A Unifying Framework for Vector-valued Manifold Regularization and Multi-view Learning - Supplementary Material

Hà Quang Minh MINH.HAQUANG@IIT.IT
Loris Bazzani LORIS.BAZZANI@IIT.IT
Vittorio Murino VITTORIO.MURINO@IIT.IT
Istituto Italiano di Tecnologia, Via Morego 30, Genova 16163, ITALY

Abstract

The Supplementary Material contains three elements. First, in Section 1, we give the proofs for all the main mathematical results in the paper. Second, in Section 2, we provide a natural generalization of our framework to the case the point evaluation operator $f(x)$ is replaced by a general bounded linear operator. Last, in Section 3, we provide an exact description of Algorithm 1 with the Gaussian or similar kernels in the degenerate case, when the kernel width $\sigma \to \infty$.

1. Proofs of Main Results

For clarity, we restate all the results in the main paper that we prove here.

Notation: the definition of $\mathbf{f}$ as given by

$$\mathbf{f} = (f(x_1), \ldots, f(x_{u+l})) \in \mathcal{W}^{u+l}, \quad (1)$$

is adopted because it is also applicable when $\mathcal{W}$ is an infinite-dimensional Hilbert space. For $\mathcal{W} = \mathbb{R}^m$,

$$\mathbf{f} = (f^1(x_1), \ldots, f^m(x_1), \ldots, f^1(x_{u+l}), \ldots, f^m(x_{u+l})).$$

This is different from (**?**), where

$$\mathbf{f} = (f^1(x_1), \ldots, f^1(x_{u+l}), \ldots, f^m(x_1), \ldots, f^m(x_{u+l})).$$

This means that our matrix M is necessarily a permutation of the matrix M in (**?**) when they give rise to the same semi-norm.

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

1.1. Proof of the Representer Theorem

Recall our general minimization problem

$$f_{\mathbf{z},\gamma} = \text{argmin}_{f\in\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} V(y_i, Cf(x_i)) + \gamma_A ||f||^2_{\mathcal{H}_K} + \gamma_I \langle \mathbf{f}, M\mathbf{f}\rangle_{\mathcal{W}^{u+l}}, \quad (2)$$

and its least square version

$$f_{\mathbf{z},\gamma} = \text{argmin}_{f\in\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} ||y_i - Cf(x_i)||^2_{\mathcal{Y}} + \gamma_A ||f||^2_{\mathcal{H}_K} + \gamma_I \langle \mathbf{f}, M\mathbf{f}\rangle_{\mathcal{W}^{u+l}}. \quad (3)$$

Theorem 1. *The minimization problem (2) has a unique solution, given by $f_{\mathbf{z},\gamma} = \sum_{i=1}^{u+l} K_{x_i} a_i$ for some vectors $a_i \in \mathcal{W}$, $1 \le i \le u+l$.*

The following is a generalization of the proof for the Representer Theorem in (**?**). Since $f(x) = K_x^* f$, the minimization problem (2) is

$$f_{\mathbf{z},\gamma} = \text{argmin}_{f\in\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} V(y_i, CK^*_{x_i} f) + \gamma_A ||f||^2_{\mathcal{H}_K} + \gamma_I \langle \mathbf{f}, M\mathbf{f}\rangle_{\mathcal{W}^{u+l}}. \quad (4)$$

Consider the operator $E_{C,\mathbf{x}} : \mathcal{H}_K \to \mathcal{Y}^l$, defined by

$$E_{C,\mathbf{x}} f = (CK^*_{x_1} f, \ldots, CK^*_{x_l} f), \quad (5)$$

with $CK^*_{x_i} : \mathcal{H}_K \to \mathcal{Y}$ and $K_{x_i} C^* : \mathcal{Y} \to \mathcal{H}_K$. For $\mathbf{b} = (b_1, \ldots, b_l) \in \mathcal{Y}^l$, we have

$$\langle \mathbf{b}, E_{C,\mathbf{x}} f\rangle_{\mathcal{Y}^l} = \sum_{i=1}^{l} \langle b_i, CK^*_{x_i} f\rangle_{\mathcal{Y}}$$

$$= \sum_{i=1}^{l} \langle K_{x_i} C^* b_i, f\rangle_{\mathcal{H}_K} = \langle \sum_{i=1}^{l} K_{x_i} C^* b_i, f\rangle_{\mathcal{H}_K}.$$

