A Robust Function Estimation in Reproducing Kernel Hilbert Space Based on Finite Dimensional Reformulations

Shinji Shimamura and Isao Yamada
Tokyo Institute of Technology, Tokyo, Japan
E-mails: {shimamura, isao}@comm.ss.titech.ac.jp

***Abstract*—The function estimation in RKHS (Reproducing Kernel Hilbert Space) from finite noisy samples is a typical ill-conditioned inverse problem, which has been discussed mainly based on infinite dimensional operator theoretic analysis. In this paper, we present equivalent finite dimensional reformulations of the problem. Thanks to our reformulations, we can apply robust estimation techniques, e.g. the reduced-rank techniques and L-curve method for suitable Tikhonov type regularization, developed originally for finite dimensional ill-conditioned inverse problems. Numerical examples show that the proposed estimations using finite dimensional techniques achieve quite robust performances in this seemingly infinite dimensional application.**

I. Introduction

The problem of estimating an unknown function with use of only finite noisy samples has been a central issue in applied mathematical sciences and technologies, e.g. best approximation theory, machine learning, pattern recognition, signal and image processing and communication systems. Many practical extensions of the so-called *Shannon's sampling theorem* have been discussed on the stage of the RKHS (Reproducing Kernel Hilbert Space)[1] defined as the vector space of all band-limited functions. This setting is mathematically convenient because the information on the function value at any specified sampling point can be incorporated simply as a linear equality constraint in the space. Among many such studies, a milestone is found in the theory of *optimal reconstruction operator* [2], where the general solution for a certain operator optimization problem is completely solved in more general scenarios. If we restrict the discussion in [2] to a simpler scenario, the theory tells us valuable information for our problem. For example, an elegant operator theoretic analysis in [2] shows that the optimal reconstruction operators offer not only the best approximation f^* of f among all possible approximations through all linear operators in noise free situation but also offers the *best linear unbiased estimate (BLUE)* of f^* even if the samples are influenced by noise (See also [3]). On the other hand, the problem of estimating an unknown function from finite noisy samples is obviously a typical constrained linear inverse problems for which the BLUE often becomes sensitive against noise due to certain ill-conditioned natures of the problems. So far the *optimal reconstruction operator* and its several extensions have been discussed through infinite dimensional operator theoretic analysis. Although some variations, e.g. [3], of the optimal reconstruction operator employ essentially certain Tikhonov type regularizations (See also [4] in a different scenario), but not necessarily designed based on existing techniques developed recently for finite dimensional inverse problems, e.g., reduced rank techniques [5], [6], [7] and L-curve method [8] which has been used extensively as an effective tuning of Tikhonov type regularization parameter.

In this paper, we first remark that the optimal reconstruction operator can be expressed equivalently in terms of finite dimensional matrix, from which we present an equivalent finite dimensional matrix expression of the squared bias and the variance achieved by any optimal reconstruction operator. These reformulations tell us that if the samples are taken so as for the kernels specified at corresponding samples to be linearly independent, the *optimal reconstruction operator* produces the unique unbiased estimator and does not have any chance to suppress further its variance caused essentially by noise even in worst case scenario such that the Gram matrix is ill-conditioned. Unfortunately, such ways of sampling have often been found in the most practical scenario where finite samples are taken equidistantly but their interval is smaller than Nyquist rate for the band-limited type RKHS.

Fortunately, thanks to our finite dimensional reformulations of the original inverse problem, even if the Gram matrix is ill-conditioned, we can apply many robust estimation techniques developed originally for finite dimensional ill-conditioned inverse problems. We propose to apply a pair of promising techniques of which the effectiveness has been confirmed in extensive applications. One is a reduced-rank estimation [5], [6], [7] and the other is a well-known L-curve method [8] for suitable Tikhonov type regularization. Numerical examples show that the proposed estimations using finite dimensional techniques achieve quite robust performances in this seemingly infinite dimensional application.

