

Extrapolation in variable RKHSs with application to the blood glucose reading

V. Naumova, S. Pereverzyev, S. Sampath

RICAM-Report 2011-04

EXTRAPOLATION IN VARIABLE RKHSs WITH APPLICATION TO THE BLOOD GLUCOSE READING

VALERIYA NAUMOVA, SERGEI V. PEREVERZYEV AND SIVANANTHAN SAMPATH

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 regularization procedure. The construction of such fully adaptive regularization algorithm is motivated by the problem of reading the blood glucose concentration of diabetic patients. We describe how proposed scheme can be used for this purpose and report the results of numerical experiments with real clinical data.

1. Introduction

In this paper we consider the problem of a reconstruction of a real-valued function $f : X \rightarrow \mathbb{R}$, $X \subset \mathbb{R}^d$, 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 $x \in \overline{co\{x_i\}}$. It is sometimes called as 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

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

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 [21].

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 [20, 6, 5, 13].

Another issue which needs to be addressed is the one of the choice of a space, which norm is used for a penalization. The impact of the chosen norm is most explicit in Tikhonov

regularization [7]. 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 there are still no general principles to advise a choice and only in a few papers [21, 17, 15, 6] 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 [17], but, as it will be seen from our discussion below, the kernel choice suggested in [17] 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 towards extrapolation.

The article is organized in the following way. 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 a possibility to use the proposed approach in diabetes therapy management, in particular, for reading blood glucose level 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 Reproducing Kernel Hilbert space \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 $K : X \times X \rightarrow \mathbb{R}$, 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 there exists a uniquely determined Hilbert space $\mathcal{H} = \mathcal{H}_K$ 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 Reproducing Kernel Hilbert spaces can be found in [1].

In the sequel we will deal with the situation when X is a compact in \mathbb{R}^d . Then, it is known [23] 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 $J_K : \mathcal{H}_K \rightarrow L_2(X)$ 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

$$J_K f = f_\xi, \quad (2)$$

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

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

$$S_{\mathbf{x}} f = \mathbf{y}, \quad (3)$$

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

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

If a kernel K has been already chosen, several methods of the general regularization theory can be applied to the equation (3), as it 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

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

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

From the Representer Theorem [21] it follows that the minimizer of (4) has the form

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

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

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

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 $\vec{y} = (y_1, y_2, \dots, y_{|\mathbf{z}|})$.

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 [20], and as it has been proven recently in [13], 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

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

Then one needs to calculate the norm

$$\begin{aligned} \sigma_{\mathcal{H}_K}^2(s) &= \|f_{\lambda_s}(x; K, \mathbf{z}) - f_{\lambda_{s-1}}(x; K, \mathbf{z})\|_K^2 \\ &:= \sum_{i=1}^{|\mathbf{z}|} \sum_{j=1}^{|\mathbf{z}|} (c_i^{\lambda_s} - c_i^{\lambda_{s-1}})(c_j^{\lambda_s} - c_j^{\lambda_{s-1}}) K(x_i, x_j), \end{aligned} \quad (7)$$

and find

$$\lambda_+ = \lambda_p : p = \arg \min \{\sigma_{\mathcal{H}_K}^2(s), s = \overline{0, \nu}\}. \quad (8)$$

2.3. The choice of the kernel from a parameterized set

But 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 [9], [5]. On the other hand, in practice we would like to have multiple kernels or parameterizations of kernels to scale well the solution to 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 problem, 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 [15] 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 worthy of notice 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 also Ω be a compact metric space and $G : \Omega \rightarrow \mathcal{K}(X)$ be an injection such that for any $x_1, x_2 \in X, w \in \Omega$ the function $w \rightarrow G(w)(x_1, x_2)$ 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 a 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 non-linear manifold.

Example

Consider $\Omega = [a, b]^3, 0 < a < b, (\alpha, \beta, \gamma) \in [a, b]^3$, and define a mapping $G : (\alpha, \beta, \gamma) \rightarrow (x_1 x_2)^\alpha + \beta e^{-\gamma(x_1 - x_2)^2}$, 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 $x_1, x_2 \in X \subset (0, \infty)$ the value $G(\alpha, \beta, \gamma)(x_1, x_2)$ continuously depends on (α, β, γ) .

Thus, kernels from the set

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

are parameterized by points of $\Omega = [a, b]^3$ in the sense described above. Once the set of kernels is fixed, one may follow [17] and select a kernel by minimizing the Tikhonov regularization functional (4) such that

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

Thus, the idea of [17] 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 $f_\lambda(\cdot; K)$.