The adjoint operator $E^*_{C,\mathbf{x}} : \mathcal{Y}^l \to \mathcal{H}_K$ is thus

$$E^*_{C,\mathbf{x}} : (b_1, \ldots, b_l) \to \sum_{i=1}^{l} K_{x_i} C^* b_i. \tag{6}$$

The operator $E^*_{C,\mathbf{x}} E_{C,\mathbf{x}} : \mathcal{H}_K \to \mathcal{H}_K$ is then

$$E^*_{C,\mathbf{x}} E_{C,\mathbf{x}} f \to \sum_{i=1}^{l} K_{x_i} C^* C K^*_{x_i} f, \tag{7}$$

with $C^*C : \mathcal{W} \to \mathcal{W}$.

Proof of Theorem 1. Denote the right handside of (2) by $I_l(f)$. Then $I_l(f)$ is coercive and strictly convex in f, and thus has a unique minimizer. Let $\mathcal{H}_{K,\mathbf{x}} = \{\sum_{i=1}^{u+l} K_{x_i} w_i : \mathbf{w} \in \mathcal{W}^{u+l}\}$. For $f \in \mathcal{H}^{\perp}_{K,\mathbf{x}}$, by the reproducing property, $E_{C,\mathbf{x}}$ satisfies

$$\langle \mathbf{b}, E_{C,\mathbf{x}} f \rangle_{\mathcal{Y}^l} = \langle f, \sum_{i=1}^{l} K_{x_i} C^* b_i \rangle_{\mathcal{H}_K} = 0,$$

for all $\mathbf{b} \in \mathcal{Y}^l$, since $C^* b_i \in \mathcal{W}$. Thus

$$E_{C,\mathbf{x}} f = (CK^*_{x_1} f, \ldots, CK^*_{x_l} f) = 0.$$

Similarly, by the reproducing property, the sampling operator $S_{\mathbf{x}}$ satisfies

$$\langle S_{\mathbf{x}} f, \mathbf{w} \rangle_{\mathcal{W}^{u+l}} = \langle f, \sum_{i=1}^{u+l} K_{x_i} w_i \rangle_{\mathcal{H}_K} = 0,$$

for all $\mathbf{w} \in \mathcal{W}^{u+l}$. Thus

$$\mathbf{f} = S_{\mathbf{x}} f = (f(x_1), \ldots, f(x_{u+l})) = 0.$$

For an arbitrary $f \in \mathcal{H}_K$, consider the orthogonal decomposition $f = f_0 + f_1$, with $f_0 \in \mathcal{H}_{K,\mathbf{x}}$, $f_1 \in \mathcal{H}^{\perp}_{K,\mathbf{x}}$. Then, because $||f_0 + f_1||^2_{\mathcal{H}_K} = ||f_0||^2_{\mathcal{H}_K} + ||f_1||^2_{\mathcal{H}_K}$, the result just obtained shows that

$$I_l(f) = I_l(f_0 + f_1) \geq I_l(f_0)$$

with equality if and only if $||f_1||_{\mathcal{H}_K} = 0$, that is $f_1 = 0$. Thus the minimizer of (2) must lie in $\mathcal{H}_{K,\mathbf{x}}$. □

1.2. Proofs for the Least Square Case

Proposition 1. *The minimization problem (3) has a unique solution $f_{\mathbf{z},\gamma} = \sum_{i=1}^{u+l} K_{x_i} a_i$, where the vectors $a_i \in \mathcal{W}$ are given by*

$$l\gamma_I \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j + C^*C(\sum_{j=1}^{u+l} K(x_i, x_j) a_j) + l\gamma_A a_i = C^* y_i, \tag{8}$$

for $1 \leq i \leq l$, and

$$\gamma_I \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j + \gamma_A a_i = 0, \tag{9}$$

for $l + 1 \leq i \leq u + l$.

