vec{f}_{\delta|X}$, then the error estimates in Theorem 1 hold with an extra term $Ch_X^{d/2-\mu}\|f-f_\delta\|_{\ell_2(X)}$ added to the upper bounds for some constant C depending on Ω, Φ, m, and ν.*

We now focus on the case that $f \in W_2^m(\Omega)$ is in the native space of Φ, and the pointwise absolute error at each data point is bounded. Following the general framework in [18, Section 8], we assume the existence of a noise level $\delta_\infty \geq 0$ relative to the Sobolev norm of f such that

$$\max_{x\in X}\left|f(x) - f_\delta(x)\right| \leq \delta_\infty \|f\|_{W_2^m(\Omega)} \tag{13}$$

for any set $X \in \Omega$. Then,

$$\|f - f_\delta\|_{\ell_2(X)} \leq n_X^{1/2}\delta_\infty\|f\|_{W_2^m(\Omega)} \leq q_X^{-d/2}\delta_\infty\|f\|_{W_2^m(\Omega)}$$

and, hence, Corollary 2 yields

$$\|s_{\delta,\lambda} - f\|_{W_2^\mu(\Omega)} \quad \leq \quad C\Big(\rho_X^d h_X^{-\mu} h_Z^m + h_X^{d/2-\mu}\lambda + \rho_Z^{m-\mu} h_Z^{m-\mu} + \rho_X^{d/2} h_X^{-\mu}\delta_\infty\Big)\|f\|_{W_2^m(\Omega)}.$$

This suggests a regularization strategy by using

$$0 \leq \lambda \leq \lambda_\delta := h_X^{-d/2}\delta_\infty = \lambda_* h_X^{-m}\delta_\infty,$$

for any allowed value of μ.

If we take $\lambda = 0$, we fail the theoretical requirement for $L_2(\Omega)$ convergence; this will be studied numerically in the next section. If we take $\lambda = \lambda_\delta$, Lemma 1 allows $0 \leq \mu \leq m$ and the density requirement on data points X is independent of trial centers Z for significantly large noise in the sense that $\delta_\infty \geq h_X^m$. In application, this means that we can take $n_Z \lesssim n_X$. Due to the presence of the ρ_Z term in the error estimates in Theorem 1, we are highly motivated to choose $Z \subset \Omega$ quasi-uniformly, or uniformly if possible. The $L^2(\Omega)$ error bound for the regularized least-squares solution of (12) with regularization parameter λ_δ will be

$$\|s_{\delta,\lambda_\delta} - f\|_{L_2(\Omega)} \quad \leq \quad C\Big(\rho_X^d h_Z^m + \rho_Z^m h_Z^m + \rho_X^{d/2}\delta_\infty\Big)\|f\|_{W_2^m(\Omega)}.$$

The presence of noise affects accuracy linearly depending on some constant that depends Ω, Φ, m, and the mesh ratio ρ_X of X. In other words, admissible noise in data is comparable to the interpolation error, i.e., $\delta_\infty = \mathcal{O}(h_Z^m)$, for RBF least-squares methods.

4. Numerical examples

Note that the native space norm of the trial function in $\mathcal{U}_{Z,\Phi,\Omega}$ in the form of (1) is given by

$$\|s\|_{\mathcal{N}_{\Phi,\Omega}} = \vec{\alpha}^T\Phi(Z,Z)\vec{\alpha}.$$

Then, it is straightforward to show that the unique solution to (6) $\big($or to (12)$\big)$ is given by the normal equation

$$\Big(\Phi(X,Z)^T\Phi(X,Z)+\lambda^2\Phi(Z,Z)\Big)\vec{\alpha}=\Phi(X,Z)^T\vec{f}_{|X}\quad\Big(\text{or }\Phi(X,Z)^T\vec{f}_{\delta|X}\Big),$$

for identifying the unknown coefficients of the approximant s_λ (or $s_{\delta,\lambda}$) in the form of (1). To avoid worsening the problem of the ill-condition, we compute a stabilized Cholesky decomposition of

$$\Big(\Phi(Z,Z)+\epsilon I_{n_Z}\Big)=LL^T,$$

for some smallest possible $\epsilon\geq 0$ so that the Cholesky algorithm does not crash due to non-positive definiteness. This stabilization is implemented to safeguard the robustness of our algorithm when a large order of smoothness m is used. With L in hand, we can recast the normal equation as

$$\begin{bmatrix}\Phi(X,Z)\\ \lambda L^T\end{bmatrix}\vec{\alpha}=\begin{bmatrix}\vec{f}_{|X}\\ 0_{n_Z}\end{bmatrix}\quad\Big(\text{or }\vec{f}_{\delta|X}\Big),$$

which becomes an overdetermined system that can be solved by standard linear solvers. If any $\epsilon>0$ is required in the Cholesky decomposition, the least-squares problem above is equivalent to adding an extra smoothness term $\epsilon\lambda^2\|\vec{\alpha}\|$ in the regularization for the sake of numerical stability.

