Extrapolation in variable RKHSs with application to the blood glucose reading

2011 Inverse Problems 27 075010

(http://iopscience.iop.org/0266-5611/27/7/075010)

Inverse Problems **27** (2011) 075010 (13pp) doi:10.1088/0266-5611/27/7/075010

Extrapolation in variable RKHSs with application to the blood glucose reading

V Naumova, S V Pereverzyev and S Sivananthan

Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergstraße 69, A-4040 Linz, Austria

E-mail: valeriya.naumova@oeaw.ac.at

Received 6 March 2011, in final form 11 May 2011
Published 14 June 2011
Online at stacks.iop.org/IP/27/075010

Abstract

In this paper we present a new scheme of a kernel adaptive regularization algorithm, where the kernel and the regularization parameter are adaptively chosen within the regularization procedure. The construction of such a fully adaptive regularization algorithm is motivated by the problem of reading the blood glucose concentration of diabetic patients. We describe how the proposed scheme can be used for this purpose and report the results of numerical experiments with real clinical data.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In this paper we consider the problem of a reconstruction of a real-valued function $f : X \to \mathbb{R}$, $X \subset \mathbb{R}^d$, from a given data set

$$\mathbf{z} = \{(x_i, y_i)\}_{i=1}^n \subset X \times \mathbb{R},$$

where it is assumed that $y_i = f(x_i) + \xi_i$, and $\xi_i = \{\xi_i\}_{i=1}^n$ is a noise vector. At this point, it should be noted that the reconstruction problem can be considered in two aspects. One aspect is to evaluate the value of a function $f(x)$ for $x \in \overline{co\{x_i\}}$, where $\overline{co\{x_i\}}$ is the closed convex hull of data points $\{x_i\}$. It is sometimes called interpolation. The other aspect is to predict the value of $f(x)$ for $x \notin \overline{co\{x_i\}}$, which is known as extrapolation.

In both aspects the reconstruction problem is ill-posed, and the task of solving it makes sense only when placed in an appropriate framework. The numerical treatment of ill-posed problems with noisy data requires the application of special regularization methods. The most popular among them is the Tikhonov method, which in the present context consists in constructing a regularized solution $f_\lambda(x)$ as a minimizer of the functional

$$T_{\lambda,r}(f) = \frac{1}{|\mathbf{z}|}\sum_{i=1}^{|\mathbf{z}|}(y_i - f(x_i))^2 + \lambda||f||^2_{W_2^r}, \tag{1}$$

where $|\mathbf{z}|$ is the cardinality of the set $\mathbf{z}$, i.e. $|\mathbf{z}| = n$, and λ is a regularization parameter, which trades off data error with smoothness measured in terms of a Sobolev space W_2^r [17].

The Tikhonov method, even in its simplest form (1), raises two issues that should be clarified before use of this scheme. One of them is how to choose a regularized parameter λ. This problem has been extensively discussed. A few selected references from the literature are [4, 5, 10, 16].

Another issue which needs to be addressed is the choice of a space, whose norm is used for penalization. The impact of the chosen norm is most explicit in Tikhonov regularization [6]. Moreover, the choice of the proper space is firmly connected to the problem of choosing a representation of the output $f_\lambda(x)$. Thus, the choice can make a significant difference in practice. Despite its significance, the second issue is much less studied. Note that there are still no general principles to advise a choice, and only in a few papers [5, 12, 13, 17] some methods for finding an appropriate space for the regularization of ill-posed problems have been proposed. Keeping in mind that a Sobolev space W_2^r used in (1) is a particular example of a reproducing kernel Hilbert space (RKHS), the above-mentioned issue is, in fact, about the choice of a kernel for an RKHS. Exactly in this context, this issue has been studied recently in [13], but, as it will be seen from our discussion below, the kernel choice suggested in [13] does not fit well for extrapolation. With this application in mind, we will propose another kernel choice rule (kernel adaptive regularization algorithm – KAR-algorithm), which is based on a splitting of a given data set $\mathbf{z}$, oriented toward extrapolation.

This paper is organized as follows. In section 2, we outline a theoretical background about RKHS, present our kernel choice criteria and prove the existence of the kernel that satisfies it. We also illustrate our approach by a few academic examples. In section 3, we discuss the possibility of using the proposed approach in diabetes therapy management, in particular, for reading blood glucose levels from subcutaneous electric current measurements, and report the results of numerical experiments.

2. Regularized interpolation and extrapolation in RKHS

2.1. Reproducing kernel Hilbert space

A RKHS $\mathcal{H}$ is a space of real-valued functions f defined on $X \subset \mathbb{R}^d$ such that for every $x \in X$, the point-wise evaluation functional $L_x(f) := f(x)$ is continuous in the topology induced by the inner product $\langle \cdot, \cdot \rangle$, which generates the norm of $\mathcal{H}$. By the Riesz representation theorem, to every RKHS $\mathcal{H}$ there corresponds a unique symmetric positive definite function $K : X \times X \to \mathbb{R}$, called the reproducing kernel of $\mathcal{H} = \mathcal{H}_K$, that has the following reproducing property: $f(x) = \langle f(\cdot), K(\cdot, x) \rangle$ for every $x \in X$ and $f \in \mathcal{H}_K$.

For any positive definite function $K(x, y)$ on X, a uniquely determined Hilbert space $\mathcal{H} = \mathcal{H}_K$ exists with an inner product $\langle \cdot, \cdot \rangle$, admitting the reproducing property. Conversely, each RKHS $\mathcal{H}$ admits a unique reproducing kernel $K(x, y)$. A comprehensive theory of RKHSs can be found in [1].

In the following we will deal with the situation when X is a compact set in $\mathbb{R}^d$. Then, it is known [19] that $\mathcal{H}_K$ is continuously embedded in the space of continuous functions $C(X)$, as well as in the space $L_2(X)$ of functions, which are square-summable on X. Moreover, the canonical embedding operator $J_K : \mathcal{H}_K \to L_2(X)$ is compact.

Now we can write the problem of a reconstruction of $f \in \mathcal{H}_K$ from noisy data $\mathbf{z}$ in the form

$$J_K f = f_\xi, \tag{2}$$

where $f_\xi \in L_2(X)$ and is such that $f_\xi(x_i) = y_i$, $i = 1, 2, \ldots, |\mathbf{z}|$.

