SPLINES WITH NON POSITIVE KERNELS

STÉPHANE CANU[1], CHENG SOON ONG[2] AND XAVIER MARY[3]

1- PSI - FRE CNRS 2645 - INSA de Rouen
76801 St Etienne du Rouvray, France
Stephane.Canu@insa-rouen.fr

2- Max Planck Institute for Biological Cybernetics
Spemannstrasse 38, 72076 Tuebingen, Germany
chengsoon.ong@tuebingen.mpg.de

3- ENSAE-CREST-LS
3 avenue Pierre Larousse, 92240 Malakoff, France
xavier.mary@ensae.fr

Non parametric regression methods can be presented in two main clusters. The one of smoothing splines methods requiring positive kernels and the other one known as Nonparametric Kernel Regression allowing the use of non positive kernels such as the Epanechnikov kernel. We propose a generalization of the smoothing spline method to include kernels which are still symmetric but not positive semi definite (they are called indefinite). The general relationship between smoothing splines, Reproducing Kernel Hilbert Spaces (RKHS) and positive kernels no longer exists with indefinite kernels. Instead the splines are associated with functional spaces called Reproducing Kernel Krein Spaces (RKKS) endowed with an indefinite inner product and thus not directly associated with a norm. Smoothing splines in RKKS have many of the interesting properties of splines in RKHS, such as orthogonality, projection and representer theorem. We show that smoothing splines can be defined in RKKS as the regularized solution of the interpolation problem. Since no norm is available in an RKKS, Tikhonov regularization cannot be defined. Instead, we propose the use of conjugate gradient type iterative methods, with early stopping as a regularization mechanism. Several iterative algorithms are collected which can be used to solve the optimization problems associated with learning in indefinite spaces. Some preliminary experiments with indefinite kernels for spline smoothing reveal the computational efficiency of this approach.

1. Inroduction

Spline functions are a widely used tool in non parametric curve estimation. The underlying theory is one of positive kernels and Reproducing Kernel Hilbert Spaces (RKHS). Thus, positivity of the kernel is both a requirement and a limitation to the use of this method. Recently, positive kernels have been popularized in the statistical learning community with Vapnik's support vector machine and more generally with kernel machines. An advantage of kernel methods pointed out in these works, is their ability to handle large data sets (typically millions of data points). In this framework, the starting point is the kernel. From the kernel, the

2

associated RKHS is built and the problem is stated in this functional space. But the solution of this problem only requires knowledge about the kernel. No explicit formulation of the RKHS is needed. Its existence is sufficient to justify the proposed algorithm.

However, in many practical problems with large data sets and a lack of models, the natural application dependent kernels are non positive ones. These problems arise in fields such as of text mining, biostatistics and astronomy. Practitioners also report good results when using indefinite kernels although, up to now, no theory was available. This paper aims at presenting such theory, a theory of non positive kernels, associated functional spaces and the equivalent splines in this framework.

This paper is organized as follows. In the first part, the functional framework associated with non positive kernels is described. It is shown that indefinite kernels are associated with Reproducing Kernel Krein Spaces (RKKS) endowed with an indefinite inner product and thus not directly associated with a norm. Apart from the norm, all the interesting properties required for learning in RKHS are preserved in RKKS. The second part deals with splines and their definition without using a norm. This definition is straightforward for interpolation splines, but care is required to define a generalization of approximation splines. Based on these definitions, the third part discusses different implementations and proposes an iterative conjugate gradient type algorithm adapted for indefinite kernel matrices.

2. Reproducing Kernel Krein Spaces

Krein spaces are indefinite inner product spaces endowed with a Hilbert topology. They can also be seen as a kind of generalized Hilbert space in the sense that their inner product is no longer positive. Before we delve into the definitions and some basic properties of Krein spaces, we give an example.

