Minimization Problems with Fourier Based Stabilizers

Šidlofová, Terezie
2004

Institute of Computer Science
Academy of Sciences of the Czech Republic

Minimization Problems with Fourier based Stabilizers

Šidlofová Terezie

Technical report No. 904

February 2004

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 051 111, fax: +420 286 585 789,
e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Minimization Problems with Fourier based Stabilizers[1]

Šidlofová Terezie[2]

Technical report No. 904

February 2004

Abstract:

We express the problem of approximating a data set $z = \{(x_i, y_i); i = 1, \ldots, N\} \subseteq \mathbb{R}^d \times \mathbb{R}$ in the form of minimizing a functional that composes of an empirical error part and a Fourier-based stabilizer. We prove existence and uniqueness of the solution. We also describe the shape of the minimizing function and show that it is in the form of a one-hidden layer feed-forward neural network with activation functions derived from the regularization part.

Keywords:
Neural networks, minimization of functionals, regularization theory, stabilizers, Fourier transform

[1] Thanks to my advisor Věra Kůrková for useful guidance. This work was partially supported by GA ČR Grant 201/02/0428.

[2] Institute of Computer Science, Academy of Sciences of CR, Pod vodárenskou věží 2, P.0. Box 5, 182 07 Prague 8, Czech Republic
Charles University, Faculty of Mathematics and Physics, Ke Karlovu 3, 121 16 Prague 2, Czech Republic, terka@cs.cas.cz

1 Introduction

Learning from data usually means to fit a function to a set of data $z = \{(x_i, y_i); i = 1, \ldots, N\} \subseteq \mathbb{R}^d \times \mathbb{R}$. The problem is what type of functions will we use for the fitting, because there are infinitely many ways to go through the given points. And even if we have a reasonable set of functions (admissible set) to pick from, there is no guarantee that the problem will have a solution and that the solution will be unique.

Typically it is not necessary that the function fits the data exactly, we approximate. Thus nice functions (smooth, continuous) come into question and we also gain generalization (see [GiJoPo95]). Some of these properties are easily expressed by the set of admissible functions, but we might have more complicated (global) external information (a-priori knowledge) about the problem and want to add it, too.

Mathematical expression of these ideas lies in formulating a functional that would among admissible functions pick the one, that is reasonably close to the data and also agrees with global property assumptions ([CuSm01], [Gi97], [PoSm03], [SchSm02], [Wa90]). Existence and uniqueness of such a solution can be secured by minimizing a functional over a corresponding set of functions.

The article deals with a stabilizer based on Fourier transform proposed in [GiJoPo95]. In [Gi97] it was stated that our problem is closely connected with Reproducing Kernel Hilbert Spaces. Taking advantage of these ideas we present construction of the admissible set (RKHS) and derive existence, uniqueness and the form of the solution of our minimization problem.

2 Preliminaries

A real or complex *Banach space* $(X, \|.\|)$ is a vector space over real or complex numbers which is complete in the topology generated by the norm $\|.\|$, defined on X. A *Hilbert space* is a Banach space in which the norm is given by an inner product $\langle ., . \rangle$, that is $\|x\| = \langle x, x \rangle^{1/2}$. Sequences of elements of spaces are denoted by $\{x_n\}$ meaning $n \in \mathbb{N}_+$, where $\mathbb{N}_+$ is the set of positive integers.

The Banach space X^* of bounded (real-valued) linear functionals on X is called the dual space. It defines *weak convergence* on X. A sequence $\{x_n\} \in X$ converges weakly to x ($x_n \rightharpoonup x$) if and only if $\lim_{n\to\infty} |f(x_n) - f(x)| = 0$ for each fixed $f \in X^*$. Let X, Y be Banach spaces and $\mathcal{F} : X \to Y$ a mapping from X to Y. We define the Gateaux derivative of $\mathcal{F}$ in f in direction h as $\mathrm{D}_f\mathcal{F}(h) = \lim_{t\to 0} \frac{\mathcal{F}(f+th)-\mathcal{F}(f)}{t}$. If the limit is uniform in h, we call the derivative the Fréchet derivative (which will be our case). We can analogously define the second and so on derivatives.

