Sampling and spectral approximation

Bertrand Gauthier

CIRM, Marseille, April 30 - May 4 2018

Cardiff University - School of Mathematics

Table of contents

Intro

General framework

- $\mathcal{X}$ a general measurable set.
- $\mathcal{H}$ a separable RKHS of real-valued functions on $\mathcal{X}$, with (measurable) reproducing kernel $K(\cdot,\cdot)$.
- $\mathcal{M}$ the set of all measures on $\mathcal{X}$, and

$$\mathcal{T}(K) = \left\{\mu \in \mathcal{M} \middle| \tau_\mu = \int_{\mathcal{X}} K(x,x)\mathrm{d}\mu(x) < +\infty\right\}.$$

Continuous inclusion and integral operator

For all $\mu \in \mathcal{T}(K)$, we have $K(\cdot,\cdot) \in L^2(\mu \otimes \mu)$; in addition, for all $h \in \mathcal{H}$, we have $h \in L^2(\mu)$ and $\|h\|^2_{L^2(\mu)} \leqslant \tau_\mu \|h\|^2_{\mathcal{H}}$. We can thus define the symmetric and positive-semidefinite integral operator T_μ on $L^2(\mu)$,

$$\forall f \in L^2(\mu), \forall x \in \mathcal{X}, T_\mu[f](x) = \int_{\mathcal{X}} K(x,t) f(t) \mathrm{d}\mu(t).$$

- $T_\mu[f] \in \mathcal{H} \subset L^2(\mu)$, and for all $h \in \mathcal{H}$, $\left(h \middle| T_\mu[f]\right)_{\mathcal{H}} = \left(h \middle| f\right)_{L^2(\mu)}$.

So what?

For a given $\mu \in \mathscr{T}(K)$, how to compute an accurate approximation of the main eigenpairs of T_μ?

Idea: "Use a quadrature", i.e., search for a discrete measure $\nu \in \mathscr{T}(K)$ supported by a small number of points, and use the spectral approximation of T_ν to approximate the one of T_μ...

Problems: *What are "good measures" ν?* ***How to design such masures?*** *What does "use the spectral approximation of T_ν to approximate the one of T_μ" mean? Can we have a money-back guarantee?*

Related work

[GS18] Bertrand Gauthier and Johan A.K Suykens. *Optimal quadrature-sparsification for integral operator approximation*. -preprint- `https://hal.archives-ouvertes.fr/hal-01416786v3`. 2018

More about T_μ

- $\mathcal{H}_{0\mu} = \left\{ h \in \mathcal{H} \middle| \|h\|_{L^2(\mu)} = 0 \right\}$ and $\mathcal{H}_\mu = \mathcal{H}_{0\mu}^{\perp_{\mathcal{H}}}$; $\rightarrow \mathcal{H} = \mathcal{H}_\mu \oplus \mathcal{H}_{0\mu}$.
- $\{\lambda_k\}_{k \in \mathbb{I}_\mu^+}$ set (at most countable) of all strictly positive eigenvalues of T_μ.
- $\{\varphi_k\}_{k \in \mathbb{I}_\mu^+}$ a set of (canonically extended) associated eigenfunctions, orthonormalised in $L^2(\mu)$, (i.e., $\varphi_k \in \mathcal{H}$, and $(\varphi_k | \varphi_{k'})_{L^2(\mu)} = \delta_{k,k'}$); $\rightarrow \{\sqrt{\lambda_k}\varphi_k\}_{k \in \mathbb{I}_\mu^+}$ o.n.b. of the subspace $\mathcal{H}_\mu$ of $\mathcal{H}$.
- The reproducing kernel $K_\mu(\cdot, \cdot)$ of $\mathcal{H}_\mu$ is given by, for all x and $t \in \mathcal{X}$,

$$K_\mu(x,t) = \sum_{k \in \mathbb{I}_\mu^+} \lambda_k \varphi_k(x) \varphi_k(t).$$

- For all $\mu \in \mathcal{T}(K)$, T_μ is an Hilbert-Schmidt op. on $L^2(\mu)$, and also on $\mathcal{H}$.

Squared-kernel discrepancy and spectral approximation

Plan

Measure of the approximation error

We denote by $\mathrm{HS}(\mathcal{H})$ the Hilbert space of all Hilbert-Schmidt operators on $\mathcal{H}$. Let μ and $\nu \in \mathcal{T}(K)$; for an o.n.b. $\{h_j\}_{j\in\mathbb{I}}$ of $\mathcal{H}$ (with $\mathbb{I}$ a general, at most countable, index set), the Hilbert-Schmidt inner product between the operators T_μ and T_ν on $\mathcal{H}$ is given by

$$\left(T_\mu | T_\nu\right)_{\mathrm{HS}(\mathcal{H})} = \sum_{j\in\mathbb{I}} \left(T_\mu[h_j] | T_\nu[h_j]\right)_{\mathcal{H}}.$$

Squared-kernel discrepancy

For μ and $\nu \in \mathcal{T}(K)$, we define

$$D_{K^2}(\mu,\nu) = \|T_\mu - T_\nu\|^2_{\mathrm{HS}(\mathcal{H})};$$

in particular,

$$D_{K^2}(\mu,\nu) = \|K\|^2_{L^2(\mu\otimes\mu)} + \|K\|^2_{L^2(\nu\otimes\nu)} - 2\|K\|^2_{L^2(\mu\otimes\nu)},$$

with $\|K\|^2_{L^2(\mu\otimes\nu)} = \int_{\mathcal{X}\times\mathcal{X}} \left(K(x,t)\right)^2 \mathrm{d}\mu(x)\mathrm{d}\nu(t) = (T_\mu|T_\nu)_{\mathrm{HS}(\mathcal{H})}$.

A property

Weighted spectral sum-of-squared-errors-type criterion

Let μ and $\nu \in \mathscr{T}(K)$ be such that $\mathcal{H}_\nu \subset \mathcal{H}_\mu$, then

$$D_{K^2}(\mu,\nu) = \sum_{k\in\mathbb{I}_\mu^+} \lambda_k \left\| T_\mu[\varphi_k] - T_\nu[\varphi_k] \right\|_{\mathcal{H}}^2,$$

and, in addition, $\sum_{k\in\mathbb{I}_\mu^+} \lambda_k \left\| T_\mu[\varphi_k] - T_\nu[\varphi_k] \right\|_{L^2(\mu)}^2 \leqslant \tau_\mu D_{K^2}(\mu,\nu)$.

Remark: the squared kernel $K^2(\cdot,\cdot) = \big(K(\cdot,\cdot)\big)^2$ is also symmetric and positive semidefinite, and is thus related to a RKHS $\mathcal{G}$. Many of the properties of the integral operators defined from $K(\cdot,\cdot)$ can be interpreted in the RKHS $\mathcal{G}$.

General remarks

- $D_{K^2}(\mu,\nu) \geqslant 0$ and $D_{K^2}(\mu,\mu) = 0$ $\rightarrow$ the "raw" minimisation of $\nu \mapsto D_{K^2}(\mu,\nu)$ on $\mathscr{T}(K)$ is of no interest (i.e., *"overall, the best approximation of T_μ is T_μ itself"*).
- For a given $n \in \mathbb{N}^*$, the search of an optimal discrete measure ν_n^* such that $D_{K^2}(\mu,\nu_n^*)$ is minimal among all measures ν_n supported by n points is in general a difficult (i.e., usually non-convex) optimisation problem on $(\mathscr{X} \times \mathbb{R}_+)^n$.