Example 2.1. *4 dimensional space-time.* Indefinite spaces were first introduced into the solution of physical problems via the 4-dimensional Minkowski space of special relativity. There we have 3 negative and one positive dimensions (see Chapter 2 of the reference[6] for the same example with 2 positive and 1 negative dimensions), and the inner product is given by

$$\langle (x_1, y_1, z_1, t_1), (x_2, y_2, z_2, t_2) \rangle = -x_1 x_2 - y_1 y_2 - z_1 z_2 + t_1 t_2$$

and it is no longer necessarily positive. In this space the vector $v = (1, 1, 1, \sqrt{3})$ is a neutral vector (such that $\langle v, v \rangle = 0$). More generally all vectors belonging to the cone $x^2 + y^2 + z^2 - t^2 = 0$ are neutral vectors (the so called light cone in relativity).

2.1. *Krein spaces*

As can be seen from Example 2.1, there are several differences between a Krein space and a Hilbert space. The main difference lies in the fact that we allow a more general inner product (detailed expositions can be found in books[4,2]).

Definition 2.1. *Inner product.* Let $\mathcal{K}$ be a vector space on the scalar field $\mathbb{R}^{\text{a}}$. An inner product $\langle .,. \rangle_{\mathcal{K}}$ on $\mathcal{K}$ is a bilinear form verifying:

- $\forall f, g \in \mathcal{K}^2, \quad \langle f, g \rangle_{\mathcal{K}} = \langle g, f \rangle_{\mathcal{K}}$ (symmetric)
- $\forall f, g, h \in \mathcal{K}^3 \quad \langle \alpha f + g, h \rangle_{\mathcal{K}} = \alpha \langle f, h \rangle_{\mathcal{K}} + \langle g, h \rangle_{\mathcal{K}}$ (linear)
- $\{ \forall g \in \mathcal{K}, \quad \langle f, g \rangle_{\mathcal{K}} = 0 \} \Rightarrow f = 0$ (nondegenerate)

An inner product is said to be: positive if $\forall f \in \mathcal{K}, \quad \langle f, f \rangle_{\mathcal{K}} \geq 0$, negative if $\forall f \in \mathcal{K}, \quad \langle f, f \rangle_{\mathcal{K}} \leq 0$, otherwise it is indefinite. A vector space $\mathcal{K}$ endowed with the inner product $\langle .,. \rangle_{\mathcal{K}}$ is called an *inner product space*. Two vectors f, g of an inner product space are said to be *orthogonal* if $\langle f, g \rangle_{\mathcal{K}} = 0$.

Definition 2.2. *Krein space.* An inner product space $(\mathcal{K}, \langle .,. \rangle_{\mathcal{K}})$ is a Krein space if there exist two Hilbert spaces $(\mathcal{H}_+, \langle .,. \rangle_{\mathcal{H}_+})$ and $(\mathcal{H}_-, \langle .,. \rangle_{\mathcal{H}_-})$ included in $\mathcal{K}$ where:

- $\forall f \in \mathcal{K}, \exists! f_+ \in \mathcal{H}_+$ and $f_- \in \mathcal{H}_-$ such that $f = f_+ + f_-$
- $\forall f, g \in \mathcal{K}^2, \langle f, g \rangle_{\mathcal{K}} = \langle f_+, g_+ \rangle_{\mathcal{H}_+} - \langle f_-, g_- \rangle_{\mathcal{H}_-}$

These spaces are then orthogonal with respect to the inner product and the Krein space may be seen as the direct difference of the two Hilbert spaces

$$\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$$

Note that such a decomposition is not unique in general.