Note that in general, $f_\xi \notin \mathrm{Range}(J_K)$, and the problem (2) is ill-posed. Moreover, noisy data $\mathbf{z}$ only allow access to a discretized version of (2), that can be written as follows:

$$S_{\mathbf{x}} f = \mathbf{y}, \tag{3}$$

where $\mathbf{x} = \{x_i\}_{i=1}^{|\mathbf{z}|}, \mathbf{y} = \{y_i\}_{i=1}^{|\mathbf{z}|}$ and $S_{\mathbf{x}} : \mathcal{H}_K \to \mathbb{R}^{|\mathbf{z}|}$ is the sampling operator $S_{\mathbf{x}} f = \{f(x_i)\}_{i=1}^{|\mathbf{z}|}$.

Observe that problem (3) inherits the ill-posedness of (2) and should be treated by means of a regularization technique.

If a kernel K has already been chosen, several methods of the general regularization theory can be applied to equation (3), as has been analyzed in [2]. In particular, as we stated in the introduction, one can apply the Tikhonov regularization. In this case the regularization estimator/predictor $f_\lambda(x) = f_\lambda(x; K, \mathbf{z})$ is constructed from (3) as the minimizer of the functional

$$T_\lambda(f) = T_\lambda(f, K, \mathbf{z}) = \frac{1}{|\mathbf{z}|} \sum_{i=1}^{|\mathbf{z}|} (y_i - f(x_i))^2 + \lambda ||f||_K^2, \tag{4}$$

where $||\cdot||_K$ is a standard norm in an RKHS $\mathcal{H}_K$.

From the Representer theorem [17], it follows that the minimizer of (4) has the form

$$f_\lambda(x; K, \mathbf{z}) = \sum_{i=1}^{|\mathbf{z}|} c_i^\lambda K(x, x_i), \tag{5}$$

where a real vector $\boldsymbol{c}_\lambda = \left(c_1^\lambda, c_2^\lambda, \ldots, c_{|\mathbf{z}|}^\lambda\right)$ of coefficients is defined as follows:

$$\boldsymbol{c}_\lambda = (\lambda |\mathbf{z}| \mathbb{I} + \mathbb{K})^{-1} \boldsymbol{y},$$

here $\mathbb{I}$ is the unit matrix of the size $|\mathbf{z}| \times |\mathbf{z}|$, $\mathbb{K} = \{K(x_i, x_j)\}_{i,j=1}^{|\mathbf{z}|}$ is the Gram matrix and $\boldsymbol{y} = (y_1, y_2, \ldots, y_{|\mathbf{z}|})$.

2.2. Adaptive parameter choice

When a kernel K is already chosen, an appropriate choice of the regularization parameter λ is crucial to ensure a good performance of the method. For example, one can use a data-driven method for choosing the regularization parameter known as the quasi-optimality criterion. It was proposed long time ago in [16] and as it has been proven recently in [10], using this criterion one potentially may achieve an accuracy of optimal order (for a given kernel K).

To apply the quasi-optimality criterion, one needs to calculate the approximations $f_\lambda(x; K, \mathbf{z})$ given by (5) for λ from a finite part of a geometric sequence

$$\Lambda_q^\nu = \{\lambda_s : \lambda_s = \lambda_0 q^s, s = 0, 1, \ldots, \nu\}, \quad q > 1. \tag{6}$$

Then one needs to calculate the norm

$$\sigma_{\mathcal{H}_K}^2(s) = ||f_{\lambda_s}(x; K, \mathbf{z}) - f_{\lambda_{s-1}}(x; K, \mathbf{z})||_K^2 \tag{7}$$

in the space $\mathcal{H}_K$, and find

$$\lambda_+ = \lambda_p : p = \arg\min \left\{\sigma_{\mathcal{H}_K}^2(s), s = 0, 1, \ldots, \nu\right\}. \tag{8}$$

Note that it is not difficult to calculate the norm in (7), since by definition we have

$$||f_{\lambda_s}(x; K, \mathbf{z}) - f_{\lambda_{s-1}}(x; K, \mathbf{z})||_K^2 = \sum_{i=1}^{|\mathbf{z}|} \sum_{j=1}^{|\mathbf{z}|} \left(c_i^{\lambda_s} - c_i^{\lambda_{s-1}}\right)\left(c_j^{\lambda_s} - c_j^{\lambda_{s-1}}\right) K(x_i, x_j).$$

2.3. The choice of the kernel from a parameterized set

It seems to be that in practice one of the most delicate and challenging issues is the choice of the kernel K. On the one hand, special cases of the single fixed kernel K have been considered in the literature [4, 7]. On the other hand, in practice we would like to have multiple kernels or parameterizations of kernels to scale well the solution of the problem of interest. This is desirable because then we can choose K from the available set of kernels, dependent on the input data, and get good performing results.

Lanckriet *et al* were among the first to emphasize the need to consider the multiple kernels or parameterizations of kernels, and not a single *a priori* fixed kernel, since practical problems often involve multiple, heterogeneous data sources. In their work [12], the authors consider the set

$$\mathcal{K}(\{K_i\}) = \left\{ K = \sum_{i=1}^{m} \beta_i K_i \right\}$$

of linear combinations of some prescribed kernels $\{K_i\}_{i=1}^m$ and propose different criteria to select the kernel from it. It is noteworthy that for some practical applications, such a set of admissible kernels is not rich enough. Therefore, more general parameterizations are also of interest.

Let us consider the set $\mathcal{K}(X)$ of all kernels (continuous, symmetric positive definite functions) defined on $X \subset \mathbb{R}^d$. Let Ω also be a compact metric space and $G : \Omega \to \mathcal{K}(X)$ be an injection such that for any $x_1, x_2 \in X$, $w \in \Omega$, the function $w \to G(w)(x_1, x_2)$ is a continuous map from Ω to $\mathbb{R}$; here $G(w)(x_1, x_2)$ is the value of the kernel $G(w) \in \mathcal{K}(X)$ at $(x_1, x_2) \in X \times X$.

Each such mapping G determines a set of kernels

$$\mathcal{K}(\Omega, G) = \{K : K = G(w), K \in \mathcal{K}(X), w \in \Omega\}$$

parameterized by elements of Ω. In contrast to $\mathcal{K}(\{K_i\})$, $\mathcal{K}(\Omega, G)$ may be a nonlinear manifold.