Nevertheless:

- If we assume that the support of ν is included in a fixed finite set of points $\mathcal{S} = \{x_k\}_{k=1}^N$ (with, in practice, N large), the squared-kernel discrepancy can be expressed as a convex quadratic function.
- Sparsity of the approximate measure can then be promoted through the introduction of an ℓ^1-type penalisation, and **the induced penalised squared-kernel-discrepancy minimisation problems consist in convex quadratic minimisation problems**.

Plan

Approximate operator

We consider two measures μ and $\nu \in \mathscr{T}(K)$, corresponding to an initial operator T_μ and an approximate operator T_ν.

Eigendecomposition of T_ν

Denote by $\{\vartheta_l\}_{l\in\mathbb{I}_\nu^+}$ the strictly positive eigenvalues of T_μ, and let $\{\psi_l\}_{l\in\mathbb{I}_\nu^+}$ be an $L^2(\nu)$-orthonormal of associated eigenfunctions, i.e., $T_\nu[\psi_l] = \vartheta_l\psi_l \in \mathcal{H}$, with $\vartheta_l > 0$ and $(\psi_l|\psi_{l'})_{L^2(\nu)} = \delta_{l,l'}$. We shall refer to the functions ψ_l as the approximate eigendirections of T_μ induced by T_ν.

Normalised approximate eigendirections

For all $l \in \mathbb{I}_\nu^+$ such that $\|\psi_l\|_{L^2(\mu)} > 0$, we introduce $\widehat{\varphi}_l = \psi_l/\|\psi_l\|_{L^2(\mu)}$.

Notice that if $\mathcal{H}_\nu \subset \mathcal{H}_\mu$, then we necessarily have $\|\psi_l\|_{L^2(\mu)} > 0$ for all $l \in \mathbb{I}_\nu^+$. If $\|\psi_l\|_{L^2(\mu)} = 0$, then $\psi_l \in \mathcal{H}_{0\mu}$ and thus $T_\mu[\psi_l] = 0$; such directions are therefore of no use in approximating the eigendirections related to the strictly positive eigenvalues of T_μ.

Remark: orthogonality test

The normalised approximate eigenfunctions $\widehat{\varphi}_l$ are by definition orthogonal in $L^2(\nu)$ and in $\mathcal{H}$, and verify $\|\widehat{\varphi}_l\|_{L^2(\mu)} = 1$. Controlling their orthogonality in $L^2(\mu)$ offer a relatively affordable way to assess their accuracy. Indeed, accurate normalised approximate eigenfunctions $\widehat{\varphi}_l$ should be almost mutually orthogonal in $L^2(\mu)$; this condition is however only a necessary condition.

It also is very instructive to try to estimate the eigenvalue, for the operator T_μ, related to an approximate eigendirection ψ_l induced by T_ν...

Approximate eigenvalues

Geometric approximate eigenvalues

For all $l \in \mathbb{I}_\nu^+$ such that $\|\psi_l\|_{L^2(\mu)} > 0$, we define

$$\begin{aligned}
\widehat{\lambda}_l^{[1]} &= 1/\|\widehat{\varphi}_l\|_{\mathcal{H}}^2 = \vartheta_l \|\psi_l\|_{L^2(\mu)}^2 = \big(\sqrt{\vartheta_l}\psi_l \big| T_\mu[\sqrt{\vartheta_l}\psi_l]\big)_{\mathcal{H}} = \big(T_\nu[\psi_l] \big| T_\mu[\psi_l]\big)_{\mathcal{H}}, \\
\widehat{\lambda}_l^{[2]} &= \big\|T_\mu[\sqrt{\vartheta_l}\psi_l]\big\|_{\mathcal{H}}, \\
\widehat{\lambda}_l^{[3]} &= \big(\widehat{\varphi}_l \big| T_\mu[\widehat{\varphi}_l]\big)_{L^2(\mu)} = \big\|T_\mu[\widehat{\varphi}_l]\big\|_{\mathcal{H}}^2 = \big(\widehat{\lambda}_l^{[2]}\big)^2 / \widehat{\lambda}_l^{[1]}, \\
\widehat{\lambda}_l^{[4]} &= \big\|T_\mu[\widehat{\varphi}_l]\big\|_{L^2(\mu)} = \big\|T_\mu[\psi_l]\big\|_{L^2(\mu)} / \big\|\psi_l\big\|_{L^2(\mu)},
\end{aligned}$$

and if $\|\psi_l\|_{L^2(\mu)} = 0$, we set $\widehat{\lambda}_l^{[1]} = \widehat{\lambda}_l^{[2]} = \widehat{\lambda}_l^{[3]} = \widehat{\lambda}_l^{[4]} = 0$.

A property

Theorem 1

For all $l \in \mathbb{I}_\nu^+$, we have $\widehat{\lambda}_l^{[1]} \leqslant \widehat{\lambda}_l^{[2]} \leqslant \widehat{\lambda}_l^{[3]} \leqslant \widehat{\lambda}_l^{[4]}$, with equality when ψ_l is an eigendirection of T_μ; in case of equality, the approximation $\widehat{\lambda}_l^{[\cdot]}$ corresponds exactly to the eigenvalue of T_μ related to the eigendirection ψ_l (in particular, equality between the four geometric approximate eigenvalues occurs as soon as two of them are equal).

In addition, for $\lambda \in \mathbb{R}$, the function

$$\lambda \mapsto \left\| \lambda \sqrt{\vartheta_l} \psi_l - T_\mu[\sqrt{\vartheta_l} \psi_l] \right\|_{\mathcal{H}}^2 = \lambda^2 - 2\lambda \widehat{\lambda}_l^{[1]} + \left(\widehat{\lambda}_l^{[2]}\right)^2$$

reaches its minimum at $\lambda = \widehat{\lambda}_l^{[1]}$. In the same way, if $\|\psi_l\|_{L^2(\mu)} > 0$ (so that the normalised approximate eigenfunction $\widehat{\varphi}_l$ is well-defined), the function

$$\lambda \mapsto \left\| \lambda \widehat{\varphi}_l - T_\mu[\widehat{\varphi}_l] \right\|_{L^2(\mu)}^2 = \lambda^2 - 2\lambda \widehat{\lambda}_l^{[3]} + \left(\widehat{\lambda}_l^{[4]}\right)^2$$

reaches its minimum at $\lambda = \widehat{\lambda}_l^{[3]}$.

Plan

Invariance of the spectral approximation (1)

Proportional approximate measures lead to the same spectral approximation of T_μ and to the same approximate kernel $K_\nu(\cdot,\cdot)$. For a given measure $\nu \in \mathcal{T}(K)$, we can thus search the value of $c \geqslant 0$ for which $D_{K^2}(\mu, c\nu)$ is minimal.