To illustrate that such an approach may fail in a prediction of the extrapolation type we use the same example as in [17], where $f(x) = 0.1(x + 2(e^{-8(\frac{4\pi}{3}-x)^2} - e^{-8(\frac{\pi}{2}-x)^2} - e^{-8(\frac{3\pi}{2}-x)^2}))$, and the given set $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^{14}$ consists of points $x_i = \frac{\pi i}{10}, i = 1, 2, \dots, 14$ 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 [17] the target function f has been chosen in such a way that it belongs to the RKHS generated by the “ideal” kernel $K_{opt}(x_1, x_2) = x_1x_2 + e^{-8(x_1-x_2)^2}$. The performance of the approximant $f_\lambda(x; K_{opt}, \mathbf{z})$ with the best λ is shown in the 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 well-estimated by $f_\lambda(x; K_{opt}, \mathbf{z})$, while for $x \notin \overline{co\{x_i\}}$ the performance of the approximant based on the K_{opt} is rather poor. Observe that the 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 $K_{opt}(x_1, x_2) = x_1x_2 + e^{-8(x_1-x_2)^2}$ used to generate the target function f .

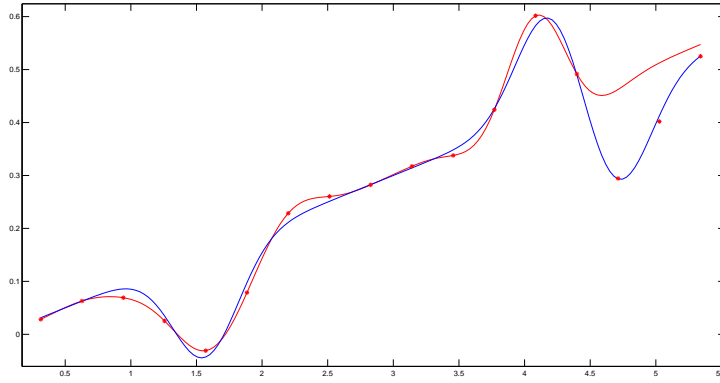


Figure 1: The performance of the approximant $f_\lambda(x; K, \mathbf{z})$ (red line) based on the kernel $K(x_1, x_2) = x_1x_2 + e^{-8(x_1-x_2)^2}$ generating the target function f .

The Figure 2 displays the performance of the approximation (5) constructed for the same data set \mathbf{z} , but with the use of the kernel $K(x_1, x_2) = (x_1x_2)^{1.9} + e^{-2.7(x_1-x_2)^2}$. As one can see, the approximation based on this kernel performs much better compared to the 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

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

as follows.

As 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

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

The approximant displayed on the 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

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

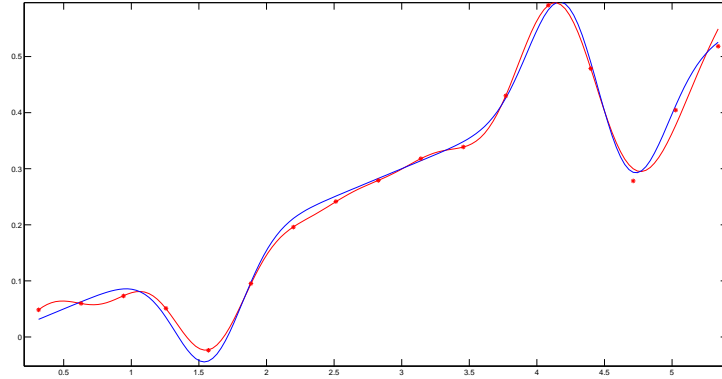


Figure 2: The performance of the approximant $f_\lambda(x; K, \mathbf{z})$ (red line) based on the kernel $K(x_1, x_2) = (x_1 x_2)^{1.9} + e^{-2.7(x_1 - x_2)^2}$, that has been chosen as a minimizer of (12).

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 worth to note 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

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

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

To construct approximant displayed on the 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

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

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

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

Since, by construction, all $\lambda(K_l) \in [\lambda_{min}, \lambda_{max}]$, one can find a subsequence $\lambda_n = \lambda(K_{l_n}), n = 1, 2, \dots$ such that $\lambda_n \rightarrow \lambda_0 \in [\lambda_{min}, \lambda_{max}]$.