Let d, k be positive integers, $\Omega \subseteq \mathbb{R}^d$. We denote by $(C(\Omega), \|.\|_C)$ the space of continuous functions on Ω with the maximum norm. C_k will denote all functions with continuous Fréchet derivative up to order k. We say that an infinitely differentiable function $f \in C_\infty$ belongs to the *Schwartz space* $\mathcal{S}(\mathbb{R}^n)$ if $pD^\alpha f$ is a bounded function for any multiindex α and any polynomial p on $\mathbb{R}^n$, (where $pD^\alpha f = \sum c_\alpha D^\alpha(f)$ and $D^\alpha(f) = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \ldots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$). For the sake of this article let us define the *normalized Lebesgue measure* m_d on $\mathbb{R}^d$ as $\mathrm{d}m_d(x) = (2\pi)^{-d/2}dx$ (see [Ru91]). The Lebesgue space $(\mathcal{L}_p(\Omega), \|.\|_p)$ of p-times integrable functions on Ω will be renormed: $\|f\|_p = \left\{\int_\Omega |f|^p \mathrm{d}m_d\right\}^{1/p}$. This will simplify the use of *Fourier transform* $\hat{f}$ of the function $f \in \mathcal{L}^1(\mathbb{R}^d)$: $\hat{f}(t) = \int_{\mathbb{R}^d} f(x)e^{-it.x}\mathrm{d}m_d$, where $t \in \mathbb{R}^d$ and $t.x = t_1x_1 + \cdots + t_dx_d$.

Let $\mathcal{F}$ be a functional $\mathcal{F} : X \to (-\infty, +\infty]$. The set $\mathrm{dom}\,\mathcal{F} = \{f \in X : \mathcal{F}(f) < +\infty\}$ is called the *domain* of $\mathcal{F}$. *Continuity* of $\mathcal{F}$ in $f \in \mathrm{dom}\,\mathcal{F}$ is defined as usual. A functional is *sequentially lower semicontinuous* if and only if the convergence of $\{f_n\}$ to f implies $\mathcal{F}(f) \le \liminf_{n\to\infty} \mathcal{F}(f_n)$. Functional $\mathcal{F}$ is *weakly sequentially lower semicontinuous* if and only if $f_n \rightharpoonup f$ implies $\mathcal{F}(f) \le \liminf_{n\to\infty} \mathcal{F}(f_n)$.

A functional $\mathcal{F}$ is *convex* on a convex set $E \subseteq \mathrm{dom}\,\mathcal{F}$ if for all $f, g \in E$ and all $\lambda \in [0, 1]$, $\mathcal{F}(\lambda f + (1-\lambda)g) \le \lambda\mathcal{F}(f) + (1-\lambda)\mathcal{F}(g)$. Functional $\mathcal{F}$ is *(strongly) quasi-convex* if for all $f, g \in E$, $f \neq g$ it holds: $\mathcal{F}\left(\frac{1}{2}f + \frac{1}{2}g\right)\ (<) \le \max\{\mathcal{F}(f), \mathcal{F}(g)\}$. Set E is *weakly sequentially compact* if any sequence in E has a weakly converging subsequence.

A symmetric real valued function $K(x, y)$ on X is *(strictly) positive definite* if for any $a_1, \ldots a_d, a_i \in$

$\mathbb{R}, \exists a_i \neq 0$, and $x_1, \dots x_d \in X$ the sum $\sum_{i,j=1}^{d} a_i a_j K(x_i, x_j)(>) \geq 0$. A *Reproducing Kernel Hilbert Space* (RKHS) $\mathcal{H}(X)$ is a Hilbert space of functions $f : X \to \mathbb{R}$ (X is a nonempty set), where for all $x \in X$ the evaluation functionals $\mathcal{F}_x : f \mapsto f(x)$, are linear and bounded (i.e. continuous). Thus by Fréchet-Riesz Theorem [Lu02, p. 19] we can define a unique *kernel* $K(.,.)$ corresponding to our RKHS as follows: $\mathcal{F}_x(f) = f(x) = \langle K(x,.), f(.)\rangle \quad \forall f \in \mathcal{H}$, ($\langle .,. \rangle$ is scalar product on $\mathcal{H}$). K is symmetric positive definite and defines a dot product (and norm) on $\mathcal{H}$. For any positive definite symmetric K we can construct an RKHS with K as a kernel ([Wa90, p.1-3]).