Definition 2.3. *Direct sum and strong topology.* Let $\mathcal{K}$ be a Krein space and $\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$ a decomposition. The *direct sum* of $\mathcal{H}_+$ and $\mathcal{H}_-$ defines a Hilbert space, denoted $|\mathcal{K}| = \mathcal{H}_+ \oplus \mathcal{H}_-$, endowed with the (positive) inner product

$$\langle f, g \rangle_{|\mathcal{K}|} = \langle f_+, g_+ \rangle_{\mathcal{H}_+} + \langle f_-, g_- \rangle_{\mathcal{H}_-}$$

with the induced norm $\|f\|_{|\mathcal{K}|}^2 := \langle f, f \rangle_{|\mathcal{K}|}$. $|\mathcal{K}|$ is the smallest Hilbert upper bound of the Krein space $\mathcal{K}$ and one defines the strong topology on $\mathcal{K}$ as the Hilbert topology of $|\mathcal{K}|$. The topology does not depend on the chosen decomposition.

Note that for all $f \in \mathcal{K} : |\langle f, f \rangle_{\mathcal{K}}| \leqslant \|f\|_{|\mathcal{K}|}^2$. $\mathcal{K}$ is said to be *Pontryagin* if it admits a decomposition with a negative part of finite dimension, that is $dim(\mathcal{H}_-) < \infty$, and *Minkowski* if it admits a decomposition with both positive and negative part of finite dimension, that is $dim(\mathcal{H}_+) < \infty$ and $dim(\mathcal{H}_-) < \infty$.

Let $\Omega \subseteq \mathbb{R}^d$ be the domain. The set of genuine functions defined on the domain Ω whose value lie in $\mathbb{R}$ is defined to be $\mathbb{R}^\Omega$. The evaluation functional tells us the value of a function at a certain point, and an RKKS is a subset of $\mathbb{R}^\Omega$ where all the evaluation functionals are continuous.

---

[a]Like Hilbert spaces, Krein spaces can be defined on $\mathbb{R}$ or $\mathbb{C}$. We use $\mathbb{R}$ in this paper.

4

Definition 2.4. *Evaluation functional.* For all $x \in \Omega$, the evaluation functional T at point x is defined as,

$$T_x : \mathcal{K} \to \mathbb{R}$$
$$f \mapsto T_x(f) = f(x)$$

Definition 2.5. *Reproducing Kernel Krein Space (RKKS).* A Krein space $(\mathcal{K}, \langle .,. \rangle_{\mathcal{K}})$ is a reproducing kernel Krein space (see chapter 7 in the reference[1]) if:

- $\mathcal{K} \subset \mathbb{R}^{\Omega}$
- $\forall x \in \Omega$, T_x is continuous on $\mathcal{K}$ endowed with its strong topology.

2.2. *From Krein spaces to kernels*

We prove an analog to the Moore-Aronszajn theorem[9], which tells us that for every kernel there is an associated Krein space, and for every Krein space, there is a unique kernel.

Proposition 2.1. Reproducing kernel of an RKKS. *Let $(\mathcal{K}, \langle .,. \rangle_{\mathcal{K}})$ be a Reproducing Kernel Krein Space and $\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$ a decomposition. Then*

- *$\mathcal{H}_+$ and $\mathcal{H}_-$ are RKHS (with kernel functions k_+ and k_-)*
- *there exists a unique symmetric function $k(x, y)$ belonging to $\mathcal{K}$ as a function of a single variable, such that: $\forall f \in \mathcal{K}, \langle f(.), k(., x) \rangle_{\mathcal{K}} = f(x)$*
- *$k = k_+ - k_-$*

Proof. Since $(\mathcal{K}, \langle .,. \rangle_{\mathcal{K}})$ is an RKKS, the evaluation functional is continuous with respect to the strong topology hence for the Hilbert topology of $(|\mathcal{K}|, \langle .,. \rangle_{|\mathcal{K}|})$. It follows that $\mathcal{H}_+$ and $\mathcal{H}_-$, as Hilbert subspaces of an RKHS, are RKHS.
Then let $f = f_+ + f_-$. The evaluation functional can be expressed as

$$\begin{aligned}
T_x(f) &= T_x(f_+) + T_x(f_-) \\
&= \langle f_+, k_+(x) \rangle_{\mathcal{H}_+} + \langle f_-, k_-(x) \rangle_{\mathcal{H}_-} \\
&= \langle f_+, k_+(x) \rangle_{\mathcal{H}_+} - \langle f_-, -k_-(x) \rangle_{\mathcal{H}_-} \\
&= \langle f, \underbrace{k_+(x) - k_-(x)}_{k(.,x)} \rangle_{\mathcal{K}}.
\end{aligned}$$