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

Let also

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

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

$$\vec{c}_n = (\lambda_n |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^n)^{-1} \vec{y}_T,$$

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 \vec{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 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 $w \rightarrow G(w)(x_1, x_2)$ 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 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 $\epsilon > 0$ there exists a natural number $k = k(\epsilon)$ 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

$$\vec{c}_0 = (\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)^{-1} \vec{y}_T,$$

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 $K^{n_k}, K^0 \in \mathcal{K}(\Omega, G)$ corresponding matrices $\mathbb{K}_T^{n_k}, \mathbb{K}_T^0$ are positive definite, for any vector $\vec{y} \in \mathbb{R}^{|\mathbf{z}_T|}$ we have

$$\begin{aligned} \|(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k})^{-1} \vec{y}\| &\leq (\lambda_{n_k} |\mathbf{z}_T|)^{-1} \|\vec{y}\|, \\ \|(\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)^{-1} \vec{y}\| &\leq (\lambda_0 |\mathbf{z}_T|)^{-1} \|\vec{y}\|. \end{aligned}$$

Therefore,

$$\begin{aligned} \|\vec{c}_0 - \vec{c}_{n_k}\| &= \|(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k})^{-1} ((\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k}) - (\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)) (\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)^{-1} \vec{y}_T\| \\ &= \|(\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k})^{-1} (\mathbb{K}_T^{n_k} - \mathbb{K}_T^0) (\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)^{-1} \vec{y}_T + \\ &\quad + (\lambda_{n_k} |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^{n_k})^{-1} (\lambda_{n_k} - \lambda_0) |\mathbf{z}_T| (\lambda_0 |\mathbf{z}_T| \mathbb{I} + \mathbb{K}_T^0)^{-1} \vec{y}_T\| \\ &\leq (\lambda_{min} |\mathbf{z}_T|)^{-2} \|\vec{y}_T\| \|\mathbb{K}_T^{n_k} - \mathbb{K}_T^0\| + (\lambda_{min} |\mathbf{z}_T|)^{-2} \|\vec{y}_T\| |\lambda_{n_k} - \lambda_0| |\mathbf{z}_T|, \end{aligned}$$

and in view of our observation that $\lambda_{n_k} \rightarrow \lambda_0$ and $\mathbb{K}_T^{n_k} \rightarrow \mathbb{K}_T^0$, we can conclude that $\vec{c}_{n_k} \rightarrow \vec{c}_0$ in $\mathbb{R}^{|\mathbf{z}_T|}$.

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

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

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

$$Q_\mu(K^0, \lambda_0, \mathbf{z}) = \lim_{k \rightarrow \infty} Q_\mu(K^{n_k}, \lambda_{n_k}, \mathbf{z}) = \inf\{Q_\mu(K, \lambda(K), \mathbf{z}), K \in \mathcal{K}(\Omega, G)\}.$$

3. Reading blood glucose level from subcutaneous electric current measurements

In this section, we discuss a possibility to use the approach described in the previous section in diabetes therapy management, in particular, for reading blood glucose level from subcutaneous electric current measurements.

Continuous Glucose Monitoring (CGM) systems provide almost in real-time indirect estimation of current blood glucose that is highly valuable for insulin therapy of diabetes. For example, needle based electrochemical sensors, such as Abbott Freestyle Navigator [24], measure electrical signal in the interstitial fluid (ISF) and return ISF glucose concentration (mg/dL) 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 [10] and references therein, [14]) that the equilibration between blood and ISF glucose is not instantaneous. As a result, CGM devices sometimes give a distorted estimation of blood glucose level, and as it has been pointed out in [10], further improvements of blood glucose reconstruction require 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 level from subcutaneous electric current measurements.

To illustrate this approach we use data sets of nine type 1 diabetic subjects studied within the framework of EU-project ‘‘DIAdvisor’’ [26] in the Montpellier University Hospital Center (CHU) and in the Padova University Hospital (UNIPD). The chosen number of data sets is consistent with earlier research [18], [10], where correspondingly 9 and 6 subjects have been studied.

In each subject, blood glucose concentration and subcutaneous electric current were measured in parallel for 3 days in hospital conditions. The blood glucose concentration was measured 30 times per day by the HemoCue glucose meter [25]. Blood samples were collected every hour during day, every 2 hours during night, every 15 minutes after meals for 2 hours. Specific sampling schedule was adopted after breakfast: 30 minutes before, mealtime, 10, 20, 30, 60, 90, 120, 150, 180, 240, 300 minutes after. Subcutaneous electric current was measured by the Abbott Freestyle Navigator every 1 minute.

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 subcutaneous electric current (ADC counts), and $y_i \in [0, 450]$ are corresponding values of blood glucose (BG) concentrations (mg/dL).