We provide some demonstrations to verify some aspects of the proven theories. All numerical tests use the standard Whittle-Matérn-Sobolev kernel

$$\Phi_m(x):=\|x\|_{\ell_2(\mathbb{R}^d)}^{m-d/2}\mathcal{K}_{m-d/2}(\|x\|_{\ell_2(\mathbb{R}^d)})\quad\text{for all }x\in\mathbb{R}^d,$$

which satisfies (4) with exact Fourier transform $(1+\|\omega\|_2^2)^{-m}$. We test $\lfloor m\rfloor>d/2$ as required in Lemma 1. Note that one can always scale the kernel by $\|x\|_2\leftarrow\epsilon\|x\|_2$ with any $\epsilon>0$. Instead of scaling, we normalize test functions in $\Omega=[-1,1]^2$ for easy comparison without fewer parameters. In particular, we consider standard test functions

$$f_1(x,y)=\texttt{peaks}(3x,3y),$$

$$f_2(x,y)=\texttt{franke}(x/2+1/2,y/2+1/2),$$

and functions in the form of $(x+y)^p$, which is in $W_2^2(\Omega)$ for any $p>3/2$, aiming to test the limit cases in theories. In our implementation, we evaluate the function and its derivatives via a smoothing by machine epsilon $\varepsilon_{mach}=2.2\times10^{-16}$ and use

$$f_3(x,y):=\max(\varepsilon_{mach},(x+y)^{3/2}).$$

All reported errors are estimated on a 51^2 uniform grid.

Example 1: Oversample ratio

The least-squares stability in Lemma 1 holds under the condition when the fill distance h_X of collocation is small enough to fulfill (7) and also (10) if $\lambda<\lambda_*=h_X^{m-d/2}$. Although these denseness requirements were not known explicitly in practice, practitioners have seldom had stability problems with RBF least-squares approximation, which hints that (7) and (10) can be fulfilled easily by some oversampling such that $n_X > n_Z$.

To test the requirement (10), we set up an unconventional asymmetric interpolation problem. We take $n_Z = 21^2$ regular nodes as trial centers Z. A set of $n_X = n_Z$ collocation points X were quasi-random generated by a Halton sequence. The resulting linear system (6) is therefore an asymmetric square system. Note that, with $\lambda = 0$, the solvability of such a system is not guaranteed by any theories. Figure 1(a) shows the resulting interpolant, which fails badly near the boundary where no collocation data is presented. The colorbar there indicates the absolute error. In Figure 1(b), the regularized fit using $\lambda = \lambda_*$ is shown. Appropriate regularization is necessary in cases of insufficient oversampling.

Next, we include more points from the Halton sequence to obtain an expanded set of $n_X = 529$ collocation points, which yields an oversample ratio of $n_X : n_Z = 1.2$. The corresponding least-squares approximants are shown in Figure 2; there is no observable difference between the unregularized and regularized fit. The former is hence preferred due to its lower computational cost. The same observations can be made for the other two test functions, and thus we omit their results from this presentation.

Example 2: h_Z–Convergence

To test convergence in the case of scattered data points, we use a Halton sequence to create the set $7^2 \le n_Z \le 21^2$ of trial centers Z. All tested sets Z are parts of the same Halton sequence with the same starting point, and hence, are nested. The oversample ratio is fixed at $n_X = 1.2\, n_Z$ and we generate the collocation sets X similarly with a different starting point. With a large enough *skip*, resulting points in X are distinct from Z. Then, we seek for the unregularized ($\lambda = 0$) and λ_*-regularized least-squares approximants by solving (6). For easy comparison between the different test functions, we report the relative $W_2^2(\Omega)$ and $L_2(\Omega)$ errors

$$\mathcal{E}_q = \frac{\|s_\lambda - f\|_{W_2^q(\Omega)}}{\|f\|_{W_2^q(\Omega)}}, \quad q \in \{0, 2\} \tag{14}$$

in Figure 3 against $h_Z \approx (4/n_Z)^{1/2}$. We point out that our proven theories do not apply to the $L_2(\Omega)$ convergence of the unregularized cases, and yet, we can see that the convergence profiles of the unregularized and regularized approximants are analogous. Under close inspection, the λ_*-regularized least-squares approximations display slightly more stable convergence behavior.