It has to be symmetric since the inner product is symmetric, but it is not necessarily positive. It is unique since the inner product is non-degenerate. □

2.3. *From kernels to Reproduicing Kernel Krein Spaces*

Let $k(x, y)$ be a symmetric real valued function, $k : \Omega \times \Omega \longrightarrow \mathbb{R}$.

Proposition 2.2. *(Theorem 8.9[8] or Theorem 2.28 p.86[7]) . The following propositions are equivalent:*

> - *there exist (at least) one RKKS with kernel k*
> - *k admits a positive decomposition, that is there exists two positive kernels k_+ and k_- such that $k = k_+ - k_-$*
> - *k is dominated by some positive kernel p that is, $p - k$ and $p + k$ are positive.*

The two last conditions are equivalent by choosing $p = k_+ + k_-$. Note that with such a kernel decomposition, the Krein space $\mathcal{K} = H_+ \ominus H_-$ is a natural choice. There is *no* bijection *but an onto mapping* between the set of RKKS and the set of generalized kernels defined in the vector space generated out of the cone of positive kernels. This is not a major problem since in the forthcomming section, our derivation only requires the existance of an RKKS.

3. Functional estimation in an RKKS

3.1. *Notations*

Assume we have a set of observations $\mathbf{x}_i, y_i, i = 1, n$ with $\mathbf{x}_i \in \Omega$ and $y_i \in \mathbb{R}$. Let $\mathbf{y} = (y_1, ..., y_n)^\top \in \mathbb{R}^n$. Let $\mathcal{H}$ be a reproducing kernel Hilbert or Krein space with kernel k. The evaluation operator T and its adjoint are defined by:

$$
\begin{aligned}
T : \mathcal{H} &\longrightarrow \mathbb{R}^n & T^\star : \mathbb{R}^n &\longrightarrow \mathcal{H} \\
f &\longmapsto (f(\mathbf{x}_1), ..., f(\mathbf{x}_n))^\top & \alpha &\longmapsto T^\star \alpha = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, x)
\end{aligned}
$$

The image of operator $T^\star$ is $\mathrm{Im}(T^\star) = \{f \in \mathcal{H} \mid \exists \alpha \in \mathbb{R}^n \text{ such that } f(x) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, x)\}$. The null space of operator T is $\mathrm{Null}(T) = \{f \in \mathcal{H} \mid f(\mathbf{x}_i) = 0, \ i = 1, n\}$. Based on the evaluation functional and its adjoint the Gram matrix $K = TT^\star$ is defined such that $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$.

Definition 3.1. *Interpolation set.* For a given set of observations $\mathbf{x}_i, y_i, i = 1, n$, the interpolation set in $\mathcal{H}$ is defined by $\mathcal{S} = \{f \in \mathcal{H} \mid Tf = y\}$.

Note that all these definitions hold for both reproducing kernel Hilbert and Krein spaces.