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

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

where

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

is a risk function introduced similar to [11] with the idea to penalize heavily the failures/delays in detection of hypoglycemia (BG below 70 mg/dL) and hyperglycemia (BG above 180 mg/dL). 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 full search over the grid of parameters $\alpha_i = 10^{-4}i$, $\beta_j = 10^{-4}j$, $\gamma_l = 10^{-4}l$, $i, j, l = 1, 2, \dots, 3 \cdot 10^4$. 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 sequel we are going to study the possibility to employ other minimization techniques in the context of the Theorem 1.

For each of the 9 subjects different kernels K have been found to construct a regularized estimator (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 clinical accuracy of constructed regularized estimator we use the original Clarke Error Grid Analysis (EGA) (see [4], [18] and references therein). In accordance with the EGA methodology, for each 9 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 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 2 days (60 HemoCue measurements) have been used as references in the 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). 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 represents 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 proposed regularized blood

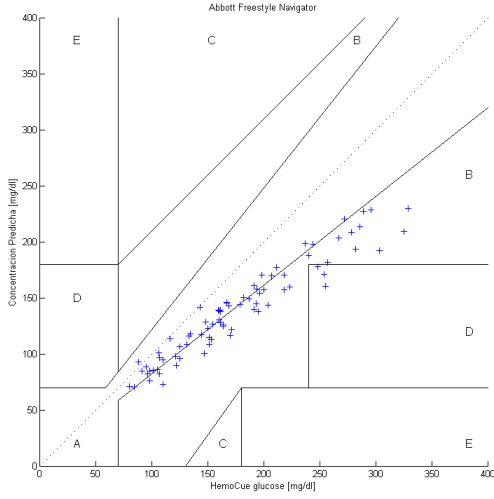


Figure 3

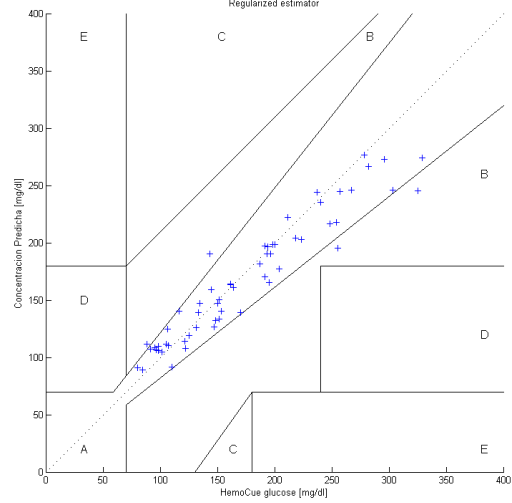


Figure 4

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

| Subject | Kernel | A | B | C | D | E |
|----------|---|-------|-------|---|------|---|
| CHU102 | $K(x_1, x_2) = (x_1 x_2)^{0.6895} + 3e^{-0.0001(x_1 - x_2)^2}$ | 85 | 15 | — | — | — |
| CHU105 | $K(x_1, x_2) = (x_1 x_2)^{0.9} + 3e^{-0.0001(x_1 - x_2)^2}$ | 87.34 | 11.29 | — | 1.27 | — |
| CHU111 | $K(x_1, x_2) = (x_1 x_2)^{0.9} + 1.7186e^{-0.0031(x_1 - x_2)^2}$ | 72.5 | 26.25 | — | 1.25 | — |
| CHU115 | $K(x_1, x_2) = (x_1 x_2)^{0.1} + 0.2236e^{-0.0011(x_1 - x_2)^2}$ | 79.49 | 20.51 | — | — | — |
| CHU116 | $K(x_1, x_2) = (x_1 x_2)^{0.8765} + 0.1674e^{-0.0007(x_1 - x_2)^2}$ | 97.44 | 2.56 | — | — | — |
| CHU119 | $K(x_1, x_2) = (x_1 x_2)^{0.6895} + 0.1e^{-0.0001(x_1 - x_2)^2}$ | 92.40 | 6.33 | — | 1.27 | — |
| CHU128 | $K(x_1, x_2) = (x_1 x_2)^{0.9} + 3e^{-0.009(x_1 - x_2)^2}$ | 87.93 | 12.07 | — | — | — |
| UNIPD202 | $K(x_1, x_2) = (x_1 x_2)^{0.9} + 3e^{-0.0031(x_1 - x_2)^2}$ | 75.64 | 23.08 | — | 1.28 | — |
| UNIPD203 | $K(x_1, x_2) = (x_1 x_2)^{0.9} + 3e^{-0.007(x_1 - x_2)^2}$ | 78.05 | 20.73 | — | 1.22 | — |
| Average | | 83.69 | 15.61 | — | 0.7 | — |

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. Comparison shows that 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.

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

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

These results allow a conclusion that in average proposed approach to reading blood glucose level from 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 regularized glucose estimator were made during 2 days assessment period. At the same time, as it was already mentioned, a recalibration of the Abbott Freestyle Navigator should sometimes be made several times per day.