Example 1. Consider $\Omega = [a, b]^3$, $0 < a < b$, $(\alpha, \beta, \gamma) \in [a, b]^3$, and define a mapping $G : (\alpha, \beta, \gamma) \to (x_1 x_2)^\alpha + \beta \exp[-\gamma(x_1 - x_2)^2]$, where $x_1, x_2 \in X \subset (0, \infty)$. It is easy to see that $G(\alpha, \beta, \gamma)$ is a positive definite as the sum of two positive definite functions of (x_1, x_2). Moreover, for any fixed $x_1, x_2 \in X$, the value $G(\alpha, \beta, \gamma)(x_1, x_2)$ continuously depends on (α, β, γ).

Thus, kernels from the set

$$\mathcal{K}(\Omega, G) = \{K : K(x_1, x_2) = (x_1 x_2)^\alpha + \beta \exp[-\gamma(x_1 - x_2)^2], (\alpha, \beta, \gamma) \in [a, b]^3\}$$

are parameterized by points of $\Omega = [a, b]^3$ in the sense described above. Note that for the extrapolation problem, there is a good reason to concentrate attention on this set of kernels, since for any γ, the summands $K_\gamma(x_1, x_2) = \exp[-\gamma(x_1 - x_2)^2]$ are the so-called universal kernels [14], which means that all continuous functions defined on a compact set $X \subset (0, \infty)$ are uniform limits of functions of the form (5), where $K = K_\gamma$ and $\{x_i\} \subset X$. At the same time, for a fixed set $\{x_i\}$ of data points and $x \notin \overline{co\{x_i\}}$, the approximation performance of functions (5) with $K = K_\gamma$ may be poor due to the fact that the values of all kernel selections $K_\gamma(x, x_i) = \exp[-\gamma(x - x_i)^2]$ are decreasing with an increase of distance from the point x to the set $\overline{co\{x_i\}}$. Then the summands $K_\alpha(x, x_i) = (x x_i)^\alpha$ serve to compensate such a decrease.

Once the set of kernels is fixed, one may follow [13] and select a kernel by minimizing the Tikhonov regularization functional (4) such that

$$K_{\text{opt}} = \arg\min\{T_\lambda(f_\lambda(\cdot; K, \mathbf{z}), K, \mathbf{z}), K \in \mathcal{K}(\Omega, G)\}.$$

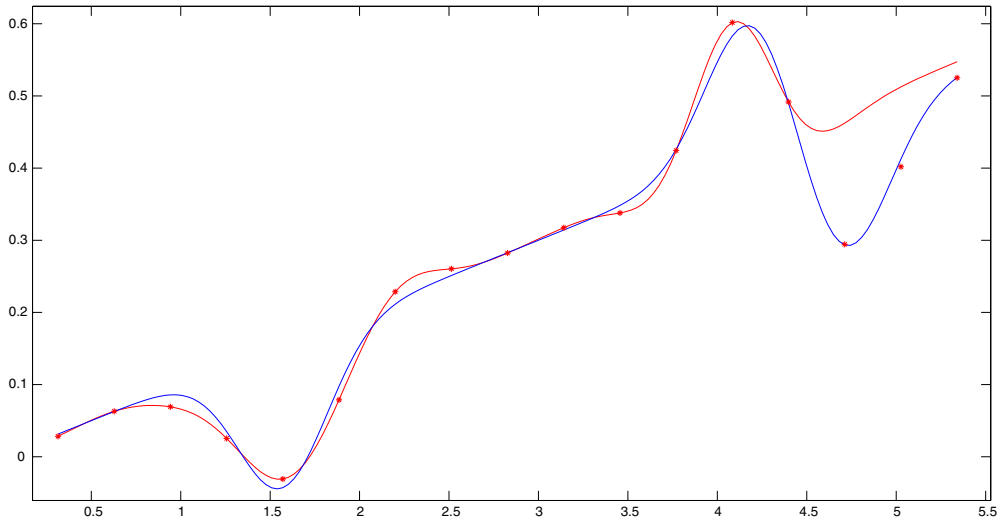

Figure 1. The performance of the approximant $f_\lambda(x; K, \mathbf{z})$ (red line) based on the kernel $K(x_1, x_2) = x_1x_2 + \exp[-8(x_1 - x_2)^2]$ generating the target function f.

Thus, the idea of [13] is to recover the kernel K generating the space $\mathcal{H}_K$, where the unknown function of interest lives, from given data, and then use this kernel for constructing the approximant $f_\lambda(\cdot; K, \mathbf{z})$.

To illustrate that such an approach may fail in a prediction of the extrapolation type, we use the same example as in [13] and [18], where $f(x) = 0.1\big(x + 2\big\{\exp\big[-8\big(\frac{4\pi}{3} - x\big)^2\big] - \exp\big[-8\big(\frac{\pi}{2} - x\big)^2\big] - \exp\big[-8\big(\frac{3\pi}{2} - x\big)^2\big]\big\}\big)$, and the given set $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^{14}$ consists of points $x_i = \frac{\pi i}{10}, i = 1, 2, \ldots, 14$ and $y_i = f(x_i) + \xi_i$, where ξ_i are random values sampled uniformly in the interval $[-0.02, 0.02]$. Note that in [13] the target function f has been chosen in such a way that it belongs to the RKHS generated by the 'ideal' kernel $K_{\text{opt}}(x_1, x_2) = x_1x_2 + \exp[-8(x_1 - x_2)^2]$. The performance of the approximant $f_\lambda(x; K_{\text{opt}}, \mathbf{z})$ with the best λ is shown in figure 1. This figure illustrates that for $x \in \overline{co\{x_i\}} = [\frac{\pi}{10}, \frac{14\pi}{10}]$, the value of $f(x)$ is estimated well by $f_\lambda(x; K_{\text{opt}}, \mathbf{z})$, while for $x \notin \overline{co\{x_i\}}$ the performance of the approximant based on the K_{opt} is rather poor. Observe that figure 1 displays the performance of the approximation (5) with the best λ. It means that the choice of the regularization parameter λ cannot improve the performance of the approximation (4), (5) given by the 'ideal' kernel, that is, the kernel $K_{\text{opt}}(x_1, x_2) = x_1x_2 + \exp[-8(x_1 - x_2)^2]$ used to generate the target function f.