3 Regularized Empirical Error Functional

The task to find an optimal solution to the setting of approximating a data set $z = \{(x_i, y_i)\}_{i=1}^N \subseteq \mathbb{R}^d \times \mathbb{R}$ by a function from a general function space X is ill-posed. A standard method to cope with ill-posed problems is to impose additional (regularization) conditions on the solution ([GiJoPo95]). These are typically things like a-priori knowledge, or some smoothness constraints. The solution f_0 has to minimize a functional $\mathcal{F} : E \to \mathbb{R}$ that is composed of the error part and the "smoothness" part:

$$\mathcal{F}(f) = \mathcal{E}_z(f) + \gamma\Phi(f),$$

where $\mathcal{E}_z$ is the error functional depending on the data $z = \{(x_i, y_i)\}_{i=1}^N \subseteq \mathbb{R}^d \times \mathbb{R}$, Φ is the regularization part — the so called stabilizer and γ is the regularization parameter giving the trade-off between the two terms of the functional to be minimized.

An error functional is usually of the form $\mathcal{E}_z(f) = \sum_{i=1}^N V(f(x_i), y_i)$. A typical example of the empirical error functional is the classical mean square error:

$$\mathcal{E}_z(f) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2.$$

In [GiJoPo95] a special stabilizer based on the Fourier Transform was proposed:

$$\Phi_G(f) = \int_{\mathbb{R}^d} \frac{|\hat{f}(s)|^2}{\hat{G}(s)} \mathrm{d}m_d(s),$$

where $\hat{G}$ is some symmetric positive function tending to zero as $\|s\| \to \infty$ (the last holds for any $G \in \mathcal{L}_1$). That means $1/\hat{G}$ is a high-pass filter.

Now we can define the functional $\mathcal{F}_G$ that is to be minimized:

$$\mathcal{F}_G(f) = \mathcal{E}_z(f) + \Phi_G(f) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2 + \gamma \int_{\mathbb{R}^d} \frac{|\hat{f}(s)|^2}{\hat{G}(s)} \mathrm{d}m_d(s).$$

4 Existence and Uniqueness of the Solution

To minimize the functional $\mathcal{F}_G$ above we need to specify the set X (of admissible functions) over which we are minimizing and thus construct a minimization problem $(\mathcal{F}_G, X)$. We will build a special set of admissible functions $\mathcal{H}$ (RKHS) and obtain existence and uniqueness of solution to the minimization problem $(\mathcal{F}_G, \mathcal{H})$.

Let us first suppose existence and show uniqueness. For this purpose we will employ Reproducing Kernel Hilbert Spaces. We build an RKHS corresponding to the regularization part of our functional (so far the only conditions on G were $G \in \mathcal{L}_1, \hat{G}$ symmetric, positive):

Let us define

$$G^\dagger(x, y) = G(x - y) = \int_{\mathbb{R}^d} \hat{G}(t) e^{it.x} e^{-it.y} \mathrm{d}m_d(t).$$

For $G^\dagger \in \mathcal{S}(\mathbb{R}^{2d})$ symmetric positive definite we obtain an RKHS $\mathcal{H}$ (using the classical construction, see [Gi97], [SchSm02],[Wa90]). $\mathcal{H} = \overline{\mathrm{span}\{G^\dagger(x,.), x \in \mathbb{R}^d\}}$, we put $\langle f, g\rangle_{\mathcal{H}} = \int_{\mathbb{R}^d} \frac{\hat{f}(s)\hat{g}^*(s)}{\hat{G}(s)} \mathrm{d}m_n(s)$ and obtain the norm $\|f\|^2_{\mathcal{H}} = \int_{\mathbb{R}^d} \frac{|\hat{f}(s)|^2}{\hat{G}(s)} \mathrm{d}m_n(s)$, where $\overline{\{\dots\}}$ denotes closure of the set $\{\dots\}$ and a^* means complex adjoint of a.

Now we will take advantage of a theorem mentioned for example in [Da71, p. 15]:

Lemma 4.1 *A strongly quasi-convex functional $\mathcal{G}$ can achieve its minimum over a convex set C at no more than one point.*

Proof: If $\mathcal{G}(f_1) = \mathcal{G}(f_2) = \inf_{f\in C} f(x)$, then $1/2f_1 + 1/2f_2 \in C$, but $\mathcal{G}(1/2f_1 + 1/2f_2) < \max\{\mathcal{G}(f_1), \mathcal{G}(f_2)\} = \inf_{f\in C} \mathcal{G}(f)$, which is a contradiction. $\square$

Now we will show strong quasi-convexity for the functional $\mathcal{F}_G$:

Lemma 4.2 *With the notation from section 3, functional $\mathcal{E}_z$ is convex and functional Φ_G is strongly quasi convex on RKHS $\mathcal{H}$. Hence, also $\mathcal{F}_G$ is strongly quasi convex on $\mathcal{H}$.*

Proof: For the first part, $\mathcal{E}_z(f)$ is a sum of N elements, each of which is a convex functional, as (real) function $z \mapsto \frac{1}{N}(z - y_i)^2$ is convex.