Test by test, the `peaks` function f_1 allows convergence faster than predicted. The `franke` function f_2 is also infinitely smooth and results in similar situations of superconvergence [19]. Most obviously, the predicted $W_2^2(\Omega)$ divergence rate for the case $m = 2$ should be $(h_X/h_Z)^{-2} \approx 1.2^{-2}$ by Corollary 1 and convergence behavior can be clearly seen in Figure 3(a)–(d). Now, we consider f_3 for a test of functions outside the native space. The $W_2^2(\Omega)$ error remains constant for all tested orders m of the kernel and sets Z, whereas the regularized least-squares approximations converge with order 2 even for kernels with higher order.

Example 3: h_X–divergence

We further investigate a possible h_X divergence *in the upper bound of the error estimate* suggested by Corollary 1 when $\mu > 0$. We focus on $\mu = 2$ and study the $W_2^2(\Omega)$ relative errors $\mathcal{E}_2$. For ease of reproduction, we only consider uniform sets Z and X. The number of trial centers is fixed at $n_Z = 21^2$ with $h_Z \approx 10^{-1}$. The number of collocations is $n_X \approx \gamma n_Z$ for $1.1 \leq \gamma \leq 10$. Because f_1 and all least-squares approximants are smooth, having some testing n_X larger than the number of error evaluation points does not affect the results. Among all three test functions, the `peaks` reveals the most noticeable trends in h_X-divergence; see Figure 4 for the results from kernel smoothness orders $m = 3$ and $m = 4$. Despite observing the trends, the rate of divergence is much below the predicted rate of -2. Although we did not further investigate larger oversampling ratio, we conjecture that h_X-divergence, if present, is illegible in practice.

Example 4: Noisy data

Using the $n_Z = 21^2$ uniform trial centers and uniform collocation points with oversample ratio $n_X = 1.2\, n_Z$ as in the last example, we test the linear divergence results with respect to noise in Corollary 2. We focus on reconstructing the first two smooth functions using kernels with $m = 2$. Noisy function values at the collocation points in X were generated by

$$f_\delta = f + \text{Unif}(-\xi,\, \xi) \quad \text{with } 10^{-5} \leq \xi \leq 15.$$

Figure 5 shows various relative errors $\mathcal{E}_q$ as in (14) for $q = 0, 1, 2$ of the least-squares reconstructions (12) of the `peaks` f_1 and `franke` f_2 functions with $\lambda = 0$ and $\lambda = \lambda_\delta$. To identify the parameter λ_∞ in (13), we must first have an estimate for $\|f\|_{W_2^m(\Omega)}$, which is a nontrivial task in general. For the sake of demonstration, we make such an estimation based on the same procedure of error evaluation. We once again see that unregularized and regularized least-squares approximations yield similar accuracy and the expected linear divergence with respect to noise can be observed.

Example 5: An application to heat equations

As mentioned in the introduction, the proposed approximation methods allow us to obtain approximate initial conditions in the trial space. From there, we can update with a method of lines or some Rothe's time stepping methods. Consider the following unforced heat equation in $\Omega = [-1, 1]^2$:

$$\begin{aligned} u_t(x, y, t) &= \Delta u(x, y, t) && \text{for } (x, y, t) \in \Omega \times (0, T], \\ u(x, y, t) &= 0 && \text{for } x \in \partial\Omega,\ t \in [0, T], \\ u(x, y, 0) &= g(x, y) && \text{for } x \in \Omega. \end{aligned} \tag{15}$$

We set the initial function as

$$g(x, y) = \cos(\pi x/2) \cos(\pi y/2) \arctan(50(x - y)),$$

which "jumps" along the line $x = y$ and is compatible with the zero boundary condition.