Figure 2 displays the performance of the approximation (5) constructed for the same data set $\mathbf{z}$, but with the use of the kernel $K(x_1, x_2) = (x_1x_2)^{1.9} + \exp[-2.7(x_1 - x_2)^2]$. As one can see, the approximation based on this kernel performs much better compared to figure 1. Note that the regularization parameter λ for this approximation has been chosen from the set Λ_q^ν by means of the quasi-optimality criterion (8). The kernel improving the approximation performance has been chosen from the set

$$\mathcal{K} = \{K(x_1, x_2) = (x_1x_2)^\alpha + \beta \exp[-\gamma(x_1 - x_2)^2], \alpha, \beta, \gamma \in [10^{-4}, 3]\} \tag{9}$$

as follows.

At first, let us split up the data $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^{|\mathbf{z}|}$ such that $\mathbf{z} = \mathbf{z}_T \cup \mathbf{z}_P$ and

$$\overline{co\{x_i : (x_i, y_i) \in \mathbf{z}_T\}} \cap \{x_i : (x_i, y_i) \in \mathbf{z}_P\} = \varnothing.$$

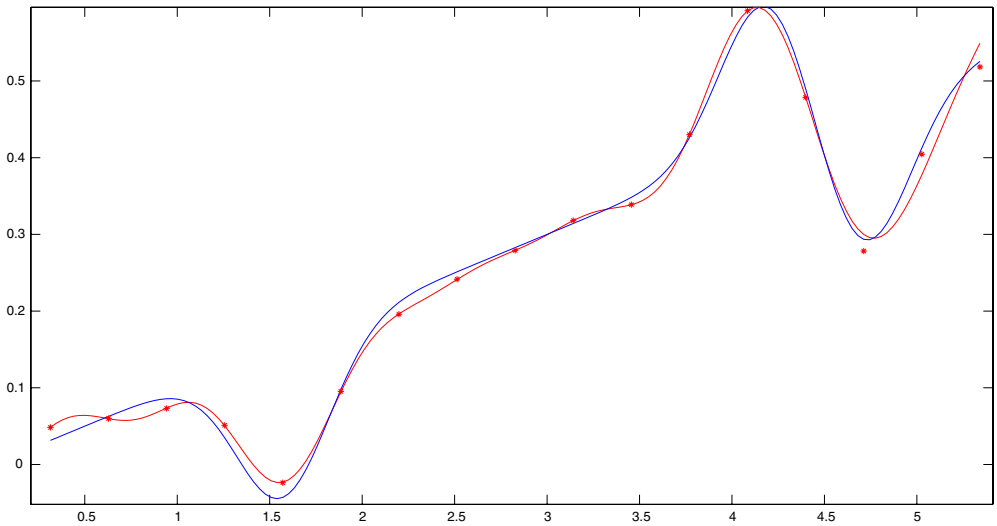

Figure 2. The performance of the approximant $f_\lambda(x; K, \mathbf{z})$ (red line) based on the kernel $K(x_1, x_2) = (x_1 x_2)^{1.9} + \exp[-2.7(x_1 - x_2)^2]$, that has been chosen as a minimizer of (12).

The approximant displayed in figure 2 corresponds to the splitting $\mathbf{z}_T = \{(x_i, y_i)\}_{i=1}^{7}$, $\mathbf{z}_P = \{(x_i, y_i)\}_{i=8}^{14}$.

For fixed $\mathbf{z}_T$ and corresponding Tikhonov-type regularization functional

$$T_\lambda(f, K, \mathbf{z}_T) = \frac{1}{|\mathbf{z}_T|} \sum_{i:(x_i, y_i)\in \mathbf{z}_T} (y_i - f(x_i))^2 + \lambda \|f\|_K^2, \tag{10}$$

we consider a rule $\lambda = \lambda(K)$ that for any $K \in \mathcal{K}(X)$ selects a regularization parameter from some fixed interval $[\lambda_{\min}, \lambda_{\max}]$, $\lambda_{\min} > 0$.

It is noteworthy again that we are interested in constructing regularized approximations of the form (5), which will reconstruct the values of the function at points inside/outside of the scope of $\mathbf{x}$. Therefore, the performance of each regularization estimator $f_\lambda(x; K, \mathbf{z}_T)$ is checked on the rest of a given data $\mathbf{z}_P$ and measured, for example, by the value of the functional

$$P(f, K, \mathbf{z}_P) = \frac{1}{|\mathbf{z}_P|} \sum_{i:(x_i, y_i)\in \mathbf{z}_P} \rho(f(x_i), y_i), \tag{11}$$

where $\rho(\cdot, \cdot)$ is a continuous function of two variables.

To construct the approximant displayed in figure 2, we take $\rho(f(x_i), y_i) = (y_i - f(x_i))^2$. However, the function $\rho(\cdot, \cdot)$ can be adjusted to the intended use of the approximant $f_\lambda(x; K, \mathbf{z}_T)$. In the next section, we present an example of such an adjustment.

Finally, the kernel of our choice is $K = K(\mathcal{K}, \mu, \lambda, \mathbf{z}; x_1, x_2)$ that minimizes the functional

$$Q_\mu(K, \lambda, \mathbf{z}) = \mu T_\lambda(f_\lambda(\cdot; K, \mathbf{z}_T), K, \mathbf{z}_T) + (1 - \mu) P(f_\lambda(\cdot; K, \mathbf{z}_T), K, \mathbf{z}_P) \tag{12}$$

over the set of admissible kernels $\mathcal{K}(\Omega, G)$. Note that the parameter μ here can be seen as a performance regulator on the sets $\mathbf{z}_T$ and $\mathbf{z}_P$. Taking μ closer to zero, we put more emphasize on the ability to extrapolate, while for $\mu > \frac{1}{2}$ we are more interested in interpolation.

The kernel choice rule based on the minimization of the functional (12) is rather general. We call this rule the kernel adaptive regularization algorithm (KAR-algorithm). Next theorem justifies the existence of the kernel and the regularization parameter that minimize the functional (12).

Theorem 1. *There are* $K^0 \in \mathcal{K}(\Omega, G)$ *and* $\lambda^0 \in [\lambda_{\min}, \lambda_{\max}]$ *such that for any parameter choice rule* $\lambda = \lambda(K)$

$$Q_\mu(K^0, \lambda^0, \mathbf{z}) = \inf\{Q_\mu(K, \lambda(K), \mathbf{z}),\ K \in \mathcal{K}(\Omega, G)\}.$$

Proof.. Let $\{K_l\} \in \mathcal{K}(\Omega, G)$ be a minimizing sequence of kernels such that

$$\lim_{l\to\infty} Q_\mu(K_l, \lambda(K_l), \mathbf{z}) = \inf\{Q_\mu(K, \lambda(K), \mathbf{z}),\ K \in \mathcal{K}(\Omega, G)\}.$$

Since, by construction, all $\lambda(K_l) \in [\lambda_{\min}, \lambda_{\max}]$, one can find a subsequence $\lambda_n = \lambda(K_{l_n})$, $n = 1, 2, \ldots$, such that $\lambda_n \to \lambda^0 \in [\lambda_{\min}, \lambda_{\max}]$.