To deal with the other functional, we observe that $\Phi_G(f) = \|f\|^2_{\mathcal{H}}$. We will prove that in any Hilbert space the norm $\|.\|$ satisfies strong quasi convexity:

$$\|\frac{1}{2}x + \frac{1}{2}y\|^2 < \max\{\|x\|^2, \|y\|^2\} \quad \forall x, y \in \mathcal{H}.$$

We will use the parallelogram law to show the fact. In any Hilbert space it holds, that $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \forall x, y \in H$, and so we get: $\frac{1}{4}\|x + y\|^2 = \frac{2}{4}(\|x\|^2 + \|y\|^2) - \frac{1}{4}\|x - y\|^2$. Hence $\|\frac{1}{2}x + \frac{1}{2}y\|^2 \le \frac{1}{2}(2 \max\{\|x\|^2, \|y\|^2\}) - \frac{1}{4}\|x - y\|^2$. Since for $x \neq y$ we have $\|x - y\|^2 > 0$ and we get: $\|\frac{1}{2}x + \frac{1}{2}y\|^2 < \max\{\|x\|^2, \|y\|^2\}$ as proposed.

So we have $\mathcal{F}_G$ a sum of a convex and a strongly quasi convex functional and so clearly $\mathcal{F}_G$ is strongly quasi convex as claimed. $\square$

Theorem 4.3 *If the problem $(\mathcal{F}_G, \mathcal{H})$ has a solution then it is unique for any $G^\dagger \in \mathcal{S}(\mathbb{R}^d)$ symmetric positive definite.*

Proof: By Lemma 4.2 we have strong quasi convexity of the problem and by Lemma 4.1 (since any space is convex) we obtain uniqueness.

$\square$

So we have proven uniqueness of the solution to the minimization problem $(\mathcal{F}_G, \mathcal{H})$. To prove existence we use two basic results of approximation theory, see [Da71, p. 7-13]:

Theorem 4.4 *A weakly sequentially lower semicontinuous functional $\mathcal{F}$ defined on a weakly sequentially compact set E has an infimum f_0 such that $\mathcal{F}(f_0) = \inf_{f\in E} \mathcal{F}(f) = \min_{f\in E} \mathcal{F}(f)$.*

Weak lower sequential semicontinuity of a functional can be secured by several means, as for example by:

Theorem 4.5 *A convex functional $\mathcal{F}$ that has first and second derivatives at all points of an open convex set E is weakly sequentially lower semicontinuous in E.*

To apply Theorem 4.5 we have to prove the derivatives of $\mathcal{F}_G$ to exist.

Theorem 4.6 *Functional $\mathcal{F}_G$ is weakly sequentially lower semicontinuous on $\mathcal{H}$.*

Proof: Let us have a look at the regularization part. We compute the first derivative:

$$\begin{aligned}
\mathrm{D}_f\Phi_G(h) &= \lim_{t\to 0}\int_{\mathbb{R}^d} \frac{\left(\int_{\mathbb{R}^d}[f(x)+th(x)]e^{-ixs}\mathrm{d}m_d(x)\right)\left(\int_{\mathbb{R}^d}[f(\check{x})+th(\check{x})]e^{-i\check{x}s}\mathrm{d}m_d(\check{x})\right)^*}{t\hat{G}(s)} \\
&\quad - \frac{\left(\int_{\mathbb{R}^d} f(x)e^{-ixs}\mathrm{d}m_d(x)\right)\left(\int_{\mathbb{R}^d} f(\check{x})e^{-i\check{x}s}\mathrm{d}m_d(\check{x})\right)^*}{t\hat{G}(s)}\mathrm{d}m_d(s) \\
&= \int_{\mathbb{R}^d} \frac{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(f(x)h(\check{x})^* + h(x)f(\check{x})^*\right)e^{-ixs}e^{i\check{x}s}\mathrm{d}m_d(x)\mathrm{d}m_d(\check{x})}{\hat{G}(s)}\mathrm{d}m_d(s) \\
&= \int_{\mathbb{R}^d} \frac{2\mathrm{Re}\left(\widehat{f}(s)\widehat{h}(s)^*\right)}{\hat{G}(s)}\mathrm{d}m_d(s)
\end{aligned}$$

where $\mathrm{D}_f\Phi_G(h)$ means the first derivative of Φ_G in f in direction h.