For demonstration purposes only, we set up a standard RBF method of lines [20] for (15), in which we treat the RBF coefficients as time dependent. Whittle-Matérn-Sobolev kernel with smoothness order $m = 3$ is used in this demonstration. We use sets interior and boundary collocation points $\mathcal{X} \subset \Omega$ and $\mathcal{Y} \subset \partial\Omega$, and simply take $Z = \mathcal{X} \cup \mathcal{Y}$ as trial centers. By firstly enforcing boundary conditions at $\mathcal{Y}$, we obtain an $n_{\mathcal{X}} \times n_{\mathcal{X}}$ ODE system, which is solved by the MATLAB ODE45 solver.

To well capture this initial condition with rapid changes, we use quasi-uniform (i.e., without refinement near the jump) data points $X \subset \overline{\Omega}$ with $h_X \approx h_Z/3$ to approximate g and, hence, obtain the initial value for the above set up ODE system. In Figure 6, we present the numerical solutions (with $n_Z \approx 600$ and $h_Z \approx 0.08$) corresponding to the unregularized ($\lambda = 0$) and λ^*-regularized approximants of initial condition g. The finite element solution (with 2577 nodes) obtained by the MATLAB PDE Toolbox is also presented for comparison. At $t = 0.25$ and 0.5, we can see that the λ^*-regularized and FEM solutions are analogous. It is obvious that the unregularized solution is not decaying as fast as the other two solutions; this is the consequence of the oscillations near the jump at the initial stage.

5. Conclusion

We provided some Sobolev error estimates for regularized RBF discrete least-squares approximation. With appropriate regularization, a least-squares stability holds for any sampling ratio of the fill distances of collocation points and trial centers. This results various $W_2^{\mu}(\Omega)$ error estimates for $\mu \geq 0$. For the unregularized least-squares approach, stability can only be shown with collocation points being sufficiently dense with respect to trial centers. This was demonstrated by a numerical example. The consequence is an extra condition on all of the proven error estimates, which only holds for $\mu > d/2$. With sufficiently dense collocation points, the unregularized least-squares method shows numerical convergence profiles that closely resemble the regularized case. It is our future work to develop new theoretical tools to extend the results to $0 \leq \mu \leq d/2$.

Acknowledgements

This work was partially supported by a Hong Kong Research Grant Council GRF Grant.

References

[1] M. D. Buhmann, Radial basis functions: Theory and implementations, Vol. 12 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2003.

[2] H. Wendland, Scattered data approximation, Vol. 17 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

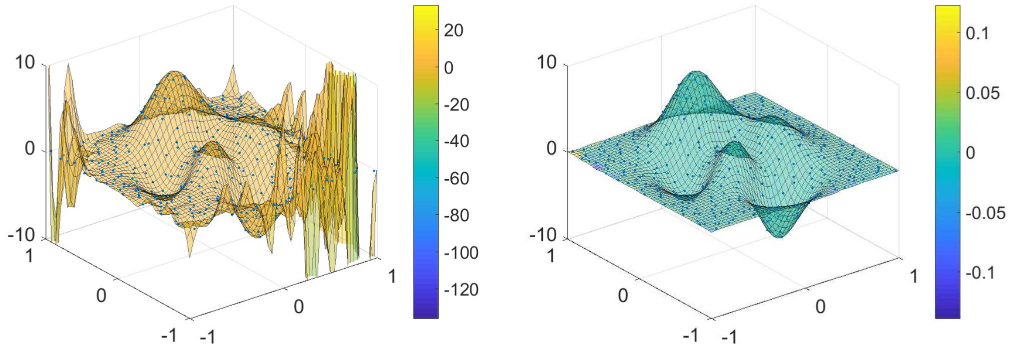
[3] G. E. Fasshauer, Meshfree approximation methods with MATLAB, Interdisciplinary Mathematical Sciences 6. Hackensack, NJ: World Scientific., 2007.

[4] G. E. Fasshauer, M. J. McCourt, Kernel-based Approximation Methods using MATLAB, Interdisciplinary Mathematical Sciences 19. Hackensack, NJ: World Scientific., 2015.

[5] F. J. Narcowich, J. D. Ward, H. Wendland, Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions, Constr. Approx. 24 (2) (2006) 175–186.