3.2. *Interpolation in an RKKS*

In this framework the minimal norm interpolating problem is the following:

Definition 3.2. *Interpolation Spline in an RKHS.* For a given set of observations $\mathbf{x}_i, y_i, i = 1, n$, the interpolation spline in a RKHS $\mathcal{H}$ is the solution of the following minimization problem:

$$
\min_{f \in \mathcal{H}} \|f\|_{\mathcal{H}}^2 \quad \text{such that: } Tf = \mathbf{y}
$$

Theorem 3.1. *Equivalent definition for splines (Theorem 58[3]). If the interpolation set associated with the interpolation spline problem is not empty then the interpolation spline in an RKHS, denoted by $\widetilde{f}$, is the unique orthogonal projection of the interpolation set on the image of $T^\star$.*

6

Proof. $\mathcal{H}$ can be decomposed in two orthogonal subspaces such that $\mathcal{H} = \mathrm{Im}(T^\star)\oplus$ $\mathrm{Null}(T)$ and $\forall f \in \mathcal{H} \ \exists f^* \in \mathrm{Im}(T^\star)$ and $f_r \in \mathrm{Null}(T)$ such that $f = f^* + f_r$. It follows that $Tf = Tf^*$ and because of the minimal norm principle, the interpolation spline verifies $f_r = 0$. The interpolation spline $\widetilde{f}$ is the unique solution of the following problem:

$$\text{find } f \in \mathrm{Im}(T^\star) \text{ such that } Tf = y$$

It verifies: $\widetilde{f} = T^\star\alpha$ where vector α is the solution of the following linear system: $K\alpha = \mathbf{y}$.

Thus, $\forall f \in \mathcal{S}$ and $\forall \mathbf{v} \in \mathbb{R}^n, \quad \mathbf{v}^\top Tf - \mathbf{v}^\top T\widetilde{f} = 0$
$$\mathbf{v}^\top Tf - \mathbf{v}^\top T\widetilde{f} = 0$$
$$\langle T^\star\mathbf{v}, f - \widetilde{f}\rangle_\mathcal{H} = 0$$

since $\widetilde{f}$ belongs to $\mathrm{Im}(T^\star)$ it is also the orthogonal projection of $\mathcal{S}$ on $\mathrm{Im}(T^\star)$. It is unique since $\mathrm{Im}(T^\star)$ is closed and convex. $\qquad\square$

This theorem gives an alternative way to define the interpolation spline without using a norm. While no norm is directly available in a Krein space, the second definition still holds and we have:

Definition 3.3. *Interpolation Spline in an RKKS.* The interpolation spline in an RKKS is the orthogonal projection of the interpolating functions on the set spanned by the kernel.

It can be computed in the same way as in the Hilbert case: first solve $K\alpha = \mathbf{y}$ and then the interpolation spline is the function $\widetilde{f}(x) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, x)$.

3.3. *Smoothing splines in an RKKS*

Smoothing splines deal with the case where the target values are known up to a certain amount of error. In this case we have $Tf = y + \varepsilon$ where ε denotes some error. The principle of smoothing splines suggests minimizing the norm of the error vector together with some regularity constraints, leading to:

Definition 3.4. *Smoothing splines in an RKHS (1).* For a given λ, a smoothing spline is the solution of the following minimization problem:

$$\mathcal{P}_\lambda: \quad \min_{f\in\mathcal{H}} \sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2 + \lambda\|f\|_\mathcal{H}^2$$

It is worth noticing that this minimization problem may be seen as a minimal decomposition in complementary spaces. Precisely, instead of decomposing $\mathcal{H} = \mathrm{Im}(T^\star) \oplus \mathrm{Im}(T^\star)^\perp$, change the inner product on $\mathrm{Im}(T^\star)$ to a new one $\langle.,.\rangle_\lambda$ such that $||f||_\mathcal{H} \leq ||f||_\lambda, \ \forall f \in \mathrm{Im}(T^\star)$. Then there exists a unique complementary space Q_λ (contractively included in $\mathcal{H}$) and a unique minimal decomposition which actually solves the problem.