Theorem 2 (part 1)

We denote by c_ν the argument of the minimum of the function $\phi : c \mapsto \phi(c) = D_{K^2}(\mu, c\nu)$, with $c \in \mathbb{R}$; we have

$$c_\nu = \frac{(T_\mu | T_\nu)_{\mathrm{HS}(\mathcal{H})}}{\|T_\nu\|^2_{\mathrm{HS}(\mathcal{H})}} = \frac{\|K\|^2_{L^2(\mu\otimes\nu)}}{\|K\|^2_{L^2(\nu\otimes\nu)}}, \text{ and } \phi\big(c_\nu\big) = \|K\|^2_{L^2(\mu\otimes\mu)} - \frac{\|K\|^4_{L^2(\mu\otimes\nu)}}{\|K\|^2_{L^2(\nu\otimes\nu)}}.$$

In particular, $T_{c_\nu \nu} = c_\nu T_\nu$ is the orthogonal projection, in $\mathrm{HS}(\mathcal{H})$, of T_μ onto the linear subspace spanned by T_ν; in addition, $\|c_\nu T_\nu - \frac{1}{2} T_\mu\|^2_{\mathrm{HS}(\mathcal{H})} = \frac{1}{4}\|K\|^2_{L^2(\mu\otimes\mu)}$, so that, in $\mathrm{HS}(\mathcal{H})$, the approximate operator $c_\nu T_\nu$ lies on a sphere centered at $\frac{1}{2}T_\mu$ and with radius $\frac{1}{2}\|T_\mu\|_{\mathrm{HS}(\mathcal{H})}$.

Invariance of the spectral approximation (2)

Theorem 2 (part 2)

Assuming that $\mathcal{H}_\nu \subset \mathcal{H}_\mu$ (for simplicity and without loss of generality), we have

$$\sum_{l\in\mathbb{I}_\nu^+} \widehat{\lambda}_l^{[1]} \left\| T_\mu[\widehat{\varphi}_l] - \widehat{\lambda}_l^{[1]} \widehat{\varphi}_l \right\|_{\mathcal{H}}^2 \leqslant \sum_{l\in\mathbb{I}_\nu^+} \widehat{\lambda}_l^{[1]} \left\| T_\mu[\widehat{\varphi}_l] - c_\nu \vartheta_l \widehat{\varphi}_l \right\|_{\mathcal{H}}^2 \leqslant D_{K^2}(\mu, c_\nu \nu),$$

$$\text{and } \sum_{l\in\mathbb{I}_\nu^+} \widehat{\lambda}_l^{[1]} \left\| T_\mu[\widehat{\varphi}_l] - \widehat{\lambda}_l^{[3]} \widehat{\varphi}_l \right\|_{L^2(\mu)}^2 \leqslant \sum_{l\in\mathbb{I}_\nu^+} \widehat{\lambda}_l^{[1]} \left\| T_\mu[\widehat{\varphi}_l] - \widehat{\lambda}_l^{[1]} \widehat{\varphi}_l \right\|_{L^2(\mu)}^2 \leqslant \tau_\mu D_{K^2}(\mu, c_\nu \nu).$$

In view of Theorem 2, in order to approximate the eigenvalues of the initial operator T_μ induced by the eigendecomposition of T_ν, we could also define the "globally rescaled" approximate eigenvalues $\{c_\nu \vartheta_l\}_{l\in\mathbb{I}_\nu^+}$; in comparison, the approximate eigenvalues $\{\widehat{\lambda}_l^{[1]}\}_{l\in\mathbb{I}_\nu^+}$ are "individually rescaled".

Quadrature-sparsification as quadratic programming

Plan

Discrete operators

We only consider measures with support included in a fixed set $\mathcal{S} = \{x_k\}_{k=1}^N$. We assume that $\mu = \sum_{k=1}^N \omega_k \delta_{x_k}$, with $\boldsymbol{\omega} > 0$, and that $\nu = \sum_{k=1}^N \upsilon_k \delta_{x_k}$, with $\boldsymbol{\upsilon} \geqslant 0$. We then have

$$D_{K^2}(\mu, \nu) = (\boldsymbol{\omega} - \boldsymbol{\upsilon})^T \mathbf{S} (\boldsymbol{\omega} - \boldsymbol{\upsilon}),$$

where $\mathbf{S}$ is the kernel matrix with i, j entry $\mathbf{S}_{i,j} = K^2(x_i, x_j) \geqslant 0$. Notice that $\mathbf{S} = \mathbf{K} * \mathbf{K}$ (Hadamard product), where $\mathbf{K}$ is the kernel matrix defined by $K(\cdot, \cdot)$ and $\mathcal{S}$, i.e., $\mathbf{K}_{i,j} = K(x_i, x_j)$.

For a given (fixed) $\boldsymbol{\omega}$, we introduce

$$D(\boldsymbol{\upsilon}) = \frac{1}{2} (\boldsymbol{\omega} - \boldsymbol{\upsilon})^T \mathbf{S} (\boldsymbol{\omega} - \boldsymbol{\upsilon}).$$

Regularised problem

Consider a **penalisation direction** $\mathbf{d} = (d_1, \cdots, d_N)^T \in \mathbb{R}^N$, with $\mathbf{d} > 0$, and a regularisation parameter $\alpha \geqslant 0$.

Regularised SKD minimisation

$$\underset{\boldsymbol{v} \in \mathbb{R}^N}{\text{minimise}} \; D_\alpha(\boldsymbol{v}) = \frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{v})^T \mathbf{S}(\boldsymbol{\omega} - \boldsymbol{v}) + \alpha \mathbf{d}^T \boldsymbol{v} \text{ subject to } \boldsymbol{v} \geqslant 0. \tag{1}$$

Since $\boldsymbol{v} \geqslant 0$, the term $\mathbf{d}^T \boldsymbol{v}$ can be interpreted as a weighted ℓ^1-type regularisation. Notice that if $\mathbf{d} = \text{diag}(\mathbf{K})$, then $\mathbf{d}^T \boldsymbol{v} = \text{trace}(T_\nu)$.

Constrained problem

Constrained SKD minimisation

For $0 \leqslant \varkappa \leqslant \mathbf{d}^T \boldsymbol{\omega}$, we can equivalently consider

$$\underset{\boldsymbol{\upsilon} \in \mathbb{R}^N}{\text{minimise}} \; D(\boldsymbol{\upsilon}) = \tfrac{1}{2}(\boldsymbol{\omega} - \boldsymbol{\upsilon})^T \mathbf{S}(\boldsymbol{\omega} - \boldsymbol{\upsilon}) \text{ subject to } \boldsymbol{\upsilon} \geqslant 0 \text{ and } \mathbf{d}^T \boldsymbol{\upsilon} = \varkappa. \quad (2)$$