Acknowledgements. This research has been performed in the course of the project “DIAdvisor” funded by the European Commission within 7-th Framework Programme. The authors gratefully acknowledge the support of the DIAdvisor - consortium.

References

- [1] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., **68** (1950), pp. 337-404.
- [2] F. Bauer, S. V. Pereverzev, and L. Rosasco, *On regularization algorithms in learning theory*, J. of Complexity, **23** (2007), pp. 52-72.
- [3] A. Caponnetto and Y. Yao, *Adaptation for Regularization Operators in Learning Theory*, Technical Report CBCL Paper 265, MIT-CSAIL-TR 2006-063.
- [4] W. L. Clarke, D. J. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, *Evaluating clinical accuracy of systems for self-monitoring of blood glucose*, Diabetes Care, **10** (1987), Issue 5, pp. 622-628.
- [5] F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bull. Amer. Math. Soc., **39**, (2001), pp. 149.
- [6] E. De Vito, S. V. Pereverzev and L. Rosasco, *Adaptive Kernel Methods Using the Balancing Principle*, Found. Comput. Math., **10**, (2010), pp. 455-479.
- [7] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 1996.
- [8] M. Eren-Oruklu, A. Cinar, L. Quinn and D. Smith, *Estimation of future glucose concentrations with subject-specific recursive linear models*, Diabetes Tech. & Therapeutics, **11** (2009), Issue 4, pp. 243-253.
- [9] T. Evgeniou, M. Pontil and T. Poggio, *Regularization Networks and Support Vector Machines*, Advances in Computational Mathematics, **13** (2000), pp. 1-50.
- [10] A. Facchinetti, G. Sparacino and C. Cobelli, *Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: Role of sensor calibration*, J. of Diabetes Sci. and Tech., **1** (2007), Issue 5, pp. 617-623.

- [11] S. Guerra, A. Facchinetti, M. Schiavon, C. Dalla Man, G. Sparacino, *New Dynamic Glucose Risk Function for Continuous Glucose Monitoring Time Series*, 3rd International Conference on Advanced Technologies & Treatments for Diabetes, Basel, 10-13 February 2010.
- [12] J. Kaipio, E. Somersalo, *Statistical inverse problems: Discretization, model reduction and inverse crimes*, J. of computational and Applied Math., **198** (2007), pp. 493-504.
- [13] S. Kindermann, A. Neubauer, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Problems and Imaging, **2**, 2008, Issue 2, pp. 291-299.
- [14] B. Kovatchev, S. Anderson, L. Heinemann, W. Clarke *Comparison of the numerical and clinical accuracy of four continuous glucose monitors*, J. of Diabetes Care, **31** (2008), Issue 6, pp. 1160-1164.
- [15] G. R. G. Lanckriet, N. Christianini, L. E. Ghaoui, P. Bartlett and M. I. Jordan, *Learning the kernel matrix with semidefinite programming*, J. Mach. Learn. Res., **5** (2004), pp. 27-72.
- [16] C. A. Micchelli, *Interpolation of scattered data: Distance matrices and conditionally positive functions*, Constr. Approxiam., **2** (1986), pp. 11-22.
- [17] C. A. Micchelli and M. Pontil, *Learning the kernel function via regularization*, J. of Machine Learning Research, **6** (2005), pp. 1099-1125.
- [18] J. Reifman, S. Rajaraman, A. Gribok and W. K. Ward, *Predictive monitoring for improved management of glucose levels*, J. of Diabetes Sci. and Tech., **1** (2007), Issue 4, pp. 478-486.
- [19] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti and C. Cobelli, *Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series*, IEEE Trans. on Biomedical Eng., **54** (2007), Issue 5, pp. 931-937.
- [20] A. N. Tikhonov and V. B. Glasko, *Use of the regularization methods in non-linear problems*, USSR Comput. Math. Math. Phys., **5** (1965), pp. 93-107.
- [21] G. Wahba, *Splines models for observational data*, Series in Applied Mathematics, **59**, SIAM, Philadelphia, 1990.
- [22] H. Wang, *Adaptive regularization algorithms in learning theory*, Master thesis, Johannes Kepler University Linz, 2008.
- [23] D.-X. Zhou, *Capacity of Reproducing Kernel Spaces in Learning Theory*, IEEE Transactions on Information Theory, **49** (2003), Issue 7, pp. 1743-1752.
- [24] <http://www.freestylenavigator.com>
- [25] <http://www.hemocue.com>
- [26] <http://www.diadvisor.eu>