Acta Appl Math (2009) 107: 339–372
DOI 10.1007/s10440-009-9474-9

Reconstructing Signals with Finite Rate of Innovation from Noisy Samples

Ning Bi · M. Zuhair Nashed · Qiyu Sun

Received: 21 August 2008 / Accepted: 5 February 2009 / Published online: 24 February 2009

Abstract A signal is said to have finite rate of innovation if it has a finite number of degrees of freedom per unit of time. Reconstructing signals with finite rate of innovation from their exact average samples has been studied in Sun (SIAM J. Math. Anal. 38, 1389–1422, 2006). In this paper, we consider the problem of reconstructing signals with finite rate of innovation from their average samples in the presence of deterministic and random noise. We develop an adaptive Tikhonov regularization approach to this reconstruction problem. Our simulation results demonstrate that our adaptive approach is robust against noise, is almost consistent in various sampling processes, and is also locally implementable.

Keywords Sampling · Signals with finite rate of innovation · Regularized least squares · Mean squared error · Wiener filter

1 Introduction

In digital signal processes, a continuous time signal is represented by its samples on a discrete set. In the classical model, the Whittaker-Shannon sampling theorem states that a continuous time signal $x(t)$ on $\mathbb{R}$ bandlimited to $[-\Omega, \Omega]$ is uniquely determined by a set of

Research of the first author was supported in part by NSFC #10631080, and the Science Foundation of Guangdong Province #04205407 and #2007B090400091.

N. Bi (✉)
Department of Scientific Computing and Computer Applications, Sun Yat-Sen University, Guangzhou, 510275, China
e-mail: mcsbn@mail.sysu.edu.cn

M.Z. Nashed · Q. Sun
Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA

M.Z. Nashed
e-mail: znashed@mail.ucf.edu

Q. Sun
e-mail: qsun@mail.ucf.edu

uniformly spaced samples $x(nT), n \in \mathbb{Z}$, taken T seconds apart with $T \leq \pi/\Omega$,

$$x(t) = \sum_{n\in\mathbb{Z}} x(nT)\mathrm{sinc}((t-nT)/T), \tag{1.1}$$

where the sinc function is defined by $\mathrm{sinc}(t) = \frac{\sin \pi t}{\pi t}$.

Over the past decade, the paradigm for representing band-limited signals in the Shannon's sampling theory has been extended to signals in shift-invariant spaces. It is well known [1, 2, 32, 33] that any signal x in the shift-invariant space

$$V_2(\phi) := \left\{ \sum_{n\in\mathbb{Z}} c(n)\phi(t-n) \,\middle|\, \sum_{n\in\mathbb{Z}} |c(n)|^2 < \infty \right\}$$

generated by a function ϕ can be reconstructed from its uniformly-spaced samples $\{x(n)\}_{n\in\mathbb{Z}}$ via the following reconstruction formula:

$$x(t) = \sum_{n\in\mathbb{Z}} x(n)S(t-n) \quad \text{for all } x \in V_2(\phi), \tag{1.2}$$

where the Fourier transforms of the reconstruction function S and the generating function ϕ are related by

$$\hat{S}(f) = \frac{\hat{\phi}(f)}{\sum_{k\in\mathbb{Z}} \hat{\phi}(f-k)}.$$

Recently, the paradigm for representing band-limited signals in the Shannon's sampling theory is further extended to represent signals with finite rate of innovation, which are neither band-limited nor live in a shift-invariant space [3, 5, 11, 14, 15, 17, 24, 28, 34]. Here a signal is said to have *finite rate of innovation* if it has a finite number of degrees of freedom per unit of time, that is, if it can be specified from a finite number of samples per unit of time [34]. The number of samples per unit to specify a signal is called the *innovative rate* of the signal. Prototypical examples of signals with finite rate of innovation include (i) band-limited signals $x(t) = \sum_{k\in\mathbb{Z}} x(kT)\mathrm{sinc}(t/T - k)$ where $T > 0$ [13, 16]; (ii) signals in a shift-invariant space $x(t) = \sum_{k\in\mathbb{Z}} c(k)\phi(t-k)$, where $(c(k))$ is a square-summable sequence and ϕ is a square-integrable function [1, 2, 32, 33]; (iii) stream of pulses $\sum_k c(k)p_k(t-t_k)$ found in GPS applications, cellular radio and ultra wide-band communication [4, 14]; (iv) bandlimited signals with additive shot noise $\sum_{k\in\mathbb{Z}} c(k)\mathrm{sinc}(t-k) + \sum_l d(l)\delta(t-t_l)$ [17]; (v) sums of bandlimited signals and non-uniform spline signals, convenient for modelling electrocardiogram signals [11].

A prototypical space $V_X(\Phi)$,

$$V_X(\Phi) := \left\{ \sum_{\lambda\in\Lambda} c(\lambda)\phi_\lambda \,\middle|\, (c_\lambda)_{\lambda\in\Lambda} \in X \right\}, \tag{1.3}$$

was introduced in [30] to model signals with finite rate of innovation, where X is a sequence space, $\Phi = (\phi_\lambda)_{\lambda\in\Lambda}$ is a family of generating functions, and Λ is an index set where each $\lambda \in \Lambda$ represents the location of the generating functions ϕ_λ. For instance, the space $V_X(\Phi)$ becomes the shift-invariant space $V_2(\phi)$ generated by a function ϕ if the space ℓ^2 of all square-summable sequences is used as the sequence space X, $\Phi = (\phi(\cdot - n))_{n\in\mathbb{Z}}$ with each generating function $\phi(\cdot - n)$ in Φ being generated by the integer shifts of the function ϕ, and $\Lambda = \mathbb{Z}$ with each index $n \in \Lambda$ being thought of as the location of the generating function

$\phi(\cdot - n)$. The space $V_X(\Phi)$ is suitable for modelling signals with finite rate of innovation, and also suitable for taking into account of natural phenomena, real reconstruction devices, as well as practical numerical implementation, see [30] for detailed discussion.

The usual assumption in sampling problems is that the samples are ideals. However, in practice, the ideal sampling is impossible to implement. A practical model considers that the samples are obtained by linear measurements. In a popular setting, for time signals in a Hilbert space H with inner product $\langle \cdot, \cdot \rangle_H$, the sample y_γ at the location $\gamma \in \Gamma$ of a time signal f is obtained by taking the inner product of f with the sampling function ψ_γ at the location γ, i.e., $y_\gamma = \langle f, \psi_\gamma \rangle_H$. We call $\Psi = (\psi_\gamma)_{\gamma \in \Gamma}$ the *average sampler*, and $Y = (y_\gamma)_{\gamma \in \Gamma}$ the *average samples*. In [28], the third named author studied the average sampling and reconstruction problem for signals in the space $V_X(\Phi)$ from their *exact* average samples Y, i.e.,

$$Y = \langle x, \Psi \rangle \quad \text{for some } x \in V_X(\Phi)$$

when each generating function ϕ_λ in Φ and each sampling functional ψ_γ in the average sampler $\Psi = (\psi_\gamma)_{\gamma \in \Gamma}$ are assumed to be *well-localized*, and the sequence space X is the space ℓ^2.

As optimal approximation of a signal based on noisy samples is one of the leading problems in approximation, statistics, and information-based complexity, in this paper we consider the reconstruction for signals in the space $V_X(\Phi)$ from *noisy* average samples Y, i.e.,

$$Y = \langle x, \Psi \rangle + \omega$$

for some signal $x \in V_X(\Phi)$ and some deterministic or random noise ω.

The paper is organized as follows. In Sect. 2, we make a few basic assumptions on the signals to be sampled and reconstructed, and on the average sampler Ψ used for sampling signals. Under these assumptions we show that the linear reconstruction procedure is the same as solving a least squares problem. Inspired by that observation, we propose to adopt the Tikhonov approach for our signal reconstruction from noisy samples. We consider reconstructing signals from noisy samples by minimizing a regularized least squares functional for deterministic noise (Sect. 3), and by minimizing the worst-case mean squares error and stochastic mean squares error for random noise (Sect. 4). In Sect. 3, we provide an explicit solution to the minimization problem of regularized least squares functional (Theorem 3.1). We emphasize that the post-filter, regularization matrix and panel parameter are three important factors in the Tikhonov approach that affect the performance of our reconstruction. Therefore in Sects. 3.2–3.4 we discuss the problem of selecting those three factors to make the Tikhonov approach adaptive to our sampling-reconstruction process, and in Sect. 3.5 we propose an *adaptive* Tikhonov approach.

In Sect. 4, we follow the work by Eldar and Unser [6] and consider reconstructing a signal with finite rate of innovation from its samples corrupted by random noise. We solve the problem by minimizing the worst-case mean squared error and stochastic mean squared error. We give explicit solutions to those minimization problems (Theorems 4.1 and 4.6), and observe that the reconstruction filter H_{WOR} in Theorem 4.1 and the Wiener filter H_{W} in Theorem 4.6 become the time-varying reconstruction filter H_{TIK} with post-filters, regularization matrix and panel parameter appropriately chosen (see Remarks 4.2 and 4.7). In Sect. 4.2, we also give a lower bound estimate for the stochastic mean squared error (Theorem 4.9) and then show that lower bound estimate can be reached *only* when the average sampler Ψ is an orthonormal basis of the prototypical spaces in which the signals live (Remark 4.10).

In Sect. 5, we include various simulations to demonstrate the performance of the approach developed in this paper for reconstructing a signal with finite rate of innovation

from its noisy samples. Our simulations show that our adaptive Tikhonov approach is robust against noise, almost consistent in various sampling processes, and locally implementable.

In this paper, we will use the following notation. Given a discrete set Λ, let $\ell^p(\Lambda)$, $1 \le p \le \infty$, be the space of all p-summable sequences on Λ with standard norm $\|\cdot\|_{\ell^p(\Lambda)}$ or $\|\cdot\|_{\ell^p}$ for short.

2 Preliminaries

We say that Λ is a *relatively-separated* subset of $\mathbb{R}^d$ if there are finitely many $\lambda \in \Lambda$ in any unit cube, more precisely

$$\sup_{t\in\mathbb{R}^d} \#\big(\Lambda \cap (t + [-1/2, 1/2]^d)\big) < \infty, \tag{2.1}$$

and a *relatively dense* subset of $\mathbb{R}^d$ if there are at least one and at most finitely many $\lambda \in \Lambda$ in any unit cube, more precisely

$$\begin{aligned} 1 &\le \inf_{t\in\mathbb{R}^d} \#\big\{\lambda \in \Lambda \cap t + [-1/2, 1/2]^d\big\} \\ &\le \sup_{t\in\mathbb{R}^d} \#\big\{\lambda \in \Lambda \cap t + [-1/2, 1/2]^d\big\} < \infty. \end{aligned} \tag{2.2}$$

2.1 Signals to Be Sampled and Reconstructed

In this paper, we make the following basic assumption on signals x to be sampled and reconstructed:

Assumption 1 Signals x live in the space

$$V_2(\Phi) := V_{\ell^2}(\Phi) = \left\{ \sum_{\lambda\in\Lambda} c(\lambda)\phi_\lambda \,\Big|\, \sum_{\lambda\in\Lambda} |c(\lambda)|^2 < \infty \right\} \tag{2.3}$$

generated by $\Phi := (\phi_\lambda)_{\lambda\in\Lambda}$, where the location set Λ of "impulse responses" ϕ_λ is a relatively-separated subset of $\mathbb{R}^d$; the "impulse responses" ϕ_λ at the location $\lambda \in \Lambda$ is enveloped by a λ-shifted function $h(\cdot - \lambda)$, i.e.,

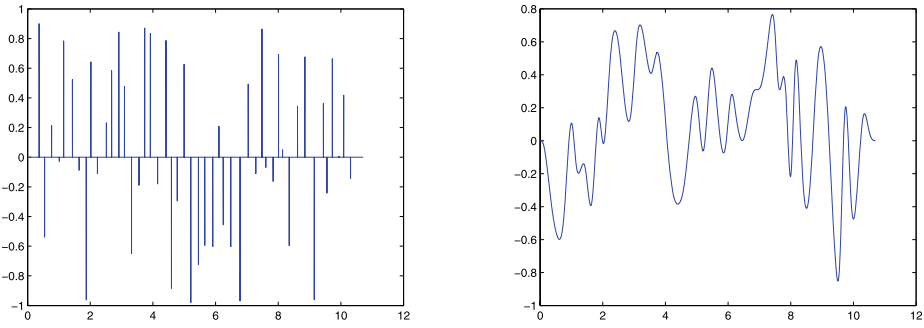
$$|\phi_\lambda(x)| \le h(x - \lambda), \tag{2.4}$$

with h having compact support or certain decay at infinity; and the family Φ of the "impulse responses" ϕ_λ is a Riesz basis of $V_2(\Phi)$, i.e., there exist positive constants A and B such that

$$A\left(\sum_{\lambda\in\Lambda} |c(\lambda)|^2\right)^{1/2} \le \left\|\sum_{\lambda\in\Lambda} c(\lambda)\phi_\lambda\right\|_2 \le B\left(\sum_{\lambda\in\Lambda} |c(\lambda)|^2\right)^{1/2}$$

for any sequence $C = (c(\lambda))_{\lambda\in\Lambda} \in \ell^2(\Lambda)$.

From the above assumption, we see that there are finitely many "impulse responses" ϕ_λ per unit time, and each "impulse response" ϕ_λ is well-localized in the neighborhood of the impulse location λ. Therefore signals x to be considered in this paper have finite rate of



(a) (b)

Fig. 1 (**a**) A Dirac stream $\sum_{n=1}^{N} c_n \delta_{\lambda_n}$ with innovative rate $1/\mu$; (**b**) A non-uniform mesh cubic spline $\sum_{n=1}^{N} d_n B_{3,\lambda_n}$ with innovative rate $1/\mu$. Here N is an integer, $\mu > 0$ is a positive density parameter, the location set $\Lambda = \{\lambda_n\}_{n=1}^{N}$ of impulse responses is iteratively defined by $\lambda_n = \lambda_{n-1} + x_n + \mu/2$ if $1 \le n \le N$ where $\lambda_0 = \mu/2$ and $x_n, 1 \le n \le N$, are numbers in $[0, \mu]$ chosen randomly, the coefficients $c_n, 1 \le n \le N$ and $d_n, 1 \le n \le N$ belong to $[-1, 1]$ and are chosen randomly, and for each $1 \le n \le N$, δ_{λ_n} is the delta pulse located in the position λ_n while B_{3,λ_n} is the normalized cubic spline with knots $\lambda_{n-2}, \lambda_{n-1}, \lambda_n, \lambda_{n+1}$ and λ_{n+2} where $\lambda_{-1} = 0, \lambda_0 = \mu/2, \lambda_{N+1} = \lambda_N + \mu, \lambda_{N+2} = \lambda_N + 2\mu$. In this figure, $N = 48$, $\mu = 0.2$, $\sum_{n=1}^{N} |c_n|^2 = 16.4930$ and $\sum_{n=1}^{N} |d_n|^2 = 13.3706$

innovation. On the other hand, most signals with finite rate of innovation can be modelled as living in some $V_2(\Phi)$ when the "impulse responses" are chosen appropriately [30]. The model signals to be considered in this paper are Dirac streams on a relatively-separated subset Λ of $\mathbb{R}$ and non-uniform mesh cubic splines with a relatively dense subset Λ of $\mathbb{R}$, see Fig. 1.

2.2 Average Samplers

In this paper, we make the following basic assumption on the average sampler $\Psi = (\psi_\gamma)_{\gamma\in\Gamma}$ to sample signals x in $V_2(\Phi)$. Here each index γ in Γ means that there is an acquisition device located at that position, and for each index γ, the impulse response of the acquisition device located at that position γ is described by the sampling functional ψ_γ.