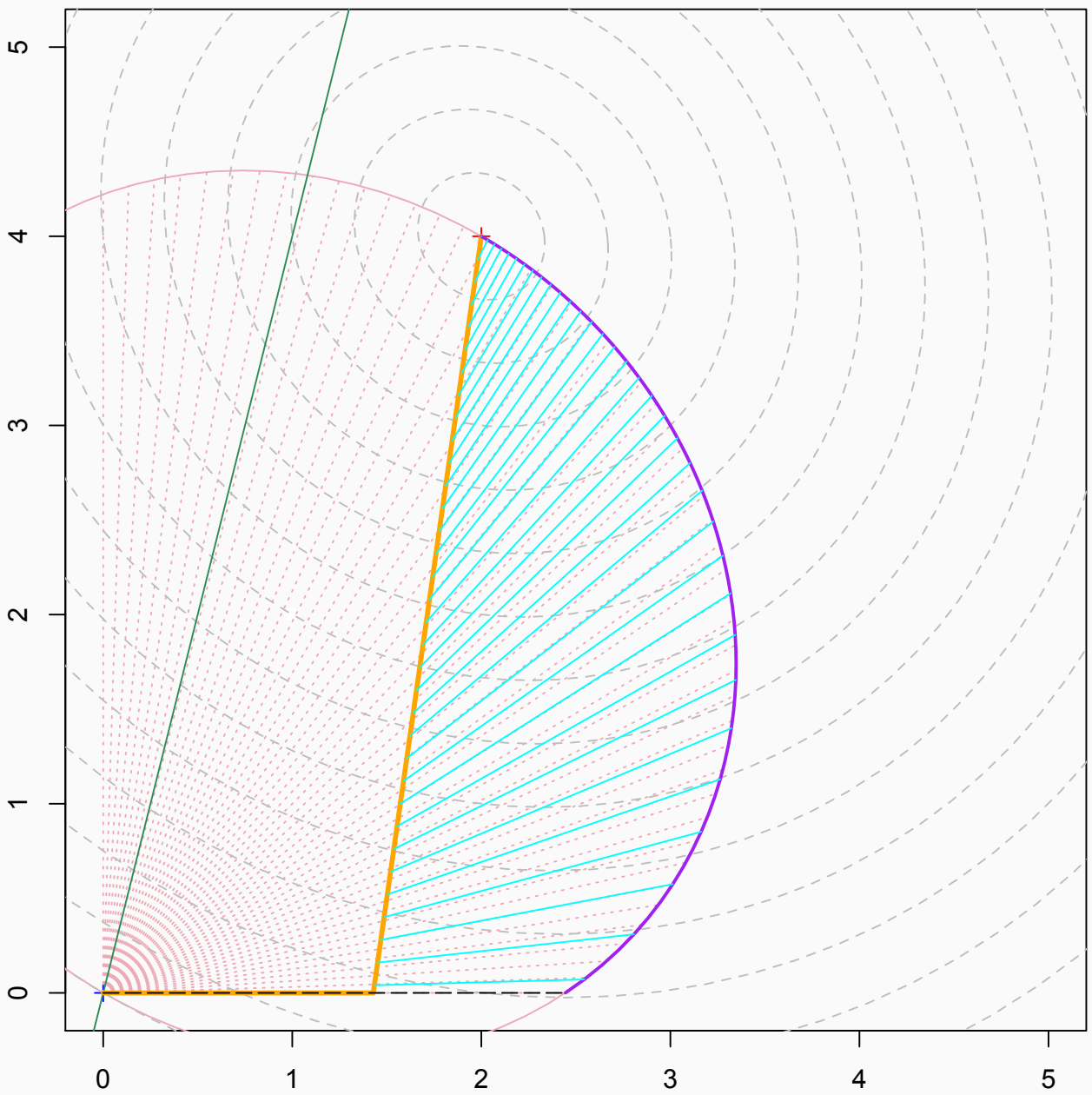
The penalised and constrained formulations are equivalent.

Penalisation and spectral approximation

No-free-lunch theorem

One can formally show that under "reasonable conditions", increasing the amount of penalisation tends to increase the sparsity of the approximate measure at the expense of monotonically reducing the overall accuracy of the induced spectral approximation.

Illustration: $N = 2$



Plan

SVM related to the regularised problem

We recall that we denote by $\mathcal{G}$ the RKHS associated with the squared kernel $K^2(\cdot,\cdot)$, and that for $\mu \in \mathcal{T}(K)$, the function $g_\mu \in \mathcal{G}$ is defined as $g_\mu(x) = \int_{\mathcal{X}} K^2(t,x)\mathrm{d}\mu(t)$, see Lemma 1. We introduce

$$\begin{aligned} \underset{g\in\mathcal{G}}{\text{minimise}} \quad & \tfrac{1}{2}\|g\|_{\mathcal{G}}^2 + (g|g_\mu)_{\mathcal{G}} \\ \text{subject to} \quad & g(x_k) \geqslant -\alpha d_k \text{ for all } k \in \{1,\cdots,N\}. \end{aligned} \tag{3}$$

Relation between the primal and dual solutions

If $\boldsymbol{v}_\alpha^*$ is a solution to (1) with $\alpha \geqslant 0$, then $g_\alpha^*(x) = \sum_{k=1}^N [\boldsymbol{v}_\alpha^* - \boldsymbol{\omega}]_k K^2(x,x_k)$ is the solution to (3).

Introducing the change of variable $\check{g} = g + g_\mu \in \mathcal{G}$, we can define

$$\underset{\check{g}\in\mathcal{G}}{\text{minimise}} \frac{1}{2}\|\check{g}\|_{\mathcal{G}}^2 \text{ subject to } \check{g}(x_k) \geqslant g_\mu(x_k) - \alpha d_k \text{ for all } k, \tag{4}$$

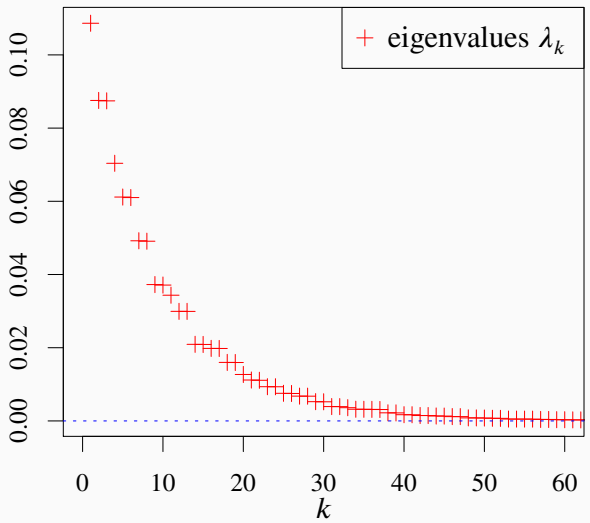
which is an equivalent formulation for (3), with solution $\check{g}_\alpha^*(x) = \sum_{k=1}^N [\boldsymbol{v}_\alpha^*]_k K^2(x,x_k) = g_{\nu_\alpha^*}(x)$.

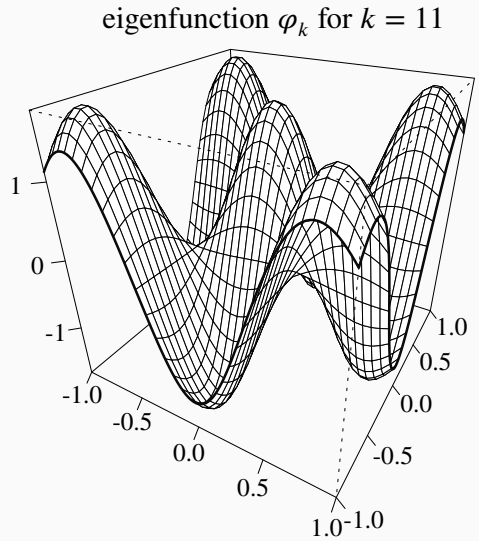
Numerical examples

Plan

Settings

- $S = \{x_k\}_{k=1}^N$ consists of the $N = 2016$ first points of a uniform Halton sequence on $[-1, 1]^2$.
- $\boldsymbol{\omega} = \mathbb{1}/N$ (we recall that $\mu = \sum_k \omega_k \delta_{x_k}$).
- Gaussian kernel $K(x, y) = \exp(-\ell \|x - y\|^2)$, where $\|x - y\|$ is the Euclidean norm on $\mathbb{R}^2$, and with $\ell = 1/0.16$.
- $\mathbf{d} = \mathbb{1}$.





How does optimal measures look?