Consider the subsequence of kernels $K^n = K_{l_n}$, $n = 1, 2, \ldots$.

Let also

$$f_{\lambda_n}(x; K^n, \mathbf{z}_T) = \sum_{i:(x_i,y_i)\in\mathbf{z}_T} c_i^n K^n(x, x_i), \quad n = 1, 2, \ldots,$$

be the minimizers of the functional (10) for $K = K^n$. From (5) we know that the vector $\boldsymbol{c_n} = (c_i^n) \in \mathbb{R}^{|\mathbf{z}_T|}$ admits a representation

$$\boldsymbol{c_n} = \left(\lambda_n |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^n\right)^{-1} \boldsymbol{y}_T,$$

where $\mathbb{I}$ is the unit matrix of the size $|\mathbf{z}_T| \times |\mathbf{z}_T|$, and the matrix $\mathbb{K}_T^n$ and the vector $\boldsymbol{y}_T$ are respectively formed by the values $K^n(x_i, x_j)$ and y_i with i, j such that $(x_i, y_i), (x_j, y_j) \in \mathbf{z}_T$. By the definition of $\mathcal{K}(\Omega, G)$, the sequence $\{K^n\} \in \mathcal{K}(\Omega, G)$ is associated with a sequence $\{w_n\} \in \Omega$ such that $K^n = G(w_n)$.

Since Ω is assumed to be a compact metric space, there is a subsequence $\{w_{n_k}\} \subset \{w_n\}$ that converges in Ω to some $w_0 \in \Omega$. Consider the kernel $K^0 = G(w_0) \in \mathcal{K}(\Omega, G)$.

Keeping in mind that for any fixed $x_1, x_2 \in X$ the function $w \to G(w)(x_1, x_2)$ is continuous on Ω, one can conclude that the entries $K^{n_k}(x_i, x_j) = G(w_{n_k})(x_i, x_j)$ of the matrices $\mathbb{K}_T^{n_k}$ converge to the corresponding entries $K^0(x_i, x_j) = G(w_0)(x_i, x_j)$ of the matrix $\mathbb{K}_T^0$. Therefore, for any $\epsilon > 0$, a natural number $k = k(\epsilon)$ exists depending only on ϵ such that for any $(x_i, y_i) \in \mathbf{z}_T$ and $k > k(\epsilon)$, we have $|K^0(x_i, x_j) - K^{n_k}(x_i, x_j)| < \epsilon$. It means that the matrices $\mathbb{K}_T^{n_k}$ converge to $\mathbb{K}_T^0$ in a standard matrix norm $\|\cdot\|$.

Consider the vector

$$\boldsymbol{c_o} = \left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)^{-1} \boldsymbol{y}_T$$

of coefficients (c_i^0) from the representation

$$f_{\lambda^0}(x; K^0, \mathbf{z}_T) = \sum_{i:(x_i,y_i)\in\mathbf{z}_T} c_i^0 K^0(x, x_i)$$

of the minimizer of the functional (10) for $K = K^0$. Since for $K^{n_k}, K^0 \in \mathcal{K}(\Omega, G)$, the corresponding matrices $\mathbb{K}_T^{n_k}, \mathbb{K}_T^0$ are positive definite, for any vector $\boldsymbol{y} \in \mathbb{R}^{|\mathbf{z}_T|}$ we have

$$\left\|\left(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}\right)^{-1} \boldsymbol{y}\right\| \leqslant (\lambda_{n_k} |\mathbf{z}_T|)^{-1} \|\boldsymbol{y}\|,$$

$$\left\|\left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)^{-1} \boldsymbol{y}\right\| \leqslant (\lambda^0 |\mathbf{z}_T|)^{-1} \|\boldsymbol{y}\|.$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{c_o} - \boldsymbol{c_{n_k}}\| &= \left\|\left(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}\right)^{-1} \left(\left(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}\right) - \left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)\right) \left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)^{-1} \boldsymbol{y}_T \right\| \\
&= \left\|\left(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}\right)^{-1} \left(\mathbb{K}_T^{n_k} - \mathbb{K}_T^0\right) \left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)^{-1} \boldsymbol{y}_T \right. \\
&\quad \left. + \left(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}\right)^{-1} (\lambda_{n_k} - \lambda^0) |\mathbf{z}_T| \left(\lambda^0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0\right)^{-1} \boldsymbol{y}_T \right\| \\
&\leqslant (\lambda_{\min} |\mathbf{z}_T|)^{-2} \|\boldsymbol{y}_T\| \left\|\mathbb{K}_T^{n_k} - \mathbb{K}_T^0\right\| + (\lambda_{\min} |\mathbf{z}_T|)^{-2} \|\boldsymbol{y}_T\| |\lambda_{n_k} - \lambda^0| |\mathbf{z}_T|,
\end{aligned}$$

and in view of our observation that $\lambda_{n_k} \to \lambda^0$ and $\mathbb{K}_T^{n_k} \to \mathbb{K}_T^0$, we can conclude that $\boldsymbol{c}_{\boldsymbol{n_k}} \to \boldsymbol{c}_0$ in $\mathbb{R}^{|\mathbf{z}_T|}$.

Now we note that for any $K \in \mathcal{K}(\Omega, G)$ and

$$f_\lambda(x; K, \mathbf{z}_T) = \sum_{j:(x_j,y_j)\in\mathbf{z}_T} c_j K(x, x_j),$$

the functional (12) can be seen as a continuous function

$$Q_\mu(K, \lambda, \mathbf{z}) = Q_\mu(\{K(x_i, x_j)\}, \lambda, \mathbf{z}, \{c_j\})$$

of λ, c_j and $K(x_i, x_j), i = 1, 2, \ldots, |\mathbf{z}|, j : (x_j, y_j) \in \mathbf{z}_T$. Therefore, summarizing our reasons, we have

$$Q_\mu(K^0, \lambda^0, \mathbf{z}) = \lim_{k\to\infty} Q_\mu(K^{n_k}, \lambda_{n_k}, \mathbf{z}) = \inf\{Q_\mu(K, \lambda(K), \mathbf{z}), K \in \mathcal{K}(\Omega, G)\}.$$

□

3. Reading blood glucose levels from subcutaneous electric current measurements

In this section, we discuss the possibility of using the approach described in the previous section in diabetes therapy management, in particular, for reading blood glucose levels from subcutaneous electric current measurements.