Assumption 2 The location set Γ of acquisition devices is a relatively-separated subset of $\mathbb{R}^d$; the impulse response ψ_γ of the acquisition device located at γ is enveloped by a γ-shifted function $g(\cdot - \gamma)$, i.e.,

$$|\psi_\gamma(x)| \le g(x - \gamma), \tag{2.5}$$

with g having compact support or certain decay at infinity; and the correlation matrix

$$A_{\Phi,\Psi} := (\langle \phi_\lambda, \psi_\gamma \rangle)_{\lambda\in\Lambda,\gamma\in\Gamma}$$

between the generator Φ and the average sampler Ψ belongs to the Schur class, that is,

$$\sup_{\lambda\in\Lambda} \sum_{\gamma\in\Gamma} |\langle \phi_\lambda, \psi_\gamma \rangle| + \sup_{\gamma\in\Gamma} \sum_{\lambda\in\Lambda} |\langle \phi_\lambda, \psi_\gamma \rangle| < \infty. \tag{2.6}$$

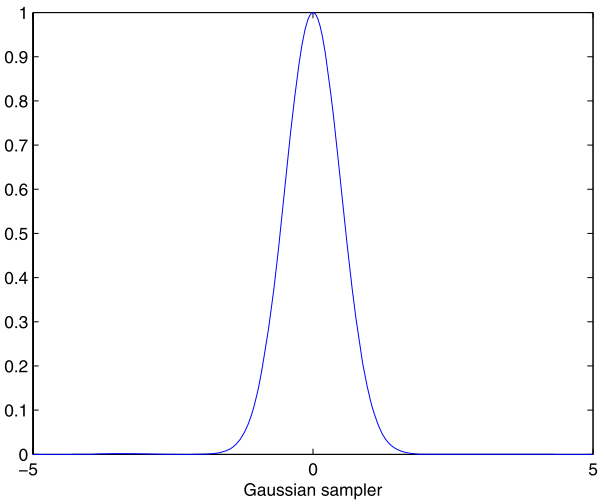


Fig. 2 Gaussian sampler $\exp(-|t|^2/2)$

We call $\Psi = (\psi_\gamma)_{\gamma\in\Gamma}$ an *average sampler* [28]. From the above assumption on the average sampler Ψ, there are finitely many acquisition devices per unit of time, each acquisition device is locally behaved, and every signal x in $V_2(\Phi)$ can be sampled by the average sampler Ψ. The model examples of the average sampler Ψ include the shifted Gaussian sampler ($\psi_\gamma(t) = \exp(-|t-\gamma|^2/(2\sigma^2))$ where $\sigma > 0$), shifted squared sinc sampler ($\psi_\gamma(t) = \mathrm{sinc}^2((t-\gamma)/T)$ where $T > 0$), the Dirac streams for ideal sampling ($\psi_\gamma(t) = \delta(t-\gamma)$), and non-uniform meshed B-spline (ψ_γ is a B-spline with knots in a fixed neighborhood of γ), see Fig. 2.

Remark 2.1 The average samplers considered in this paper include the case of multiple sampling devices located at the same position. For instance, if Γ is the set of all positions where at least one acquisition devices are located, and the impulse responses of the acquisition devices located at the position $\gamma \in \Gamma$ are described by the sampling functionals $\psi_\gamma^1, \ldots, \psi_\gamma^{k(\gamma)}$ respectively, where $1 \le k(\gamma)$, then the new index set $\tilde{\Gamma} := \bigcup_{\gamma\in\Gamma}\{\gamma + j\epsilon(\gamma), 1 \le j \le k(\gamma)-1\}$ and the new average sampler $\tilde{\Psi} = (\tilde{\psi}_{\tilde{\gamma}})_{\tilde{\gamma}\in\tilde{\Gamma}}$ are the location set and the average sampler considered in this paper, where $\epsilon(\gamma) \in [0,1)^d$ is chosen so that $\{\gamma + j\epsilon(\gamma), 0 \le j \le k(\gamma)-1\} \cap \{\gamma' + j\epsilon(\gamma'), 0 \le j \le k(\gamma')-1\} = \emptyset$ for all $\gamma \ne \gamma'$, and $\tilde{\psi}_{\tilde{\gamma}} = \psi_\gamma^j$ if $\tilde{\gamma} = \gamma + j\epsilon(\gamma) \in \tilde{\Gamma}$ for some $\gamma \in \Gamma$ and $0 \le j \le k(\gamma)-1$.

2.3 Sampling and Reconstructing

We take a signal $x \in V_2(\Phi)$ and write

$$x = C^T\Phi \tag{2.7}$$

by Assumption 1, where $C = (c(\lambda))_{\lambda\in\Lambda} \in \ell^2(\Lambda)$. Then the average sample Y of the signal x via the average sampler Ψ is given by

$$Y := \langle x, \Psi\rangle = A_{\Psi,\Phi}C. \tag{2.8}$$

The average sample of the Dirac streams and the non-uniform mesh cubic spline in Fig. 1 via the average samplers in Fig. 2 is given in Fig. 3.

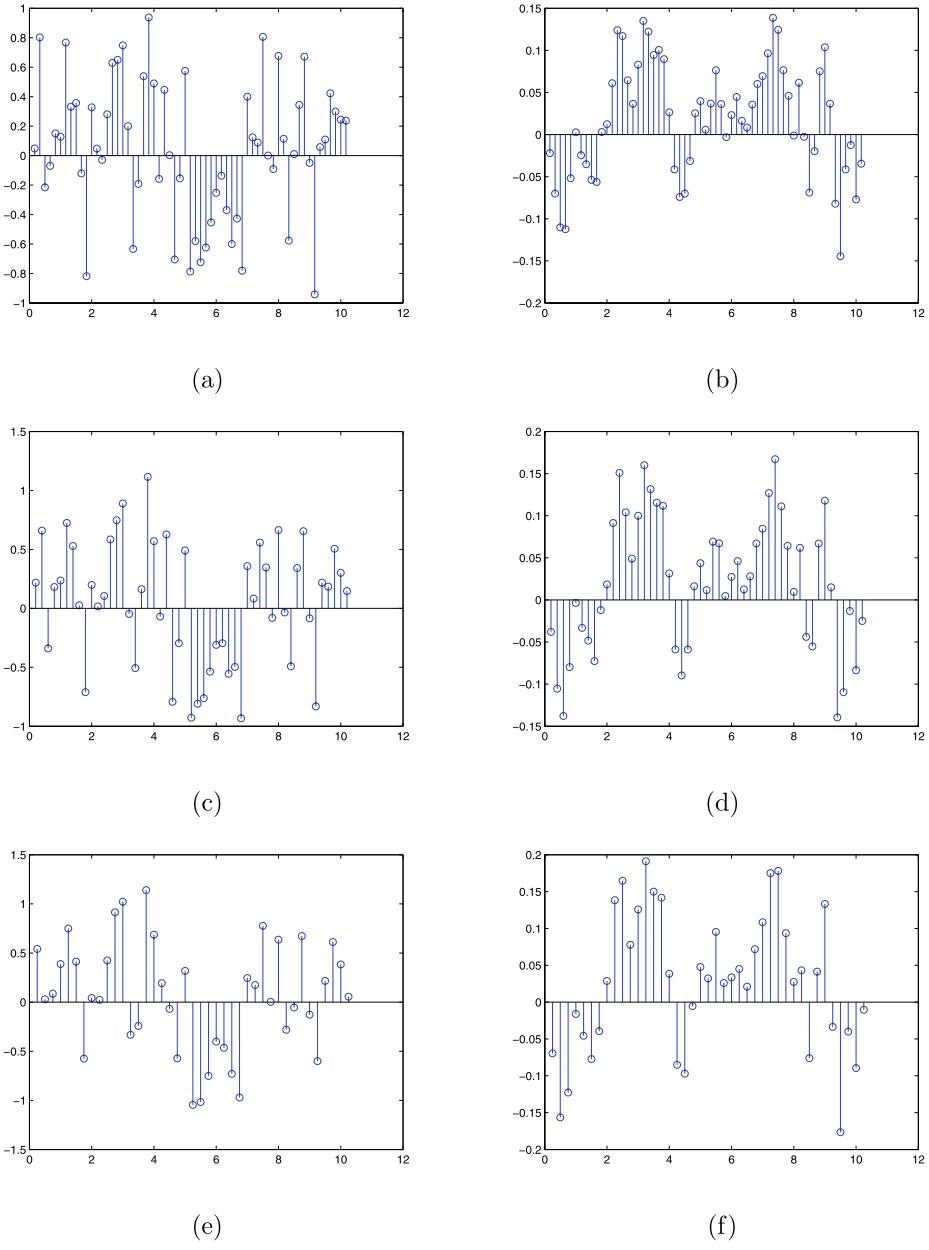


Fig. 3 On the *left hand side* is the sampling data $Y_{dg} := \{\sum_{n=1}^{N} c_n \exp(-2|\lambda_n - kT|^2/T^2) | 1 \le k \le [\frac{(\lambda_N+2\mu)}{T}]+1\}$ of the Dirac streams $x = \sum_{n=1}^{N} c_n \delta_{\lambda_n}$ in Fig. 1, while on the *right hand side* is the sampling data $Y_{cg} := \{\sum_{n=1}^{N} c_n \int_{\mathbb{R}} e^{-2|t-kT|^2/T^2} B_{3,\lambda_n}(t)dt | 1 \le k \le [\frac{(\lambda_N+2\mu)}{T}]+1\}$ of the non-uniform mesh cubic spline $x = \sum_{n=1}^{N} c_n B_{3,\lambda_n}$ in Fig. 1. The average samplers from *top* to *bottom* are generated by the T-shifted Gaussian sampler $\exp(-2|t-kT|^2/T^2)$ with the ratios between the sampling rate $1/T$ of Gaussian sampling devices and the innovative rate $1/\mu$ of the given signal being $6/5, 1, 4/5$ respectively

By our basic assumptions on the generator Φ and the average sampler Ψ, the average sample Y is a square-summable sequence (i.e. $Y \in \ell^2(\Gamma)$), and the reconstruction of a signal $x \in V_2(\Phi)$ from the exact average sample Y reduces to solving the linear system

$$A_{\Psi,\Phi} C = Y. \tag{2.9}$$

2.4 Reconstructing a Signal from Its Exact Average Samples and Solving a Least Squares problem

We say that the average sampler Ψ is *stable* on the space $V_2(\Phi)$ if the *average sampling operator* S,

$$S: \ V_2(\Phi) \ni x \longmapsto \langle x, \Psi \rangle \in \ell^2(\Gamma), \tag{2.10}$$

is bounded and has bounded left-inverse [2, 28]. For a stable average sampler Ψ, we see that any signal $x \in V_2(\Phi)$ is uniquely determined by its average sample $Y := \langle x, \Psi \rangle$, and

$$x = Y^T H_{\text{Ideal}} \Phi \tag{2.11}$$

for some time-varying reconstruction filter

$$H_{\text{Ideal}} := A_{\Psi,\Phi} (A_{\Phi,\Psi} A_{\Psi,\Phi})^{-1}. \tag{2.12}$$

We observe from (2.12) that the time-varying reconstruction filter H_{Ideal} is the pseudo-inverse of the correlation matrix $A_{\Psi,\Phi}$, and therefore for the case that the average sample is obtained exactly, the reconstruction of a signal in $V_2(\Phi)$ from average sample Y reduces to solving the least squares problem:

$$x = \text{argmin}_{\tilde{x} \in V_2(\Phi)} \|Y - \langle \tilde{x}, \Psi \rangle\|_{\ell^2}^2. \tag{2.13}$$

This observation inspires us to follow the above fashion in Sects. 3 and 4, and to consider reconstructing a signal in $V_2(\Phi)$ from the given noisy average sample by minimizing the (regularized) least squares functional in the deterministic setting for the noise, and by minimizing certain mean squared error (MSE) in the random setting for the noise.

3 Time-Varying Reconstruction Filters: Deterministic Setting

In this section, we investigate the problem of reconstructing a signal $x \in V_2(\Phi)$ from a given noisy average sample $Y = (y_\gamma)_{\gamma \in \Gamma} \in \ell^2(\Gamma)$ via the Tikhonov approach,

$$x = \text{argmin}_{\tilde{x} \in V_2(\Phi)} \epsilon_{\text{TIK}}(\tilde{x}), \tag{3.1}$$

where the regularized least-squares functional $\epsilon_{\text{TIK}}(\tilde{x})$ is given by

$$\epsilon_{\text{TIK}}(\tilde{x}) := \sum_{p=1}^{P} \|\mathcal{L}_p (Y - \tilde{Y})\|_{\ell^2}^2 + \alpha C^T A C, \tag{3.2}$$

$\mathcal{L}_p$, $1 \le p \le P$, are time-varying post-filters,

$$\tilde{Y} = \langle \tilde{x}, \Psi \rangle := (\langle \tilde{x}, \psi_\gamma \rangle)_{\gamma \in \Gamma} \tag{3.3}$$

represents the average samples derived from the signal $\tilde{x} := C^T\Phi \in V_2(\Phi)$, α is a nonnegative panel parameter, and the regularization matrix A is positive semi-definite, i.e., $C^TAC \geq 0$ for all $C \in \ell^2$. The reader may refer to [6, 7, 18–21, 26, 27] and references therein for some aspects of the theory and applications of (regularized) least squares problems.

In the first subsection below we solve the above reconstruction problem when post-filters, regularization matrices and panel parameters are known. In the next three subsections we discuss the problem of selecting post-filters, regularization matrices and panel parameters, which are three key factors that affect the performance of our reconstruction process. In the last subsection, we propose an adaptive Tikhonov approach for reconstructing a signal with finite rate of innovation from its samples corrupted by deterministic noises. The simulations in Sect. 5 show that the adaptive Tikhonov approach is robust against noise and is almost consistent irrespective of the value of the ratio between the sampling rate of sampling process and the innovative rate of signals to be sampled. More importantly the adaptive approach can be implemented locally and then can be used in real-time reconstruction where sample data in the future are unknown a priori, or in the partial reconstruction where partial sample data are missing.

3.1 Solution of the Regularized Least Squares Problem

The following theorem gives an explicit solution to the regularized least squares problem (3.2) where the post-filters, the regularization matrix and the panel parameter are given.

Theorem 3.1 *Let $\mathcal{L}_p$, $1 \leq p \leq P$, be time-varying post-filters, $\alpha \geq 0$ be a nonnegative parameter, the regularization matrix A be positive semi-definite, Φ be a Riesz basis of the space $V_2(\Phi)$, and Ψ be the average sampler that satisfies (2.6). Set $L = \sum_{p=1}^{P} \mathcal{L}_p^*\mathcal{L}_p$. Assume that $A_{\Phi,\Psi}LA_{\Psi,\Phi} + \alpha A$ has bounded inverse. Then*

(i) *For any given average sample $Y \in \ell^2(\Gamma)$, there exists a unique signal $x \in V_2(\Phi)$ that minimizes the regularized least-squares functional in (3.2).*

(ii) *For any given average sample $Y \in \ell^2(\Gamma)$, the signal $x \in V_2(\Phi)$, that minimizes the regularized least squares in (3.2), depends linearly on the noisy average sample Y. Moreover,*

$$x = Y^T H_{\mathrm{TIK}}\Phi \tag{3.4}$$

where the time-varying reconstruction filter H_{TIK} is given by

$$H_{\mathrm{TIK}} := LA_{\Psi,\Phi}(A_{\Phi,\Psi}LA_{\Psi,\Phi} + \alpha A)^{-1}. \tag{3.5}$$

(iii) *If Y is obtained by average sampling $x_0 := C_0^T\Phi \in V_2(\Phi)$ (i.e. $Y = \langle x_0, \Psi\rangle$), then the signal x_0 is the minimizer of the regularized least squares ϵ_{TIK} in (3.2) if and only if $\alpha AC_0 = 0$.*

Proof The theorem can be proved in a standard way. For the completeness of this paper, we include a proof here.

Take any $\tilde{x} = C^T\Phi \in V_2(\Phi)$ with $C^T \neq Y^T H_{\mathrm{TIK}}$. Noticing that correlation matrices $A_{\Psi,\Phi}$ and $A_{\Phi,\Psi}$ are related by taking the matrix transfer operation, i.e. $(A_{\Psi,\Phi})^T = A_{\Phi,\Psi}$, we have

$$\epsilon_{\mathrm{TIK}}(\tilde{x}) = \sum_{p=1}^{P} \|\mathcal{L}_p(Y - A_{\Psi,\Phi}C)\|_{\ell^2}^2 + \alpha C^TAC$$

$$
\begin{aligned}
&= (Y^T - C^T A_{\Phi,\Psi}) L (Y - A_{\Psi,\Phi} C) + \alpha C^T A C \\
&= Y^T (I - H_{\mathrm{TIK}} A_{\Phi,\Psi}) L (I - A_{\Psi,\Phi} H_{\mathrm{TIK}}^T) Y + \alpha Y^T H_{\mathrm{TIK}} A H_{\mathrm{TIK}}^T Y \\
&\quad + (C^T - Y^T H_{\mathrm{TIK}}) (A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha A)(C - H_{\mathrm{TIK}}^T Y) \\
&> \epsilon_{\mathrm{TIK}}^{\mathrm{opt}},
\end{aligned}
$$

where $\epsilon_{\mathrm{TIK}}^{\mathrm{opt}}$ is the value of the regularized least squares functional associated with the signal $x = Y^T H_{\mathrm{TIK}} \Phi$. This proves the first two conclusions.

For $Y = \langle x_0, \Psi \rangle$ for some $x_0 = C_0^T \Phi \in V_2(\Phi)$, it follows from (3.4) and (3.5) that the reconstructed signal x is given by

$$
x = C_0^T A_{\Phi,\Psi} L A_{\Psi,\Phi} (A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha A)^{-1} \Phi.
$$

Therefore, $x = x_0$ if and only if $C_0^T A_{\Phi,\Psi} L A_{\Psi,\Phi} (A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha A)^{-1} = C_0^T$ if and only if $\alpha C_0^T A = 0$ if and only if $\alpha A C_0 = 0$. The third conclusion is proved. □

Remark 3.2 If A is a positive definite matrix, i.e., there exists $m > 0$ such that $C^T A C \geq m\, C^T C$ for all $C \in \ell^2$, then we can write the time-varying filter H_{TIK} in (3.5) as the following two equivalent formulations:

$$
H_{\mathrm{TIK}} = L A_{\Psi,\Phi} A^{-1/2} (A^{-1/2} A_{\Phi,\Psi} L A_{\Psi,\Phi} A^{-1/2} + \alpha I)^{-1} A^{-1/2} \tag{3.6}
$$

and

$$
H_{\mathrm{TIK}} = L^{1/2} (L^{1/2} A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} L^{1/2} + \alpha I)^{-1} L^{1/2} A_{\Psi,\Phi} A. \tag{3.7}
$$

The above two equivalent formulations will be used later to find the optimal panel parameter α, see Sect. 3.4 for details.

Remark 3.3 If the cardinality of the set Φ is large, it will be costly and numerically unstable to find the time-varying reconstruction filter H_{TIK} by computing the inverse of the matrix $A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha A$. As our purpose is to reconstruct the signal x from its noisy sample Y or equivalently to find the left action of the filter H_{TIK} on Y^T, we may circumvent the inversion problem as usual by solving the following linear system

$$
(A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha A) C = A_{\Phi,\Psi} L Y \tag{3.8}
$$

numerically. In our average sampling and reconstruction problem, the generator Φ of the space $V_2(\Phi)$ and the average sampler Ψ usually have polynomial decay (or subexponential decay, or exponential decay) at infinity, i.e.

$$
\sup_{x \in \mathbb{R}^d, \lambda \in \Lambda} u(x - \lambda) |\phi_\lambda(x)| < \infty \tag{3.9}
$$

and

$$
\sup_{x \in \mathbb{R}^d, \gamma \in \Gamma} u(x - \gamma) |\psi_\gamma(x)| < \infty, \tag{3.10}
$$

where u is the polynomial weight $(1 + |x|)^\alpha$ with $\alpha > d$ for the polynomial decay case, the subexponential weight $\exp(\beta |x|^\delta)$ with $0 < \delta < 1, 0 < \beta < \infty$ for the subexponential decay case, and the exponential weight $\exp(\epsilon |x|)$ with $0 < \epsilon < \infty$ for the exponential decay case.

Thus the correlation matrices $A_{\Psi,\Phi}$ and $A_{\Phi,\Psi}$ are in the Gohberg-Baskakov-Sjöstrand class $\mathcal{C}_{\infty,u}(\Gamma,\Lambda)$ and $\mathcal{C}_{\infty,u}(\Lambda,\Gamma)$ with polynomial (resp. subexponential or exponential) weight u respectively, where the Gohberg-Baskakov-Sjöstrand class of infinite matrices [25, 29] is defined by

$$\mathcal{C}_{\infty,u}(\Lambda,\Gamma) := \Big\{(a(\lambda,\gamma))_{\lambda\in\Lambda,\gamma\in\Gamma} : \sup_{\lambda\in\Lambda,\gamma\in\Gamma} |a(\lambda,\gamma)|u(\lambda,\gamma) < \infty\Big\}.$$

Therefore by the Wiener's lemma [12, 25, 29], the time-varying reconstruction filter H_{TIK} is also a matrix in the Gohberg-Baskakov-Sjöstrand class with polynomial (resp. subexponential or exponential) weight (hence H_{TIK} is a numerically sparse matrix) if the generator Φ of the signal space $V_2(\Phi)$ and the average sampler Ψ have polynomial decay (resp. subexponential or exponential decay) at infinity, and the time-varying post-filters $\mathcal{L}_p$, $1 \le p \le P$, and the regularized matrix A belong to the Gohberg-Baskakov-Sjöstrand class $\mathcal{C}_{\infty,u}(\Gamma,\Gamma)$ and $\mathcal{C}_{\infty,u}(\Lambda,\Lambda)$ with polynomial (resp. subexponential or exponential) weight u respectively. Our demonstration confirms that theoretical result and shows that the number of numerically large entries in the time-varying reconstruction filter H_{TIK} depends almost linearly on the size of the filter, and the numerically nonzero entries of the matrices $A_{\Phi,\Psi}LA_{\Psi,\Phi}+\alpha A$ and its inverse lie in a fixed band around the diagonal. From that observation on the time-varying reconstruction filter H_{TIK} and the fact that the matrix $A_{\Phi,\Psi}LA_{\Psi,\Phi}+\alpha A$ is positive definite, well known effective numerical methods, such as finite section method and conjugate gradient algorithm with preconditioning, may be useful to solve the linear system (3.8) efficiently [8–10, 22].

3.2 Post-Filters in Tikhonov Approach

The post-filters $\mathcal{L}_p$, $1 \le p \le P$, are usually determined (or selected) by the reconstruction problem including the pattern of deterministic noises. For instance, in the global reconstruction problem where all (noisy) average samples are given, the post-filters $\mathcal{L}_p$, $1 \le p \le P$, are usually selected either to be *all-pass* filters in the sense that

$$\mathcal{L} : \ell^2(\Gamma) \in c \longmapsto (\mathcal{L}_1(c),\ldots,\mathcal{L}_P(c))^T \in (\ell^2(\Gamma))^P$$

is bounded and has bounded inverse, or to be *preconditioners* in the sense that $\mathcal{L}_p$, $1 \le p \le P$, are diagonal matrices with diagonal entries forming a sequence of weights $(U_\gamma)_{\gamma\in\Gamma}$, where $U_\gamma > 0$ [8, 26, 27].

On the other hand, in the local reconstruction problem where only average samples on a finite set $\tilde{\Gamma} \subset \Gamma$ are given (or could be used such as in real-time implementation), the post-filters $\mathcal{L}_p$, $1 \le p \le P$, should be *local* in the sense that

$$\mathcal{L}_p P_{\tilde{\Gamma}} = \mathcal{L}_p, \quad 1 \le p \le P,$$

where $P_{\tilde{\Gamma}}$ is the projection operator from $\ell^2(\Gamma)$ to $\ell^2(\tilde{\Gamma})$ defined by

$$(P_{\tilde{\Gamma}}c)(\gamma) = \begin{cases} c(\gamma) & \text{if } \gamma \in \tilde{\Gamma} \\ 0 & \text{if } \gamma \notin \tilde{\Gamma} \end{cases} \quad \text{for } c = (c(\gamma))_{\gamma\in\Gamma}.$$

For those post-filters satisfying the above local property, the average samples outside the finite set $\tilde{\Gamma}$ will not be used in the minimization problem based on the regularized least squares.

Given a finite set $\tilde{\Gamma} \subset \Gamma$, let $Y_{\tilde{\Gamma}}$ denote the given sample on $\tilde{\Gamma}$, $\tilde{\Psi} = (\psi_\gamma)_{\gamma\in\tilde{\Gamma}}$ be the subfamily of sampling functionals ψ_γ located in some position $\gamma \in \tilde{\Gamma}$, and $\mathcal{L}_{p,\tilde{\Gamma}}$ be the submatrix obtained by taking all γ-rows of the post-filter $\mathcal{L}_p$ with $\gamma \in \tilde{\Gamma}$. Given a finite set $\tilde{\Gamma} \subset \Gamma$, we let $\tilde{\Lambda}$ be the set of all $\lambda \in \Lambda$ such that the supports of the generating function ϕ_λ and the sampling functional ψ_γ located in the position γ have nonempty intersection for some $\gamma \in \tilde{\Gamma}$. Let also $\tilde{\Phi} = (\phi_\lambda)_{\lambda\in\tilde{\Lambda}}$ be a subfamily of functions ϕ_λ whose supports having nonempty intersection to that of some sampling functionals in $\tilde{\Psi}$. Then in the local reconstruction problem where only average samples on a finite set $\tilde{\Gamma} \subset \Gamma$ are given (or can be used) we can show that minimizing the regularized least-squares functional in (3.2) is the same as solving another minimization problem of *small size*, if we further assume that the generating functions ϕ_λ and the average sampling functionals ψ_γ are supported in a fixed compact set and if the regularization matrix A is a diagonal matrix $\mathrm{diag}(a_\lambda)_{\lambda\in\Lambda}$. Precisely we need only to minimize the following least-squares functional

$$\epsilon_{\mathrm{LTIK}}(\tilde{x}) := \sum_{p=1}^{P} \|\mathcal{L}_{p,L}(Y_L - \tilde{Y}_L)\|_{\ell^2}^2 + \alpha C^T \tilde{A} C, \tag{3.11}$$

where $\tilde{Y}_L = \langle \tilde{x}, \tilde{\Psi}\rangle := (\langle \tilde{x}, \psi_\gamma\rangle)_{\gamma\in\tilde{\Gamma}}$ represents the average samples derived from the signal $\tilde{x} = C^T\tilde{\Phi}$ and the regularization matrix $\tilde{A} = \mathrm{diag}(a_\lambda)_{\lambda\in\tilde{\Lambda}}$ is the diagonal matrix with the index in the diagonal elements included in $\tilde{\Lambda}$. The solution of the above minimization problem is

$$x = Y_L^T \tilde{L} A_{\tilde{\Psi},\tilde{\Phi}} (A_{\tilde{\Phi},\tilde{\Psi}} \tilde{L} A_{\tilde{\Psi},\tilde{\Phi}} + \alpha\tilde{A})^{-1}\tilde{\Phi}, \tag{3.12}$$

where $\tilde{L} = \sum_{p=1}^{P} \mathcal{L}_{p,L}^* \mathcal{L}_{p,L}$. The above argument shows that the Tikhonov approach can be implemented *locally* when local post-filters are selected and the regularization matrix is a diagonal matrix.

3.3 Regularization Matrix in Tikhonov Approach

The regularization matrix A in the regularized least squares is adjusted accordingly to the various sampling and reconstruction situations such as in the reconstruction of signals with finite duration or with low frequency. In most applications, the regularization matrix A in the Tikhonov approach is given by a regularizer R on the space $V_2(\Phi)$ in the sense that

$$C^T A C = \langle R\tilde{x}, R\tilde{x}\rangle = C^T A_{R\Phi,R\Phi} C \tag{3.13}$$

where $\tilde{x} = C^T\Phi \in V_2(\Phi)$. Typical examples of the regularizer R in the average sampling problems are (i) $(Rx)(t) = x(t)w(t)$ for some weight w in the time domain, such as $w(t) = (1+|t|)^\alpha, \alpha \geq 0$, and (ii) $\widehat{Rx}(f) = \hat{x}(f)\hat{w}(f)$ for some weights $\hat{w}$ in the frequency domain, such as $Rx = x''$. The first type of regularity operator can be used to reconstruct signals with fast decay in the time domain, while the second one with low frequency in the frequency domain. For instance, for the spline case, Φ is a family of splines on non-uniform meshes, we usually take m-th derivative $Rx = x^{(k)}$ for some $k \geq 0$ as the regularity operator [23, 35].

3.4 Panel Parameter in Tikhonov Approach

For better performance of the reconstruction process, it could be crucial to select an appropriate panel parameter α so that the regularized least squares minimization problem is

consistent (or almost consistent), the average sample of the reconstructed signal is "close" to the noisy sample, the reconstruction signal has certain regularity, and the reconstruction process is numerically stable and robust against noise. Here, the consistency of the regularized least squares minimization problem means that for the average sampler Y obtained by average sampling a signal $x_0 \in V_2(\Phi)$, the signal x that minimizes the regularized least squares has its average samples the same as the average samples of the original signal x_0 [33]. We use the above consistency instead of the more popular "strong" consistency where the reconstruction signal is the same as the original signal for considering undersampling problem where the reconstruction process is not unique in general.

In this subsection, we introduce three quantities to measure the sampling error, regularization of the reconstructed signal, and numerical stability of the reconstruction process respectively. For this purpose, we always assume that A is a positive definite in this subsection.

Let

$$e_q(\alpha) = \|L^{1/2}(Y - \langle x, \Psi\rangle)\|_{\ell^q}^q \tag{3.14}$$

be the ℓ^q sampling error between the noisy sampling data Y and the sampling data $\langle x, \Psi\rangle$ of the reconstructed signal $x = Y^T H_{\mathrm{TIK}}\Phi$, where $1 \le q \le 2$ and $L = \sum_{p=1}^{P} \mathcal{L}_p^* \mathcal{L}_p$. We note that

$$\lim_{\alpha\to\infty} e_q(\alpha) = \|L^{1/2}Y\|_{\ell^q}^q. \tag{3.15}$$

For $q = 2$, the sampling error $e_2(\alpha)$ is the first term in the regularized least squares $\epsilon_{\mathrm{TIK}}(x)$, and has the following explicit expression

$$\begin{aligned} e_2(\alpha) &= Y^T LY - Y^T LA_{\Psi,\Phi} B_\alpha \\ &\quad \times (A_{\Phi,\Psi} LA_{\Psi,\Phi} + 2\alpha A) B_\alpha A_{\Phi,\Psi} LY \end{aligned} \tag{3.16}$$

by (3.4) and (3.5), where

$$B_\alpha := (A_{\Phi,\Psi} LA_{\Psi,\Phi} + \alpha A)^{-1}. \tag{3.17}$$

Therefore

$$\frac{de_2(\alpha)}{d\alpha} = 2\alpha Y^T LA_{\Psi,\Phi} B_\alpha AB_\alpha AB_\alpha A_{\Phi,\Psi} LY \ge 0 \tag{3.18}$$

and

$$e_2(\alpha) \text{ is an increasing function on } (0, \infty). \tag{3.19}$$

Our numerical results show that $e_1(\alpha)$ is also increasing function. Thus

Claim 1 *The sampling data of the reconstructed signal has smaller sampling error to the observed noisy sampling data for smaller panel parameter* $\alpha \ge 0$.

Let

$$r(\alpha) := Y^T H_{\mathrm{TIK}} AH_{\mathrm{TIK}}^T Y = Y^T LA_{\Psi,\Phi} B_\alpha AB_\alpha A_{\Phi,\Psi} LY \tag{3.20}$$

be the second term in the regularized least squares functional $\epsilon_{\mathrm{TIK}}(x)$ to measure the regularity of the reconstruction signal $x = Y^T H_{\mathrm{TIK}}\Phi$. Then

$$\lim_{\alpha\to+\infty} r(\alpha) = 0 \tag{3.21}$$

and

$$r(\alpha) \text{ is a decreasing function on } (0, \infty), \tag{3.22}$$

as

$$\frac{dr(\alpha)}{d\alpha} = -2Y^T L A_{\Psi,\Phi} B_\alpha A B_\alpha A B_\alpha A_{\Phi,\Psi} L Y < 0. \tag{3.23}$$

Therefore

Claim 2 *The reconstruction signal has better regularity for larger panel parameter* $\alpha \geq 0$.

Define

$$c_1(\alpha) = \|(A^{-1/2} A_{\Phi,\Psi} L A_{\Psi,\Phi} A^{-1/2} + \alpha I)^{-1}\| \, \|A^{-1/2} A_{\Phi,\Psi} L A_{\Psi,\Phi} A^{-1/2}\|, \tag{3.24}$$

$$c_2(\alpha) = \|(L^{1/2} A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} L^{1/2} + \alpha I)^{-1}\| \, \|L^{1/2} A_{\Psi,\Phi} A A_{\Phi,\Psi} L^{1/2}\|, \tag{3.25}$$

and

$$c(\alpha) = \min(c_1(\alpha), c_2(\alpha)), \tag{3.26}$$

where $\|\cdot\|$ is the operator norm of a matrix. Because of the equivalent formulations (3.6) and (3.7) of the time-varying reconstruction filter H_{TIK}, we use $c(\alpha)$ to measure the numerical stability of the reconstruction process. We notice that

$$c(\alpha) \text{ is a decreasing function on } (0, \infty), \tag{3.27}$$

with

$$\lim_{\alpha \to +\infty} c(\alpha) = 0. \tag{3.28}$$

Hence

Claim 3 *The process of obtaining the reconstructed signal from the observed noisy sampling data has better numerical stability for larger panel parameter* α.

The above claims suggest a trade-off to select the panel parameter α.

3.5 Adaptive Tikhonov Approach

In this subsection, we use the three claims in the last subsection to select the panel parameter α appropriately and then develop an adaptive Tikhonov approach for reconstructing a signal with finite rate of innovation from its noisy samples.