As a matter of fact, depending on the value of parameter λ, the smoothing spline is not a single solution but rather a sequence of solutions. The sequence may be defined as a decreasing sequence of Hilbert norms on $\mathrm{Im}(T^\star)$ *i.e.*

$$\lambda < \lambda' \Rightarrow ||f||_{\lambda'} \leq ||f||_\lambda, \ \forall f \in \mathrm{Im}(T^\star)$$

This sequence can also be defined by using a sequence of embedded sets $\mathcal{H}_\lambda$:

Definition 3.5. *Smoothing splines in an RKHS (2).* Let c_λ be a sequence of increasing positive real numbers. Then for $\mathcal{H}_\lambda = \left\{ f \in \mathcal{H} \mid ||f||_\mathcal{H}^2 \leq c_\lambda \right\}$ a smoothing spline is the sequence of solutions of the following minimization problem:

$$\mathcal{P}_\lambda : \quad \min_{f \in \mathcal{H}_\lambda} \sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2$$

Note that for c_λ large enough, the solution to the problem $\mathcal{P}_\lambda$ is also the interpolation spline $\widetilde{f}$. Thus the sequence $\widetilde{f}_\lambda$ of solutions of problems $\mathcal{P}_\lambda$ converges towards $\widetilde{f}$. The definition of this sequence depends only on the definition of the sequence of embedded spaces. This suggest the following analogous definition:

Definition 3.6. *Smoothing splines in an RKKS.* Let $\mathcal{K}_\mu$ be a sequence of embedded reproducing kernel Krein subspaces converging towards $\mathrm{Im}(T^\star)$ when μ increases towards ∞ (we have $\mu_1 < \mu_2 \Rightarrow \mathcal{K}_{\mu_1} \subseteq \mathcal{K}_{\mu_2}$). Then a smoothing spline is the sequence of solutions of the following minimization problem:

$$\mathcal{P}_\mu : \quad \min_{f \in \mathcal{K}_\mu} \sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2$$

The remaining problem is to define the sequence of embedded spaces. In this framework the smoothing effect is due to an early stopping in the regularization path.

4. Implementation of smoothing splines in an RKKS

4.1. *Three different regularization strategies*

There exists three different ways to build a sequence of embedded spaces. These subspaces can be defined by limiting their size (it is the penalization approach), by explicitly using more and more generating functions (such as basis functions) or through an iterative process building implicitly at each iteration a more complex solution. We will show that, for smoothing splines in an RKKS, the latter is preferable.

The first regularization principle is the classical penalization approach also known as Tikhonov regularization. But since no norm is available in a Krein space a different variational principle has to be used. Instead of looking for a minimum the problem can be restated using the stabilization principle. The stabilizer of some functional is the point where the gradient vanishes.

$$\mathcal{P}_\lambda : \mathop{\mathrm{stabilize}}_{f \in \mathcal{K}} ||Tf - \mathbf{y}||^2 + \lambda \langle f, f \rangle_\mathcal{K}$$

8

This problem is the same as the one of the Hilbertian case but also presents a major drawback. Just like the RKHS spline the solution of this problem is given by the solution of: $(K + \lambda I)\alpha = \mathbf{y}$. But in the RKKS case, the matrix K is no longer positive definite and this strategy, because it increases all eigenvalues by a factor λ, can lead to a singular system. An interesting way to bypass this problem would be the one of complementary spaces. However the building of a decreasing sequence of indefinite inner products on $\text{Im}(T^\star)$ is far more difficult than in the positive case.

The second regularizing strategy consists in building explicit subspaces. Since the interpolating solution lies in $\text{Im}(T^\star)$ the truncated spectral factorization approach suggests using the eigenvectors of the Gram matrix K. At each step the following problem has to be solved

$$\mathcal{P}_\ell : \min_{\alpha \in \mathcal{K}_\ell} \|K\alpha - \mathbf{y}\|^2$$

where $\mathcal{K}_\ell = \text{span}\{\phi_1, ..., \phi_\ell\}$, ϕ_ℓ being the ℓth eigenvector of K. But this technique is not computationally suitable since it requires the computation of the spectrum of K.