Continuous glucose monitoring (CGM) systems provide, almost in real-time, an indirect estimation of current blood glucose that is highly valuable for the insulin therapy of diabetes. For example, needle-based electrochemical sensors, such as Abbott Freestyle Navigator [20], measure electrical signal in the interstitial fluid (ISF) and return ISF glucose concentration (mg dL^{-1}) exploiting some internal calibration procedure. This ISF glucose reading is taken as an estimate of current blood glucose concentration. At the same time, a recalibration of Abbott CGM-sensors should sometimes be made several times per day.

On the other hand, it is known (see [8] and references therein, [11]) that the equilibration between blood and ISF glucose is not instantaneous. As a result, CGM devices sometimes give a distorted estimation of blood glucose levels, and as pointed out in [8], further improvements of blood glucose reconstruction require a more sophisticated procedure than the standard calibration by which ISF glucose is determined in CGM systems, such as Abbott Freestyle Navigator.

In this section, we consider how the approach based on the minimization of the functional (12) can be adapted for reading blood glucose levels from subcutaneous electric current measurements.

To illustrate this approach, we use data sets of nine type 1 diabetic subjects studied within the framework of the EU-project 'DIAdvisor' [22] in the Montpellier University Hospital Center (CHU) and the Padova University Hospital (UNIPD). The chosen number of data sets is consistent with earlier research [8, 15], where correspondingly nine and six subjects have been studied.

In each subject, blood glucose concentration and subcutaneous electric current were measured in parallel for three days in hospital conditions. The blood glucose concentration was measured 30 times per day by the HemoCue glucose meter [21]. Blood samples were collected every hour during day, every 2 h during night, every 15 min after meals for 2 h. A specific sampling schedule was adopted after breakfast: 30 min before mealtime, 10, 20, 30, 60, 90, 120, 150, 180, 240, 300 min after. Subcutaneous electric current was measured by the Abbott Freestyle Navigator every 1 min.

For each subject a data set $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^{30}$ has been formed by data collected during the first day. Here $x_i \in [1, 1024]$ are the values of the subcutaneous electric current (ADC

counts), and $y_i \in [0, 450]$ are the corresponding values of blood glucose (BG) concentrations (mg dL^{-1}).

Then for each subject the corresponding data set $\mathbf{z}$ has been used for choosing a kernel from the set (9) in the way described above. For this purpose, the data set $\mathbf{z}$ has been split into two parts, namely $\mathbf{z}_P = \{(x_i, y_i)\}$, $|\mathbf{z}_P| = 4$, is formed by two minimum and two maximum values of x_i; $\mathbf{z}_T = \mathbf{z}\backslash\mathbf{z}_P$. The idea behind such a splitting is that we try to incorporate more data in the construction of the estimator $f_\lambda(\cdot; K, \mathbf{z}_T)$, and that is why $|\mathbf{z}_T| > |\mathbf{z}_P|$, but at the same time, we try to test the ability of the estimator to extrapolate to extreme cases from observed data.

Then the kernel for each subject has been chosen as the approximate minimizer of the functional (12), where $\mu = 0.1$ and $\lambda = \lambda(K)$ is given by the quasi-optimality criterion (6)–(8) with $\lambda_0 = 1.01 \times 10^{-4}$, $q = 1.01$. Moreover, in (12) the functional $P(f, K, \mathbf{z}_P)$ has been adjusted to the considered problem as follows:

$$P(f, K, \mathbf{z}_P) = \frac{1}{|\mathbf{z}_P|} \sum_{i:(x_i, y_i)\in\mathbf{z}_P} |y_i - f(x_i)| \cdot (|r(y_i) - r(f(x_i))| \cdot r(f(x_i))^{-1} + 1), \tag{13}$$

where

$$r(x) = \begin{cases} 100, & \text{if } x < 20\,(\text{mg dL}^{-1}), \\ 10 \cdot (1.509[(\ln(x))^{1.084} - 5.3811])^2, & \text{if } x \geqslant 20\,(\text{mg dL}^{-1}) \wedge x \leqslant 70\,(\text{mg dL}^{-1}), \\ 1, & \text{if } x \geqslant 82\,(\text{mg dL}^{-1}) \wedge x \leqslant 170\,(\text{mg dL}^{-1}), \\ 10 \cdot (1.509[(\ln(x))^{1.084} - 5.3811])^2, & \text{if } x \geqslant 180\,(\text{mg dL}^{-1}) \wedge x \leqslant 600\,(\text{mg dL}^{-1}), \\ 100, & \text{if } x > 600\,(\text{mg dL}^{-1}), \\ \text{linear interpolation}, & \text{otherwise} \end{cases} \tag{14}$$

is a risk function introduced in a similar way as [9] with the idea to penalize heavily the failures/delays in detection of hypoglycemia (BG below 70 mg dL^{-1}) and hyperglycemia (BG above 180 mg dL^{-1}). The minimization of the functional $Q_\mu(K, \lambda(K), \mathbf{z})$ of the form (12), (13) on the set (9) has been performed by a full search over the grid of parameters $\alpha_i = 10^{-4}i$, $\beta_j = 10^{-4}j$, $\gamma_l = 10^{-4}l$, $i, j, l = 1, 2, \ldots, 3 \times 10^4$. Of course, the application of the full search method in finding the minimum of (12), (13) is computationally intensive, but in the present context it can be performed off-line. In the following, we are going to study the possibility of employing other minimization techniques in the context of theorem 1.

For each of the nine subjects, different kernels K have been found to construct a regularized esitimator (4), (5) of the blood glucose concentration that, starting from a raw electric signal $x \in [1, 1024]$, returns a blood glucose concentration $y = f(K, \lambda(K), \mathbf{z}, x)$, where $\lambda = \lambda(K)$ has been chosen from (6) in accordance with the quasi-optimality criterion (8).

To quantify the clinical accuracy of the constructed regularized estimator, we use the original Clarke error grid analysis (EGA) (see [3, 15] and references therein). In accordance with the EGA methodology, for each of the nine subjects the available blood glucose values obtained in the HemoCue meter have been compared with the estimates of the blood glucose $y = f(K, \lambda(K), \mathbf{z}, x)$. Here x is a subcutaneous current value at the moment when the corresponding HemoCue measurement was executed. Since HemoCue measurements made during the first day have been used for constructing $f(K, \lambda(K), \mathbf{z}, x)$, only the data from the other two days (60 HemoCue measurements) have been used as references in Clarke's analysis.

In this analysis each pair (reference value, estimated/predicted value) identifies a point in the Cartesian plane, where the positive quadrant is subdivided into five zones, A to E, of varying degrees of accuracy and inaccuracy of glucose estimations (see figure 3, for example).

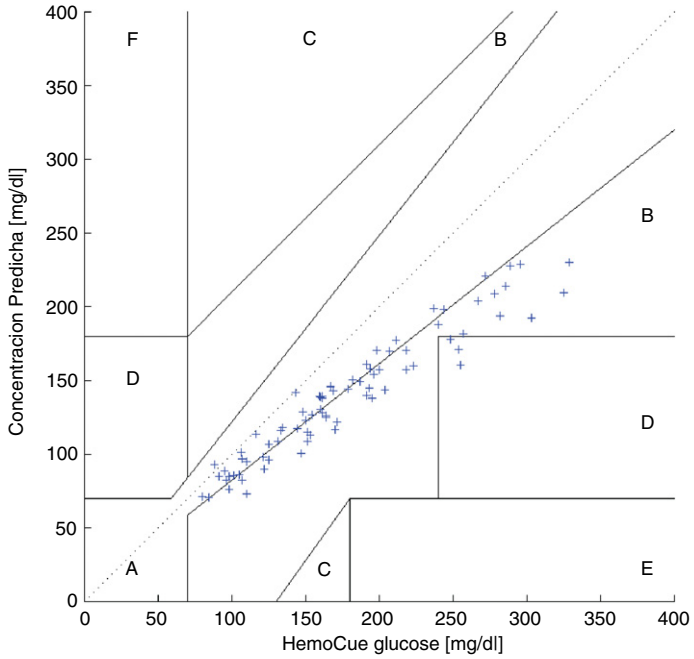

Figure 3. EGA for the Abbott Freestyle Navigator.

Points in zones A and B represent accurate or acceptable glucose estimations. Points in zone C may prompt unnecessary corrections that could lead to a poor outcome. Points in zones D and E represent a dangerous failure to detect and treat. In short, the more points that appear in zones A and B, the more accurate the estimator/predictor is in terms of clinical utility.

A representative Clarke error grid (subject CHUP128) for the proposed regularized blood glucose estimator is shown in figure 4.

Figure 3 illustrates the results of EGA for blood glucose estimations determined from the internal readings of the Abbott Freestyle Navigator calibrated according to the manufacturer's instruction for the same subject and reference values. A comparison shows that the regularized estimator is more accurate, especially if we compare the percentage of data in zone D produced by the proposed estimator and Abbott glucose meter.

The respective kernels, which were chosen by the proposed algorithm, and the results of EGA for all subjects are summarized in table 1. Table 2 presents results of EGA for Abbott Freestyle Navigator readings.

These results allow a conclusion that on average the proposed approach to reading blood glucose levels from the subcutaneous electric current is more accurate than estimations given by the Abbott Freestyle Navigator on the basis of the standard calibration procedure. We would like to stress that no recalibrations of the regularized glucose estimator were made during the two day assessment period. At the same time, as already mentioned, a recalibration of the Abbott Freestyle Navigator should sometimes be made several times per day.

Remark 1. It may be interesting to compare, once again, the proposed approach with the kernel choice rule suggested in [13]. In the present context, the latter one corresponds to the minimization of the functional (12), where $\mu = 1$ and the set $\mathbf{z}_T$ contains the whole data sample $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^{30}$ collected from a patient during the first day (i.e.

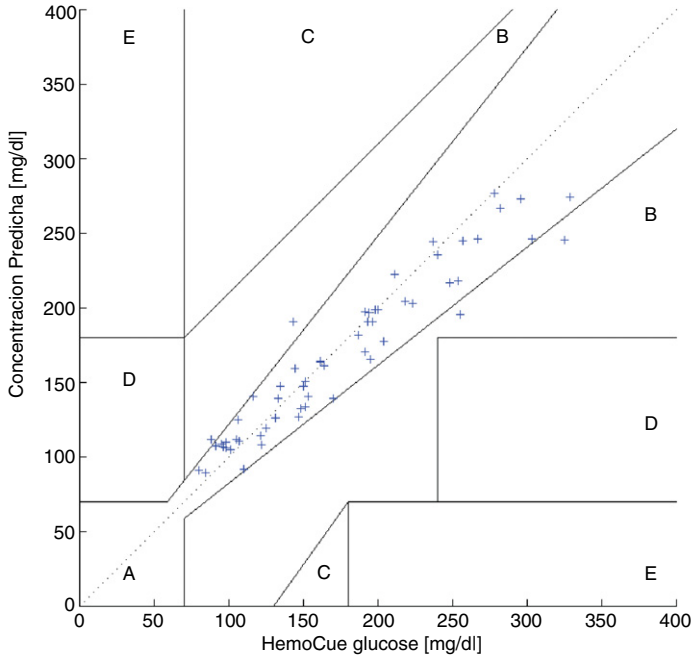

Figure 4. EGA for the regularized estimator.

Table 1. Kernels that have been found for patients in accordance with KAR-algorithm, and percentages of points in EGA-zones for the corresponding estimators.