The following is a generalization of the proof for Proposition 1 in (**?**). We have

$$f_{\mathbf{z},\gamma} = \operatorname{argmin}_{f \in \mathcal{H}_K} \frac{1}{l} \sum_{i=1}^{l} ||y_i - CK^*_{x_i} f||^2_{\mathcal{Y}} + \gamma_A ||f||^2_K + \gamma_I \langle \mathbf{f}, M\mathbf{f} \rangle_{\mathcal{W}(u+l)}. \tag{10}$$

With the operator $E_{C,\mathbf{x}}$, (10) is transformed into the minimization problem

$$f_{\mathbf{z},\gamma} = \operatorname{argmin}_{f \in \mathcal{H}_K} \frac{1}{l} ||E_{C,\mathbf{x}} f - \mathbf{y}||^2_{\mathcal{Y}^l} + \gamma_A ||f||^2_K + \gamma_I \langle \mathbf{f}, M\mathbf{f} \rangle_{\mathcal{W}^{u+l}}. \tag{11}$$

Proof of Proposition 1. By the Representer Theorem, (3) has a unique solution. Differentiating (11) and setting the derivative to zero gives

$$(E^*_{C,\mathbf{x}} E_{C,\mathbf{x}} + l\gamma_A I + l\gamma_I S^*_{\mathbf{x},u+l} M S_{\mathbf{x},u+l}) f_{\mathbf{z},\gamma} = E^*_{C,\mathbf{x}} \mathbf{y}.$$

By definition of the operators $E_{C,\mathbf{x}}$ and $S_{\mathbf{x}}$, this is

$$\sum_{i=1}^{l} K_{x_i} C^* C K^*_{x_i} f_{\mathbf{z},\gamma} + l\gamma_A f_{\mathbf{z},\gamma} + l\gamma_I \sum_{i=1}^{u+l} K_{x_i} (M\mathbf{f}_{\mathbf{z},\gamma})_i = \sum_{i=1}^{l} K_{x_i} C^* y_i,$$

which we rewrite as

$$f_{\mathbf{z},\gamma} = -\frac{\gamma_I}{\gamma_A} \sum_{i=1}^{u+l} K_{x_i} (M\mathbf{f}_{\mathbf{z},\gamma})_i + \sum_{i=1}^{l} K_{x_i} \frac{C^* y_i - C^* C K^*_{x_i} f_{\mathbf{z},\gamma}}{l\gamma_A}.$$

This shows that there are vectors a_i's in $\mathcal{W}$ such that

$$f_{\mathbf{z},\gamma} = \sum_{i=1}^{u+l} K_{x_i} a_i.$$

We have $f_{\mathbf{z},\gamma}(x_i) = \sum_{j=1}^{u+l} K(x_i, x_j) a_j$, and

$$(M\mathbf{f}_{\mathbf{z},\gamma})_i = \sum_{k=1}^{u+l} M_{ik} \sum_{j=1}^{u+l} K(x_k, x_j) a_j = \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j.$$

Also $K_{x_i}^* f_{\mathbf{z},\gamma} = f_{\mathbf{z},\gamma}(x_i) = \sum_{j=1}^{u+l} K(x_i, x_j)a_j$. Thus for $1 \leq i \leq l$:

$$a_i = -\frac{\gamma_I}{\gamma_A} \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j + \frac{C^* y_i - C^* C(\sum_{j=1}^{u+l} K(x_i, x_j) a_j)}{l\gamma_A},$$

which gives the formula

$$l\gamma_I \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j + C^* C(\sum_{j=1}^{u+l} K(x_i, x_j) a_j) + l\gamma_A a_i = C^* y_i.$$