II. Preliminaries

A. Reproducing Kernel Hilbert Space

Let $\mathcal{H}_K$ be a real Hilbert space of a class of real valued functions defined on $\mathcal{D} \subset \mathbb{R}^n$. If there is a function $K : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that

1) $\forall \hat{\mathbf{x}} \in \mathcal{D},\ K(\cdot, \hat{\mathbf{x}}) \in \mathcal{H}_K$
2) $\forall \hat{\mathbf{x}} \in \mathcal{D}, \forall f \in \mathcal{H}_K,\ f(\hat{\mathbf{x}}) = \langle f(\cdot), K(\cdot, \hat{\mathbf{x}}) \rangle_{\mathcal{H}_K}$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ stands for the inner product of the $\mathcal{H}_K$, the Hilbert space $\mathcal{H}_K$ is called a *Reproducing Kernel Hilbert*

Space (RKHS) with its *Reproducing Kernel* K (See [1] for mathematical properties of the RKHS).

Example 1: (Band-limited type reproducing kernel [9])
The closed subspace

$$\mathcal{H}_{K^\Omega} := \left\{ f \in L^2(-\infty,\infty) \mid \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt = 0, \quad \forall\omega \notin [-\Omega,\Omega] \right\}$$

of the Hilbert space $L^2(-\infty,\infty)$ is an RKHS with its reproducing kernel

$$K^\Omega(x,\hat{x}) = \frac{\sin\Omega(x-\hat{x})}{\pi(x-\hat{x})}, \quad (x,\hat{x} \in \mathbb{R}). \tag{1}$$

Any function $f \in \mathcal{H}_{K^\Omega}$ is said to be band-limited with bandwidth Ω because its Fourier transform of f is vanished outside the interval $[-\Omega,\Omega]$. The RKHS $\mathcal{H}_{K^\Omega}$ has been the main stage of the Shannon's sampling theorem and its many extensions. A simplest as well as most typical set of samples is equidistant point sampling at higher than Nyquist rate, i.e.,

$$x_k = x_1 + (k-1)\delta \quad (k \in \mathbb{N}),$$

where $0 < \delta < \pi/\Omega$. In this case, the corresponding Gram matrix $G^\Omega := \left[K^\Omega(\mathbf{x}_i,\mathbf{x}_j)\right] \in \mathbb{R}^{\ell\times\ell}$ is positive definite but becomes very ill-conditioned for large ℓ because its singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_\ell (> 0)$ follow $C(\ell) = \frac{\sigma_1}{\sigma_\ell} \backsimeq \frac{\rho}{\ell^{1/4}} e^{\frac{\gamma\ell}{2}}$, where ρ and γ do not change significantly.

B. An Optimal Reconstruction Operator as a Function Estimation from Noisy Samples

Given finite noisy samples $\{(\mathbf{x}_i, y_i) \mid i = 1,\cdots,\ell\} \subset \mathbb{R}^n \times \mathbb{R}$ observed as

$$\mathbf{y} := (y_1,\ldots,y_\ell)^t = (f(\mathbf{x}_1),\ldots,f(\mathbf{x}_\ell))^t + \boldsymbol{n} \tag{2}$$

$$A(f) := (f(\mathbf{x}_1),\ldots,f(\mathbf{x}_\ell))^t \text{ and } \boldsymbol{n} := (n_1,\ldots,n_\ell)^t \tag{3}$$

where f is the unknown function to be estimated, n_i is a zero-mean additive noise and the operator $A : \mathcal{H}_K \to \mathbb{R}^\ell$ is a linear operator called the *sampling operator* defined as in (3). Our goal is to give a good estimate of the function $f \in \mathcal{H}_K$ from the noisy sample data $\{(\mathbf{x}_i, y_i) \mid i = 1,\cdots,\ell\}$. This problem has been studied extensively as a valuable practical extension of so called the *Shannon's sampling theorem*. A milestone along this direction is found in the theory of *optimal reconstruction operator* [2], where the general solution for a certain operator optimization problem is explicitly presented in more general scenarios. If we restrict the discussion in [2] to a special case: $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_K$ and $A_1 = I$ (see [2] for the definitions of $\mathcal{H}_1$, $\mathcal{H}_2$ and A_1), the theory is applicable to our problem, i.e., function estimation in RKHS with use of noisy finite samples. In this setting, Theorem 1 in [2] guarantees the existence of an optimal linear operator $X_{opt} : \mathbb{R}^\ell \to \mathcal{H}_K$ that not only achieves $X_{opt}A = P_{\mathcal{R}(A^*)}$ but also guarantees that $X_{opt}(\boldsymbol{y})$ is the best linear unbiased estimate (called also the minimum variance unbiased linear estimate) of $f^* := P_{\mathcal{R}(A^*)}(f)$ (See also [3] on this interpretation), where $P_{\mathcal{R}(A^*)}$ denotes the orthogonal projection onto the range subspace $\mathcal{R}(A^*)$ of the adjoint operator A^* of the sampling operator A. An elegant operator theoretic analysis in [2] shows that the subspace $\mathcal{R}(A^*)$ is the largest one among all possible range spaces achieved by any linear operator $X : \mathbb{R}^\ell \to \mathcal{H}_K$, i.e., $\mathcal{R}(XA) \subset \mathcal{R}(A^*) = \mathcal{R}(X_{opt}A)$, hence $X_{opt}(\boldsymbol{y}) = P_{\mathcal{R}(A^*)}(f)$ is the best approximation of f among all possible approximations through all linear operators in noise free situation.

III. A Robust Function Estimation via Finite Dimensional Reformulation

A. Limitation of Optimal Reconstruction Operator

Note that the function estimation problem touched in the previous section allows us to use only finite information to determine a point in the infinite dimensional space $\mathcal{H}_K$. This simple observation suggests that we can reformulate the estimation problem in terms of finite dimensional vector space $\mathbb{R}^\ell$.

We start our discussion with the following simple observation:

$$\mathcal{R}(A^*) = \mathcal{M} := \operatorname{span}(K(\cdot,\mathbf{x}_1),\cdots,K(\cdot,\mathbf{x}_\ell)) \subset \mathcal{H}_K.$$

which is confirmed simply by

$$\langle A(f),\boldsymbol{\alpha}\rangle_{\mathbb{R}^\ell} = \sum_{i=1}^{\ell}\alpha_i f(\mathbf{x}_i) = \sum_{i=1}^{\ell}\alpha_i\langle f, K(\cdot,\boldsymbol{x}_i)\rangle_{\mathcal{H}_K}$$

$$= \left\langle f, \sum_{i=1}^{\ell}\alpha_i K(\cdot,\boldsymbol{x}_i)\right\rangle_{\mathcal{H}_K} = \langle f, A^*(\boldsymbol{\alpha})\rangle_{\mathcal{H}_K}$$

$$(\forall f \in \mathcal{H}_K, \forall\boldsymbol{\alpha} := (\alpha_1,\ldots,\alpha_\ell)^t \in \mathbb{R}^\ell).$$

From this observation, we see that the optimal linear operator $X_{opt} : \mathbb{R}^\ell \to \mathcal{H}_K$ in the previous section satisfies

$$X_{opt}A(f) = P_{\mathcal{M}}(f) = \sum_{i=1}^{\ell}\beta_i K(\cdot,\mathbf{x}_i) \in \mathcal{M},$$

for some $\boldsymbol{\beta} := (\beta_1,\ldots,\beta_\ell)^t \in \mathbb{R}^\ell$, where the orthogonal projection theorem in Hilbert space ensures the unique existence of $P_{\mathcal{M}}(f)$ as the best approximation of f in $\mathcal{M}$ hence the existence of such a $\boldsymbol{\beta}$ (Note: the uniqueness of $\boldsymbol{\beta}$ is guaranteed only when $\{K(\cdot,\mathbf{x}_i)\}_{i=1}^{\ell}$ are linearly independent). Moreover, by using the well-known fact $\mathcal{M}^\perp = \mathcal{R}(A^*)^\perp = \mathcal{N}(A)$, we can also observe that the model (2) is reduced to

$$\begin{aligned}\boldsymbol{y} &= A(f) + \boldsymbol{n} = A\left(P_{\mathcal{M}}(f) + P_{\mathcal{N}(A)}(f)\right) + \boldsymbol{n} \\ &= A\left(P_{\mathcal{M}}(f)\right) + \boldsymbol{n} = \sum_{i=1}^{\ell}\beta_i A\left(K(\cdot,\mathbf{x}_i)\right) + \boldsymbol{n} \\ &= \sum_{i=1}^{\ell}\beta_i\left(K(\mathbf{x}_1,\mathbf{x}_i),\ldots,K(\mathbf{x}_\ell,\mathbf{x}_i)\right)^t + \boldsymbol{n},\end{aligned} \tag{4}$$

which suggests that $\boldsymbol{y}$ does not contain any effective information for the component $P_{\mathcal{M}^\perp}(f)$. The component $P_{\mathcal{M}^\perp}(f)$ can be estimated only with some additional a priori knowledge on f not through simple sampling at $\boldsymbol{x}_i$ $(i = 1,2,\ldots,\ell)$.

B. Finite Dimensional Reformulation of An Optimal Reconstruction Operator

Assume the standard situation: $\mathcal{R}(A) = \left\{(f(\boldsymbol{x}_1),\ldots,f(\boldsymbol{x}_\ell))^t \in \mathbb{R}^\ell \mid f \in \mathcal{H}_K\right\} = \mathbb{R}^\ell$. In this case, any linear operator $X : \mathbb{R}^\ell \to \mathcal{H}_K$ satisfying $XA = P_{\mathcal{R}(A^*)}$ has its own range space $\mathcal{R}(X) = \mathcal{R}(XA) = \mathcal{R}\left(P_{\mathcal{R}(A^*)}\right) = \mathcal{R}(A^*) = \mathcal{M}$, hence there exists $\Phi := (\phi_1,\ldots,\phi_\ell)^t \in \mathbb{R}^{\ell\times\ell}$ such that

$$X(\boldsymbol{y}) = \sum_{i=1}^{\ell} K(\cdot,\boldsymbol{x}_i)\phi_i^t \boldsymbol{y} \quad (\boldsymbol{y}\in\mathbb{R}^\ell).$$

This fact shows clearly that any design problem for the operator X is equivalent to that for the matrix Φ without loss of any generality. Moreover, since the best possible approximation of f by linear operator under noise free situation is

$$P_{\mathcal{M}}(f) = \sum_{i=1}^{\ell} \beta_i K(\cdot,\boldsymbol{x}_i),$$

our best achievable goal is the minimization of

$$\begin{aligned} J_{mse}(X) &:= E\left(X(\boldsymbol{y}) - P_{\mathcal{M}}(f)\right)^2 \\ &= E\left[\sum_{i=1}^{\ell} K(\cdot,\boldsymbol{x}_i)\phi_i^t\boldsymbol{y} - P_{\mathcal{M}}(f)\right]^2 \\ &=: \widehat{J}_{mse}(\Phi). \end{aligned}$$

If we use the Gram matrix $G := [K(\mathbf{x}_i,\mathbf{x}_j)] \in \mathbb{R}^{\ell\times\ell}$, we have alternative expression:

$$\begin{aligned} \widehat{J}_{mse}(\Phi) &= \left\{(\Phi G - I)\boldsymbol{\beta}\right\}^t G(\Phi G - I)\boldsymbol{\beta} \\ &\quad + \sum_i\sum_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)\phi_i^t E(\boldsymbol{n}\boldsymbol{n}^t)\phi_j, \end{aligned}$$

where the 1st term expresses the squared bias of the estimate $X(\boldsymbol{y})$ of $P_{\mathcal{M}}(f)$, and the 2nd term expresses its variance. Finally, we observe that the design of an *optimal reconstruction operator* [2] is reduced to the finite dimensional constrained optimization problem (on the matrix Φ):

$$\begin{cases} \text{minimize} & \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} K(\boldsymbol{x}_i,\boldsymbol{x}_j)\phi_i^t E(\boldsymbol{n}\boldsymbol{n}^t)\phi_j \\ \text{subject to} & \left\{(\Phi G - I)\boldsymbol{\beta}\right\}^t G(\Phi G - I)\boldsymbol{\beta} = 0. \end{cases}$$

Remark that under the most typical situation where G is a positive definite matrix (See Example 1), we have only one choice $\Phi = G^{-1}$ in order to satisfy the unbiasedness. In this case, the *optimal reconstruction operator* offers $G^{-1}\boldsymbol{y}$ as the estimate of $\boldsymbol{\beta}$, i.e.,

$$\sum_{i=1}^{\ell} K(\cdot,\boldsymbol{x}_i){\boldsymbol{e}_i^{(\ell)}}^t G^{-1}\boldsymbol{y} \approx \sum_{i=1}^{\ell} K(\cdot,\boldsymbol{x}_i)\beta_i = P_{\mathcal{M}}(f),$$

where $\boldsymbol{e}_i^{(\ell)}$ denotes the i-th vector of the canonical basis of $\mathbb{R}^\ell$.

Unfortunately, as seen in the Example 1, the matrix G is often very ill-conditioned, hence the *optimal reconstruction operator* determined with $\Phi = G^{-1}$ in this case becomes very sensitive against noise.

C. Robust Function Estimation by Finite Dimensional Techniques

In this subsections, we propose a pair of robust function estimators by applying finite dimensional techniques [5], [6], [7], [8] applicable to estimation of β in (4) which is usually ill-conditioned inverse problems.

Assume that the noise $\mathbf{n}\in\mathbb{R}^\ell$ in (2) is a zero mean random vector with its positive definite covariance matrix $E(\mathbf{n}\mathbf{n}^t) = \sigma^2 Q \in \mathbb{R}^{\ell\times\ell}$. Let the singular value decomposition (SVD) of G:

$$G = U\Sigma V^t = \sum_{i=1}^{\mathrm{rank}(G)} \sigma_i \mathbf{u}_i\mathbf{v}_i^t \tag{5}$$

where $U = (\mathbf{u}_1,\cdots,\mathbf{u}_\ell)\in\mathbb{R}^{\ell\times\ell}, V = (\mathbf{v}_1,\cdots,\mathbf{v}_\ell)\in\mathbb{R}^{\ell\times\ell}$ are orthogonal matrices and $\Sigma\in\mathbb{R}^{\ell\times\ell}$ contains on its main diagonal the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\mathrm{rank}(G)}$ of G and $0'$s elsewhere. We also assume that these pieces of information are available to estimate $\boldsymbol{\beta}$.

1) Reduced-Rank Techniques: In this simple scenario, we propose to use, as a simplest example of MV-PURE estimator [7],

$$\Phi_r := \tilde{V}_r\tilde{V}_r^t(\tilde{G}^t\tilde{G})^{-1}\tilde{G}^t Q^{-1/2}, \tag{6}$$

where $r \leq \mathrm{rank}(G), \tilde{V}_r = (\tilde{\mathbf{v}}_1,\cdots,\tilde{\mathbf{v}}_r)$ and $\tilde{V} := (\tilde{\mathbf{v}}_1,\cdots,\tilde{\mathbf{v}}_\ell)$ is given by the SVD: $\tilde{G} := Q^{-1/2}G = \tilde{U}\tilde{\Sigma}\tilde{V}^t$, and approximate $\boldsymbol{\beta}$ by $\Phi_r(\boldsymbol{y})$. This estimator was proposed in [6] as a direct generalization of Marquardt's reduced-rank estimator [5]. Obviously, Φ_r satisfies $\mathrm{rank}(\Phi_r) \leq r$ and can eliminate the influence of very small singular values $\sigma_{r+1},\cdots,\sigma_\ell$. We will present a simple criterion for selection of r in Section IV (8).

2) L-curve method for Tikhonov regularization: A Tikhonov type regularization of the optimal reconstruction operator has been proposed for example in [3], which inherently introduces bias to make the operator to be more robust against noise. However the design of the regularization parameter in [3] is not made based on well-known techniques e.g. L-curve method [8] which has been widely used mainly for finite dimensional inverse problems. We propose a Tikhonov type regularization with L-curve method for the estimation of $\boldsymbol{\beta}$. By using the SVD of G (5), Tikhonov's regularization estimator Φ_α is derived as follows:

$$\Phi_\alpha := \sum_{i=1}^{\mathrm{rank}(G)} \frac{\sigma_i}{\sigma_i^2+\alpha}\mathbf{v}_i\mathbf{u}_i^t. \tag{7}$$

The parameter α is determined by the algorithm shown in [10].

IV. Numerical Examples

In this section, we present numerical examples for the RKHS with its kernel $K^{2\pi}(\mathbf{x},\hat{\mathbf{x}}) = \frac{\sin 2\pi(\mathbf{x}-\hat{\mathbf{x}})}{\pi(\mathbf{x}-\hat{\mathbf{x}})}$ in Example 1. The unknown function to be estimated is given by $f(\mathbf{x}) = \mathrm{sinc}(\mathbf{x}) = \frac{\sin\pi\mathbf{x}}{\pi\mathbf{x}}$. Noisy samples are given equidistantly as $\left\{(x_i, y_i = f(x_i)+n_i) \mid x_i = -10 + \frac{20}{\ell-1}i \quad (i=0,\ldots,\ell-1)\right\}$ where n_i is white Gaussian noise of which the covariance matrix is $\sigma^2\mathrm{I}_\ell$. The accuracy of the estimate $\hat{f}$ is measured by

$$\|f-\hat{f}\|_{\max} := \max_{x\in[-10,10]} |f(x)-\hat{f}(x)|$$

$$\|f - \hat{f}\|_2 := \left\{ \int_{-10}^{10} |f(x) - \hat{f}(x)|^2 \, dx \right\}^{1/2}$$

We compared the performances of the function estimators based on finite dimensional estimation techniques for $\boldsymbol{\beta} \in \mathbb{R}^\ell$ in III-A. Fig.1, Fig.2 and Fig.3 depict the experimental result. 'Reduced-Rank' denotes the Reduced-Rank Estimator (6), where rank r is chosen by following rule:

$$\mathrm{argmax}_r \left(\frac{\sqrt{\sigma_1^2 + \cdots + \sigma_r^2}}{\sqrt{\sigma_1^2 + \cdots + \sigma_\ell^2}} < 0.8 \right) \tag{8}$$

'L-curve' uses the Tikhonov's regularization method (7) with its regularization parameter designed by L-curve method [10]. 'BLUE' denotes the function estimated through BLUE for $\boldsymbol{\beta} \in \mathbb{R}^\ell$. Fig.1 and Fig.2 demonstrate the effectiveness of the proposed techniques for a pair of criteria in particular for large number of samples are used. We also demonstrate the potential of the Reduced-Rank technique in Fig.3 by showing the ideal performance achievable with globally optimal rank.