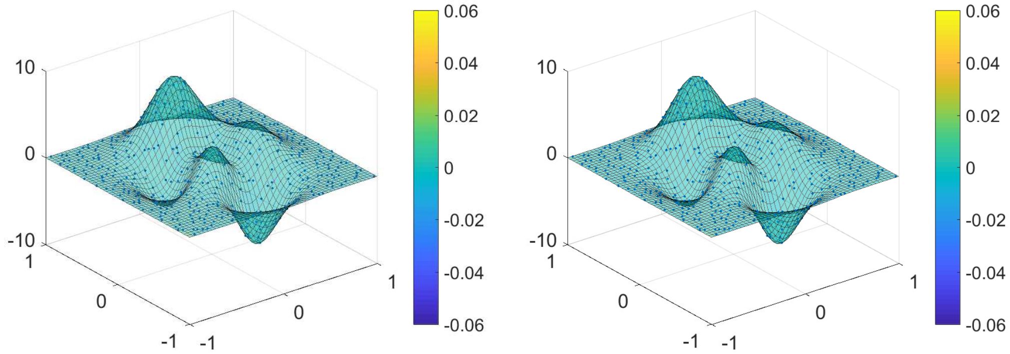
[6] J. D. Ward, Least squares approximation using radial basis functions: an update, in: M. Neamtu, E. B. Saff (Eds.), Advances in Constructive Approaximation: Vanderbilt2003, Nashboro Press, Brentwood, TN, 2004, pp. 499–508.

[7] F. J. Narcowich, J. D. Ward, H. Wendland, Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting, Math. Comp. 74 (250) (2005) 743–763.
[8] Q. T. L. Gia, F. J. Narcowich, J. D. Ward, H. Wendland, Continuous and discrete least-squares approximation by radial basis functions on spheres, J. Approx. Theory 143 (2006) 124–133.
[9] L. C. Evans, Partial differential equations, Graduate studies in mathematics, American Mathematical Society, Providence (R.I.), 1998, reimpr. avec corrections : 1999, 2002.
[10] M. Scheuerer, R. Schaback, M. Schlather, Interpolation of spatial data–a stochastic or a deterministic problem?, European J. Appl. Math. 24 (2013) 601–629.
[11] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, J. Royal Stat. Soc 58 (1) (1996) 267–288.
[12] A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[13] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. Royal Stat. Soc: Series B 67 (2) (2005) 301–320.
[14] J. K. Tay, J. Friedman, R. Tibshirani, Principal component-guided sparse regression, arXiv preprint arXiv:1810.04651.
[15] A. Iske, Approximation, Springer-Lehrbuch Masterclass, Springer Spektrum, 2018.
[16] R. A., M. C. L. de S., J. J. T., An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing, Numer. Math. 107 (2) (2007) 181–211.
[17] K. C. Cheung, L. Ling, R. Schaback, H^2–convergence of least-squares kernel collocation methods, SIAM J. Numer. Anal. 56 (1) (2018) 614–633.
[18] R. Schaback, All well–posed problems have uniformly stable and convergent discretizations, Numer. Math. 132 (2016) 597–630.
[19] R. Schaback, Superconvergence of kernel-based interpolation, J. Approx. Theory 235 (2018) 1–19.
[20] Y. C. Hon, R. Schaback, M. Zhong, The meshless kernel-based method of lines for parabolic equations, comput.. Math. Appl. 68 (12A) (2014) 2057–2067.



(a) Unregularized (b) Regularized

Figure 1: Example 1. Least-squares approximants obtained by $n_Z = 441$ uniformly placed set of trial center Z and $n_X = n_Z$ quasi-random Halton set of collocation X with (a) no regularization $\lambda = 0$ and (b) regularization parameter λ_*.



(a) Unregularized (b) Regularized

Figure 2: Example 1. Least-squares approximants obtained by the setup in Figure 1 but with oversampling $n_X = 1.2\, n_Z$.