Now we compute the second derivative:

$$\mathrm{DD}_f\Phi_G(h,k) = \lim_{t\to 0}\int_{\mathbb{R}^d} \frac{2\mathrm{Re}\left(\widehat{f+tk}(s)\widehat{h}^*(s)\right)}{t\hat{G}(s)}\mathrm{d}m_d(s) -$$
$$\frac{2\mathrm{Re}\left(\widehat{f}(s)\widehat{h}^*(s)\right)}{t\hat{G}(s)}\mathrm{d}m_d(s) = \int_{\mathbb{R}^d}\frac{2\mathrm{Re}\left(\widehat{k}(s)\widehat{h}^*(s)\right)}{\hat{G}(s)}\mathrm{d}m_d(s)$$

where $\mathrm{DD}_f\Phi_G(h,k)$ is the second derivative of Φ_G in f in directions h,k.

Now we will need also the error part derivative (recall the error part is of the form $\mathcal{E}_z(f) = \frac{1}{N}\sum_{i=1}^N (f(x_i)-y_i)^2$):

$$\mathrm{D}_f\mathcal{E}_z(h) = \frac{1}{N}\lim_{t\to 0}\frac{\sum_{i=1}^N (f(x_i)+th(x_i)-y_i)^2 - \sum_{i=1}^N (f(x_i)-y_i)^2}{t} =$$
$$= \frac{1}{N}\sum_{i=1}^N (2f(x_i)h(x_i) - 2h(x_i)y_i)$$

The second derivative is:

$$\mathrm{DD}_f\mathcal{E}_z(h,k) = \frac{1}{N}\lim_{t\to 0}\frac{\sum_{i=1}^N (2(f+tk)(x_i)h(x_i) - 2h(x_i)y_i)}{t} -$$
$$\frac{\sum_{i=1}^N (2f(x_i)h(x_i) - 2h(x_i)y_i)}{t} = \frac{1}{N}\sum_{i=1}^N 2k(x_i)h(x_i)$$

By Theorem 4.5 $\mathcal{F}_G$ is weakly sequentially lower semicontinuous. □

Theorem 4.7 *The problem $(\mathcal{F}_G, E)$ has a solution for any $G^\dagger \in \mathcal{S}(\mathbb{R}^d)$ symmetric positive definite, $E \in \mathcal{H}$ bounded weakly closed.*

Proof: Every bounded weakly closed subset of a reflexive space is weakly sequentially compact (see [LuMa95]). Since any Hilbert space is reflexive, we obtain the second condition of theorem 4.4 and using 4.6 the first condition comes and we conclude. □

5 The Form of the Solution

We can describe the shape of the solution using a well known fact from mathematical analysis, see for example [GiJoPo95]:

Theorem 5.1 *Let the functional $\mathcal{F}$ defined on a set E in a Banach space X be minimized at a point $f_0 \in E$, with f_0 an interior point in the norm topology. If $\mathcal{F}$ has a derivative $\mathrm{D}\mathcal{F}_{f_0}$ at f_0, then $\mathrm{D}\mathcal{F}_{f_0} = 0$.*

The existence and uniqueness of the solution have been proven, so we can use Theorem 5.1 to derive the form of the solution. Similar results have been sketched in [GiJoPo95] but without taking advantage of RKHS.

Theorem 5.2 *Let $G^\dagger : \mathbb{R}^{2d} \to \mathbb{R}$ be a positive definite symmetric function from $\mathcal{L}_1$ and let $G : \mathbb{R}^d \to \mathbb{R}$ have symmetric positive Fourier transform (with the notation from section 3). Then the unique minimizing function $f_0 \in \mathcal{H}$ of the problem $(\mathcal{F}_G, \mathcal{H})$ is of the form*

$$f_0(x) = \sum_{i=1}^{N} c_i G(x - x_i),$$

where x_i are the data points.