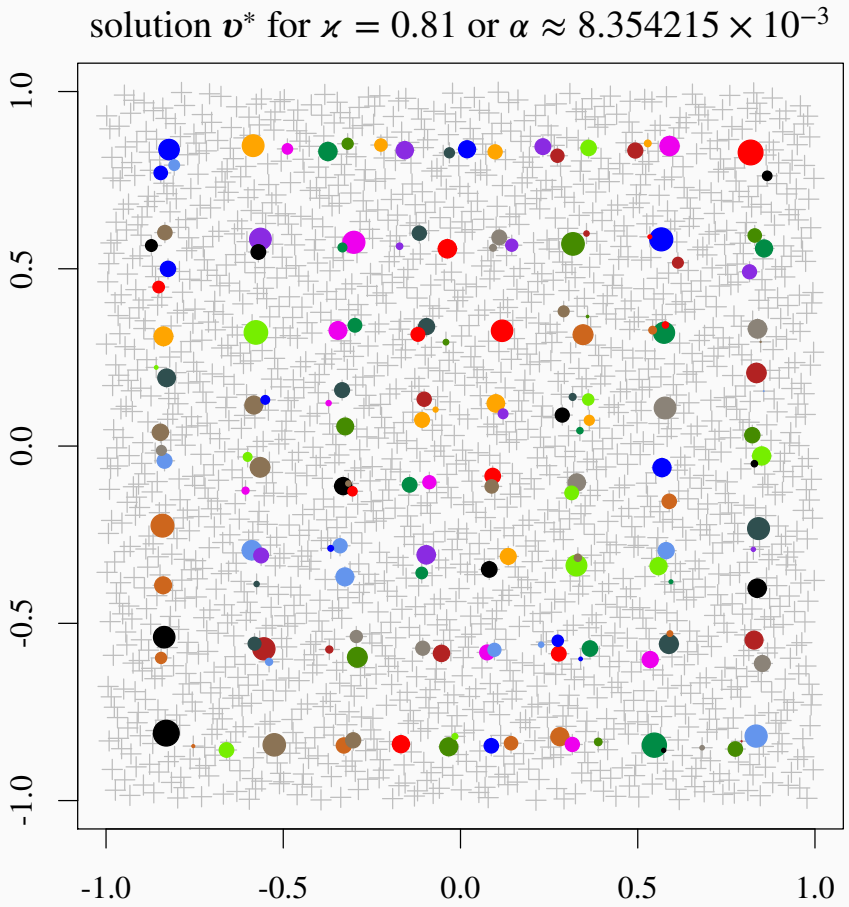


Figure 1: Graphical representation (Gaussian kernel, $\boldsymbol{\omega} = \mathbb{1}/N$ and $\mathbf{d} = \mathbb{1}$) of the solution υ^* to problem (2) with $\varkappa = 0.81$, or equivalently, to problem (1) with $\alpha \approx 8.354215 \times 10^{-3}$. The grey crosses represent the points in $\mathcal{S}$ and the filled dots are the strictly positive components of υ^* (surface being proportional to υ_k^*).

Orthogonality test

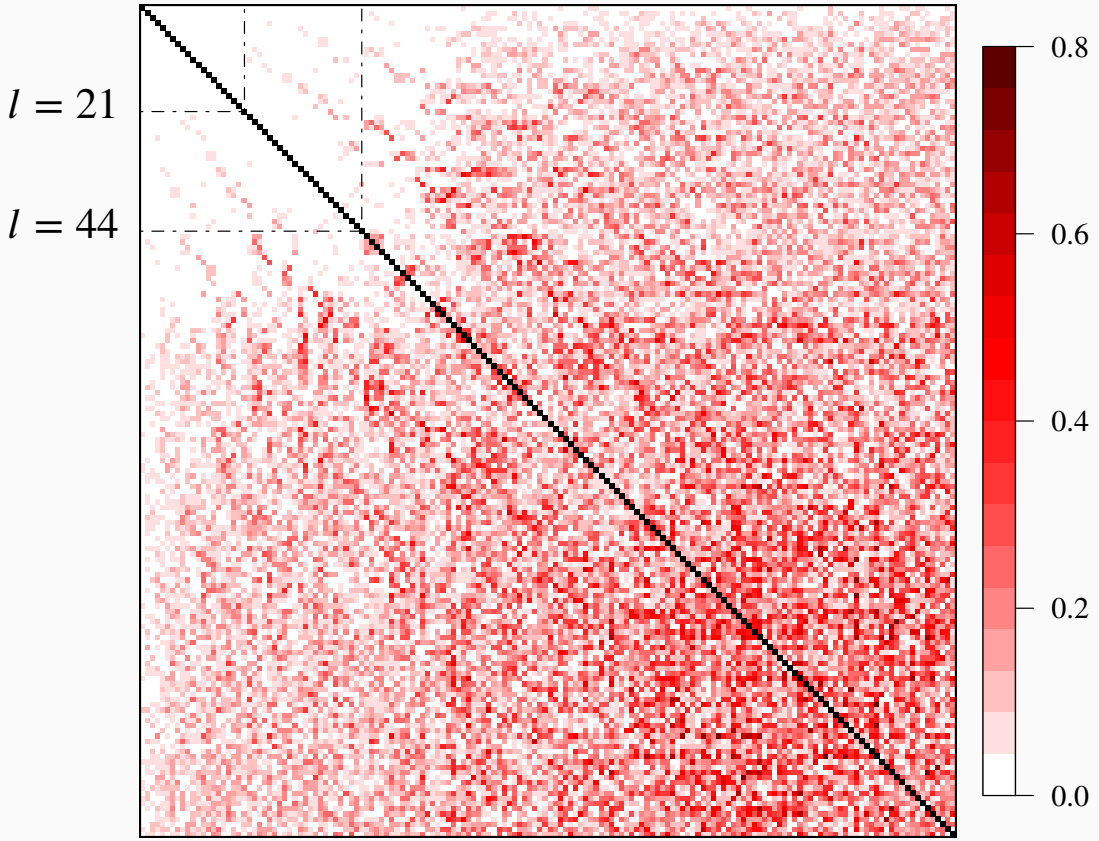


Figure 2: Graphical representation of the matrix with l, l' entry $\left|\left(\widehat{\varphi}_l \middle| \widehat{\varphi}_{l'}\right)_{L^2(\mu)}\right|$ for the 160 normalised approximate eigendirections induced by the solution $\boldsymbol{\upsilon}^*$ presented in Figure 1 (i.e., $\varkappa = 0.81$).