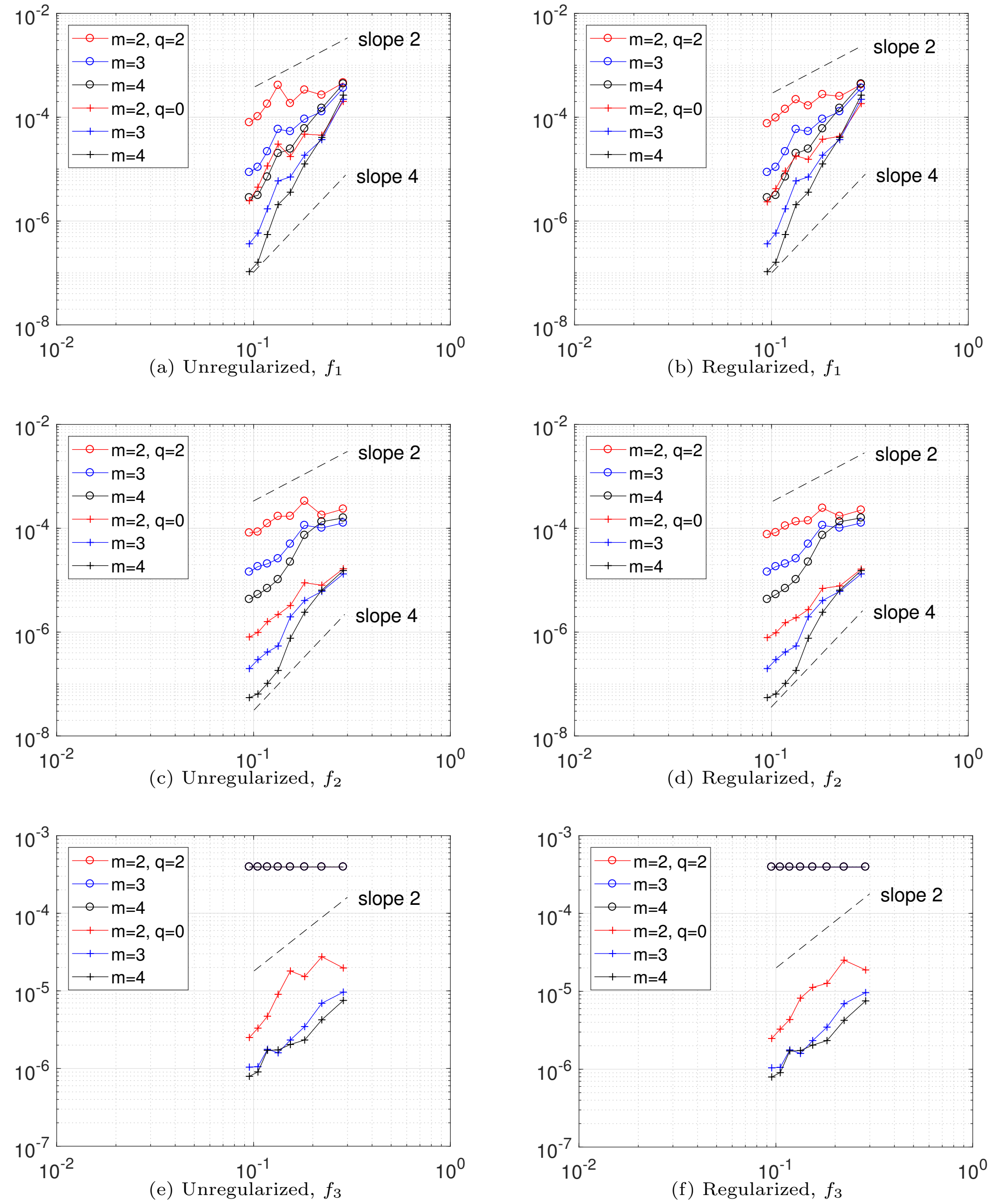
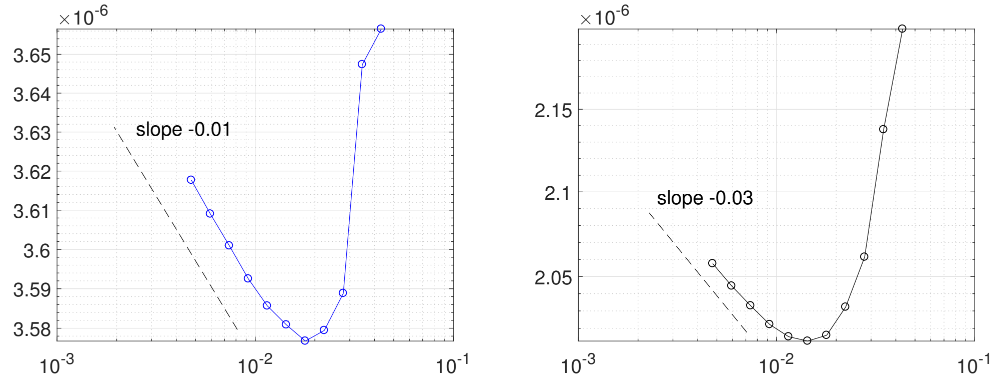
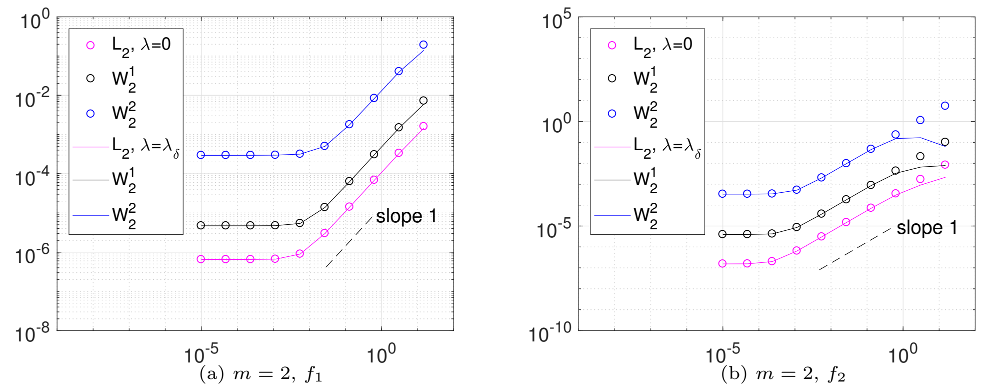