Before we apply the three claims in the last subsection, we first normalize the measurements there. In fact, we normalize the sampling error $e_q(\alpha)$ by negative logarithm of the relative sampling error,

$$E_q(\alpha) = -\log_{10} \frac{e_q(\alpha)}{e_q(+\infty)}, \quad 1 \leq q \leq 2;$$

the regularization measurement $r(\alpha)$ by logarithm of the ratio between the regularization measurement $r(\alpha)$ and the energy of the noisy sample Y,

$$R(\alpha) := \log_{10}\left(\frac{r(\alpha)}{Y^T L Y} \times \frac{\|A_{\Phi,\Psi} A_{\Psi,\Phi}\| \|L\|}{\|A\|}\right); \tag{3.29}$$

and numerical stability constant $c(\alpha)$ by its logarithm

$$C(\alpha) = \log_{10} c(\alpha). \tag{3.30}$$

In Fig. 4, we present a simulation result of the logarithm $E_q(\alpha)$ of the relative sampling error between the given noisy sample data and the sample data of the reconstruction signal, the logarithm $R(\alpha)$ of relative regularization measurement of the reconstruction signal, and the logarithm $C(\alpha)$ of numerical stability number of the reconstruction process. If we use the popular signal-to-noise ratio (SNR) in decibels, 10 times the logarithm of the power ratio between the exact sample Y_e and the additive noise $Y - Y_e$, then the SNRs of the above additive noise are 24.7397db, 24.3544db, 24.8322db for figures on the right hand side from top to bottom and 24.7701db, 23.9798db, 24.5314db figures on the left hand side from top to bottom. Figure 4 confirms the monotonicity of the functions $E_1(\alpha)$, $E_2(\alpha)$, $R(\alpha)$, $C(\alpha)$ (except the functions $E_1(\alpha)$ and $E_2(\alpha)$ in Fig. 4(e) for small α due to the numerical instability) and the side-effect of over-regularization for large panel parameter α.

Now let us select the panel parameter α. To judge the reasonability of the reconstruction procedure via regularization, we require that the normalized regularization measurement $R(\alpha)$ should satisfy

$$R(\alpha) \le \log_{10} R \tag{3.31}$$

where R is a positive constant. From our simulation, we let $R = 10^2$.

For the stability of the reconstructed procedure, it is reasonable to require that

$$C(\alpha) \le \log_{10} S \tag{3.32}$$

where S is a positive constant. From our simulation, we let $S = 10^3$ for reconstructing Dirac streams and $S = 10^6$ for reconstructing cubic splines.

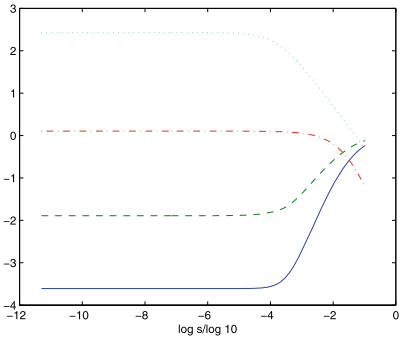
The entries in the matrix $A_{\Psi,\Phi}$ are usually obtained by numerical quadrature, which leads to an error e_0. As impulse responses of the original signal and the sampling devices at each location are localized, the error of each entry of the matrix $A_{\Phi,\Psi} L A_{\Psi,\Phi}$ in (3.6) and the matrix $A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi}$ in (3.7) obtained by numerical quadrature is controlled by a scalar multiple of $e_0 \|L A_{\Psi,\Phi}\|$ and $e_0 \|A_{\Psi,\Phi} A^{-1}\|$ respectively. So it is reasonable to have certain regularization,

$$\alpha \ge \operatorname{sinf} := e_0 \|L A_{\Psi,\Phi} A^{-1}\|. \tag{3.33}$$

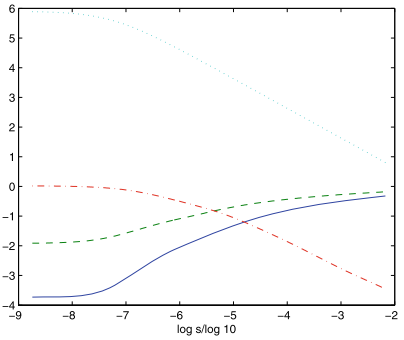
From our simulation, we will let $e_0 = 10^{-9}$ for Dirac stream signal and $e_0 = 10^{-7}$ for cubic spline signal.

The sampling error $e_2(\alpha)$ between the noisy sampling data Y and the sampling data $\langle x, \Psi\rangle$ of the reconstructed signal $x = Y^T H_{\mathrm{TIK}} \Phi$ should not exceed certain percentage p of the energy of the noisy data $Y^T L Y$, i.e.,

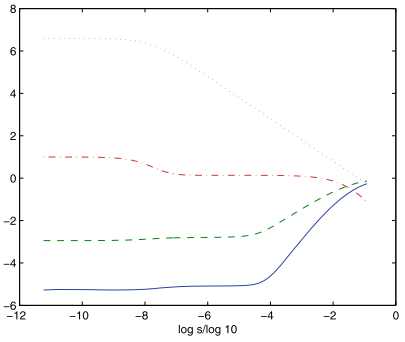
$$E_2(\alpha) \ge -\log_{10} p \tag{3.34}$$



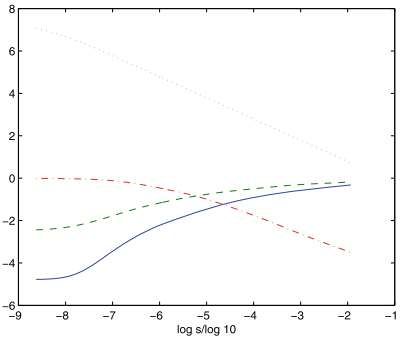
(a)



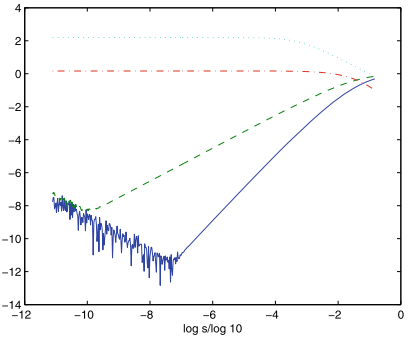
(b)



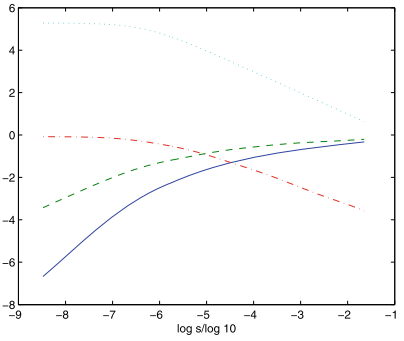
(c)



(d)



(e)



(f)

Fig. 4 The functions $E_1(10^s)$ (*dashed line*), $E_2(10^s)$ (*solid line*), $R(10^s)$ (*dash-dotted line*), $C(10^s)$ (*dotted line*) of the Dirac streams in Fig. 1 (the *left hand side*) and of the nonuniform cubic spline in Fig. 1 (the *right hand side*). The average sampler Ψ from top to bottom is generated by the T-shifted Gaussian function $\exp(-2|t-kT|^2/T^2)$, $1 \le k \le [\frac{\lambda_N+2\mu}{T}]+1$, with $\mu/T = 6/5, 1, 4/5$ respectively. The noisy data Y are given by $Y = Y_e + \sigma_0 \max(|Y_e|)(r(\gamma))_{\gamma\in\Gamma}$ where the noise level σ_0 is 0.05, $r(\gamma), \gamma \in \Gamma$ are numbers in $[-1, 1]$ chosen randomly and where the sampling data Y_e for figures on the left and right hand sides are the sampling data Y_{dg} and Y_{cg} given in Fig. 3 respectively. The postfilter is L_g in (5.3), and the regularization matrix is A_d in (5.1) for figures on the *left hand side* and A_c in (5.2) for figures on the *right hand side*

otherwise either the sampling data Y is too noisy or our reconstruction procedure via regularization is over-regularized. Therefore by (3.16) we have

$$Y^T L A_{\Psi,\Phi} B_\alpha (2\alpha A + A_{\Phi,\Psi} L A_{\Psi,\Phi}) B_\alpha A_{\Phi,\Psi} L Y \ge (1-p) Y^T L Y.$$

This together with

$$2\alpha^{-1} A^{-1} \ge B_\alpha (2\alpha A + A_{\Phi,\Psi} L A_{\Psi,\Phi}) B_\alpha$$

implies that the panel parameter α should satisfy

$$\alpha \le \text{ssup} := 2(1-p)^{-1} \|L^{1/2} A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} L^{1/2}\|. \tag{3.35}$$

From our simulation, we will let $p = 10\%$.

So we propose to select *the smallest* $\alpha \in [\text{sinf}, \text{ssup}]$ *satisfying* (3.31) *and* (3.32) *as the panel parameter* α_{op} *in the Tihonov approach for reconstructing signals with finite rate of innovation from noisy samples*. In our simulation, we will use the following simple algorithm on the logarithmic scale of the panel parameter to determine the adaptive parameter α_{op}:

Panel Parameter Algorithm:

1. Input noisy sample data Y and correlation matrix $A_{\Psi,\Phi}$.
2. Input postfilter L, and regularization matrix A.
3. Input sinf in (3.33) and ssup in (3.35).
4. Input N_0, R, S (In our simulation, $N_0 = 30$, $R = 100$ and $S = 1000$ for reconstructing Dirac streams and $S = 10^6$ for reconstructing cubic splines)
5. Define $a = \log_{10} \text{sinf}$, $b = \log_{10} \text{ssup}$.
   For $k = 1$ to N_0
   $m = (a+b)/2$, $p2 = R(10^m)$, $p3 = C(10^m)$;
   if $p2 \le \log_{10} R$ and $p3 \le \log_{10} S$
   $a = a, b = m$;
   else
   $a = m, b = b$;
   end
6. $\alpha_{op} = m$.
7. Input Φ.
8. Define the reconstruction signal x_{re} from noisy sample data Y by

$$x_{re} = Y^T L A_{\Psi,\Phi} (A_{\Phi,\Psi} L A_{\Psi,\Phi} + \alpha_{\text{op}} A)^{-1} \Phi.$$

4 Time-Varying Reconstruction Filters: Random Setting

Following the work by Eldar and Unser [6], we introduce in this section two design criteria for reconstructing a signal $\tilde{x} \in V_2(\Phi)$ from the noisy average sampling data $Y = (y_\gamma)_{\gamma\in\Gamma}$ of the following form:

$$y_\gamma = \langle x_0, \psi_\gamma \rangle + w_\gamma, \quad \gamma \in \Gamma \tag{4.1}$$

for some $x_0 \in V_2(\Phi)$ and discrete additive noise $w = (w_\gamma)_{\gamma\in\Gamma}$ with zero mean and known covariance matrix W.

4.1 Worst-Case MSE

Given a reconstruction filter H, a positive number δ_0, a positive definite regularization matrix A, and a linear functional $\mathcal{M}$ on $V_2(\Phi)$, we define the *worst-case mean squared error* $\epsilon_{\mathrm{WOR}} := \epsilon_{\mathrm{WOR}}(H, \delta_0, A, \mathcal{M})$ by

$$\epsilon_{\mathrm{WOR}} = \max_{x \in V_2(\Phi), \|x\|_A \le \delta_0} \mathbf{E}\{|\mathcal{M}(x - \tilde{x})|^2\} \tag{4.2}$$

where $\|x\|_A := \sqrt{C^T A C}$ for $x = C^T \Phi \in V_2(\Phi)$,

$$\tilde{x} = (\langle x, \Psi^T \rangle + w^T) H \Phi \tag{4.3}$$

is the reconstructed signal from the noisy sample $Y = \langle x, \Psi \rangle + w$ via a reconstruction filter H, and $w = (w_\gamma)_{\gamma \in \Gamma}$ is additive noise with zero mean and with covariance matrix W.

The next result gives the solution to the reconstruction problem based on the minimization of the worst-case mean squared error.

Theorem 4.1 *Let $\delta_0 > 0$, Λ and Γ be finite sets, A be a positive definite matrix, $\mathcal{M}$ a linear functional on $V_2(\Phi)$, Φ a Riesz basis of $V_2(\Phi) \subset L^2$, and Ψ an average sampler that satisfies (2.6). Assume that the additive noise $w = (w_\gamma)_{\gamma \in \Gamma}$ has zero mean and covariance matrix W, and that $\delta_0^2 A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} + W$ has bounded inverse. Then the reconstruction filter*

$$H_{\mathrm{WOR}} := \delta_0^2 \big(\delta_0^2 A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} + W\big)^{-1} A_{\Psi,\Phi} A^{-1} \tag{4.4}$$

minimizes the worst-case MSE $\epsilon_{\mathrm{WOR}}(H, \delta_0, A, \mathcal{M})$ in (4.2), that is,

$$H_{\mathrm{WOR}} = \mathrm{argmin}_H \epsilon_{\mathrm{WOR}}(H, \delta_0, A, \mathcal{M}). \tag{4.5}$$

Proof For any $x = C^T A^{-1/2} \Phi$ with $\|C\|_{\ell^2} \le \delta_0$,

$$\tilde{x} = (C^T A^{-1/2} A_{\Phi,\Psi} + \omega^T) H \Phi, \tag{4.6}$$

where H is the reconstruction filter, and w is the additive noise with zero mean and covariance matrix W. Then we obtain

$$\begin{aligned}
&\epsilon_{\mathrm{WOR}}(H, \delta_0, A, \mathcal{M}) \\
&\quad = \max_{\|C\|_{\ell^2} \le \delta_0} \mathbf{E}\big(\big|C^T A^{-1/2}(A_{\Phi,\Psi} H - I)(\mathcal{M}\Phi) + w^T H(\mathcal{M}\Phi)\big|^2\big) \\
&\quad = \max_{\|C\|_{\ell^2} \le \delta_0} |C^T A^{-1/2}(A_{\Phi,\Psi} H - I)(\mathcal{M}\Phi)|^2 + (\mathcal{M}\Phi)^T H^T W H(\mathcal{M}\Phi) \\
&\quad = \delta_0^2 (\mathcal{M}\Phi)^T (H^T A_{\Psi,\Phi} - I) A^{-1} (A_{\Phi,\Psi} H - I)(\mathcal{M}\Phi) + (\mathcal{M}\Phi)^T H^T W H(\mathcal{M}\Phi) \\
&\quad = \epsilon_{\mathrm{WOR}}(H_{\mathrm{WOR}}, \delta_0, A, \mathcal{M}) + (\mathcal{M}\Phi)^T (H - H_{\mathrm{WOR}})^T \\
&\qquad \times \big(\delta_0^2 A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} + W\big)(H - H_{\mathrm{WOR}})(\mathcal{M}\Phi) \\
&\quad \ge \epsilon_{\mathrm{WOR}}(H_{\mathrm{WOR}}, \delta_0, A, \mathcal{M}).
\end{aligned}$$

Hence the conclusion (4.4) follows. □

Remark 4.2 The construction filter H_{WOR} in the worst-case MSE coincides with the construction filter H_{TIK} when the covariance matrix W and the post-filters $\mathcal{L}_p, 1 \le p \le P$, satisfy $WL = \delta_0^2 \alpha I$, where $L = \sum_{p=1}^{P} \mathcal{L}_p^* \mathcal{L}_p$. From the above observation, we see that for the case that average samples are corrupted by white Gaussian noise with zero mean and variance σ_0^2, our approach could have better performance if post-filters is trivial (i.e. $L = I$) and σ_0^2/δ_0^2 is selected as the panel parameter α.

Remark 4.3 The model example of the linear functional $\mathcal{M}$ in the worst-case MSE is the evaluation functional at time t (i.e. $\mathcal{M}\Phi = \Phi(t)$) when Φ is continuous [6, 31]. Our simulation shows that the reconstruction filter H_{WOR} does not minimize the following worst-case weighted MSE

$$\tilde{\epsilon}_{\mathrm{WOR}} = \max_{x \in V_2(\Phi), \|x\|_A \le \delta_0} \int_{\mathbb{R}^d} \mathbf{E}\{|x(t) - \tilde{x}(t)|^2\} d\mu(t), \tag{4.7}$$

where $\mu(t)$ is a probability measure on $\mathbb{R}^d$. But observing from Theorem 4.1 that the reconstruction filter H_{WOR} is independent of the linear functional $\mathcal{M}$, we have that the reconstruction filter H_{WOR} also minimizes the following linearized worst-case weighted MSE,

$$\epsilon_{\mathrm{WOR}} = \int_T \max_{x \in V_2(\Phi), \|x\|_A \le \delta_0} \mathbf{E}\{|x(t) - \tilde{x}(t)|^2\} d\mu(t), \tag{4.8}$$

where $\mu(t)$ is a probability measure.

Remark 4.4 For the case that Ψ is an *interpolating average sampler* on $V_2(\Phi)$, i.e. there exist two positive constants A and B such that

$$A\|c\|_{\ell^2} \le \|A_{\Phi,\Psi} c\|_{\ell^2} \le B\|c\|_{\ell^2} \quad \text{for all } c \in \ell^2(\Gamma), \tag{4.9}$$

the operator norm $\|H_{\mathrm{WOR}}\|$ of the reconstruction filter H_{WOR} can be bounded by a constant independent of the parameter δ_0 and the covariance matrix W because

$$\begin{aligned} \|H_{\mathrm{WOR}}\| &\le \delta_0^2 \left\| \left(\delta_0^2 A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi} + W\right)^{-1} \right\| \|A_{\Psi,\Phi}\| \|A^{-1}\| \\ &\le \left\| (A_{\Psi,\Phi} A^{-1} A_{\Phi,\Psi})^{-1} \right\| \|A_{\Psi,\Phi}\| \|A^{-1}\| < \infty. \end{aligned}$$

Hence the reconstruction process from average samples via the reconstruction filter H_{WOR} is uniformly stable.

For the case that Ψ is a stable average sampler on $V_2(\Phi)$, i.e. (2.10) holds, we may not be able to find an upper bound of the operator norm $\|H_{\mathrm{WOR}}\|$ that is independent of the parameter δ_0 and the covariance matrix W. For instance for $\Phi = \{\chi_{[0,1]}\}$, $\Psi = \{\chi_{[0,1/2]}, \chi_{[1/2,1]}\}$, $A = 1$, and $W = \begin{pmatrix} (\delta-1)^2 & \delta^2-1 \\ \delta^2-1 & (\delta+1)^2 \end{pmatrix}$, the reconstruction filter $H_{\mathrm{WOR}} := \begin{pmatrix} 1+\delta \\ 1-\delta \end{pmatrix}$, has its operator norm tending to infinity as $\delta \to +\infty$. But as the additive noise w is the white Gaussian noise with zero mean and variance σ_0 (that is, $W = \sigma_0^2 I$), the operator norm $\|H_{\mathrm{WOR}}\|$ of the reconstruction filter H_{WOR} can be bounded by a constant independent of the parameter δ_0 and the variance σ_0, since in that case the reconstruction filter H_{WOR} can also be written as

$$H_{\mathrm{WOR}} = \delta_0^2 A_{\Psi,\Phi} (\delta_0^2 A^{-1} A_{\Phi,\Psi} A_{\Psi,\Phi} + \sigma_0^2 I)^{-1} A^{-1}. \tag{4.10}$$

Remark 4.5 The assumption that both Λ and Γ are finite sets is a technical assumption in Theorem 4.1. From the proof of Theorem 4.1, we see that the filter H_{WOR} is still the

reconstruction filter that minimizes the worst-case MSE ϵ_{WOR} if the finite-cardinality assumption on the sets Λ and Γ in Theorem 4.1 is replaced by some technical conditions on the noise w, the random signal x, and the reconstruction filter H in the minimization procedure, and the filter H_{WOR} in (4.4). For instance, we may assume that the random signal $x = C^T\Phi \in V_2(\Phi)$, the random noise $w = (w_\gamma)_{\gamma\in\Gamma}$, the construction filter H in the minimization problem and the optimal reconstruction filter H_W satisfy the following conditions: (i) $c_\lambda \in [-A_\lambda, A_\lambda]$, $\lambda \in \Lambda$ for some sequence $A_\lambda, \lambda \in \Lambda$ with $\sum_{\lambda\in\Lambda} |A_\lambda|^2 < \infty$ where $C = (c_\lambda)_{\lambda\in\Lambda}$, (ii) $w_\gamma \in [-B_\gamma, B_\gamma]$ for some sequence $B_\gamma, \gamma \in \Gamma$ with $\mathbf{B}^T|H|, \mathbf{B}^T|H_{\mathrm{WOR}}| \in \ell^2(\Lambda)$ where $\mathbf{B} = (B_\gamma)_{\gamma\in\Gamma}$, the (γ, λ) entries of the matrices $|H|$ and $|H_{\mathrm{WOR}}|$ are the moduli of the one of the matrices H and H_{WOR} respectively. From the above assumption, we see that the *sparseness* of the reconstruction filter H_{WOR} is an extremely important requirement in the reconstruction process where the average samples are corrupted by random noise. As discussed before (Remark 3.3), the reconstruction filter H_{WOR} has polynomial decay (resp. subexponential or exponential decay) if the generator Φ of the space $V_2(\Phi)$ in which signals live, the average sampler Ψ, the regularized matrix A, and the covariance matrix W have polynomial decay (resp. subexponential or exponential decay).

4.2 Stochastic MSE

Given a regularization matrix A and a reconstruction filter H, we define the stochastic mean squared error (MSE) $\epsilon_{\mathrm{MSE}} := \epsilon_{\mathrm{MSE}}(H, A)$ by

$$\epsilon_{\mathrm{MSE}} := \mathbf{E}\{(C - \tilde{C})^T A(C - \tilde{C})\}, \tag{4.11}$$

where $x := C^T\Phi$ is a zero-mean random process in $V_2(\Phi)$ with covariance function $\mathbf{cov}(x(t), x(t'))$, $\tilde{x} := \tilde{C}^T\Phi = Y^T H\Phi$ is the reconstruction signal from the noisy average sampler $Y = \langle x, \Psi\rangle + w$ via the reconstruction filter H, and $w = (w_\gamma)_{\gamma\in\Gamma}$ is an additive zero-mean noise independent of x and with covariance matrix W. In the case that the regularization matrix A in (4.11) is given by a regularity operator R on the space $V_2(\Phi)$ as in (3.13), the mean squared error ϵ_{MSE} in (4.11) becomes $\mathbf{E}\{\|R(x - \tilde{x})\|^2\}$.

Theorem 4.6 *Let Λ, Γ be finite sets, A be a positive definite matrix, Φ be a Riesz basis of $V_2(\Phi)$, and Ψ be an average sampler that satisfies (2.6). Assume that x is a zero-mean random process in $V_2(\Phi)$ with covariance function $\mathbf{cov}(x(t), x(t'))$, $w = (w_\gamma)_{\gamma\in\Gamma}$ is an additive zero-mean noise that is independent of x and has covariance matrix W, and $A_{\Psi,\Phi}A_{\Phi,\Phi}^{-1}\mathbf{Cov}(x)A_{\Phi,\Phi}^{-1}A_{\Phi,\Psi} + W$ has bounded inverse where*

$$\mathbf{Cov}(x) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{cov}(x(t), x(t'))\Phi(t)(\Phi(t'))^T dtdt'. \tag{4.12}$$

Then the Wiener filter H_W defined by

$$H_{\mathrm{W}} = (A_{\Psi,\Phi}A_{\Phi,\Phi}^{-1}\mathbf{Cov}(x)A_{\Phi,\Phi}^{-1}A_{\Phi,\Psi} + W)^{-1}A_{\Psi,\Phi}A_{\Phi,\Phi}^{-1}\mathbf{Cov}(x)A_{\Phi,\Phi}^{-1}, \tag{4.13}$$

is the reconstruction filter that minimizes the functional ϵ_{MSE} in (4.11), i.e.,

$$H_{\mathrm{W}} = \mathrm{argmin}_H \epsilon_{\mathrm{MSE}}(H, A). \tag{4.14}$$

Proof For any $x \in V_2(\Phi)$, we write $x = C^T\Phi$ for some $C \in \ell^2(\Lambda)$. One may then verify that the covariance matrix associated with the random sequence C and the covariance function

associated with the random signal x are related as follows:

$$\mathbf{cov}(C) = (A_{\Phi,\Phi})^{-1}\mathbf{Cov}(x)(A_{\Phi,\Phi})^{-1}. \tag{4.15}$$

For $x = C^T\Phi \in V_2(\Phi)$, the reconstruction signal $\tilde{x} := \tilde{C}^T\Phi$ from the noisy samples via the reconstruction filter H is given by

$$\tilde{x} = (C^T A_{\Phi,\Psi} + w^T)H\Phi,$$

where w is additive noise with zero mean and covariance matrix W. Therefore

$$\begin{aligned}
\epsilon_{\mathrm{MSE}} &= \mathbf{E}\{(C-\tilde{C})^T A(C-\tilde{C})\} \\
&= \mathbf{E}\big\{C^T(A_{\Phi,\Psi}H - I)A(H^T A_{\Psi,\Phi} - I)C\big\} + \mathbf{E}\big\{w^T HAH^T w\big\} \\
&= \mathbf{trace}\big(\mathbf{cov}(C)(A_{\Phi,\Psi}H - I)A(H^T A_{\Psi,\Phi} - I)\big) + \mathbf{trace}\big(WHAH^T\big) \\
&= \mathbf{trace}\big((H^T A_{\Psi,\Phi} - I)\mathbf{cov}(C)(A_{\Phi,\Psi}H - I)A\big) + \mathbf{trace}\big(H^T WHA\big) \\
&= \epsilon_{\mathrm{MSE}}^{\mathrm{opt}} + \mathbf{E}(w^T(H - H_{\mathrm{W}})A(H - H_{\mathrm{W}})^T w) \\
&\quad + \mathbf{E}(C^T A_{\Phi,\Psi}(H - H_{\mathrm{W}})A(H - H_W)^T A_{\Psi,\Phi}C) \\
&\geq \epsilon_{\mathrm{MSE}}^{\mathrm{opt}},
\end{aligned} \tag{4.16}$$

where $\mathbf{trace}(B)$ is the trace of a square matrix B, and $\epsilon_{\mathrm{MSE}}^{\mathrm{opt}}$ is the MSE associated with the reconstruction filter H_W. The conclusion in Theorem 4.6 then follows. □

Remark 4.7 The Wiener filter (4.13) in the stochastic MSE is independent of the regularization matrix A in (4.11), and coincides with the reconstruction filter (4.4) in the minimax MSE if the regularization matrix A in (4.3) is chosen so that

$$\mathbf{cov}(x(t), x(t')) = \delta_0^2(\Phi(t))^T A^{-1}\Phi(t'). \tag{4.17}$$

Remark 4.8 Similar to the worst-case MSE, we define the linearized mean squared error (LMSE)

$$\epsilon_{\mathrm{LMSE}} = \mathbf{E}\{|\mathcal{M}(x - \tilde{x})|^2\}, \tag{4.18}$$

where $\mathcal{M}$ is a linear functional with $\mathcal{M}\Phi \in \ell^2(\Lambda)$, x is zero-mean random process in $V_2(\Phi)$ with covariance function $\mathbf{cov}(x(t), x(t'))$, $\tilde{x}$ is the reconstruction signals from the noisy average sampler $Y = \langle x, \Psi\rangle + w$ via the reconstruction filter H, $w = (w_\gamma)_{\gamma\in\Gamma}$ is an additive zero-mean noise independent of x and with covariance matrix W. One may then verify that the Wiener filter H_{W} in (4.13) is also the reconstruction filter that minimizes the linearized MSE ϵ_{LMSE} in (4.18).

The next result gives a lower bound estimate of the stochastic MSE ϵ_{MSE} for identically distributed random signal $x \in V_2(\Phi)$ and noise w with zero means.

Theorem 4.9 *Let Λ and Γ be finite sets, $\Phi = (\phi_\lambda)_{\lambda\in\Lambda}$ be an orthonormal basis of its span $V_2(\Phi)$, and $\Psi = (\psi_\gamma)_{\gamma\in\Gamma}$ be a stable average sampler on $V_2(\Phi)$, i.e. (2.10) holds. Assume that x is an identically distributed random signal in $V_2(\Phi)$ that has zero mean and covariance function*

$$\mathbf{cov}(x(t), x(t')) = \sigma_0^2\Phi(t)^T\Phi(t') \tag{4.19}$$

for some $\sigma_0 > 0$, and that w is the identically distributed random noise that has zero mean and the covariance matrix

$$W = \sigma_1^2 I \tag{4.20}$$

for some $\sigma_1 \geq 0$. Then the stochastic MSE ϵ_{MSE} in (4.11) satisfies

$$\epsilon_{\mathrm{MSE}} \geq \frac{N^2\sigma_0^2\sigma_1^2}{\|\Psi\|^2\sigma_0^2 + N\sigma_1^2} + (M-N)\sigma_0^2 \tag{4.21}$$

where $\|\Psi\| = (\sum_{\gamma\in\Gamma}\|\psi_\gamma\|_2^2)^{1/2}$, and M and N are the cardinality of the sets Λ and Γ respectively. The inequality (4.21) becomes an equality if and only if the Wiener filter H_{W} is used as the reconstruction filter and $\Psi \subset V_2(\Phi)$ is a tight frame for $V_2(\Phi)$, i.e., there exists a positive constant B such that

$$\sum_{\gamma\in\Gamma} |\langle f, \psi_\gamma\rangle|^2 = B\|f\|_2^2 \quad \textit{for all } f \in V_2(\Phi).$$

Proof By (4.16), (4.19) and (4.20), the stochastic MSE ϵ_{MSE} in (4.11) is larger than or equal to

$$\begin{aligned}
&\mathbf{trace}\big(\mathbf{cov}(C)(A_{\Phi,\Psi}H_{\mathrm{W}} - I)A(H_{\mathrm{W}}^T A_{\Psi,\Phi} - I)\big) + \mathbf{trace}\big(W H_{\mathrm{W}} A H_{\mathrm{W}}^T\big)\\
&\quad = \mathbf{trace}\big(-\mathbf{cov}(C)A_{\Phi,\Psi}(A_{\Psi,\Phi}\mathbf{cov}(C)A_{\Phi,\Psi} + W)^{-1}A_{\Psi,\Phi}\mathbf{cov}(C)A + \mathbf{cov}(C)A\big)\\
&\quad = \sigma_0^2\sigma_1^2\mathbf{trace}\big((\sigma_0^2 A_{\Psi,\Phi}A_{\Phi,\Psi} + \sigma_1^2 I)^{-1}\big),
\end{aligned} \tag{4.22}$$

where we have used the fact that $\mathbf{trace}(AB) = \mathbf{trace}(BA)$ for any matrices A and B. Let λ_k, $1 \leq k \leq N$, are eigenvalues of the matrix $A_{\Phi,\Psi}A_{\Psi,\Phi}$. Noting that all eigenvalues of the matrix $A_{\Phi,\Psi}A_{\Psi,\Phi}$ are positive by the stability of the average sampler Ψ, and that the sets of nonzero eigenvalues of the matrices $A_{\Psi,\Phi}A_{\Phi,\Psi}$ and $A_{\Phi,\Psi}A_{\Psi,\Phi}$ are the same, we then have that $N \leq M$ and $\lambda_1, \ldots, \lambda_N, 0, \ldots, 0$ are all eigenvalues of the matrix $A_{\Psi,\Phi}A_{\Phi,\Psi}$. This together with (4.22) implies that

$$\begin{aligned}
\epsilon_{\mathrm{MSE}} &= \sigma_0^2\sigma_1^2\sum_{k=1}^{N}(\sigma_0^2\lambda_k + \sigma_1^2)^{-1} + (M-N)\sigma_0^2\\
&\geq N\sigma_0^2\sigma_1^2\left(\frac{\sigma_0^2\sum_{\gamma\in\Gamma}\|P\psi_\gamma\|_2^2}{N} + \sigma_1^2\right)^{-1} + (M-N)\sigma_0^2\\
&\geq \frac{N^2\sigma_0^2\sigma_1^2}{\|\Psi\|^2\sigma_0^2 + N\sigma_1^2} + (M-N)\sigma_0^2,
\end{aligned} \tag{4.23}$$

where we have the facts that

$$\sum_{k=1}^{N}\lambda_k = \mathbf{trace}(A_{\Phi,\Psi}A_{\Psi,\Phi}) = \sum_{\gamma\in\Gamma}\|P\psi_\gamma\|^2 \leq \sum_{\gamma\in\Gamma}\|\psi_\gamma\|^2 = \|\Psi\|^2,$$

to obtain the first and second inequalities, and P is the project operator from L^2 to $V_2(\Phi)$.

From the above argument, we see that all inequalities in (4.22) and (4.23) become equalities if and only if the Wiener filter H_{W} in (4.13) is used as the reconstruction filter and all

eigenvalues of the matrix $A_{\Phi,\Psi}A_{\Psi,\Phi}$ are equal. Therefore the second conclusion in Theorem 4.9 follows. □

Remark 4.10 We say that the average sampler $\Psi = \{\psi_\gamma\}_{\gamma\in\Gamma}$ is *normalized* if $\|\psi_\gamma\|_2 = 1, \gamma \in \Gamma$. In this case, the lower bound estimate in (4.21) becomes

$$\epsilon_{\mathrm{MSE}} \geq \frac{N^2\sigma_0^2\sigma_1^2}{M\sigma_0^2 + N\sigma_1^2} + (M-N)\sigma_0^2 \geq \frac{N\sigma_0^2\sigma_1^2}{\sigma_0^2+\sigma_1^2},$$

where we have used the fact that $M \geq N$ to obtain the second inequality. This indicates that an orthonormal basis of the space $V_2(\Phi)$ (for instance the generator Φ) is the **best** normalized average sampler on that space $V_2(\Phi)$.

5 Numerical Simulations

In this section, we present the results of some numerical simulations for reconstructing a signal with finite rate of innovation from the given noisy samples.

5.1 Adaptive Tikhonov Approach

In the following simulation, we use the adaptive Tikhonov approach developed in Sect. 3 to reconstruct Dirac streams and cubic splines from their corrupted average samples. For this purpose, we use the following regularization matrix

$$A_d = \mathrm{diag}\bigg(\frac{2}{\lambda_2-\lambda_1}, \frac{\lambda_3-\lambda_1}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)}, \ldots, \frac{\lambda_N-\lambda_{N-2}}{(\lambda_{N-1}-\lambda_{N-2})(\lambda_N-\lambda_{N-1})}, \frac{2}{\lambda_N-\lambda_{N-1}}\bigg) \tag{5.1}$$

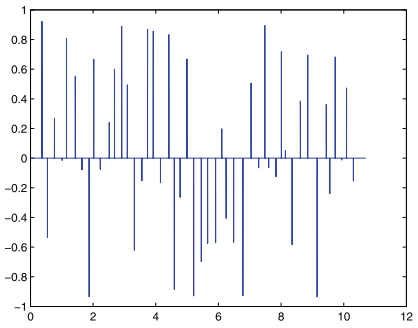
for the Dirac streams $x = \sum_{n=1}^N c_n\delta_{\lambda_n}$, and the following regularization matrix

$$A_c = \bigg(\int_{\mathbb{R}} B_{3,\delta_n}(t)B_{3,\delta_m}(t)dt\bigg)_{1\leq n,m\leq N} + \bigg(\int_{\mathbb{R}} B''_{3,\delta_n}(t)B''_{3,\delta_m}(t)dt\bigg)_{1\leq n,m\leq N}, \tag{5.2}$$

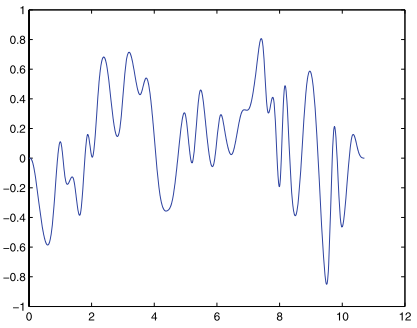
for the nonuniform mesh cubic spline signal $y = \sum_{n=1}^N c_n B_{3,\lambda_n}$. Also for the sampling functional $\Psi = \{\psi_{\gamma_n}\}_{\gamma_n\in\Gamma}$ with $\gamma_1 < \gamma_2 < \cdots < \gamma_K$ we will use the following preconditioner:

$$L_g := \sum_{p=1}^P \mathcal{L}_p^*\mathcal{L}_p = \mathrm{diag}(\tilde{\gamma}_1, \ldots, \tilde{\gamma}_K) \tag{5.3}$$

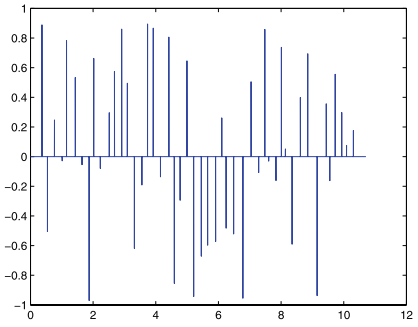
where $\tilde{\gamma}_1 = (\gamma_2-\gamma_1)^{-1}$, $\tilde{\gamma}_k = (\gamma_{k+1}-\gamma_{k-1})^{-1}$ if $2 \leq k \leq K-1$, and $\tilde{\gamma}_K = (\gamma_K-\gamma_{K-1})^{-1}$. In Fig. 5, we present the reconstructed signals from noisy sampling data via the adaptive Tikhonov approach developed in Sect. 3, i.e., the Tikhonov approach with the panel parameter α_{op} being obtained in the Optimal Parameter Algorithm.



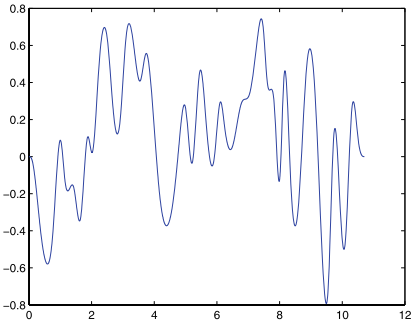
(a)



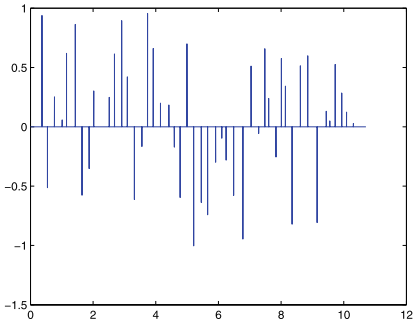
(b)



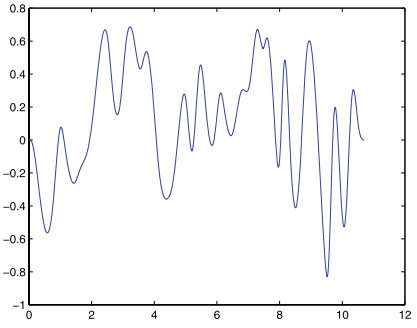
(c)



(d)



(e)



(f)

Fig. 5 The reconstructed Dirac streams $\sum_{n=1}^{N} \tilde{c}_n \delta_{\lambda_n}$ (the *left hand side*) and the nonuniform mesh cubic spline $\sum_{n=1}^{N} \tilde{d}_n B_{3,\lambda_n}$ (the *right hand side*) via the adaptive Tikhonov approach. The sampling functionals $\Psi := (\psi_\gamma)$ from *top* to *bottom* are the T-shifted Gaussian sampler $\exp(-2|t - kT|^2/T^2)$, $1 \le k \le [\frac{\lambda_N + 2\mu}{T}] + 1$, with the ratios between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x being $6/5, 1, 4/5$ respectively. The noisy data Y are given in Fig. 4, the postfilter is the matrix L_g in (5.3), and the regularization matrix is the matrix A_d in (5.1) for figures on the *left hand side* and the matrix A_c in (5.2) for figures on the *right hand side* respectively

The difference between the reconstructed signal $\tilde{x} := \sum_{n=1}^{N} \tilde{c}_n \delta_{\lambda_n}$ and the original Dirac stream $x = \sum_{n=1}^{N} c_n \delta_{\lambda_n}$, and the difference between the sample $\langle \tilde{x}, \Psi \rangle$ of the constructed signal $\tilde{x}$ and the given noisy sample Y are presented on the left hand side and on the right hand side of Fig. 6 respectively. In the noiseless sampling environment (i.e. $\sigma_0 = 0$ in Fig. 4), the ℓ^2-norm $(\sum_{n=1}^{N} |\tilde{c}_n - c_n|^2)^{1/2}$ of the difference for the reconstructed signal $\tilde{x} = \sum_{n=1}^{N} \tilde{c}_n \delta_{\lambda_n}$ and the original Dirac stream $x = \sum_{n=1}^{N} c_n \delta_{\lambda_n}$ is $1.0784 \times 10^{-8}, 0.5637, 1.7062$ for the ratios between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x being $6/5, 1, 4/5$ respectively. In the same noiseless sampling environment, our simulation shows that the ℓ^2 norm of the difference between the sample $\langle \tilde{x}, \Psi \rangle$ of the reconstructed signal $\tilde{x}$ and the given noiseless sample $Y = \langle x, \Psi \rangle$ is $1.1563 \times 10^{-9}, 0.0093, 3.2993 \times 10^{-7}$, while the ℓ^2 norm of the given noiseless sample $Y = \langle x, \Psi \rangle$ is $3.7050, 3.7265, 3.6214$ for the ratio between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x being $6/5, 1, 4/5$ respectively.

The difference between the reconstructed nonuniform mesh cubic spline $\tilde{y} = \sum_{n=1}^{N} \tilde{d}_n B_{3,\lambda_n}$ and the original nonuniform mesh cubic spline $y = \sum_{n=1}^{N} d_n B_{3,\lambda_n}$, and the difference between the sample $\langle \tilde{y}, \Psi \rangle$ of the constructed signal $\tilde{y}$ and the given noisy sample Y are presented on the left hand side and on the right hand side of Fig. 7 respectively. In the noiseless environment (i.e. $\sigma_0 = 0$ in Fig. 4), the ℓ^2-norm $(\sum_{n=1} |\tilde{d}_n - d_n|^2)^{1/2}$ of the difference between the reconstructed nonuniform mesh cubic spline $\tilde{y} = \sum_{n=1}^{N} \tilde{d}_n B_{3,\lambda_n}$ and the original nonuniform mesh cubic spline $y = \sum_{n=1}^{N} d_n B_{3,\lambda_n}$ is given by $0.0162, 0.6101, 1.5942$ for the ratio between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x being $6/5, 1, 4/5$ respectively. The ℓ^2 norm of the difference between the sample data $\langle \tilde{y}, \Psi \rangle$ of the reconstructed cubic spline $\tilde{y}$ and the given noiseless sample data $\langle y, \Psi \rangle$ is $3.3688 \times 10^{-4}, 0.0075, 2.6447 \times 10^{-4}$; and the ℓ^2 norm of the given noiseless sample data $\langle y, \Psi \rangle$ is $0.5466, 0.5835, 0.6286$ for the ratio between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x being $6/5, 1, 4/5$ respectively.

From the above demonstration, we notice that our reconstruction procedure is almost consistent in the noiseless sampling procedure, and robust against noise. We also observe that the procedure has better performance for the oversampling case than for the undersampling case, and has much smaller error between the sampling data of our reconstructed signal and the given noisy sampling data than the error between the reconstructed signal and the original signal.

5.2 Local Implementation of Adaptive Tikhonov Approach

In the following simulation, we want to demonstrate that our adaptive Tikhonov approach can be implemented locally. The setting in the simulation is as follows: the nonuniform mesh cubic spline $x = \sum_{n=1}^{N} d_n B_{3,\lambda_n}$ in Fig. 1 is taken as the original signal, the average sampler $\Psi = \{\delta_{kT}\}_{k=1}^{K}$ is generated by the delta functional on the set $\Gamma = \{kT\}_{k=1}^{K}$ with $T = \frac{5}{6}\mu$ and $K = [(\lambda_N + 2\mu)/T] + 1$, the exact sample Y_{cd} is given by

$$Y_{cd} = \left\{ \sum_{n=1}^{N} d_n B_{3,\lambda_n}(kT) | 1 \le k \le K \right\}, \tag{5.4}$$

and the noisy sample Y is generated by

$$Y = Y_{cd} + \sigma_0 \max(|Y_{cd}|)(r(\gamma))_{\gamma \in \Gamma} \tag{5.5}$$

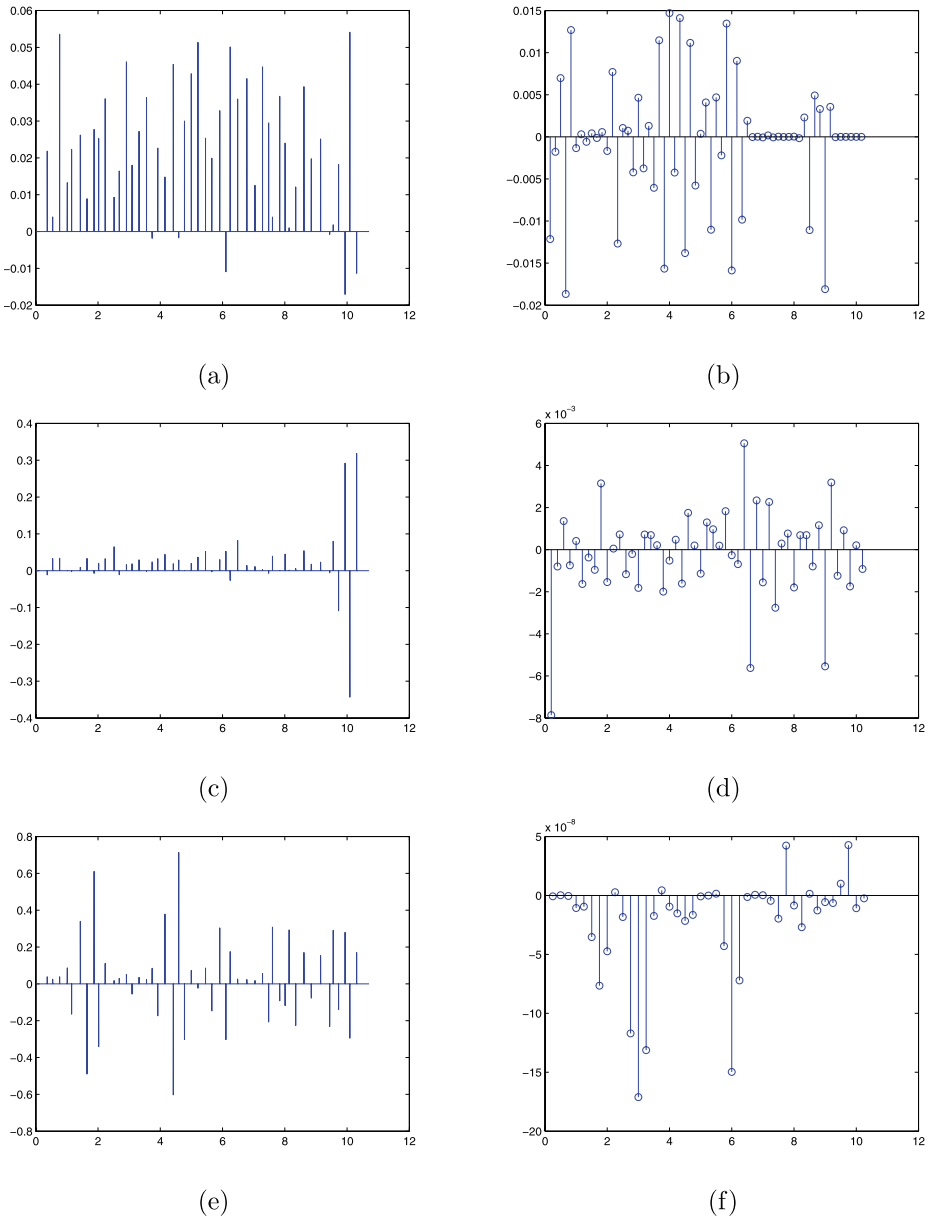


Fig. 6 The difference between the reconstructed Dirac streams $\tilde{x} = \sum_{n=1}^{N} \tilde{c}_n \delta_{\lambda_n}$ in Fig. 5 and the original Dirac streams $x = \sum_{n=1}^{N} \tilde{c}_n \delta_{\lambda_n}$ (the *left hand side*), and the difference between the sample $\langle \tilde{x}, \Psi \rangle$ of the reconstructed Dirac streams $\tilde{x}$ and the given noisy data Y in Fig. 5 (the *right hand side*). From *top* to *bottom*, the ℓ^2-norms $(\sum_{n=1}^{N} |\tilde{c}_n - c_n|^2)^{1/2}$ of the difference of three signals are 0.1992, 0.6009, 1.7326 respectively, while the ℓ^2-norm $(\sum_{n=1}^{N} |c_n|^2)^{1/2}$ of the original signal is 4.0612; the ℓ^2-norms of the difference $\langle \tilde{x}, \Psi \rangle - Y$ of three sample data are $0.0589, 0.0154, 3.2593 \times 10^{-7}$ respectively, and the ℓ^2-norms of the given noisy samples are 3.7185, 3.7392, 3.6461 respectively

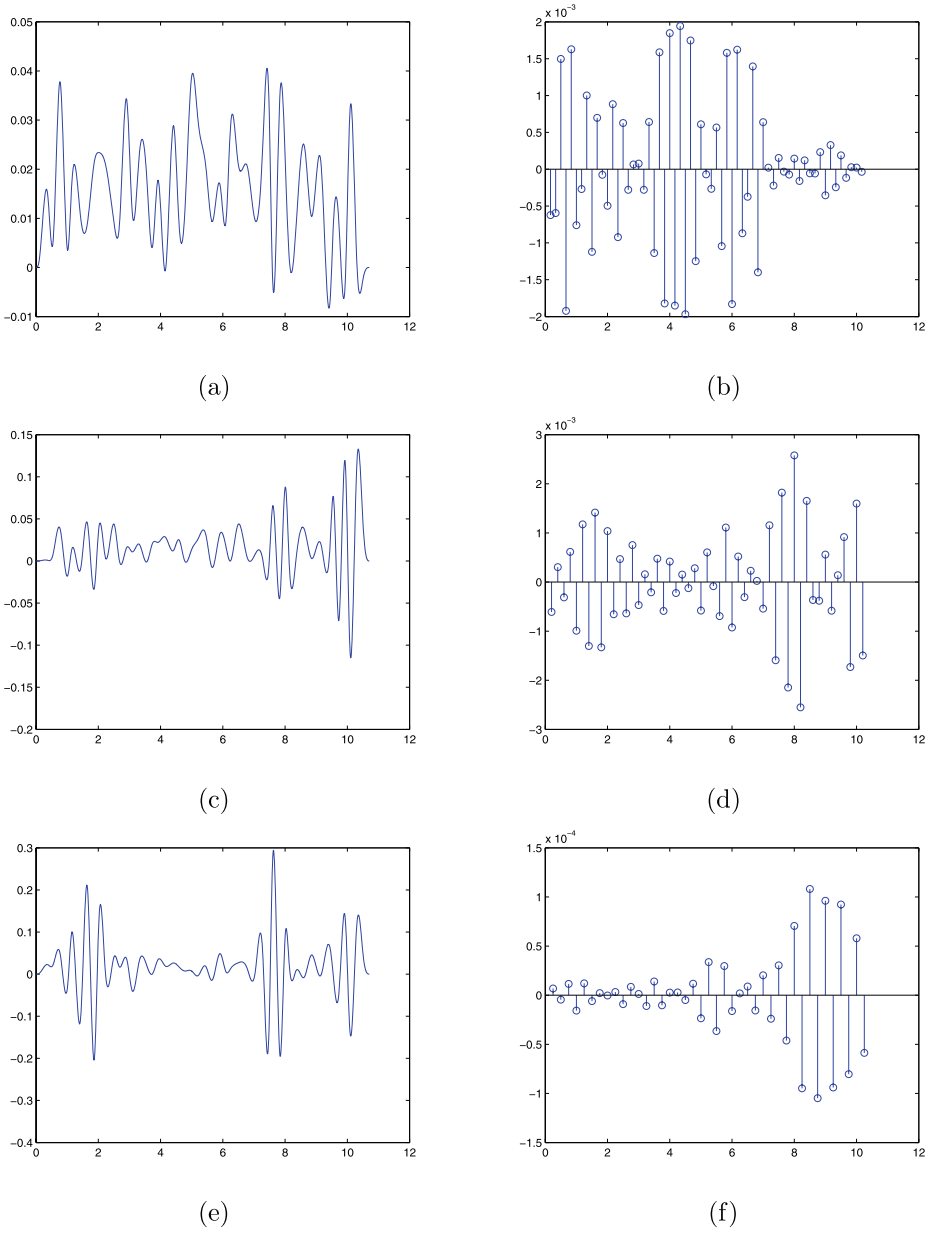


Fig. 7 The difference between the reconstructed cubic spline $\tilde{y} = \sum_{n=1}^{N} \tilde{d}_n B_{3,\lambda_n}$ in Fig. 5 and the original cubic spline $y = \sum_{n=1}^{N} d_n B_{3,\lambda_n}$ (the *left hand side*), and the difference between sampling data $\langle \tilde{y}, \Psi \rangle$ of the reconstructed cubic spline $\tilde{y}$ and the given noisy data Y (the *right hand side*). On the *left hand side* from *top* to *bottom*, the ℓ^2-norms $(\sum_{n=1} |\tilde{d}_n - d_n|^2)^{1/2}$ of the difference of three signals are 0.2056, 0.7243, 1.6116 while the ℓ^2-norm $(\sum_{n=1}^{N} |d_n|^2)^{1/2}$ of the original cubic spline signal y is 3.6566. On the *right hand side* from *top* to *bottom*, the ℓ^2-norms of the difference $\langle \tilde{y}, \Psi \rangle - Y$ of three sample data are 0.0076, 0.0073, 0.0003, and the ℓ^2-norm of the original samples Y are 0.5568, 0.5907, 0.6365, respectively

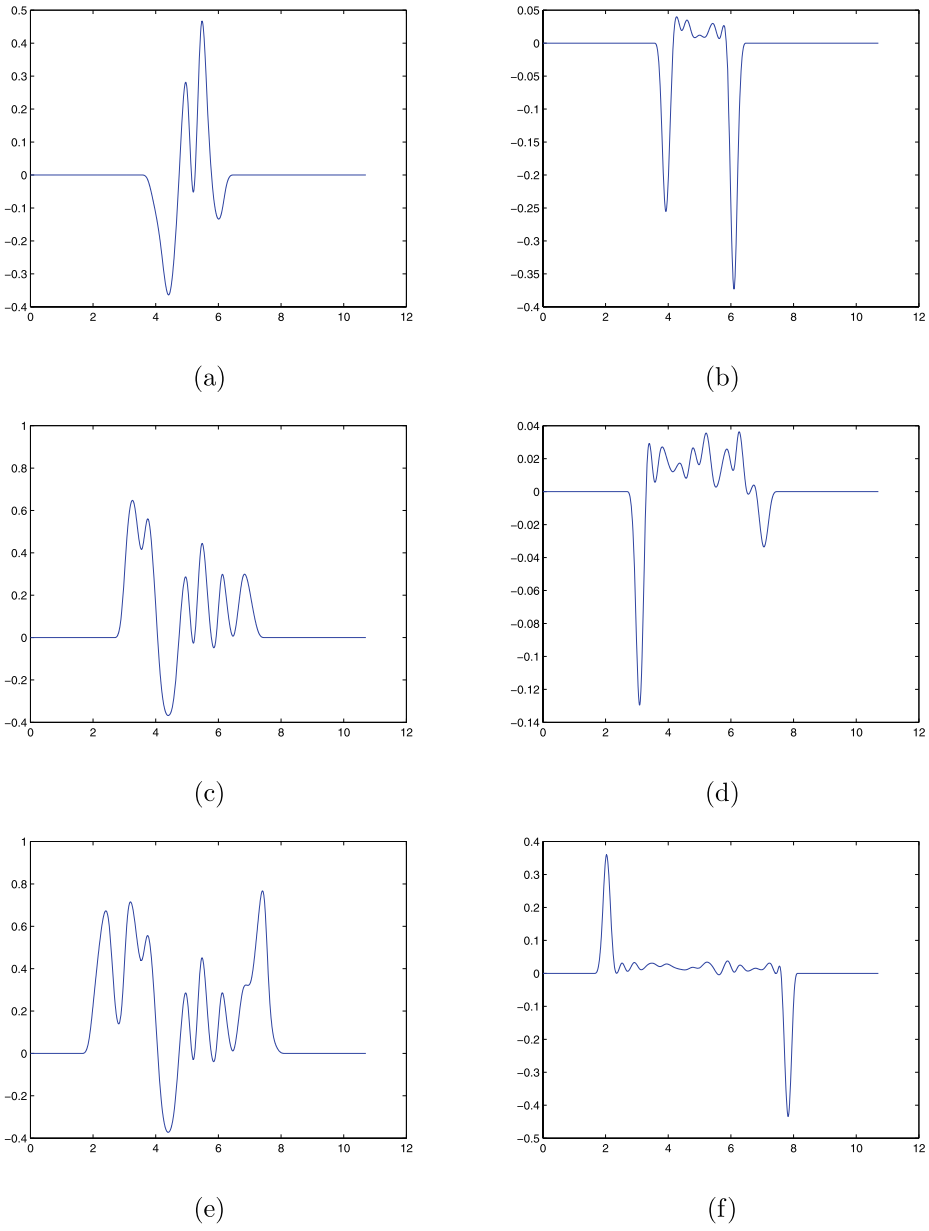


Fig. 8 The locally-reconstructed cubic spline $\sum_{\lambda_n \in \Lambda_{k_0,R}} \tilde{d}_{n,R} B_{3,\lambda_n}$ (the *left hand side*) and its difference to the original signal $\sum_{\lambda_n \in \Lambda_{k_0,R}} d_{n,R} B_{3,\lambda_n}$ (the *right hand side*) from its noisy sample data $Y_{k_0,R}$ located on $(k_0 - R)T, \ldots, (k_0 + R)T$, where $R = 5, 10, 15$ from *top* to *bottom*. Here $N = 50$ and $k_0 = 15$. The local error measurement for the difference of the locally-reconstructed cubic spline and the original spline is given in Fig. 9

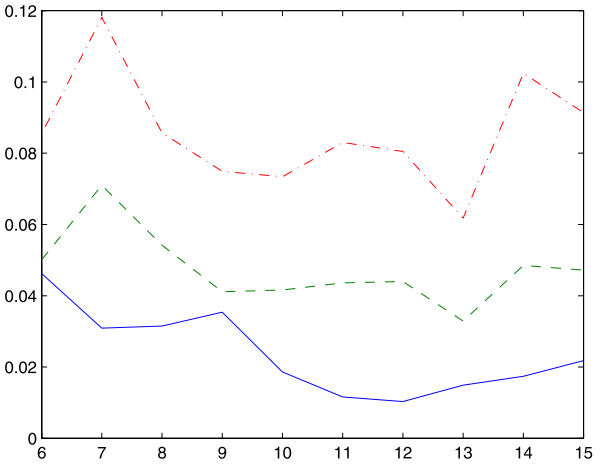


Fig. 9 The error measurement $D(R) = (\sum_{\lambda_n \in \Lambda_{k_0,5}} \times |d_{n,R} - d_n|^2)^{1/2}$ between the locally-reconstructed cubic spline $\sum_{\lambda_n \in \Lambda_{k_0,R}} \tilde{d}_{n,R} B_{3,\lambda_n}$ and the original signal $\sum_{\lambda_n \in \Lambda_{k_0,R}} d_{n,R} B_{3,\lambda_n}$ from noisy samples with the noise level σ_0 in Fig. 4 being $0, 0.025, 0.05$ are presented by *solid line*, *dashed line* and *dash-dotted line* respectively

where the noise level σ_0 is 0.05 and $r(\gamma), \gamma \in \Gamma$ are numbers in $[-1, 1]$ chosen randomly, the post-filter L is given by

$$L := \sum_{p=1}^{P} \mathcal{L}_p^* \mathcal{L}_p = \text{diag}(0, \ldots, 0, \tilde{\gamma}_{k_0-R}, \ldots, \tilde{\gamma}_{k_0+R}, 0, \ldots, 0) \tag{5.6}$$

where $\tilde{\gamma}_k, 1 \le k \le K$, are defined in (5.3), and the regularization matrix A is determined by (5.2).

Now let us define the localized versions of the sample Y, the post-filter L, the regularization matrix A, the generator Φ, the average sampler Ψ. Given $1 \le k_0 \le N$ and $R < \min(k_0, N - k_0)$, define $\Gamma_{k_0,R} = \{kT\}_{k=k_0-R}^{k_0+R}$ and $\Lambda_{k_0,R} = \{\lambda_n\}_{n=n_0}^{n_1}$ where n_0 is the index such that $\lambda_{n_0+2} > (k_0 - R)T$ and $\lambda_{n_0+1} \le (k_0 - R)T$ and n_1 is the index such that $\lambda_{n_0-2} \le (k_0 + R)T$ and $\lambda_{n_0-1} > (k_0 + R)T$. Using these two index sets, we define $\Psi_R = \{\delta_\mu\}_{\mu \in \tilde{\Gamma}_R}$ as the local version of the average sampler Ψ, $\Phi_{k_0,R} = \{B_{3,\lambda_n}\}_{\lambda_n \in \Lambda_{k_0,R}}$ as the local version of the generator Φ, $L_{k_0,R} = \text{diag}(\tilde{\gamma}_{k_0-R}, \ldots, \tilde{\gamma}_{k_0+R})$ as the local version of the post-filter L in (5.6), the matrix $A_{K_0,R}$ obtained by taking all λ_n-columns and rows of the regularization matrix A_c in (5.2) where $\lambda_n \in \Lambda_{k_0,R}$ as the local version of the regularization matrix A_c, and the sequence $Y_{k_0,R} = (y(\gamma))_{\gamma \in \Gamma_{k_0,R}}$ as the local version of the noisy sample $Y = (y(\gamma))_{\gamma \in \Gamma}$ in (5.5).

We will use the reconstruction formula (3.12) to reconstruct signal from its noisy sampling data $Y_{k_0,R}$ obtained at locations $\gamma_{k_0-R}, \ldots, \gamma_{k_0+R}$, and the panel parameter α is selected by the Panel Parameter Algorithm with the noisy sample Y, the post filter L, the regularization matrix A, the generator Φ and the average sampler Ψ being replaced by their local versions $Y_{k_0,R}, L_{k_0,R}, A_{k_0,R}, \Phi_{k_0,R}, \Psi_{k_0,R}$ respectively.

We observe that this regularization matrix A in this simulation is not a diagonal matrix, but we still use (3.12) to reconstruct signal from local sample data because it is the solution that minimizes the regularized least squares functional $\epsilon_{\text{TIK}}(\tilde{x})$ with $\tilde{x} \in V_2(\Phi_{k_0,R})$ (instead of its superspace $V_2(\Phi)$).

In Fig. 8 we let $N = 50$ and $k_0 = 15$. On the left hand side of Fig. 8 we present the reconstruction cubic spline $\sum_{\lambda_n \in \Lambda_{k_0,R}} \tilde{d}_{n,R} B_{3,\lambda_n}$ for $R = 5, 10, 15$ (from top to bottom) from noisy sampling data $Y_{k_0,R}$. On the right hand side of Fig. 8 is the difference $\sum_{\lambda_n \in \Lambda_{k_0,R}} (\tilde{d}_{n,R} -$

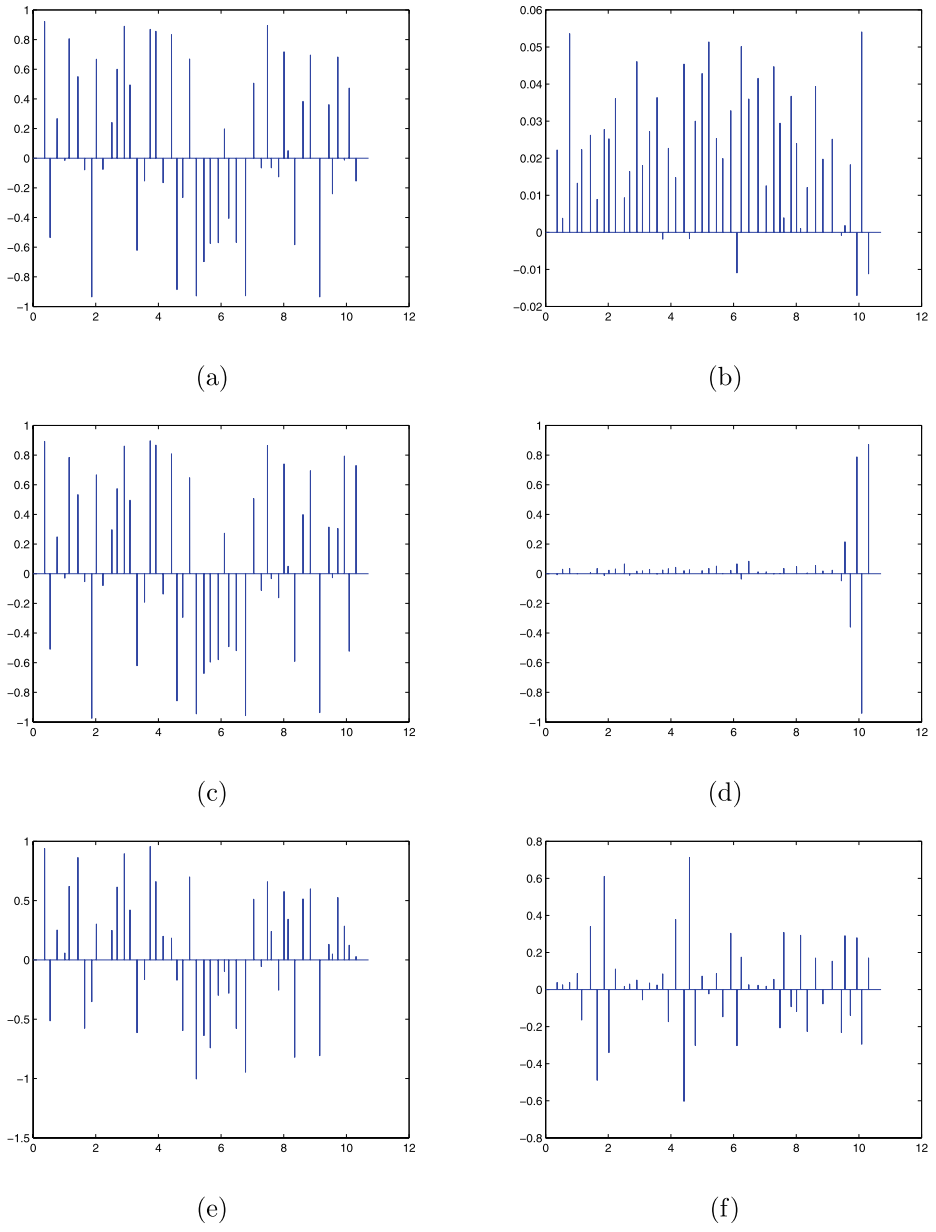


Fig. 10 The reconstructed Dirac streams $\tilde{x}$ (the *left hand side*) obtained by applying the filter H_{WOR} in (4.4), and the difference between the reconstructed Dirac streams and the original Dirac streams (the *right hand side*). From top to bottom the ratios between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x are $6/5, 1, 4/5$ respectively. The average sampler Ψ and the regularization matrix A are the same as in Fig. 4, the noisy sample Y in (4.1) is the same as the noisy sample in Fig. 4, and the covariance matrix W for the additive noises is $\frac{2}{3}\sigma_0^3 \max(|Y_{dg}|)^3 I$ with $\sigma_0 = 0.05$ and Y_{dg} given in Fig. 4

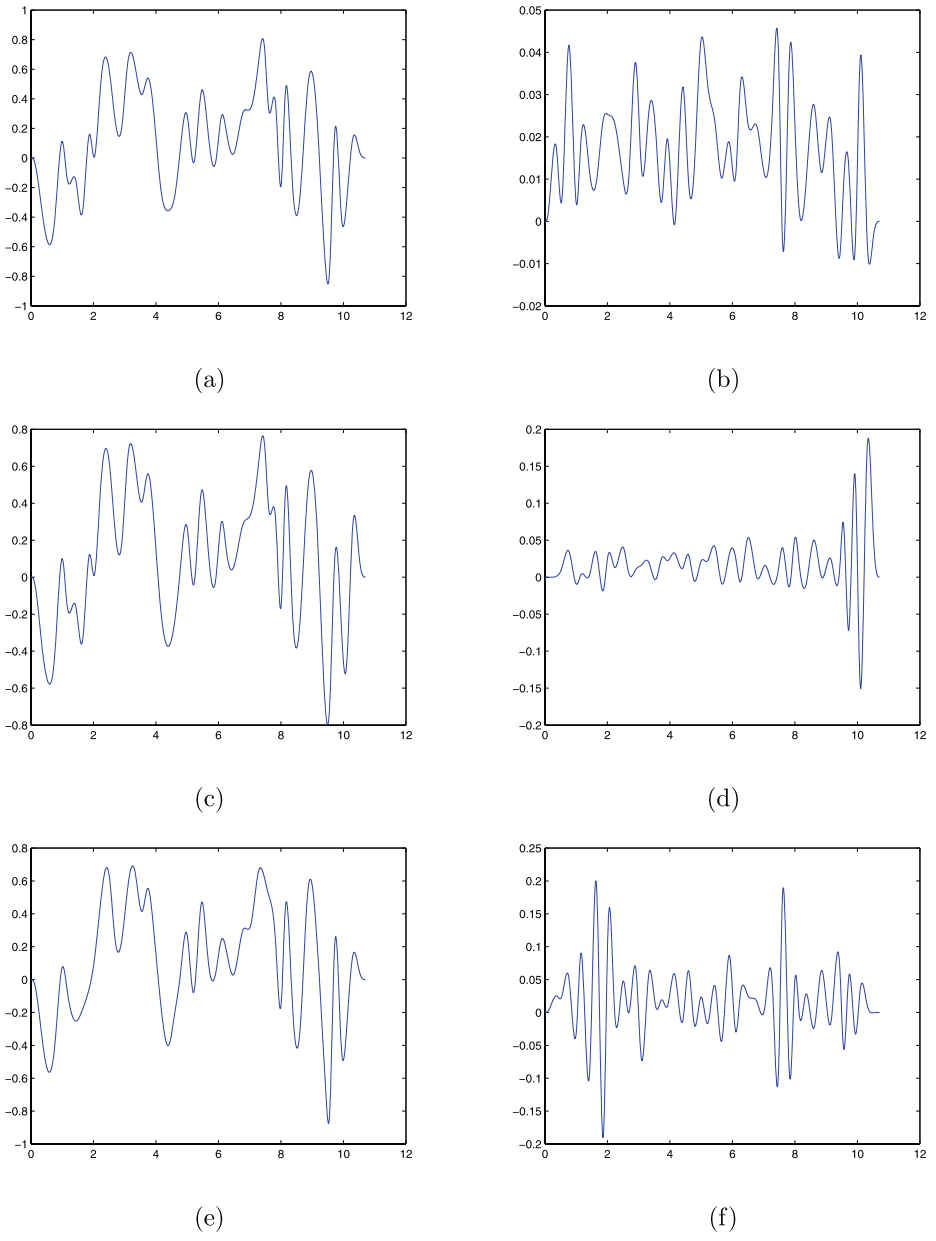


Fig. 11 The reconstructed cubic spline $\tilde{y}$ (the *left hand side*) obtained by applying the filter H_{WOR} in (4.4), and the difference between the reconstructed cubic spline and the original cubic spline (the *right hand side*). From *top* to *bottom* the ratios between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x are $6/5, 1, 4/5$ respectively. The average sampler Ψ is the same as in Fig. 4, the regularization matrix A is given in (3.13) with the regularizer R being the identity operator, the noisy sample Y in (4.1) is the same as the noisy sample in Fig. 4, and the covariance matrix W for the additive noises is $\frac{2}{3}\sigma_0^3 \max(|Y_{cg}|)^3$ with $\sigma_0 = 0.05$ and Y_{cg} given in Fig. 4

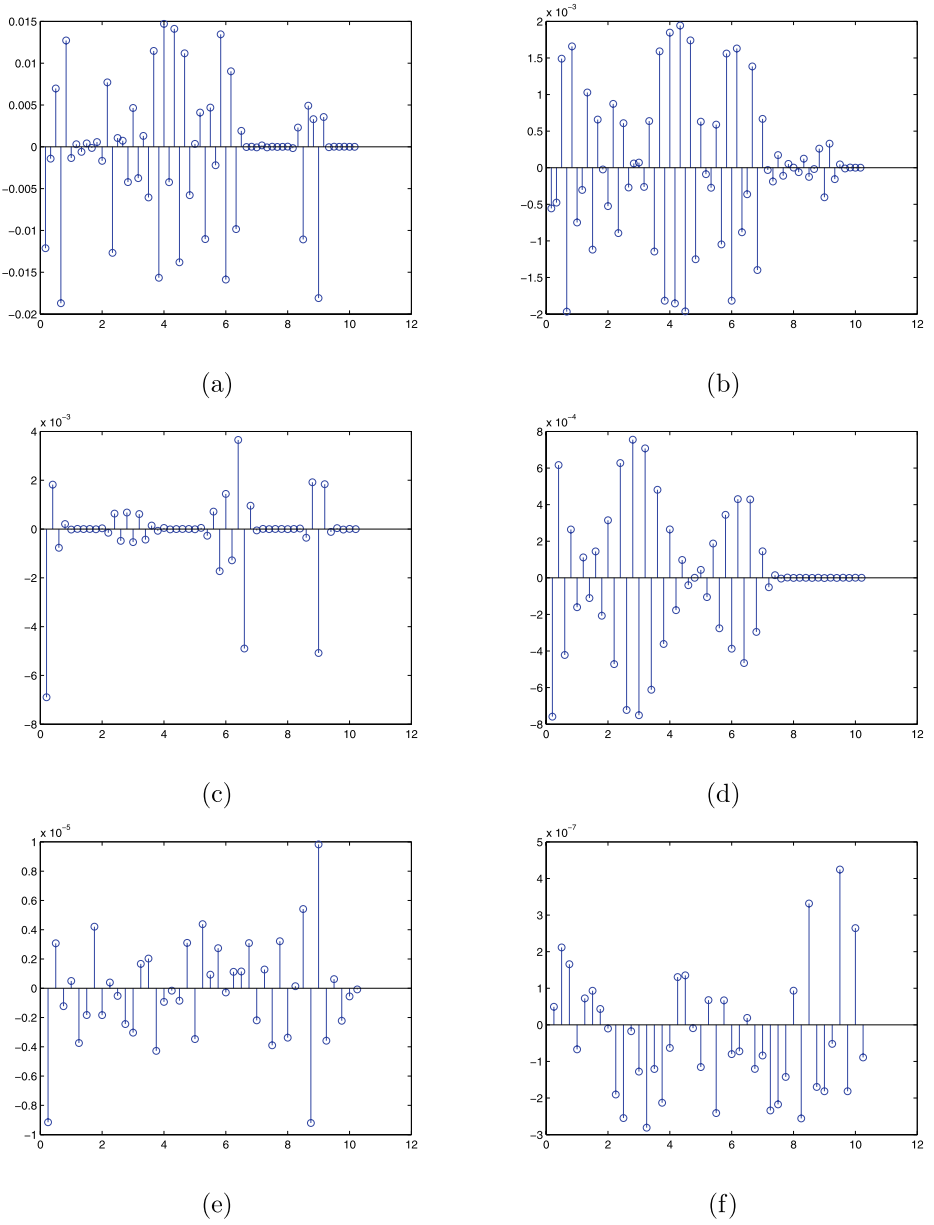


Fig. 12 The difference $(e_1(\gamma))_{\gamma\in\Gamma}$ between sampling data $\langle \tilde{x}, \Psi\rangle$ of the reconstructed Dirac stream $\tilde{x}$ in Fig. 10 and the given noisy data Y (the *left hand side*), and the difference $(e_2(\gamma))_{\gamma\in\Gamma}$ between sampling data $\langle \tilde{y}, \Psi\rangle$ of the reconstructed cubic spline $\tilde{y}$ in Fig. 11 and the given noisy data Y (the *right hand side*). From *top* to *bottom* the ratios between sampling rate $1/T$ of Gaussian sampling devices and innovative rate $1/\mu$ of signal x are $6/5, 1, 4/5$ respectively. From *top* to *bottom*, the errors $(\sum_{\gamma\in\Gamma}|e_1(\gamma)|^2)^{1/2}$ for reconstructing Dirac streams are 0.0621, 0.0228, 0.0217, respectively, while the errors $(\sum_{\gamma\in\Gamma}|e_2(\gamma)|^2)^{1/2}$ for reconstructing cubic splines are 0.0081, 0.0039, 0.0036, respectively

$d_n)B_{3,\lambda_n}$ of the reconstruction cubic spline and the original cubic spline with $R = 5, 10, 15$ (from top to bottom).

We will use

$$D(R) := \Big(\sum_{\lambda_n \in \Lambda_{k_0,5}} |d_{n,R} - d_n|^2 \Big)^{1/2}$$

to measure the coefficient difference between the reconstructed cubic spline and the original cubic spline near the position λ_n close to the center $k_0 T$ of the given sampling location, see Fig. 9.

From Figs. 8 and 9, we see that there is some boundary effect in our local reconstruction but the boundary effect fades away in the center of the sampling location as we have more sampling data around that center. This, in the other words, means that our adaptive Tikhonov approach can be implemented locally and could be possibly used in real-time implementation.

5.3 Tikhonov Approach for Random Noise

In Figs. 10, 11, and 12, we present the reconstructed Dirac streams and cubic splines obtained by applying the filter H_{WOR} in (4.4), the difference between the reconstructed signal and the original signal, and the difference between sampling data of the reconstructed signal and the given noisy sample data.

From the above simulation, we observe that the reconstructed signal from our reconstruction procedure have better approximation to the original for the oversampling case than for the undersampling case, while on the other hand the sampling data of the reconstructed signal has smaller error to the given sampling data for the undersampling case than for the oversampling case. The reason behind this observation could be that there is much less freedom for the oversampling case than for the undersampling case to recovery signals from the given sample.

Acknowledgements The authors would like to thank Professor Akram Aldroubi and the reviewer for their suggestions to the improvement.

References

1. Aldroubi, A., Gröchenig, K.: Nonuniform sampling and reconstruction in shift-invariant spaces. SIAM Rev. **43**, 585–620 (2001)
2. Aldroubi, A., Sun, Q., Tang, W.-S.: Convolution, average sampling and a Calderon resolution of the identity for shift-invariant spaces. J. Fourier Anal. Appl. **11**, 215–244 (2005)
3. Blu, T., Dragotti, P.L., Veterli, M., Marziliano, P., Coulot, L.: Sparse sampling of signal innovations: theory, algorithms and performance bounds. IEEE Signal Process. Mag. **31**, 31–40 (2008)
4. Cramer, R.J.-M., Scholtz, R.A., Win, M.Z.: Evaluation of an ultra wide-band propagation channel. IEEE Trans. Antennas Propag. **50**, 561–569 (2002)
5. Dragotti, P.L., Vetterli, M., Blu, T.: Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strans-Fix. IEEE Trans. Signal Process. **55**, 1741–1757 (2007)
6. Eldar, Y., Unser, M.: Nonideal sampling and interpolation from noisy observations in shift-invariant spaces. IEEE Trans. Signal Process. **54**, 2636–2651 (2006)
7. Eldar, Y.C., Dvorkind, T.G.: A minimum squared-error framework for generalized sampling. IEEE Trans. Signal Process. **54**, 2155–2167 (2006)
8. Feichtinger, H.G., Gröchenig, K., Strohmer, T.: Efficient numerical methods in non-uniform sampling theory. Numer. Math. **69**, 423–440 (1995)
9. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)

10. Gröchenig, K., Schwab, H.: Fast local reconstruction methods for nonuniform sampling in shift-invariant spaces. SIAM J. Matrix Anal. Appl. **24**, 899–913 (2003)
11. Hao, Y., Marziliano, P., Vetterli, M., Blu, T.: Compression of ECG as signal with finite rate of innovation. In: 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, China, September 2005
12. Jaffard, S.: Proprietes des matrices "bien localisees" pres de leur diagonale et quelques applications. Ann. Inst. Henri Poincaré Anal. Non-Lineaire **7**, 61–76 (1990)
13. Jerri, J.A.: The Shannon sampling theorem—its various extensions and applications: A tutorial review. Proc. IEEE **65**, 1565–1596 (1977)
14. Kusuma, J., Maravic, I., Vetterli, M.: Sampling with finite rate of innovation: Channel and timing estimation for UWB and GPS. In: IEEE Conference on Communication 2003, Achorage, AK (2003)
15. Maravic, I., Vetterli, M.: Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. IEEE Trans. Signal Process. **53**, 2788–2805 (2005)
16. Marks II, R.J.: Introduction to Shannon Sampling and Interpolation Theory. Springer, Berlin (1991)
17. Marziliano, P., Vetterli, M., Blu, T.: Sampling and exact reconstruction of bandlimited signals with shot noise. IEEE Trans. Inform. Theory **52**, 2230–2233 (2006)
18. Nashed, M.Z., Wahba, G.: Generalized inverses in reproducing kernel spaces: an approach to regularization of linear operator equations. SIAM J. Math. Anal. **5**, 974–987 (1974)
19. Oraizi, H.: Application of the method of least squares to electromagnetic engineering problems. IEEE Antennas Propag. Mag. **48**, 50–74 (2006)
20. Rabinovich, V.S., Roch, S.: Reconstruction of input signals in time-varying filters. Numer. Funct. Anal. Optim. **27**, 697–720 (2006)
21. Ramani, S., De Ville, D.V., Unser, M.: Nonideal sampling and regularization theory. IEEE Trans. Signal Process. **56**, 1055–1070 (2008)
22. Schaback, R., Wendland, H.: Kernel techniques: from machine learning to meshless methods. Acta Numer. 1–97 (2006)
23. Schumaker, L.: Spline Functions: Basic Theory. Wiley, New York (1981)
24. Shukla, P., Dragotti, P.L.: Sampling schemes for multidimensional signals with finite rate of innovation. IEEE Trans. Signal Process. **55**, 3670–3686 (2007)
25. Sjöstrand, J.: An algebra of pseudodifferential operators. Math. Res. Lett. **1**, 185–192 (1994)
26. Smale, S., Zhou, D.-X.: Shannon sampling and function reconstruction from point values. Bull. Am. Math. Soc. **41**, 279–305 (2004)
27. Smale, S., Zhou, D.-X.: Shannon sampling II: Connection to learning theory. Appl. Comput. Harmon. Anal. **19**, 285–302 (2005)
28. Sun, Q.: Non-uniform average sampling and reconstruction of signals with finite rate of innovation. SIAM J. Math. Anal. **38**, 1389–1422 (2006)
29. Sun, Q.: Wiener's lemma for infinite matrices. Trans. Am. Math. Soc. **359**, 3099–3123 (2007)
30. Sun, Q.: Frames in spaces with finite rate of innovation. Adv. Comput. Math. **28**, 301–329 (2008)
31. Tarczynski, A.: Sensitivity of signal reconstruction. IEEE Signal Process. Lett. **4**, 192–194 (1997)
32. Unser, M.: Sampling—50 years after Shannon. Proc. IEEE **88**, 569–587 (2000)
33. Unser, M., Aldroubi, A.: A general sampling theory for nonideal acquisition devices. IEEE Trans. Signal Process. **42**, 2915–2925 (1994)
34. Vetterli, M., Marziliano, P., Blu, T.: Sampling signals with finite rate of innovation. IEEE Trans. Signal Process. **50**, 1417–1428 (2002)
35. Wahba, G.: Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59. SIAM, Philadelphia (1990)