Similarly, for $l+1 \leq i \leq u+l$,

$$a_i = -\frac{\gamma_I}{\gamma_A} \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j,$$

which is equivalent to

$$\gamma_I \sum_{j,k=1}^{u+l} M_{ik} K(x_k, x_j) a_j + \gamma_A a_i = 0.$$

This completes the proof. □

Proposition 2.

$$(\mathbf{C}^* \mathbf{C} J_l^{\mathcal{W},u+l} K[\mathbf{x}] + l\gamma_I M K[\mathbf{x}] + l\gamma_A I)\mathbf{a} = \mathbf{C}^* \mathbf{y}, \quad (12)$$

where $\mathbf{a} = (a_1, \ldots, a_{u+l})$, $\mathbf{y} = (y_1, \ldots, y_{u+l})$ are considered as column vectors in $\mathcal{W}^{u+l}$ and $\mathcal{Y}^{u+l}$, respectively, and $y_{l+1} = \cdots = y_{u+l} = 0$.

Proof of Proposition 2. This is straightforward to obtain from Proposition 1 using the operator-valued matrix formulation described in the main paper. □

Proposition 3. *For $C = \mathbf{c}^T \otimes I_P$, $\mathbf{c} \in \mathbb{R}^m$, $M_W = L \otimes I_P$, $M_B = I_{u+l} \otimes (M_m \otimes I_P)$, the system of linear equations (12) in Proposition 2 is equivalent to*

$$BA = Y_C, \quad (13)$$

where

$$B = \left((J_l^{u+l} \otimes \mathbf{c}\mathbf{c}^T) + l\gamma_B (I_{u+l} \otimes M_m) + l\gamma_W L\right) G[\mathbf{x}] + l\gamma_A I_{(u+l)m}, \quad (14)$$

which is of size $(u+l)m \times (u+l)m$, A is the matrix of size $(u+l)m \times P$ such that $\mathbf{a} = \mathrm{vec}(A^T)$, and Y_C is the matrix of size $(u+l)m \times P$ such that $\mathbf{C}^*\mathbf{y} = \mathrm{vec}(Y_C^T)$. $J_l^{u+l} : \mathbb{R}^{u+l} \to \mathbb{R}^{u+l}$ is a diagonal matrix of size $(u+l) \times (u+l)$, with the first l entries on the main diagonal being 1 and the rest being 0.

Proof of Proposition 3. Recall some properties of the Kronecker tensor product:

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \quad (15)$$

$$(A \otimes B)^T = A^T \otimes B^T, \quad (16)$$

and

$$\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B). \quad (17)$$

Thus the equation

$$AXB = C \quad (18)$$

is equivalent to

$$(B^T \otimes A)\mathrm{vec}(X) = \mathrm{vec}(C). \quad (19)$$

In our context,$\gamma_I M = \gamma_B M_B + \gamma_W M_W$, which is

$$\gamma_I M = \gamma_B I_{u+l} \otimes M_m \otimes I_P + \gamma_W L \otimes I_P.$$

$$\mathbf{C}^* = I_{u+l} \otimes C^*.$$

Using the property stated in Equation (16), we have for $C = \mathbf{c}^T \otimes I_P$,

$$\mathbf{C}^* = I_{u+l} \otimes \mathbf{c} \otimes I_P \in \mathbb{R}^{Pm(u+l) \times P(u+l)}, \quad (20)$$

$$C^* C = (\mathbf{c} \otimes I_P)(\mathbf{c}^T \otimes I_P) = (\mathbf{c}\mathbf{c}^T \otimes I_P).$$

So then

$$\mathbf{C}^*\mathbf{C} = (I_{u+l} \otimes \mathbf{c}\mathbf{c}^T \otimes I_P). \quad (21)$$

$$J_l^{\mathcal{W},u+l} = J_l^{u+l} \otimes I_m \otimes I_P. \quad (22)$$

It follows that

$$\mathbf{C}^*\mathbf{C} J_l^{\mathcal{W},u+l} = (J_l^{u+l} \otimes \mathbf{c}\mathbf{c}^T \otimes I_P). \quad (23)$$

Then

$$\mathbf{C}^*\mathbf{C} J_l^{\mathcal{W},u+l} K[\mathbf{x}] = (J_l^{u+l} \otimes \mathbf{c}\mathbf{c}^T) G[\mathbf{x}] \otimes I_P.$$

$$\gamma_I M K[\mathbf{x}] = (\gamma_B I_{u+l} \otimes M_m + \gamma_W L) G[\mathbf{x}] \otimes I_P.$$

Consider again now the system

$$(\mathbf{C}^*\mathbf{C} J_l^{\mathcal{W},u+l} K[\mathbf{x}] + l\gamma_I M K[\mathbf{x}] + l\gamma_A I)\mathbf{a} = \mathbf{C}^*\mathbf{y}.$$

The left hand side is

$$(B \otimes I_P)\mathrm{vec}(A^T),$$

where $\mathbf{a} = \mathrm{vec}(A^T)$, A is of size $(u+l)m \times P$ and

$$B = \left((J_l^{u+l} \otimes \mathbf{c}\mathbf{c}^T) + l\gamma_B (I_{u+l} \otimes M_m) + l\gamma_W L\right) G[\mathbf{x}] + l\gamma_A I_{(u+l)m}.$$

Then we have the linear system

$$(B \otimes I_P)\mathrm{vec}(A^T) = \mathrm{vec}(Y_C^T),$$

which, by properties (18) and (19), is equivalent to

$$A^T B^T = Y_C^T \Longleftrightarrow BA = Y_C.$$

This completes the proof. □

Remark 1. The vec operator is implemented by the flattening operation (:) in MATLAB. To compute the matrix Y_C^T, note that by definition

$$\mathrm{vec}(Y_C^T) = \mathbf{C}^*\mathbf{y} = (I_{u+l} \otimes C^*)\mathbf{y} = \mathrm{vec}(C^*Y),$$

where Y is the $P \times (u+l)$ matrix with the ith column being y_i, with

$$\mathbf{y} = \mathrm{vec}(Y).$$

Note that Y_C^T and C^*Y in general are not the same: Y_C^T is of size $P \times (u+l)m$, whereas C^*Y is of size $Pm \times (u+l)$.

2. Learning with General Bounded Linear Operators

The present framework generalizes naturally beyond the point evaluation operator

$$f(x) = K_x^* f.$$

Let $\mathcal{H}$ be a separable Hilbert space of functions on $\mathcal{X}$. We are *not* assuming that the functions in $\mathcal{H}$ are defined pointwise or with values in $\mathcal{W}$, rather we assume that $\forall x \in \mathcal{X}$, there is a bounded linear operator

$$E_x : \mathcal{H} \to \mathcal{W}, \quad ||E_x|| < \infty, \tag{24}$$

with adjoint $E_x^* : \mathcal{W} \to \mathcal{H}$. Consider the minimization

$$f_{\mathbf{z},\gamma} = \mathrm{argmin}_{\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} V(y_i, CE_{x_i}f) + \gamma_A ||f||^2_{\mathcal{H}} + \gamma_I \langle \mathbf{f}, M\mathbf{f}\rangle_{\mathcal{W}^{u+l}}, \quad \text{where} \quad \mathbf{f} = (E_{x_i}f)_{i=1}^{u+l}, \tag{25}$$

and its least square version

$$f_{\mathbf{z},\gamma} = \mathrm{argmin}_{\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} ||y_i - CE_{x_i}f||^2_{\mathcal{Y}} + \gamma_A ||f||^2_{\mathcal{H}} + \gamma_I \langle \mathbf{f}, M\mathbf{f}\rangle_{\mathcal{W}^{u+l}}. \tag{26}$$

Following are the corresponding Representer Theorem and Proposition stating the explicit solution for the least square case. When $\mathcal{H} = \mathcal{H}_K$, $E_x = K_x^*$, we recover Theorem 1 and Proposition 1, respectively.

Theorem 2. *The minimization problem (25) has a unique solution, given by $f_{\mathbf{z},\gamma} = \sum_{i=1}^{u+l} E_{x_i}^* a_i$ for some vectors $a_i \in \mathcal{W}$, $1 \le i \le u+l$.*

Proposition 4. *The minimization problem (26) has a unique solution $f_{\mathbf{z},\gamma} = \sum_{i=1}^{u+l} E_{x_i}^* a_i$, where the vectors $a_i \in \mathcal{W}$ are given by*

$$l\gamma_I \sum_{j,k=1}^{u+l} M_{ik} E_{x_k} E_{x_j}^* a_j + C^*C(\sum_{j=1}^{u+l} E_{x_i}E_{x_j}^* a_j) + l\gamma_A a_i = C^* y_i, \tag{27}$$

for $1 \le i \le l$, and

$$\gamma_I \sum_{j,k=1}^{u+l} M_{ik} E_{x_k} E_{x_j}^* a_j + \gamma_A a_i = 0, \tag{28}$$

for $l+1 \le i \le u+l$.

The reproducing kernel structures come into play through the following.

Lemma 1. *Let $E : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{W})$ be defined by*

$$E(x,t) = E_x E_t^*. \tag{29}$$

Then E is a positive definite operator-valued kernel.

Proof of Lemma 1. For each pair $(x,t) \in \mathcal{X} \times \mathcal{X}$, the operator $E(x,t)$ satisfies

$$E(t,x)^* = (E_t E_x^*)^* = E_x E_t^* = E(x,t).$$

For every set $\{x_i\}_{i=1}^N$ in $\mathcal{X}$ and $\{w_i\}_{i=1}^N$ in $\mathcal{W}$,

$$\sum_{i,j=1}^{N} \langle w_i, E(x_i,x_j)w_j\rangle_{\mathcal{W}} = \sum_{i,j=1}^{N} \langle w_i, E_{x_i}E_{x_j}^* w_j\rangle_{\mathcal{W}}$$

$$= \sum_{i,j=1}^{N} \langle E_{x_i}^* w_i, E_{x_j}^* w_j\rangle_{\mathcal{H}} = ||\sum_{i=1}^{N} E_{x_i}^* w_i||^2_{\mathcal{H}} \ge 0.$$

Thus E is an $\mathcal{L}(\mathcal{W})$-valued positive definite kernel. $\square$

Proofs of Theorem 2 and Proposition 4. These are entirely analogous to those of Theorem 1 and Proposition 1, respectively. Instead of the sampling operator $S_{\mathbf{x}}$, we consider the operator $E_{\mathbf{x}} : \mathcal{H} \to \mathcal{W}^l$, with

$$E_{\mathbf{x}} f = (E_{x_i} f)_{i=1}^l, \tag{30}$$

with the adjoint $E_{\mathbf{x}}^* : \mathcal{W}^l \to \mathcal{H}$ given by

$$E_{\mathbf{x}}^* \mathbf{b} = \sum_{i=1}^{l} E_{x_i}^* b_i. \tag{31}$$

for all $\mathbf{b} = (b_i)_{i=1}^l \in \mathcal{W}^l$. The operator $E_{C,\mathbf{x}} : \mathcal{H} \to \mathcal{Y}^l$ is now defined by

$$E_{C,\mathbf{x}} f = (CE_{x_1}f, \dots, CE_{x_l}f). \tag{32}$$

The adjoint $E_{C,\mathbf{x}}^* : \mathcal{Y}^l \to \mathcal{H}$ is

$$E_{C,\mathbf{x}}^* \mathbf{b} = \sum_{i=1}^{l} E_{x_i}^* C^* b_i, \tag{33}$$

for all $\mathbf{b} \in \mathcal{Y}^l$, and $E_{C,\mathbf{x}}^* E_{C,\mathbf{x}} : \mathcal{H} \to \mathcal{H}$ is

$$E_{C,\mathbf{x}}^* E_{C,\mathbf{x}} f = \sum_{i=1}^{l} E_{x_i}^* C^* C E_{x_i} f. \tag{34}$$

We then apply all the steps in the proofs of Theorem 1 and Proposition 1 to get the desired results. $\square$

Remark 2. We stress that in general, the function $f_{\mathbf{z},\gamma}$ is *not* defined pointwise, which is the case in the following example. Thus one cannot make a statement about $f_{\mathbf{z},\gamma}(x)$ for all $x \in \mathcal{X}$ without additional assumptions.

Example 1. (**?**) $\mathcal{X} = [0,1]$, $\mathcal{H} = L^2(\mathcal{X})$, $\mathcal{W} = \mathbb{R}$. Let $G : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be continuous and

$$E_x f = \int_0^1 G(x,t)f(t)dt. \tag{35}$$

for $f \in \mathcal{H}$. One has the reproducing kernel

$$E_x E_t^* = E(x,t) = \int_0^1 G(x,u)G(t,u)du. \tag{36}$$

3. The Degenerate Case

This section deals with the Gaussian kernel $k(x,t) = \exp\left(-\frac{||x-t||^2}{\sigma^2}\right)$ when $\sigma \to \infty$ and other kernels with similar behavior. We show that the matrix A in Proposition 3 has an analytic expression. This can be used to verify the correctness of an implementation of our algorithm. At $\sigma = \infty$, for each pair (x,t), we have

$$K(x,t) = I_{Pm}, \tag{37}$$

and

$$f_{\mathbf{z},\gamma}(x) = \sum_{i=1}^{u+l} K(x_i,x)a_i = \sum_{i=1}^{u+l} a_i. \tag{38}$$

Thus $f_{\mathbf{z},\gamma}$ is a constant function. Let us examine the form of the coefficients a_i's for the case

$$C = \frac{1}{m}\mathbf{e}_m^T \otimes I_P.$$

We have

$$G[\mathbf{x}] = \mathbf{e}_{u+l}\mathbf{e}_{u+l}^T \otimes I_m.$$

For $\gamma_I = 0$, we have

$$B = \frac{1}{m^2}(J_l^{u+l} \otimes \mathbf{e}_m\mathbf{e}_m^T)(\mathbf{e}_{u+l}\mathbf{e}_{u+l}^T \otimes I_m) + l\gamma_A I_{(u+l)m},$$

which is

$$B = \frac{1}{m^2}(J_l^{u+l}\mathbf{e}_{u+l}\mathbf{e}_{u+l}^T \otimes \mathbf{e}_m\mathbf{e}_m^T) + l\gamma_A I_{(u+l)m}.$$

Equivalently,

$$B = \frac{1}{m^2}(J_{ml}^{(u+l)m}\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T) + l\gamma_A I_{(u+l)m}.$$

The inverse of B in this case has a closed form:

$$B^{-1} = \frac{I_{(u+l)m}}{l\gamma_A} - \frac{J_{ml}^{(u+l)m}\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T}{l^2 m\gamma_A(m\gamma_A+1)}, \tag{39}$$

where we have used the identity

$$\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T J_{ml}^{(u+l)m}\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T = ml\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T. \tag{40}$$

We have thus

$$A = B^{-1}Y_C = \left(\frac{I_{(u+l)m}}{l\gamma_A} - \frac{J_{ml}^{(u+l)m}\mathbf{e}_{(u+l)m}\mathbf{e}_{(u+l)m}^T}{l^2 m\gamma_A(m\gamma_A+1)}\right)Y_C. \tag{41}$$

Thus in this case we have an analytic expression for the coefficient matrix A, as we claimed.

References

Minh, H.Q. and Sindhwani, V. Vector-valued manifold regularization. In *ICML*, 2011.

Rosenberg, D., Sindhwani, V., Bartlett, P., and Niyogi, P. A kernel for semi-supervised learning with multi-view point cloud regularization. *IEEE Sig. Proc. Mag.*, 26(5):145–150, 2009.

Wahba, G. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667, 1977.