Figure 3: Example 2. $W_2^q(\Omega)$-Relative error, $\mathcal{E}_q$ for $q \in \{0, 2\}$, with respect to $h_Z = (4/n_Z)^{1/2}$ of the unregularized and λ_*-regularized least-squares approximant.



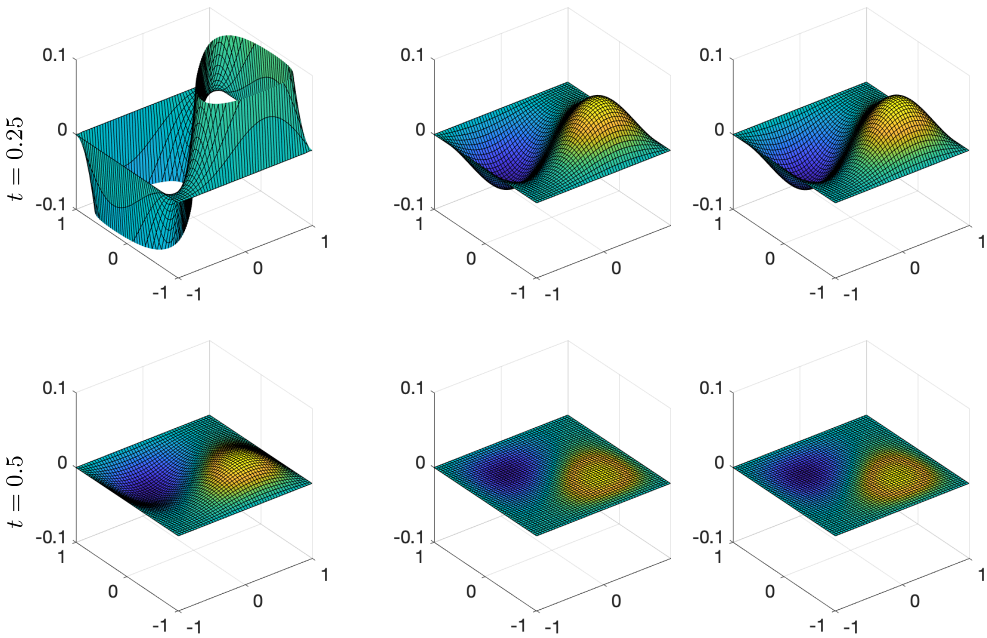
(a) $m = 3$ (b) $m = 4$

Figure 4: Example 3. $W_2^2(\Omega)$ error profiles of least-squares approximation using the same uniform $n_Z = 21^2$ trial centers against different sets of uniformly distributed collocation points under oversampling ratio.



(a) $m = 2,\ f_1$ (b) $m = 2,\ f_2$

Figure 5: Example 4. Error profiles of least-squares approximation based on noisy data against maximum plointwise noise $\xi = \max_X |f_(x) - f_\delta(x)|$.



(a) Unregularized IC (b) λ^*-Regularized IC (c) FEM

Figure 6: Example 5. Snapshots at time $t = 0.25$ and 0.5 of numerical solutions for a heat equation obtained by (a) the method of lines with unregularized and (b) λ^*-regularized initial conditions (IC), and (c) finite element method.