Proof: We have existence and uniqueness of f_0 from section 4. The derivative of $\mathcal{F}_G$ in f in direction h is:

$$\mathrm{D}\mathcal{F}_{G\,f}(h) = 2\frac{1}{N}\sum_{i=1}^{N}\Big(f(x_i)h(x_i) - h(x_i)y_i\Big) + \gamma\int_{\mathbb{R}^d}\frac{\widehat{f}(s)\widehat{h}(s)^* + \widehat{f}^*(s)\widehat{h}(s)}{\hat{G}(s)}\mathrm{d}m_d(s)$$

Now we put $\mathrm{D}\mathcal{F}_{G\,f_0}(h) = 0$ (f_0 is the minimizing function). This has to hold for all h. Let us take $h(s) = G(s - x)$. We obtain (using symmetricity of G and $\hat{G}$):

$$0 = \mathrm{D}\mathcal{F}_{G\,f_0}(h) = 2\frac{1}{N}\sum_{i=1}^{N}\Big(f_0(x_i)G(x_i - x) - G(x_i - x)y_i\Big) + 2\gamma f_0(x)$$

and thus we have

$$\gamma f_0(x) = \frac{1}{N}\sum_{i=1}^{N} G(x - x_i)(f_0(x_i) - y_i).$$

So we see that the solution must be in the form:

$$f_0(x) = \sum_{i=1}^{N} c_i G(x - x_i).$$

□

The solution derived is very nice, since it resembles a neural network with G as the activation functions shifted to the data points x_i. The problem of the number of hidden units being too large to be implemented can be solved by variable basis approximation using the obtained shape of the activation functions (see [KuSa03]).

6 Conclusion

We have derived existence and uniqueness of the solution to the problem of finding a function close to the given data and simultaneously reasonably smooth (in terms of its Fourier transform). We showed that the solution is in the form of a one-hidden-layer feedforward neural network with activation functions depending on the form of the stabilizer.

The drawback of this approach is that the obtained neural network has too many hidden units (as much as the number of data). This problem can be dealt with by variable basis approximation limiting the number of hidden units. Nice approximation properties have been proven. This is unfortunately out of the scope of this article, so we kindly ask the reader to refer for example to work [KuSa02].

Bibliography

[CuSm01] Cucker F., Smale S. (2001). *On the Mathematical Foundations of Learning.* Bulletin of the American Mathematical Society **3**9, 1-49.

[Da71] Daniel J. W. (1971). *The Approximate Minimization of Functionals.* Prentice-Hall, Inc.

[Gi97] Girossi F. (1998). *An Equivalence between Sparse Approximation and Support Vector Machines.* Neural Computation **10**, 1455-1480, MIT. (A.I. Memo No. 1606, MIT, 1997)

[GiJoPo95] Girossi F., Jones M., Poggio T. (1995). *Regularization Theory and Neural Networks Architectures.* Neural Computation, **7**, 219-269.

[Ku03] Kůrková V. (2003). *High-dimensional Approximation by Neural Networks.* In Advances in Learning Theory: Methods, Models and Applications, Stuykens J. et. al. Ed., 69-88, Amsterdam, IOS Press.

[KuSa02] Kůrková V., Sanguinetti M. (2002). *Error Estimates for Approximate Optimization by the Extended Ritz Method.* Research Report ICS-2002-882, Institute of Computer Science, Prague, submitted to SIAM Journal on Optimization.

[KuSa03] Kůrková V., Sanguinetti M. (2003). *Learning with generalization capability by kernel methods with bounded complexity.* Research Report ICS-2003-901, Institute of Computer Science, Prague, submitted to Journal of Complexity.

[Lu02] Lukeš J. (2002). *Zápisky z funkcionální analýzy.* Karolinum, UK Praha.

[LuMa95] Lukeš J., Malý J. (1995). *Measure and Integral.* Matfyzpress, Praha, 1995.

[PoSm03] Poggio T., Smale S. (2003). *The Mathematics of Learning: Delaing with Data.* Notices of the AMS **5**0, 5, 536-544.

[Ru91] Rudin W. (1991). *Functional Analysis*,. 2nd Edition, McGraw-Hill, NY.

[SchSm02] Schölkopf B., Smola A. J. (2002). *Learning with Kernels.* MIT Press, Cambridge, Massachusetts.

[Wa90] Wahba G. (1990). *Spline Models for Observational Data.* Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia.