A note on SVM estimators in RKHS for the deconvolution problem

Sungho Lee[1,a]

[a]Department of Statistics, Daegu University, Korea

Abstract

In this paper we discuss a deconvolution density estimator obtained using the support vector machines (SVM) and Tikhonov's regularization method solving ill-posed problems in reproducing kernel Hilbert space (RKHS). A remarkable property of SVM is that the SVM leads to sparse solutions, but the support vector deconvolution density estimator does not preserve sparsity as well as we expected. Thus, in section 3, we propose another support vector deconvolution estimator (method II) which leads to a very sparse solution. The performance of the deconvolution density estimators based on the support vector method is compared with the classical kernel deconvolution density estimator for important cases of Gaussian and Laplacian measurement error by means of a simulation study. In the case of Gaussian error, the proposed support vector deconvolution estimator shows the same performance as the classical kernel deconvolution density estimator.

Keywords: deconvolution, ill-posed problem, kernel density estimator, regularization, reproducing kernel Hilbert space (RKHS), support vector machines (SVM)

1. Introduction

The support vector machines (SVM) was developed in the 1960s by Vapnik and co-workers (Vapnik and Lerner, 1963; Vapnik and Chervonenkis; 1964). It was initially designed to solve pattern recognition problems. Later the SVM was extended to classification, regression and real valued function estimation (e.g. Vapnik, 1995). In the early 1990s the SVM was proposed for classification in the context of Vapnik's learning theory. The use of the SVM grew rapidly among computer scientists due to its success with real world data analysis problems. In the 1990s it became clear to statistician that the SVM with the kernel trick was a solution to optimization problems in a reproducing kernel Hilbert space (RKHS)). Reproducing kernel Hilbert spaces (Aronszajn, 1950; Wahba, 1990) became prominent when their relation to the SVM was clear (Moguerza and Munoz, 2006; Wahba, 2006). Reproducing kernel Hilbert spaces (RKHS) were explicitly introduced in learning theory by Girosi (1998). In general it is quite difficult to find useful function spaces that are not RKHS.

The problem of measurements being contaminated with noise exists in many different fields (e.g. Mendelsohn and Rice, 1982; Stefanski and Carroll, 1990; Zhang, 1992). This deconvolution problem can be stated as follows. Let X and Z be independent random variables with density functions $f(x)$ and $q(z)$, respectively, where $f(x)$ is unknown and $q(z)$ is known. One observes a sample of random variables $Y_i = X_i + Z_i$, $i = 1, 2, \ldots, n$. The objective is to estimate the density function $f(x)$ where $g(y)$ is the convolution of $f(x)$ and $q(z)$, $g(y) = (f * q)(y) = \int_{-\infty}^{\infty} f(y - z)q(z)dz$. The most popular approach to this deconvolution problem has been to estimate $f(x)$ using a kernel estimator and Fourier

---

[1] Department of Statistics, Daegu University, Kyungbuk 38453, Korea. E-mail: shlee1@daegu.ac.kr

transform (e.g. Carroll and Hall, 1988; Fan, 1991). Kernel density estimation is widely considered the most popular approach to density deconvolution; however, other alternatives have been proposed (e.g. Hall and Qiu, 2005; Hazelton and Turlach, 2009; Mendelsohn and Rice, 1982; Pensky and Vidakovic, 1999).

The support vector regression method to density estimation was discussed by Weston *et al.* (1999) and support vector density estimation in RKHS based on Phillips' residual method (Phillips, 1962) was discussed by Mukherjee and Vapnik (1999). The SVM in RKHS can be applied to the deconvolution problem with the support vector density estimation and Fourier transform (Lee, 2010, 2012). In this paper we discuss deconvolution estimators obtained by support vector density estimation and Fourier transform in the framework of Tikhonov regularization method in RKHS and propose a support vector deconvolution estimator which leads to very sparse solution.

2. RKHS and SVM

The support vector regression algorithm computes a nonlinear function in the space of the input data R^m by using a linear function in high dimensional feature space $\mathfrak{F}$ with a dot product. The functions take the form $f(x) = \omega \cdot \Phi(x) + b$ with $\Phi : R^m \to \mathfrak{F}$ and $\omega \in \mathfrak{F}$. In order to estimate $f(x)$ in the deconvolution problem by applying the support vector regression method (Vapnik, 1995), we first take $g(y) = \omega \cdot \Phi(x) + b$ from a training set $\{(y_1, G_n(y_1)), \ldots, (y_n, G_n(y_n))\}$, where $G_n(y)$ is the empirical distribution function. Then we try to minimize the empirical risk function $R_{emp}(G)$ with a complexity term $\|\omega\|^2$:

$$R_{reg}(G) = R_{emp}(G) + \frac{\lambda}{2}\|\omega\|^2 = \frac{1}{n}\sum_{i=1}^{n}|G(y_i) - G_n(y_i)|_\epsilon + \frac{\lambda}{2}\|\omega\|^2, \qquad (2.1)$$

where Vapnik's ϵ-insensitive loss function $|G(y) - G_n(y)|_\epsilon$ is described by

$$|G(y) - G_n(y)|_\epsilon = \begin{cases} |G(y) - G_n(y)| - \epsilon, & \text{for } |G(y) - G_n(y)| \geq \epsilon, \\ 0, & \text{otherwise.} \end{cases} \qquad (2.2)$$

Equation (2.1) can be minimized by solving the quadratic programming problem formulated in terms of dot products in feature space $\mathfrak{F}$. Then applying the Fourier inversion formula, we can obtain $\hat{f}(x)$ (Lee and Taylor, 2008). This method can be constructed in the framework of regularization theory in RKHS.

A (real) RKHS H is a Hilbert space of real-valued functions f on an interval T with the property that, for each $t \in T$, the evaluation functional L_t, $L_t : f \to f(t)$, is a bounded linear functional. Then, by Riesz representation theorem, for each $t \in T$ there exists a unique element $K_t \in H$ such that for each $f \in H$, $L_t(f) = f(t) = < K_t, f >$. The function defined by $K_u(v) = K(u, v) = < K_u, K_v >$ for $u, v \in T$ is called the reproducing kernel. Conversely, a positive definite kernel K can define a unique RKHS with the following Moore-Aronszajn theorem. This result is important because it gives us a construction of an RKHS in terms of its reproducing kernel. Let X be some set, for example a subset of R^n or R^n itself. A kernel is a symmetric function $K : X \times X \to R$. A kernel K is called positive definite if its associated kernel matrix $\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0$ for any $n \in N$, and any $x_1, x_2, \ldots, x_n \in X$ and $c_1, c_2, \ldots, c_n \in R$. Alternatively, if $\iint_{X \times X} K(x, z)u(x)u(z)dxdz \geq 0$, for all $u \in L_1(X)$, then kernel K is called integrally positive definite. Two definitions are equivalent for continuous kernel K.

Theorem 1. (Aronszajn, 1950) *For every positive definite kernel K on $X \times X$ there is a unique RKHS H_K on X with K as its reproducing kernel.*

We now present Mercer's theorem (Mercer, 1909) briefly, which is one of the fundamental mathematical results underlying learning theory with kernels. In the SVM literature, the kernel is called Mercer kernel and used as a linear function in high dimensional feature space $\mathfrak{F}$ with a dot product. The Mercer's theorem is as follows (e.g. see Scholkopf and Smola (2002) for the details): A positive definite kernel $K : X \times X \rightarrow R$, with X a closed subset of R^n has the expansion $K(x, y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(y)$, where the convergence is in $L_2(X, \mu)$ and $\phi_i \in L_2(X, \mu)$, $i = 1, 2, \ldots$, are the orthonormal eigenfunctions of the integral equation $\int_X K(x, y) \phi(x) d\mu(x) = \lambda \phi(y)$.

We can now define the RKHS as the space of functions spanned by the eigenfunctions of the integral operator defined by the kernel. Let us consider the set of functions,

$$H_K = \left\{ f | f(x) = \sum_{r=1}^{\infty} \omega_r \phi_r(x) \quad \text{and} \quad \|f\|_{H_K} < \infty \right\},$$

where the inner product is defined as $< f, g >= \sum_{i=1}^{\infty} \omega_i \omega_i' / \lambda_i$ and the RKHS norm $\|f\|_{H_K}$ is defined as $\|f\|_{H_K}^2 = \sum_{i=1}^{\infty} \omega_i^2 / \lambda_i$. Then we have a RKHS H_K with its reproducing kernel K. Assume further that $\iint_{X \times X} K(x, y)^2 dx dy < \infty$. Then kernel K will have a countable sequence of eigenvalues and eigenfunctions, and hence a natural feature map that arises from Mercer's theorem is

$$y \rightarrow \Phi(y) = \left(\sqrt{\lambda_1} \phi_1(y), \ldots, \sqrt{\lambda_i} \phi_i(y), \ldots \right) \quad \text{and} \quad K(x, y) =< \Phi(x), \Phi(y) > .$$

We can apply these properties of RKHS to the density estimation. Notice that for the Gaussian kernel on the infinite real line, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x-y)^2/\sigma^2} dx dy = \infty$ and hence Mercer's theorem does not apply. However, it turns out that one can derive a Mercer expansion for the Gaussian kernel with the orthonormal functions in $L_2(R, \rho)$, $\rho(x) = (\alpha / \sqrt{\pi}) \cdot e^{-\alpha^2 x^2}$, $\alpha > 0$ (Rasmussen and Williams, 2006).

Eigenvalues are not necessary countable and Mercer theorem does not apply if the kernel is defined over an unbounded domain. Feature maps do not necessary arise from Mercer's theorem. Let us consider a special class of kernels that is widely used in practice. The kernel is translation invariant, or $K(x, y) = K(x - y)$. This implies that we will have to consider Fourier hypothesis spaces and all these spaces will be defined via Fourier transform. The eigenvalue problem for translation invariant kernels is $\int_{-\infty}^{\infty} K(x - y) \phi(x) dx = \lambda \phi(y)$, and hence by the convolution theorem the Fourier transform of the equation is $\tilde{K}(\omega) \tilde{\phi}(\omega) = \lambda \tilde{\phi}(\omega)$, where $\tilde{\phi}(\omega) = \int_{-\infty}^{\infty} \phi(x) e^{-iwx} dx$. The following Bochner's theorem relates the Fourier transform of a kernel to it being positive definite.

Theorem 2. (Bochner, 1959) *A function $K(x - y)$ is positive definite if and only if it is the Fourier transform of a symmetric, positive function $\tilde{K}(\omega)$ decreasing to at infinity.*

If we consider a positive definite function $K(x - y)$ and define, in the Fourier domain, the inner product $< f(x), g(x) >_{H_K} = 1/(2\pi) \int_R \tilde{f}(\omega) \tilde{g}(\omega)^* / \tilde{K}(\omega) d\omega$, the subspace H_K of L_2 space of the functions f is a RKHS with the norm $\|f\|_{H_K}^2 (= 1/(2\pi) \int_R |\tilde{f}(\omega)|^2 / \tilde{K}(\omega) d\omega < \infty)$. The well-known Gaussian kernel represents an inner product in feature space $\mathfrak{F}$. Let $C_0(R)$ denote the set of continuous functions on R that vanish at infinity. Then the reproducing kernel Hilbert space H_K (Vert and Vert, 2006) associated with the normalized Gaussian kernel $K(x, y) = (\sigma \sqrt{2\pi})^{-1} e^{-(x-y)^2/2\sigma^2}$ is

$$H_K = \left\{ f \in C_0(R) : f \in L_1(R) \quad \text{and} \quad \int_R \left| \tilde{f}(\omega) \right|^2 e^{\frac{\sigma^2 \omega^2}{2}} d\omega < \infty \right\}$$

and since $\int_R |\tilde{f}(\omega)|^2 e^{\sigma^2 \omega^2/2} d\omega = \sum_{k=0}^{\infty} \sigma^{2k}/(2^k k!) \int_R \omega^{2k} |\tilde{f}(\omega)|^2 d\omega$, the space H_K for the Gaussian kernel consists of the set of square integrable functions whose derivatives of all orders are square integrable. When sample observations are contaminated by normally distributed errors, the kernel K that has compactly supported Fourier transform, $\tilde{K}(\omega) = (1 - \omega^2)^3 I_{[-1,1]}(\omega)$, is used in common. Since $\tilde{K}(\omega)(= (1 - \omega^2)^3 I_{[-1,1]}(\omega))$ is also a symmetric, positive function decreasing to 0 at infinity, the RKHS H_K associated with the kernel K,

$$K(x, y) = K(x - y) = \frac{48 \cos(x - y)}{\pi(x - y)^4} \left(1 - \frac{15}{(x - y)^2}\right) - \frac{144 \sin(x - y)}{\pi(x - y)^5} \left(2 - \frac{5}{(x - y)^2}\right),$$

is a subspace of $L_2(R)$ with norm, $\|f\|_{H_K}^2 = 1/(2\pi) \int_R |\tilde{f}(\omega)|^2/\tilde{K}(\omega) d\omega < \infty$.

3. SVM deconvolution estimators in RKHS

The problem of approximating a density function from data can be dealt with an ill-posed problem (e.g. Vapnik, 1995). A density function $f(x)$ will be defined as the solution of the equation $\int_{-\infty}^{x} f(t)dt = F(x)$, where $F(x)$ is a distribution function and hence the problem of solving the linear operator equation $Af = F$ with approximation $\hat{F}_n(x)$ of unknown distribution function $F(x)$ is ill-posed. A classical solution to this problem is the regularization theory (Tikhonov and Arsenin, 1977). Given a training sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, the Tikhonov regularization method that we will consider here is defined by an optimization problem over RKHS:

$$f^* = \arg \min_{f \in H_K} \left\{ \frac{1}{n} \sum_{i=1}^{n} V(f(x_i), y_i) + \frac{\lambda}{2} \|f\|_{H_K} \right\},$$

where V is a loss function, $\|f\|_{H_K}$ is a norm in RKHS H_K defined by the positive definition function K and λ is the regularization parameter. In RKHS H_K defined by the positive definite kernel K, The solution to the Tikhonov regularization method has a very compact representation by the generalized Represener theorem (refer to Scholkopf *et al.* (2001) for the details) as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i), \quad \alpha_i \in R, \ 1 \leq i \leq n.$$

Now we discuss a SVM deconvolution estimator by Method I (Mukherjee and Vapnik, 1999; Lee, 2012) within the framework of Tikhonov regularization method in RKHS H_K as ill-posed problems and propose a new SVM deconvolution estimator by Method II. Let X and Z be independent random variables with density functions $f(x)$ and $q(z)$ respectively, where $f(x)$ is unknown and $q(z)$ is known. Assume that we observe a random sample $Y_1, Y_2, \ldots, Y_n$ from a density $g(y) \in H_K$, where $Y_i = X_i + Z_i$, $i = 1, 2, \ldots, n$ and let $\{y_1, y_2, \ldots, y_n\}$ be a training data set.

Method I.

Step 1. Choose a positive definite kernel K and ϵ-insensitive loss function which consists of distribution function $G(y)$ of Y and empirical distribution function $G_n(y)$ of a random sample $Y_1, Y_2, \ldots, Y_n$.

Step 2. Find a density estimator $\hat{g}(y)$ to satisfy the following Tikhonov minimization problem by applying Represener theorem on RKHS H_K and SVM regression method (e.g. Smola and

Scholkopf, 2003; Vapnik, 1995):

$$\hat{g} = \arg\min_{g \in H_K} \left\{ \frac{1}{n} \sum_{i=1}^{n} |G(y_i) - G_n(y_i)|_\epsilon + \frac{\lambda}{2} \|g\|_{H_K}^2 \right\},$$

where $g(y, \omega) = \sum_{i=1}^{n} \omega_i K_h(y_i, y)$, $\omega_i \geq 0$, $\sum_{i=1}^{n} \omega_i = 1$, $\|g\|_{H_K}^2 = \sum_{i,j=1}^{n} \omega_i \omega_j K_h(y_i, y_j)$, $G_n(y) = (1/n) \sum_{i=1}^{n} I(y_i \leq y)$, $G(y) = \int_{-\infty}^{y} \sum_{i=1}^{n} \omega_i K_h(y_i, t) dt$, and the coefficients ω_i's of $\hat{g}(\omega)$ can be found by solving the following problem

$$\max_{\alpha, \alpha^*, \eta, \eta^*} \left(\min_{\omega, \xi, \xi^*} L(\omega, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) \right)$$

$$= \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_i \omega_j K_h(y_i, y_j) + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) - \sum_{i=1}^{n} (\eta_i \xi_i + \eta_i^* \xi_i^*)$$

$$- \sum_{i=1}^{n} \alpha_i \left(\epsilon + \xi_i - G_n(y_i) + \sum_{j=1}^{n} \omega_j \int_{-\infty}^{y_i} K_h(y, y_j) dy \right)$$

$$- \sum_{i=1}^{n} \alpha_i^* \left(\epsilon + \xi_i^* + G_n(y_i) - \sum_{j=1}^{n} \omega_j \int_{-\infty}^{y_i} K_h(y, y_j) dy \right),$$

where $\omega, \xi, \xi^*, \alpha, \alpha*, \eta, \eta^* \geq 0$.

Then the coefficients ω_i's can be found by solving the following quadratic programming problem and applying the equation $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$:

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^t R^t \Gamma_h^{-1} R(\alpha - \alpha^*) - \sum_{i=1}^{n} G_n(y_i)(\alpha_i - \alpha_i^*) + \epsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*),$$

$$0 \leq \alpha_i^*, \ \alpha_i \leq C, \ i = 1, \ldots, n,$$

where $\Gamma_h = [K_h(y_i, y_j)]_{n \times n}$, $R = [r_{ij}]_{n \times n}$, $r_{ij} = \int_{-\infty}^{y_j} K_h(y, y_i) dy = \int_{-\infty}^{y_j} (1/h_n) K((y - y_i)/h_n) dy$.

Step 3. Find density estimator $\hat{f}(x)$ by applying the Fourier inversion formula,

$$\hat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\tilde{\hat{g}}(\xi)}{\tilde{q}(\xi)} e^{i\xi x} \, d\xi, \quad \text{where} \ \tilde{\hat{g}}(\xi) = \int_{-\infty}^{\infty} \sum_{i=1}^{n} \omega_i K_h(y_i, y) e^{-i\xi y} \, dy.$$

A remarkable property of SVM is that ϵ-insensitive loss function (2.2) leads to sparse solutions. Unfortunately, the support vector deconvolution density estimator by Method I does not preserve sparsity in terms of coefficients ω, $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$ as expected because a sparse decomposition in α and α^* is spoiled by $\Gamma_h^{-1} R$, which is not in general diagonal. However it is still attractive in the sense that some coefficients in $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$ are very close to zero.

Now we propose another support vector method which has a more sparse solution than Method I. The following Method II uses the support vector Regression and Tikhonov regularization method in RKHS. The difference between two methods depends on components of ϵ-insensitive loss function. In Method I, ϵ-insensitive loss function is defined by distribution function $G(y)$ and empirical distribution function $G_n(y)$, which cause it to lose sparse solution. In Method II, ϵ-insensitive loss function is

defined by density function $g(y)$ and a classical kernel density estimator $\hat{g}_k$, which leads to accomplish sparse solution. In kernel density estimation, $\hat{g}_k(y)$ converge to $g(y)$ in probability if g is continuous at y and $h_n \to 0$, $nh_n \to \infty$ as $n \to \infty$. Now we will consider the problem of approximating based on the set of data, $\{(y_1, \hat{g}_k(y_1)), (y_2, \hat{g}_k(y_2)), \ldots, (y_n, \hat{g}_k(y_n))\}$, by applying the support vector Regression and Tikhonov regularization method in RKHS.

Method II.

Step 1. Choose a positive definite kernel K and ϵ-insensitive loss function which consists of density function $g(y)$ and a classical kernel density estimator $\hat{g}_k(y)$ with the kernel K.

Step 2. Find a density estimator $\hat{g}(y)$ to satisfy the following Tikhonov regularization problem by applying Representer theorem on RKHS H_K and SVM regression method:

$$\hat{g} = \arg\min_{g \in H} \left\{ \frac{1}{n} \sum_{i=1}^{n} |g(y_i) - \hat{g}_k(y_i)|_\epsilon + \frac{\lambda}{2} \|g\|^2_{H_K} \right\},$$

where $\hat{g}_k(y) = (1/nh_n) \sum_{j=1}^{n} K((y - y_j)/h_n)$ and the coefficients ω_i's can be found by solving the following problem

$$\max_{\alpha, \alpha^*, \eta, \eta^*} \left(\min_{\omega, \xi, \xi^*} L\left(\omega, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*\right) \right)$$

$$= \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_i \omega_j K_h\left(y_i, y_j\right) + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) - \sum_{i=1}^{n} (\eta_i \xi_i + \eta_i^* \xi_i^*)$$

$$- \sum_{i=1}^{n} \alpha_i \left(\epsilon + \xi_i - \hat{g}_k(y_i) + \sum_{j=1}^{n} \omega_j K_h\left(y_i, y_j\right) \right) - \sum_{i=1}^{n} \alpha_i^* \left(\epsilon + \xi_i^* + \hat{g}_k(y_i) - \sum_{j=1}^{n} \omega_j K_h\left(y_i, y_j\right) \right),$$

where $\omega, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^* \geq 0$.

Then the coefficients ω_i's can be found by solving the following quadratic programming problem and applying the equation $\omega = \alpha - \alpha^*$:

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^t \Gamma_h (\alpha - \alpha^*) - \sum_{i=1}^{n} \hat{g}_k(y_i)(\alpha_i - \alpha_i^*) + \epsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*),$$

where $0 \leq \alpha_i^*, \alpha_i \leq C$, $i = 1, \ldots, n$, $\Gamma_h = [K_h(y_i, y_j)]_{n \times n}$.

Step 3. Find density estimator $\hat{f}(x)$ by applying the Fourier inversion formula,

$$\hat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\tilde{\hat{g}}(\xi)}{\tilde{q}(\xi)} e^{i\xi x} \, d\xi, \quad \text{where } \tilde{\hat{g}}(\xi) = \int_{-\infty}^{\infty} \sum_{i=1}^{n} \omega_i K_h(y_i, y) e^{-i\xi y} \, dy.$$

This setting of Method II preserve sparsity in terms of the coefficients ω because $\omega = \alpha - \alpha^*$ and α_i, α_i^* vanish for $|g(y_i) - \hat{g}_k(y_i)| < \epsilon$. However, Method I does not preserve sparsity because $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$ and a sparse decomposition in α and α^* is spoiled by $\Gamma_h^{-1} R$.

4. Simulation and discussion

The asymptotic properties of the kernel density estimator in deconvolution problems depend strongly on error distribution. Following the work of Fan (1991), two types of error distributions can be considered: ordinary smooth and supersmooth distributions. Normal, mixture normal, and Cauchy distributions are supersmooth, therefore, the Fourier transform $\tilde{q}(\xi)$ ($= \int_{-\infty}^{\infty} e^{-i\xi z} q(z) dz$) of $q(z)$ satisfies

$$d_0 |\xi|^{-\beta_0} \exp\left(-\frac{|\xi|^{\beta}}{\gamma}\right) \leq |\tilde{q}(\xi)| \leq d_1 |\xi|^{-\beta_1} \exp\left(-\frac{|\xi|^{\beta}}{\gamma}\right) \quad \text{as } \xi \to \infty,$$

for some positive constants d_0, d_1, β, γ and constants β_0 and β_1. Gamma and double exponential distributions are ordinary smooth, that is, the Fourier transform $\tilde{q}(\xi)$ of $q(z)$ satisfies

$$d_0 |\xi|^{-\beta} \leq |\tilde{q}(\xi)| \leq d_1 |\xi|^{-\beta}, \quad \text{as } \xi \to \infty,$$

for some positive constants d_0, d_1, β. Thus in the classical deconvolution literature, normal and double exponential distributions have been typically selected and investigated as error distributions. In this section we compare the performance of three different deconvolution density estimators when measurement errors are Laplacian or normal: classical kernel deconvolution estimators and support vector deconvolution density estimators based on Method I and II. Target distributions are selected from distribution functions used in Hazelton and Turlach (2009).

When measurement errors are double exponential, the classical kernel deconvolution density estimator with Gaussian kernel is evaluated as

$$\hat{f}(x) = \frac{1}{2\pi n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{e^{i\xi(x-y_j)} \tilde{K}(\sigma_h \xi)}{\tilde{q}(\xi)} \, d\xi$$

$$= \frac{1}{n\sqrt{2\pi}\,\sigma_h} \sum_{j=1}^{n} e^{-\frac{(x-y_j)^2}{2\sigma_h^2}} \left\{1 - \frac{\sigma_Z^2}{\sigma_h^2}\left(\left(\frac{x-y_j}{\sigma_h}\right)^2 - 1\right)\right\}, \tag{4.1}$$

where $\tilde{K}(\sigma_h \xi) = e^{-0.5\sigma_h^2 \xi^2}$, $q(z) = (1/2\sigma_Z)e^{-|z|/\sigma_Z}$, and $\tilde{q}(\xi) = (1 + \sigma_Z \xi^2)^{-1}$.

The support vector deconvolution density estimator with Gaussian kernel is evaluated as
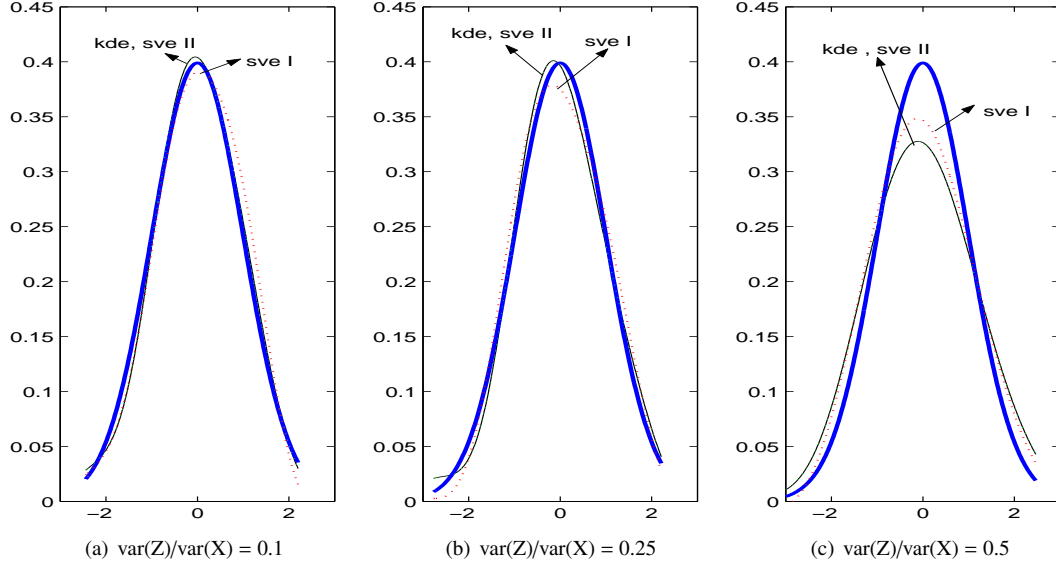
$$\hat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{j=1}^{n} \frac{\omega_j \tilde{K}_h(y_j, \xi) e^{i\xi x}}{\tilde{q}(\xi)} \, d\xi$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_h} \sum_{j=1}^{n} \omega_j \, e^{-\frac{(x-y_j)^2}{2\sigma_h^2}} \left\{1 - \frac{\sigma_Z^2}{\sigma_h^2}\left(\left(\frac{x-y_j}{\sigma_h}\right)^2 - 1\right)\right\}. \tag{4.2}$$

When measurement errors are normal, in order to avoid problems of integrability, a kernel K that has compactly supported Fourier transform $\tilde{K}(\omega) = (1 - \omega^2)^3 I_{[-1,1]}(\omega)$ is used in common. Using this kernel $K(x) = \{(48\cos x)/(\pi x^4)\}(1 - 15/x^2) - \{(144\sin x)/(\pi x^5)\}(2 - 5/x^2)$, a kernel deconvolution density estimator $\hat{f}(x)$ in the presence of normal measurement error can be calculated as:

$$\hat{f}(x) = \frac{1}{2\pi n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{e^{i\xi(x-y_j)} \tilde{K}(h_n \xi)}{\tilde{q}(\xi)} \, d\xi$$

$$= \frac{1}{n\pi h_n} \sum_{j=1}^{n} \int_{0}^{1} \left(1 - \xi^2\right)^3 \cos\left(\xi\left(\frac{x-y_j}{h_n}\right)\right) e^{\frac{\sigma_Z^2 \xi^2}{2h_n^2}} \, d\xi, \tag{4.3}$$

Table 1: The bandwidth used in the simulation when $q(z)$ is Gaussian

$f(x)$	$N(0,1)$			$0.5N(-2.5,1) + 0.5N(2.5,1)$			$\frac{2}{3}N(0,1) + \frac{1}{3}N(0,0.04)$		
var(Z)/var(X)	0.10	0.25	0.50	0.10	0.25	0.50	0.10	0.25	0.50
bandwidth($= h_n$)	0.20	0.23	0.35	0.40	0.50	0.65	0.15	0.22	0.25

(a) var(Z)/var(X) = 0.1   (b) var(Z)/var(X) = 0.25   (c) var(Z)/var(X) = 0.5

Figure 1: $f(x)$ is $N(0,1)$ and $q(z)$ is Gaussian.

where $\tilde{K}(\xi) = (1-\xi^2)^3 I_{[-1,1]}(\xi)$, $q(z) = (\sqrt{2\pi}\sigma_Z)^{-1}e^{-z^2/2\sigma_Z^2}$, and $\tilde{q}(\xi) = e^{-\sigma_Z^2\xi^2/2}$.

The support vector deconvolution density estimator is evaluated as

$$\hat{f}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{j=1}^{n}\frac{\omega_j\tilde{K}_h\left(y_j,\xi\right)e^{i\xi x}}{\tilde{q}(\xi)}\,d\xi$$

$$= \frac{1}{2\pi}\sum_{j=1}^{n}\omega_j\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\left(1-h_n^2\xi^2\right)^3 e^{-i\xi y_j}e^{\frac{\sigma_Z^2\xi^2}{2}}e^{i\xi x}\,d\xi$$

$$= \frac{1}{\pi h_n}\sum_{j=1}^{n}\omega_j\int_0^1\left(1-\xi^2\right)^3\cos\left(\xi\left(\frac{x-y_j}{h_n}\right)\right)e^{\frac{\sigma_Z^2\xi^2}{2h_n^2}}\,d\xi. \tag{4.4}$$

All of the Figures 1–6 show plots of the classical kernel deconvolution estimates and support vector deconvolution estimates when 100 points are randomly generated respectively from a target distribution $f(x)$ and a noise distribution, normal distribution $q(z)$ with mean zero or double exponential distribution $q(z)$ with mean zero. The measurement error variance is set at low (= var(Z)/var(X) = 0.1), moderate (= var(Z)/var(X) = 0.25), and high levels (= var(Z)/var(X) = 0.5) as shown in Hazelton and Turlach (2009). The exact probability density function $f(x)$ is shown in bold line. For the support vector deconvolution estimates, Gunn's program (Gunn, 1998) and MATLAB 6.5 were used. In kernel density estimation the choice of kernel is not crucial, but the choice of bandwidth is very important. In the following figures, a rule of thumb bandwidth as an initial value (Fan, 1991, 1992; Wang and
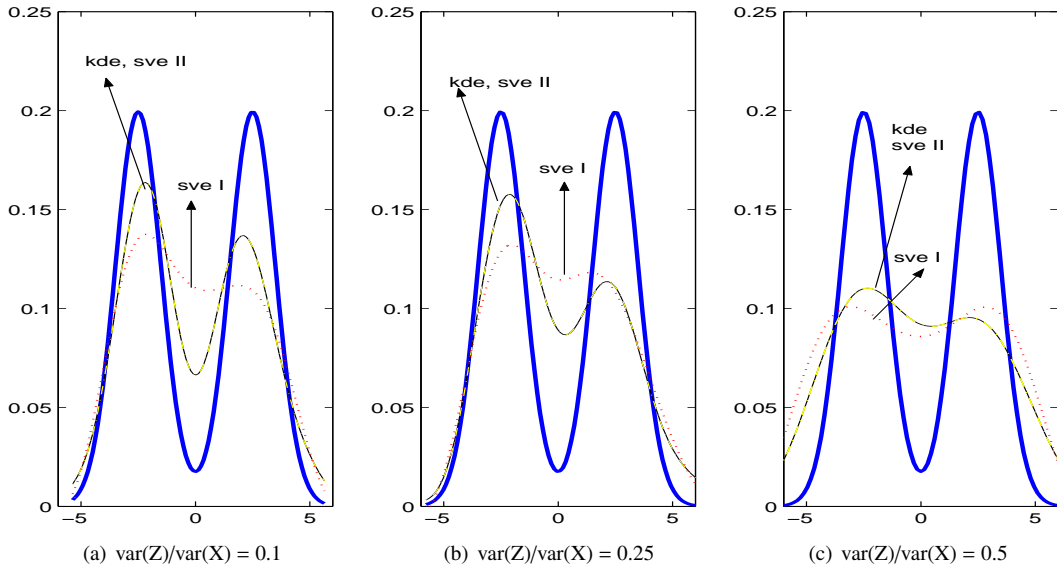
(a) var(Z)/var(X) = 0.1   (b) var(Z)/var(X) = 0.25   (c) var(Z)/var(X) = 0.5

Figure 2: $f(x)$ is $0.5N(-2.5, 1) + 0.5N(2.5, 1)$ and $q(z)$ is Gaussian.

(a) var(Z)/var(X) = 0.1   (b) var(Z)/var(X) = 0.25   (c) var(Z)/var(X) = 0.5
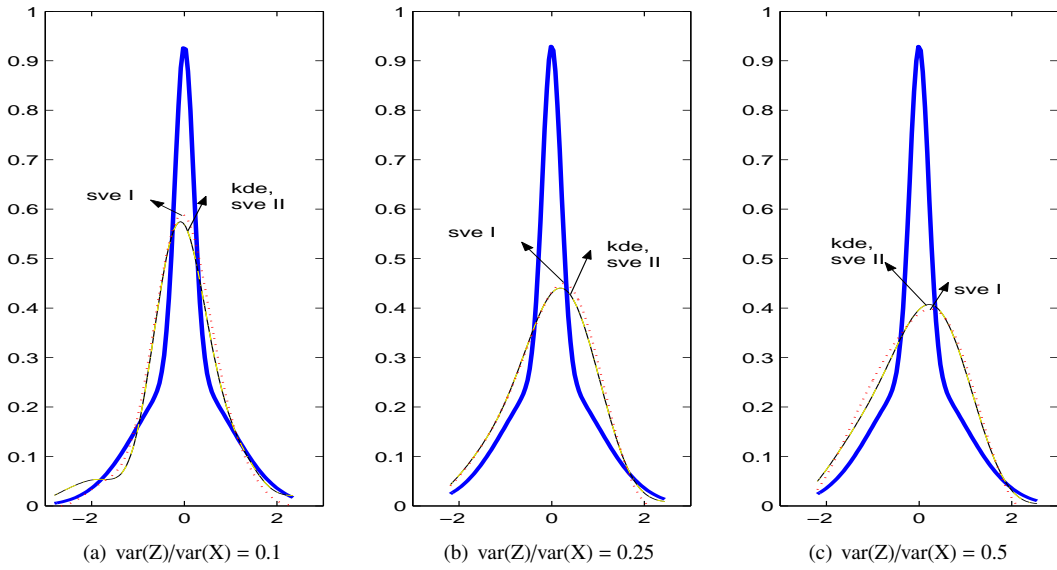
Figure 3: $f(x)$ is $2/3N(0, 1) + 1/3N(0, 0.04)$ and $q(z)$ is Gaussian.

Wang, 2011) is used: for Laplacian errors, $\sigma_h = (5\sigma_Z^4/n)^{1/9}$ and for Gaussian errors, $h_n = \sigma_Z(\ln n)^{-0.5}$. Figures 1–3 presents the simulation study when $q(z)$ is Gaussian. The bandwidth($= h_n$) used for kernel deconvolution estimates (4.3) and support vector deconvolution estimates (4.4) in the simulation is in Table 1. All the deconvolution estimates used the same bandwidth. The parameters (ϵ, C) used for all the support vector deconvolution estimates are $(0.05, \infty)$. In the figures kde, sve I and sve II represent kernel deconvolution estimates, support vector deconvolution estimates by Method I, and support

Table 2: The bandwidth($= \sigma_h$) used in the simulation when $q(z)$ is Laplacian

$f(x)$	$N(0,1)$			$0.5N(-2.5,1) + 0.5N(2.5,1)$			$\frac{2}{3}N(0,1) + \frac{1}{3}N(0,0.04)$		
var(Z)/var(X)	0.10	0.25	0.50	0.10	0.25	0.50	0.10	0.25	0.50
kde	0.50	0.60	0.70	0.75	0.90	1.20	0.35	0.50	0.55
sve I	0.95	0.95	0.95	1.10	1.15	1.20	0.80	0.80	0.90
sve II	0.80	0.80	0.80	1.10	1.15	1.40	0.60	0.60	0.90

(a) var(Z)/var(X) = 0.1   (b) var(Z)/var(X) = 0.25   (c) var(Z)/var(X) = 0.5
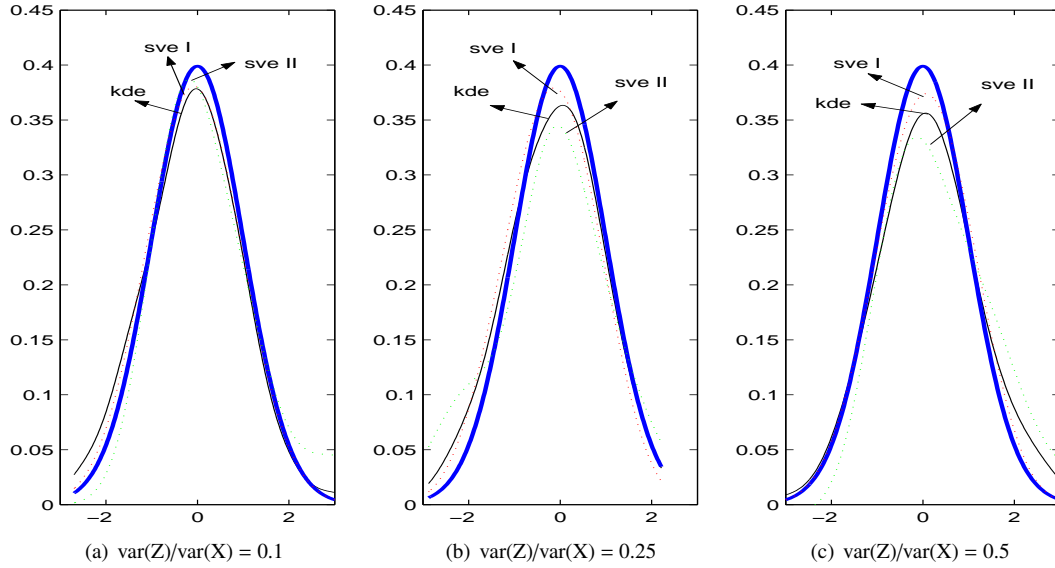
Figure 4: $f(x)$ is $N(0,1)$ and $q(z)$ is Laplasian.

vector deconvolution estimates by Method II respectively. In the case of Gaussian error, all support vector deconvolution estimates by Method II show the same figures as the usual kernel deconvolution estimates as Figures 1–3 indicates. In Figures 1–3, the number of data (support vectors) used for sve II is 10 to 15 among 100 random numbers and the number of data used for sve I is 40 to 60.

Figures 4–6 presents a simulation study when $q(z)$ is double exponential distribution with mean zero. The bandwidth($= \sigma_h$) used for kernel deconvolution estimator (4.1) and support vector deconvolution estimator (4.2) in the simulation is in Table 2. The parameters (ϵ, C) used for the support vector deconvolution estimates by Method I and the support vector deconvolution estimates by Method II are $(0.05, \infty)$ and $(0.05, 0.01)$ respectively. In the case of Laplacian error, support vector deconvolution density estimates by Method II show very similar figures to the usual kernel deconvolution density estimates as Figures 4–6 indicate. In Figures 4–6, the number of data (support vectors) used for sve II is 5 to 15 among 100 random numbers. However the number of data used for sve I is not sparse as $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$ indicates.

5. Concluding remarks

Three different deconvolution density estimators were discussed in this paper when the sample observations are contaminated by Gaussian or double exponentially distributed errors. The support vector deconvolution density estimators by Method II lead to very sparse solution and the number of support
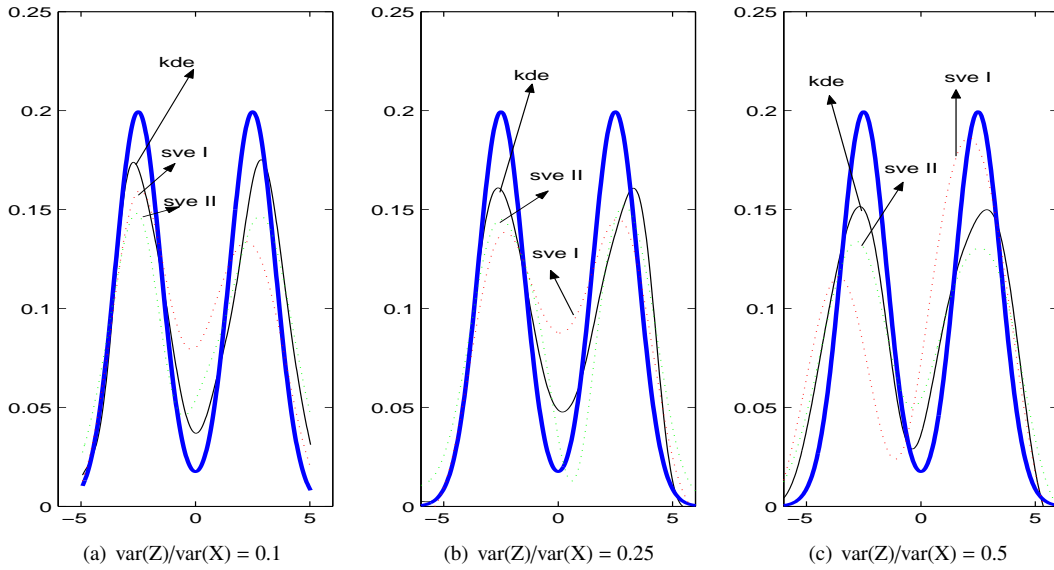
Figure 5: $f(x)$ is $0.5N(-2.5, 1) + 0.5N(2.5, 1)$ and $q(z)$ is Laplasian.

(a) var(Z)/var(X) = 0.1   (b) var(Z)/var(X) = 0.25   (c) var(Z)/var(X) = 0.5
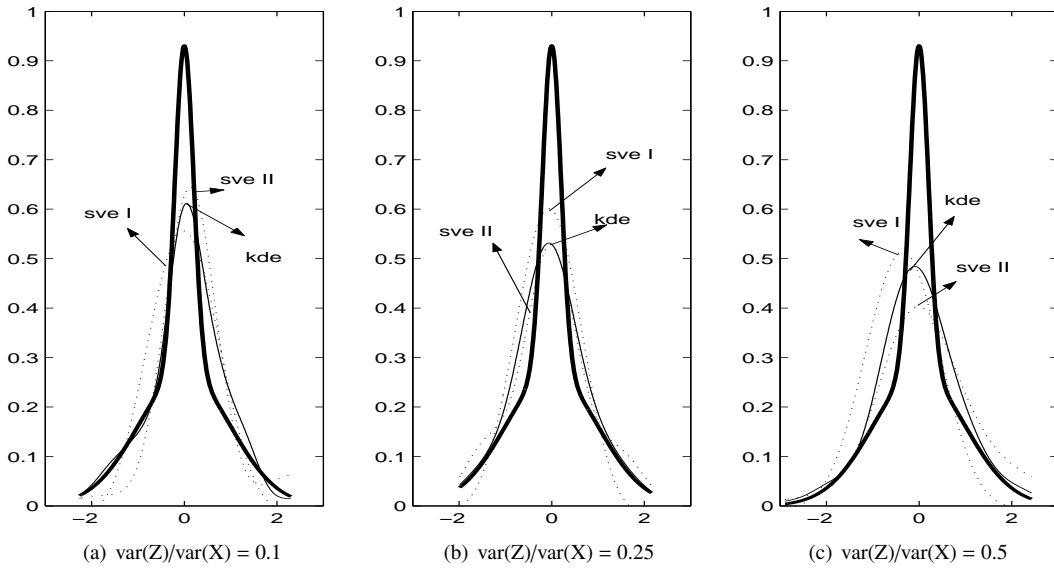
Figure 6: $f(x)$ is $2/3N(0, 1) + 1/3N(0, 0.04)$ and $q(z)$ is Laplasian.

vectors is 10 to 15 among 100 random numbers in the simulation. Furthermore, it is as good as the usual kernel deconvolution density estimators even though the simulation study conducted is limited. In the case of Gaussian error, the support vector deconvolution density estimates by Method II show the same figures as the kernel deconvolution density estimates. The support vector deconvolution density estimator by Method I is also attractive in the sense that some coefficients in $\omega = \Gamma_h^{-1} R(\alpha - \alpha^*)$ are close to zero.

Acknowledgements

References

Aronszajn N (1950). Theory of reproducing kernels, *Transactions of the American Mathematical Society*, **68**, 337–404.

Bochner S (1959). *Lectures on Fourier Integral*, Princeton University Press, Princeton, New Jersey.

Carroll RJ and Hall P (1988). Optimal rates of convergence for deconvoluting a density, *Journal of the American Statistical Association*, **83**, 1184–1886.

Fan J (1991). On the optimal rates of convergence for nonparametric deconvolution problem, *Annals of Statistics*, **19**, 1257–1272.

Fan J (1992). Deconvolution with supersmooth distribution, *Canadian Journal of Statistics*, **20**, 159–169.

Girosi F (1998). An equivalence between sparse approximation and support vector machines, *Neural Computation*, **10**, 1455–1480.

Gunn SR (1998). *Support vector machines for classification and regression*, Technical report, University of Southampton.

Hall P and Qiu P (2005). Discrete-transform approach to deconvolution problems, *Biometrika*, **92**, 135–148.

Hazelton ML and Turlach BA (2009). Nonparametric density deconvolution by weighted kernel estimators, *Statistics and Computing*, **19**, 217–228.

Lee S (2010). A support vector method for the deconvolution problem, *Communications of the Korean Statistical Society*, **17**, 451–457.

Lee S (2012). A note on deconvolution estimators when measurements errors are normal, *Communications of the Korean Statistical Society*, **19**, 517–526.

Lee S and Taylor RL (2008). A note on support vector density estimation for the deconvolution problem, *Communications in Statistics: Theory and Methods*, **37**, 328–336.

Mendelsohn J and Rice R (1982). Deconvolution of microfluorometric histograms with B splines, *Journal of the American Statistical Association*, **77**, 748–753.

Mercer J (1909). Functions of positive and negative type and their connection with the theory of integral equations, *Philosophical Transactions of the Royal Society of London*, **A 209**, 415–446.

Moguerza JM and Munoz A (2006). Support vector machines with applications, *Statistical Science*, **21**, 322–336.

Mukherjee S and Vapnik V (1999). Support vector method for multivariate density estimation, In , *Proceedings in Neural Information Processing Systems*, 659–665.

Pensky M and Vidakovic B (1999). Adaptive wavelet estimator for nonparametric density deconvolutoin, *Annals of Statistics*, **27**, 2033–2053 .

Phillips DL (1962). A technique for the numerical solution of integral equations of the first kind, *Journal of the Association for Computing Machinery*, **9**, 84–97.

Rasmussen CE and Williams CKI (2006). *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.

Scholkopf B, Herbrich R, and Smola AJ (2001). A generalized representer theorem, *Computational Learning Theory*, Lecture Notes in Computer Science, **2111**, 416–426.

Scholkopf B and Smola AJ (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA.

Smola AJ and Scholkopf B (2003). A tutorial on support vector regression, *Statistics and Computing*, **14**, 199–222.

Stefanski L and Carroll RJ (1990). Deconvoluting kernel density estimators, *Statistics*, **21**, 169–184.

Tikhonov AN and Arsenin VY (1977). *Solution of Ill-posed Problems*, W. H. Winston, Washington.

Vapnik V (1995). *The Nature of Statistical Learning Theory*, Springer Verlag, New York.

Vapnik V and Chervonenkis A (1964). A note on one class of perceptrons, *Automation and Remote Control*, **25**, 103–109.

Vapnik V and Lerner L (1963). Pattern recognition using generalized portrait method, *Automation and Remote Control*, **24**, 774–780.

Vert R and Vert J (2006). Consistency and convergence rates of one-class SVMs and related algorithms, *Journal of Machine Learning Research*, **7**, 817–854.

Wahba G (1990). *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, **59**, SIAM, Philadelphia.

Wahba G (2006). Comments to support vector machines with applications by J.M. Moguerza and A. Munoz, *Statistical Sciences*, **21**, 347–351.

Wang X and Wang B (2011). Deconvolution estimation in measurement error models: The R package decon, *Journal of Statistical Software*, **39**, i10.

Weston J, Gammerman A, Stitson M, Vapnik V, Vovk V, and Watkins C (1999). Support vector density estimation. In Scholkopf, B. and Smola, A., editors, *Advances in Kernel Methods-Support Vector Learning*, 293–306, MIT Press, Cambridge, MA.

Zhang HP (1992). On deconvolution using time of flight information in positron emission tomography, *Statistica Sinica*, **2**, 553–575.