The third regularization strategy consists in implicitly building the solution subspace by using an iterative approach. There are two well known iterative approaches: gradient iterations and Krylov subspace techniques. Gradient iterations (also known as Landweber-Friedman iterations) propose for a given stepsize ρ:

$$\alpha_\ell = \alpha_{\ell-1} - \rho\nabla_\alpha(\|K\alpha_{\ell-1} - \mathbf{y}\|^2).$$

Krylov subspace (also known as conjugate gradient type) methods propose to build a sequence of iteration polynomials q_ℓ to get $\alpha_\ell = q_{\ell-1}(K)\mathbf{y}$. When possible, the later is preferable because it converges faster for the same computational complexity as the fixed stepsize gradient. Furthermore, there exists Krylov subspace approaches dealing with non positive Gram matrices such as the Minimal Residual II (MR II) algorithm.

4.2. MR II: a Krylov subspace algorithm for indefinite matrix

The MR II algorithm[5] is an iterative procedure computing the following sequence:

$$\alpha_\ell = \alpha_{\ell-1} - \rho\mathbf{d}_\ell$$

where ρ is the stepsize and $\mathbf{d}_\ell$ the descent direction. The sequence converges towards the solution of the linear interpolating system $K\alpha = \mathbf{y}$. The MR II algorithm is also a minimal residual algorithm for symmetric indefinite linear systems with starting vector $K\mathbf{y}$ instead of $\mathbf{y}$. By minimal residual we mean that it minimizes at each step the squared residual error:

$$\min_{\alpha \in \mathbf{R}^n} \|K\alpha - \mathbf{y}\|^2.$$

At each step, coefficient ρ is chosen to minimize this residual, *i.e.* such that the differential of the cost function vanishes. Descent directions $\mathbf{d}_\ell$ are built to be K^2

conjugates to decouple the cost function. Thus we have:

$$\alpha = \sum_{i=1}^{n} a_i \mathbf{d}_i \quad \text{with} \quad \mathbf{d}_i^\top K K \mathbf{d}_j = 0, i \neq j$$

These directions $\mathbf{d}_\ell$ are determined based on a three term recurrence formula (a standard Lanczos process). This kind of algorithm belongs to the family of Krylov subspace algorithms since it admits a polynomial interpretation:

$$\alpha_\ell = q_{\ell-1}(K)\mathbf{y} \qquad r_\ell = p_\ell(K)\mathbf{y}$$
$$r_\ell = \mathbf{y} - K\alpha_\ell \Longrightarrow p_\ell(K) = I - K q_{\ell-1}(K)$$

where $q_{\ell-1}$ and p_ℓ are two families of orthogonal polynomials is some sense. For regularization purposes[b] the Krylov space over which the residual is minimized is chosen as follows:

$$\alpha_\ell \in \text{span}\{K\mathbf{y}, K^2\mathbf{y}, ..., K^{\ell-1}\mathbf{y}\}$$

Algorithm 1 summarizes MR II. Note that for practical reasons it is preferable to compute another vector sequence storing $K\mathbf{d}_\ell$ omitted here for brevity and clarity.

Algorithm 1 : *Kernel Spline MR II*

---

1. $(\alpha_1, \mathbf{r}_1, \mathbf{d}_0, \mathbf{d}_{-1}, c, \ell)$ ← initialize $(0, \mathbf{y}, K\mathbf{y}/c, 0, (\mathbf{y}^\top K^4 \mathbf{y})^{\frac{1}{2}}, 1)$
while SolutionReached ==FALSE
    2. Compute next solution
        2.1 Compute the stepsize $\rho \leftarrow \mathbf{r}_\ell^\top K \mathbf{d}_{\ell-1}$
        2.2 Compute the next solution $\alpha_{\ell+1} \leftarrow \alpha_\ell + \rho \mathbf{d}_{\ell-1}$
    3. Update the residual $\mathbf{r}_{\ell+1} \leftarrow \mathbf{y} - K\alpha_{\ell+1}$
    4. Compute a new descent direction
        4.1 Compute the stepsize $b \leftarrow \mathbf{d}_{\ell-1}^\top K^3 \mathbf{d}_{\ell-1}$
        4.2 Update the direction $\mathbf{d}_\ell \leftarrow K\mathbf{d}_{\ell-1} - b\mathbf{d}_{\ell-1} - c\mathbf{d}_{\ell-2}$
        4.3 Compute the stepsize $c \leftarrow (\mathbf{d}_\ell^\top K^2 \mathbf{d}_\ell)^{\frac{1}{2}}$
        4.4 Normalize the direction $\mathbf{d}_\ell \leftarrow \frac{\mathbf{d}_\ell}{c}$
    5. Counting the number of iterations $\ell \leftarrow \ell + 1$
    if $\|\mathbf{r}\| < \varepsilon$ or $\ell = n$
        6. SolutionReached ← TRUE
    end if
end while

MRII algorithm is not *the only* solution of the approximation spline in indefinite reproducing kernel space but one among several others. It presents the advantage of being very fast and regularizing. Its main drawback is that it is not flexible

---

[b]Regularizing properties of this method are detailed in chapter 6[5].

10

regarding the choice of the regularization parameter. Because the regularization process is controlled through an early stopping mechanism, it may happen that the "truth" is lying between two iterations. But since the RKKS framework is fruitful, many other algorithms remain to be found.

5. Conclusion

In this paper we have shown how to do splines using non positive kernels. A functional framework has been designed together with new definitions of what splines are. These definitions are coherent with usual splines in reproducing kernel Hilbert spaces (RKHS) and can be applied to non positive kernels.

Regarding the functional framework, we show that reproducing kernel spaces are not necessarily Hilbertian and reproducing kernel Krein spaces (RKKS) can be built from an indefinite kernel in almost the same way as RKHS is built from positive kernels. In this framework, the representation of the evaluation functional through kernels still holds. The main difference between RKHS and RKKS is that no norm is induced by the inner product in the latter structure. So usual splines were revisited to be defined without norms. Interpolating splines are seen as an orthogonal projection while approximation splines are defined as a regularized family of solutions to the interpolation problem. A solution to this problem has been proposed through the use of MR II, an efficient iterative conjugate gradient technique, suitable for indefinite matrices. Note that large problems can be solved using this algorithm since its complexity is $\mathcal{O}(kn^2)$ where k is the number of iteration needed.

We have reasons to believe that this complexity can be practically reduced. The strategy to achieve such a result would be the same as in SVM, to consider a regularization mechanism involving a sparse representation (a lot of components of the solution vector set to 0). Note also that this framework could be extended to non symmetric kernels[7] and to the pattern recognition case.

References

1. D. Alpay The Schur algorithm, reproducing kernel spaces and system theory, volume 5, SMF/AMS Texts and Monographs of the Société mathématique de France (2001)
2. T.Ya. Azizov and I.S. Iokhvidov, Linear operators in spaces with an indefinite metric, Wiley (1989)
3. A. Berlinet and C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Kluwer Academic Pub (2004)
4. J. Bognar, Indefinite Inner Product Spaces, Springer-Verlag (1974)
5. M. Hanke Conjugate Gradient Type Methods for Ill-Posed Problems, Pitman Research Notes in Mathematics 327, (1995)
6. B. Hassibi, A. H. Sayed, and T. Kailath, Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories, SIAM, Philadelphia (1999)
7. X. Mary, Hilbert subspaces, subdualities and applications, PhD thesis, Laboratory PSI, INSA Rouen (2003)
8. J. Rovnyak, Methods of Krein space operator theory, Toeplitz lectures, Interpolation theory, systems theory, and related topics, Oper. Theory Adv. Appl., vol. 134, Birkhauser, Basel, 2002, pp. 31–66, (1999)
9. G. Wahba, Spline Models for Observational Data, SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics, volume 59 (1990)