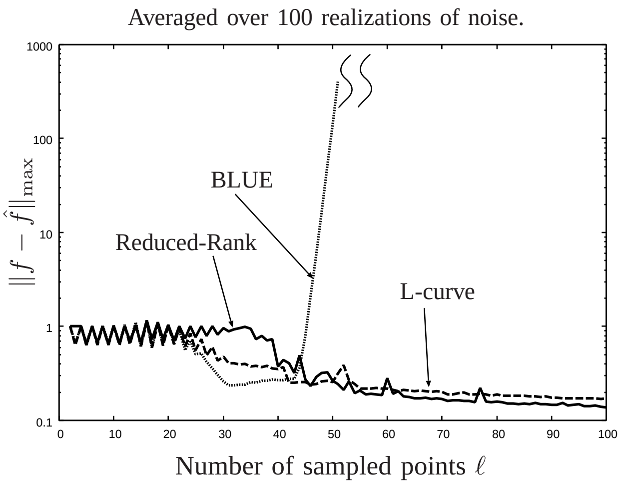

Fig. 1. $\sigma^2 = 0.01$

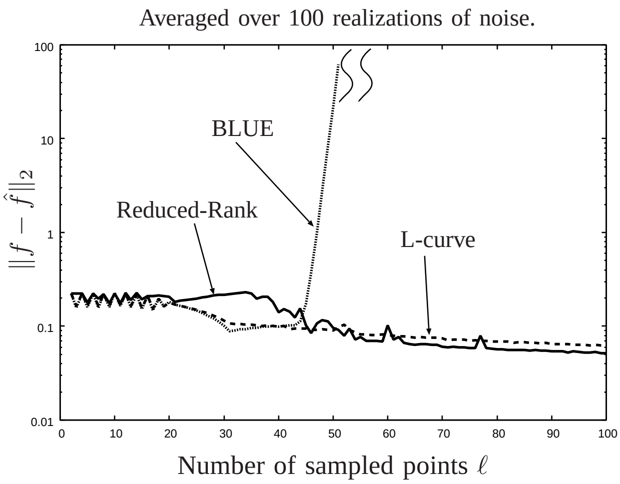

Fig. 2. $\sigma^2 = 0.01$

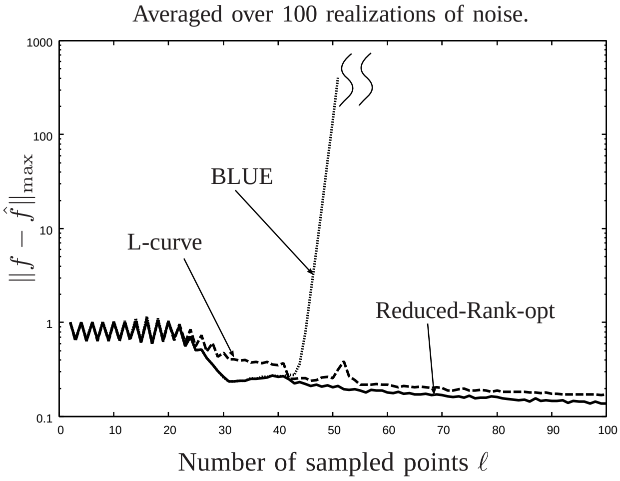

Fig. 3. $\sigma^2 = 0.01$. 'Reduced-Rank-opt' uses Reduced-Rank Estimator, where rank is chosen optimally among all possible ranks.

ACKNOWLEDGMENT

The authors would like to express their deep gratitude to Prof.K. Sakaniwa of the Tokyo Institute of Technology for helpful discussions.

REFERENCES

[1] N. Aronszajn, "Theory of Reproducing Kernels" *Transactions of the American Mathematical Society* vol.68, pp. 337-404, 1950.

[2] H. Ogawa, and A. Hirabayashi, "Sampling Theorem with Optimum Noise Suppression" *Sampling Theory in Signal and Image Processing* vol.6, No.2, pp. 167-184,May 2007.

[3] A. Tanaka, M. Sugiyama, H. Imai, M. Kudo, and M. Miyakoshi, "Model Selection Using a Class of Kernels with an Invariant Metric," *Joint IAPR International Workshops on Syntactical and Structural Pattern Recognition and Statistical Pattern Recognition Hong Kong,China*, Aug. 2006.

[4] E. Oja, and H. Ogawa, "Parametric Projection Filter for Image and Signal Restoration" *IEEE Trans. on Acoustics, Speech and Signal Processing* ASSP-34 pp.1643-1653, 1986.

[5] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol.12, pp.591-612, 1970.

[6] J. S. Chipman, "Linear restrictions, rank reduction, and biased estimation in linear regression," *Linear Algebra Appl.*, vol.289, pp.55-74, 1999.

[7] T. Piotrowski, and I. Yamada, "MV-PURE Estimator: Minimum-Variance Pseudo-Unbiased Reduced-Rank Estimator: for Linearly Constrained Ill-Conditioned Inverse Problems," *IEEE Trans. on Signal Process*, vol.56, pp.3408-3423, Aug. 2008.

[8] P. C. Hansen, "Analysis of Discrete Ill-Posed Problems by Means of the L-Curve" *SIAM Rev*, vol.34,Issue4,pp.561-580,Dec. 1992.

[9] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-V:the discrete case" *Bell Syst. Tech.J* vol.57 pp.1371-1430, 1978.

[10] P. C. Hansen, Regularization Tools Version3.1 (for MATLAB Version 6.0.) [Online]. Available: http://www2.imm.dtu.dk/pch/Regutools/Software.zip