Subject	Kernel	*A*	*B*	*C*	*D*	*E*
CHU102	$K(x_1,x_2)=(x_1x_2)^{0.6895}+3\exp[-0.0001(x_1-x_2)^2]$	85	15	–	–	–
CHU105	$K(x_1,x_2)=(x_1x_2)^{0.9}+3\exp[-0.0001(x_1-x_2)^2]$	87.34	11.29	–	1.27	–
CHU111	$K(x_1,x_2)=(x_1x_2)^{0.9}+1.7186\exp[-0.0031(x_1-x_2)^2]$	72.50	26.25	–	1.25	–
CHU115	$K(x_1,x_2)=(x_1x_2)^{0.1}+0.2236\exp[-0.0011(x_1-x_2)^2]$	79.49	20.51	–	–	–
CHU116	$K(x_1,x_2)=(x_1x_2)^{0.8765}+0.1674\exp[-0.0007(x_1-x_2)^2]$	97.44	2.56	–	–	–
CHU119	$K(x_1,x_2)=(x_1x_2)^{0.6895}+0.1\exp[-0.0001(x_1-x_2)^2]$	92.40	6.33	–	1.27	–
CHU128	$K(x_1,x_2)=(x_1x_2)^{0.9}+3\exp[-0.009(x_1-x_2)^2]$	87.93	12.07	–	–	–
UNIPD202	$K(x_1,x_2)=(x_1x_2)^{0.9}+3\exp[-0.0031(x_1-x_2)^2]$	75.64	23.08	–	1.28	–
UNIPD203	$K(x_1,x_2)=(x_1x_2)^{0.9}+3\exp[-0.007(x_1-x_2)^2]$	78.05	20.73	–	1.22	–
Average		83.69	15.61	–	0.7	–

$\mathbf{z}_P = \varnothing$). It turns out that for all considered patient data sets $\mathbf{z} = \mathbf{z}_T$, the same kernel $K(x_1,x_2) = (x_1x_2)^{0.9} + 3\exp[-0.009(x_1-x_2)^2]$ was found as the minimizer of the corresponding functionals (12) over the set (9). Table 3 presents results of EGA for the regularized blood glucose estimators $f(K,\lambda(K),\mathbf{z},x)$ based on this kernel. Comparing tables 2 and 3, we can conclude that these estimators perform at the level of the Abbott Freestyle Navigator (except they have more erroneous estimations in zones C and E than the commercial devices). At the same time, the comparison of tables 1–3 shows that on average the estimators based on the proposed KAR-algorithm are more accurate than the others considered here.

Table 2. Percentage of points in EGA-zones for the Abbott Freestyle Navigator.

Subject	*A*	*B*	*C*	*D*	*E*
CHU102	93.83	6.17	–	–	–
CHU105	92.5	5	–	2.5	–
CHU111	85.9	12.82	–	1.28	–
CHU115	94.81	5.19	–	–	–
CHU116	86.84	10.53	–	2.63	–
CHU119	83.54	16.46	–	–	–
CHU128	48.98	44.9	–	6.12	–
UNIPD202	89.19	8.11	–	2.7	–
UNIPD203	76	21.33	–	2.67	–
Average	83.51	14.5	–	1.99	–

Table 3. Percentages of points in EGA-zones for the regularized blood glucose estimators based on the kernel, chosen in accordance with [13].

Subject	*A*	*B*	*C*	*D*	*E*
CHU102	76.82	21.95	–	–	1.22
CHU105	79.82	18.81	–	1.37	–
CHU111	82.50	15	1.25	1.25	–
CHU115	83.33	16.67	–	–	–
CHU116	92.30	7.70	–	–	–
CHU119	77.22	22.78	–	–	–
CHU128	87.93	12.07	–	–	–
UNIPD202	67.95	29.49	–	2.56	–
UNIPD203	79.26	15.85	3.66	1.22	–
Average	80.79	17.81	0.55	0.71	0.14

Acknowledgments

The authors thank the anonymous reviewers for useful comments and a careful review of the paper. This research has been performed in the course of the project 'DIAdvisor' funded by the European Commission within the 7th Framework Programme. The authors gratefully acknowledge the support of the 'DIAdvisor' – consortium.

References

[1] Aronszajn N 1950 Theory of reproducing kernels *Trans. Am. Math. Soc.* **68** 337–404
[2] Bauer F, Pereverzev S V and Rosasco L 2007 On regularization algorithms in learning theory *J. Complexity* **23** 52–72
[3] Clarke W, Cox D, Gonder Frederick L, Carter and Pohl S 1987 Evaluating clinical accuracy of systems for self-monitoring of blood glucose *J. Diabetes Care* **10** 622–28
[4] Cucker F and Smale S 2001 On the mathematical foundations of learning *Bull. Am. Math. Soc.* **39** 1–49
[5] De Vito E, Pereverzev S V and Rosasco L 2010 Adaptive kernel methods using the balancing principle *Found. Comput. Math.* **10** 455–79
[6] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[7] Evgeniou T, Pontil M and Poggio T 2000 Regularization networks and support vector machines *Adv. Comput. Math.* **13** 1–50

[8] Facchinetti A, Sparacino G and Cobelli C 2007 Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: role of sensor calibration *J. Diabetes Sci. Technol.* **1** 617–23
[9] Guerra S, Facchinetti A, Schiavon M, Dalla Man C and Sparacino G 2010 A new dynamic risk measure for continuous glucose monitoring time series *3rd Int. Conf. on Advanced Technologies and Treatments for Diabetes (ATTD) (10–13 February 2010)* (Birkhauser: Basel)
[10] Kindermann S and Neubauer A 2008 On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization *Inverse Problems Imaging* **2** 291–9
[11] Kovatchev B, Anderson S, Heinemann L and Clarke W 2008 Comparison of the numerical and clinical accuracy of four continuous glucose monitors *J. Diabetes Care* **31** 1160–4
[12] Lanckriet G R G, Christianini N, Ghaoui L E, Bartlett P and Jordan M I 2004 Learning the kernel matrix with semidefinite programming *J. Mach. Learn. Res.* **5** 27–72
[13] Micchelli C A and Pontil M 2005 Learning the kernel function via regularization *J. Machine Learning Res.* **6** 1099–125
[14] Micchelli C A, Xu Y and Zhang H 2006 Universal Kernels *J. Machine Learning Res.* **7** 2651–67
[15] Reifman J, Rajaraman S, Gribok A and Ward W K 2007 Predictive monitoring for improved management of glucose levels *J. Diabetes Sci. Technol.* **1** 478–86
[16] Tikhonov A N and Glasko V B 1965 Use of the regularization methods in non-linear problems *USSR Comput. Math. Math. Phys.* **5** 93–107
[17] Wahba G *Spline Models for Observational Data* (*Series in Applied Mathematics* vol 59) (Philadelphia: SIAM)
[18] Wang H 2008 Adaptive regularization algorithms in learning theory *Master Thesis* Johannes Kepler University Linz
[19] Zhou D-X 2003 Capacity of reproducing kernel spaces in learning theory *IEEE Trans. Inf. Theory* **49** 1743–52
[20] http://www.freestylenavigator.com
[21] http://www.hemocue.com
[22] http://www.diadvisor.eu