Geometric approximate eigenvalues and spectral-ratio test

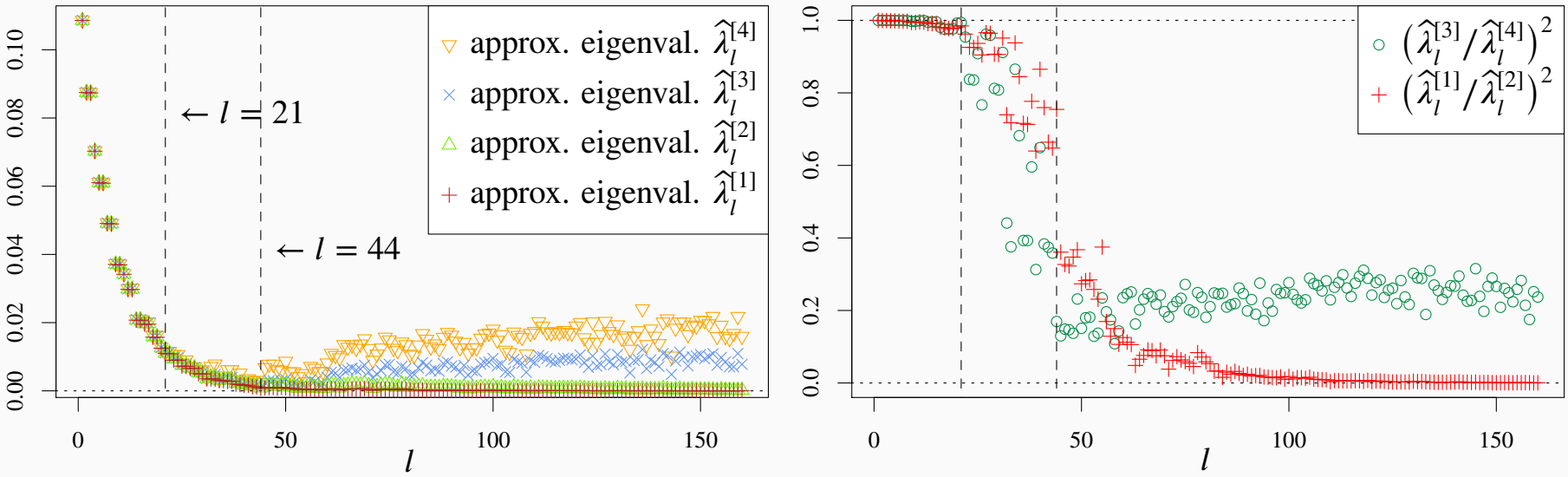


Figure 3: Approximate eigenvalues $\widehat{\lambda}_l^{[1]}$, $\widehat{\lambda}_l^{[2]}$, $\widehat{\lambda}_l^{[3]}$ and $\widehat{\lambda}_l^{[4]}$ induced by the solution υ^* presented in Figure 1 (left); ratios $\left(\widehat{\lambda}_l^{[1]}/\widehat{\lambda}_l^{[2]}\right)^2$ and $\left(\widehat{\lambda}_l^{[3]}/\widehat{\lambda}_l^{[4]}\right)^2$ highlighting the accuracy of the approximate eigendirections ψ_l of T_μ (right).

About the approximate eigenvalues

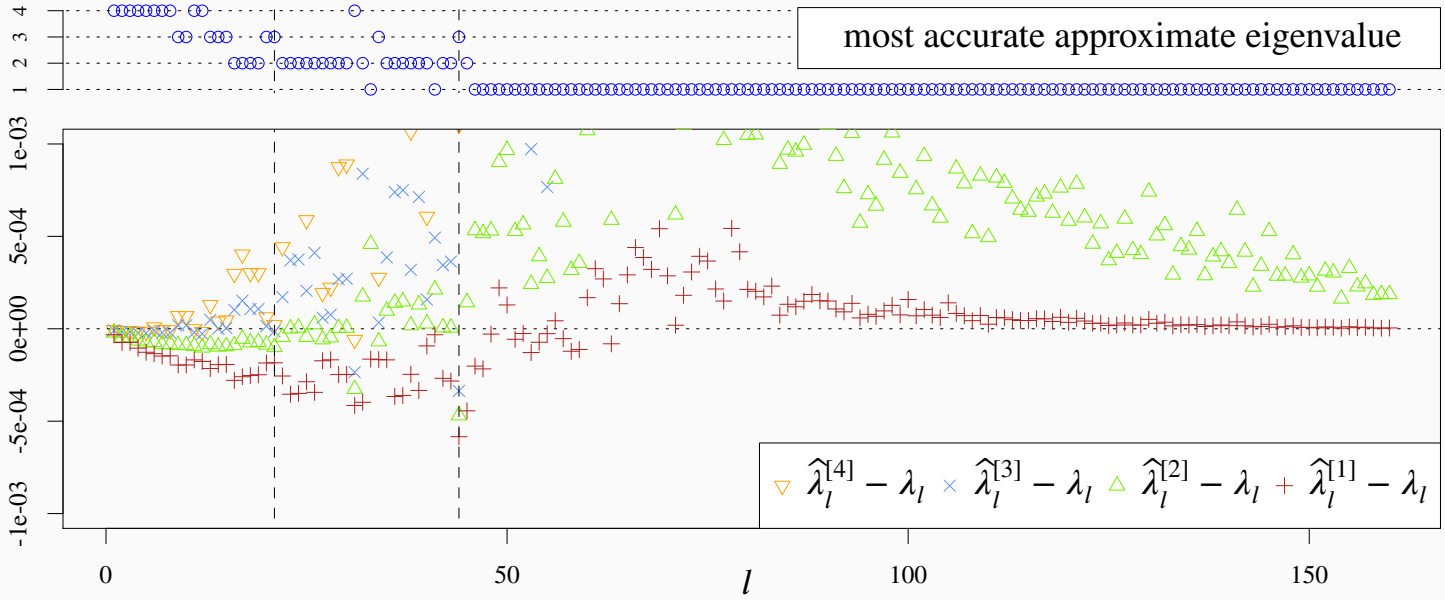


Figure 4: Errors $\widehat{\lambda}_l^{[\cdot]} - \lambda_l$ for the geometric approximate eigenvalues induced by the solution v^* presented in Figure 1 (bottom), and indication of the most accurate (smallest absolute error) approximation among $\widehat{\lambda}_l^{[1]}$, $\widehat{\lambda}_l^{[2]}$, $\widehat{\lambda}_l^{[3]}$ and $\widehat{\lambda}_l^{[4]}$ (top).

Error on the eigenvectors

Table 1: Approximation error $\|\widehat{\varphi}_l - \varphi_l\|^2_{L^2(\mu)}$, with $1 \leqslant l \leqslant 20$, for the normalised approximate eigendirections induced by the solution υ^* presented in Figure 1 (i.e., $\varkappa = 0.81$); the values of l grouped together correspond to pairs of eigendirections related to the approximation of an eigensubspace of dimension two.

l	1	2 and 3		4	5 and 6		7 and 8		9 and 10	
$\widehat{\lambda}_l^{[1]}$	0.10861	0.08747	0.08737	0.07028	0.06103	0.06089	0.04907	0.04895	0.03706	0.03692
$\|\widehat{\varphi}_l - \varphi_l\|^2_{L^2(\mu)}$	0.00017	0.00035	0.00035	0.00056	0.00054	0.00120	0.00115	0.00117	0.00245	0.00243
l	11	12 and 13		14 and 15		16 and 17		18 and 19		20
$\widehat{\lambda}_l^{[1]}$	0.03418	0.02976	0.02971	0.02073	0.02070	0.01954	0.01954	0.01573	0.01571	0.01251
$\|\widehat{\varphi}_l - \varphi_l\|^2_{L^2(\mu)}$	0.00196	0.00128	0.00448	0.00438	0.00456	0.00773	0.00685	0.00843	0.00830	0.00711

Regularisation path

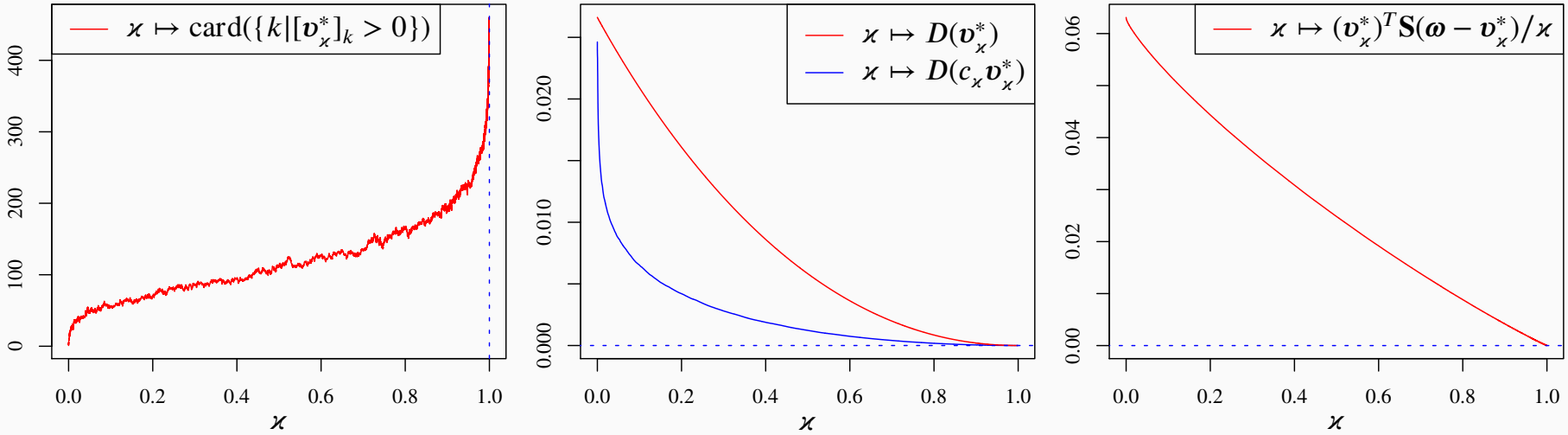


Figure 5: For the two-dimensional example (Gaussian kernel, $\boldsymbol{\omega} = \mathbb{1}/N$ and $\mathbf{d} = \mathbb{1}$), graphical representation of the $12\,818$ first events of the regularisation path related to problem (2) for increasing $\varkappa$; number of strictly positive components of $\boldsymbol{v}_\varkappa^*$ as function of $\varkappa$ (left); graph of $\varkappa \mapsto D(\boldsymbol{v}_\varkappa^*)$ and $\varkappa \mapsto D(c_\varkappa \boldsymbol{v}_\varkappa^*)$ (middle), and relation between $\varkappa$ and the parameter α of problem (1) (right).

Greedy pairwise merging

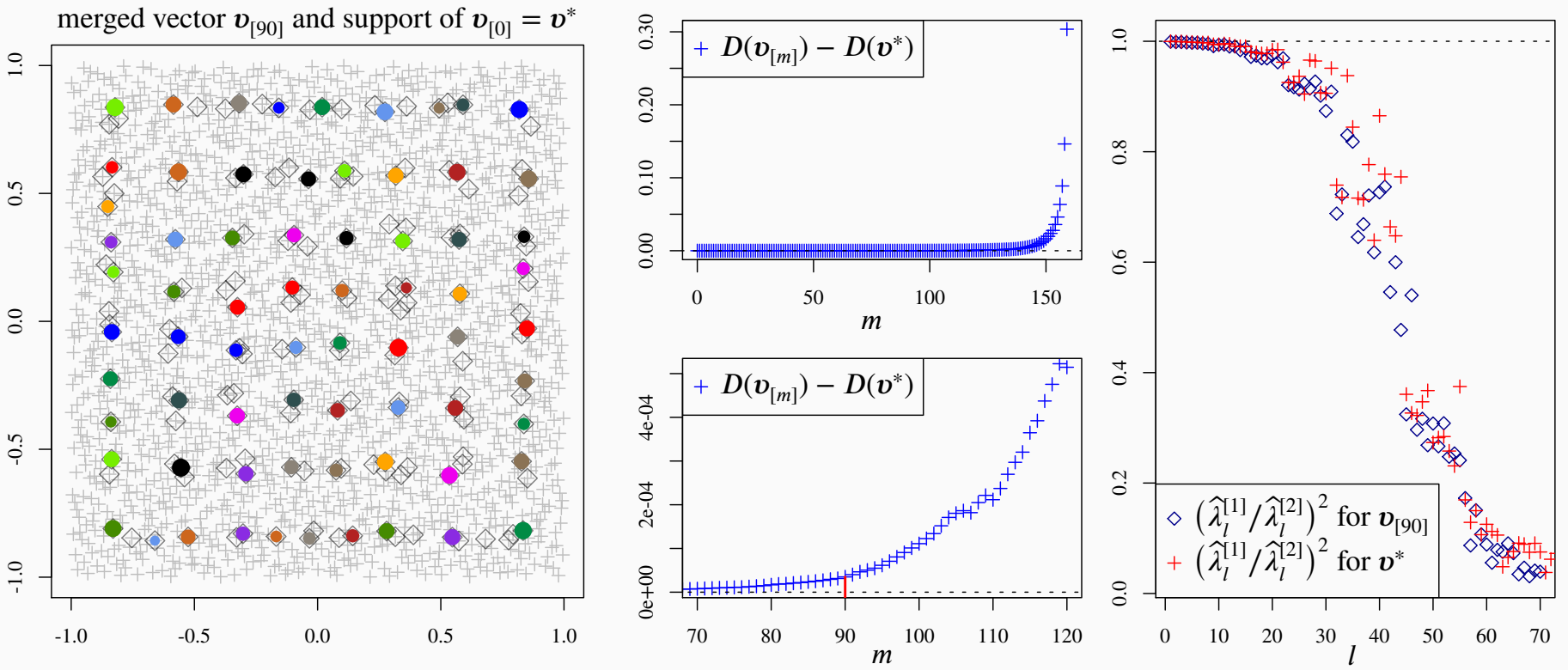


Figure 6: Merged measure $\boldsymbol{v}_{[90]}$ obtained after 90 iterations of the strong-pairwise-merging strategy applied to the solution $\boldsymbol{v}^*$ presented in Figure 1; the grey diamonds indicate the support of $\boldsymbol{v}^*$ (left). Increase of the cost $D(\cdot)$ induced by each merging iteration, for the whole 159 iterations (top-middle), and zoom around the 90-th iteration (bottom-middle). Representation of the ratios $\left(\hat{\lambda}_l^{[1]}/\hat{\lambda}_l^{[2]}\right)^2$ obtained from the merged vector $\boldsymbol{v}_{[90]}$ and comparison with the same ratios for the solution $\boldsymbol{v}^*$ (right).

Penalisation direction

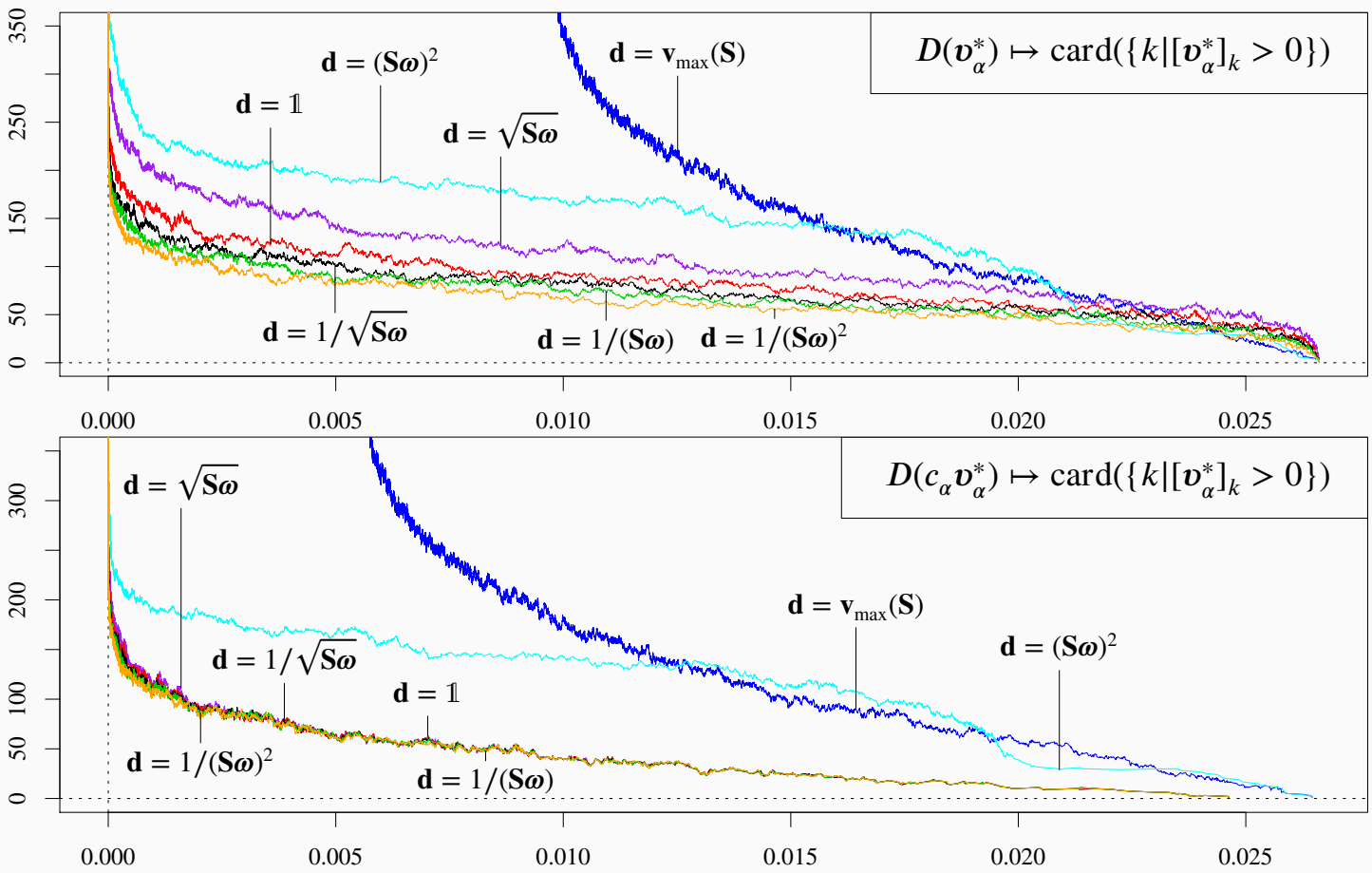


Figure 7: Number of strictly positive components of the solution $\boldsymbol{v}_\alpha^*$ to problem (1) as function of the squared-kernel discrepancy $D(\boldsymbol{v}_\alpha^*)$ (top), and of the conic squared-kernel discrepancy $D(c_\alpha \boldsymbol{v}_\alpha^*)$ (bottom) for various penalisation vectors $\mathbf{d}$; all the curves have been obtained thanks to the regularisation-path strategy.

Plan

Settings

- $\mathcal{S} = \{x_k\}_{k=1}^N$ consists of the $N = 500\,000$ test points (in $\mathbb{R}^{18}$) of the standardised UCI-SUSY dataset.
- $\boldsymbol{\omega} = \mathbb{1}/N$.
- Gaussian kernel $K(x, y) = \exp(-\ell \|x - y\|^2)$ with $\ell = 1/0.4$.
- $\mathbf{d} = \mathbb{1}$.
- Computations (CPU) performed on a 2015 desktop endowed with an Intel Core i7-4790 processor with 16 GB of RAM; "full C" implementation.

Computation of the dual distortion term $\mathbf{S}\boldsymbol{\omega}$ $\quad\rightarrow\quad$ 5 665.6 seconds.

We compute an approximate solution (vertex-exchange strategy) for the constrained problem (2) with $\varkappa = 0.3$; we perform four consecutive batches of $50\,000$ iterations each, the solver being initialised at $\widetilde{\boldsymbol{\upsilon}} = \mathbf{e}_1$. After $200\,000$ iterations (i.e., at the end of the fourth batch), the obtained approximate solution $\widehat{\boldsymbol{\upsilon}}^*$ verifies $D(\widehat{\boldsymbol{\upsilon}}^*) = 3.931629 \times 10^{-5}$ and has $n = 20\,664$ strictly positive components.

Kernelised VEX

Table 2: Information relative to the approximate solutions to problem (2) with $\varkappa = 0.3$ returned by the VEX strategy for four consecutive batches of $50\,000$ iterations (the solver is initialised at a vertex of the polytope); for each batch, execution time, total number of iterations, Frank-Wolfe error bound ϵ and number n of strictly positive components of the approximate solution.

	batch 1	batch 2	batch 3	batch 4
time (in sec.)	$1\,148.7$	$1\,158.3$	$1\,158.5$	$1\,159.1$
total nb. of it.	$50\,000$	$100\,000$	$150\,000$	$200\,000$
ϵ	3.1413×10^{-7}	6.5477×10^{-8}	2.7049×10^{-8}	7.0928×10^{-9}
n	$19\,721$	$20\,619$	$20\,693$	$20\,674$

To enhance sparsity, we perform a weak-pairwise merging of the approximate solution $\widehat{\boldsymbol{v}}^*$; the computation of $20\,673$ merging iterations took 78.86 seconds. The merged solution $\boldsymbol{v}_{[13674]}$ is supported by $7\,000$ points and $D(\boldsymbol{v}_{[13674]}) = D(\widehat{\boldsymbol{v}}^*) + 5.271960 \times 10^{-7}$ (i.e., increase of only 1.34%).

Spectral approximation (1)

- Computing the 300 first normalised approximate eigenvectors $\widehat{\mathbf{v}}_l$ of $\mathbf{KW}$ (with $\mathbf{W} = \mathrm{diag}(\boldsymbol{\omega})$) induced by $\boldsymbol{\upsilon}_{[13674]}$ (i.e., $\widehat{\mathbf{v}}_l \in \mathbb{R}^N$ is the vector corresponding to $\widehat{\varphi}_l$) took $3\,278.2$ seconds (time for canonical extension and rescaling), and we thus also obtain the approximate eigenvalues $\widehat{\lambda}_l^{[1]}$.
- For l and $l' \in \{1, \cdots, 300\}$, we have $\max_{l \neq l'} |(\widehat{\varphi}_l | \widehat{\varphi}_{l'})_{L^2(\mu)}| \approx 0.003734$, so that we can expect the approximations $\widehat{\varphi}_l$ to be relatively accurate.
- To access precisely their accuracy, we compute $T_\mu[\widehat{\varphi}_l]$ (i.e., $\mathbf{KW}\widehat{\mathbf{v}}_l$) for these 300 first approximate eigendirections; this operation took $191\,622.3$ seconds (i.e., around 53 hours).

Spectral approximation (2)

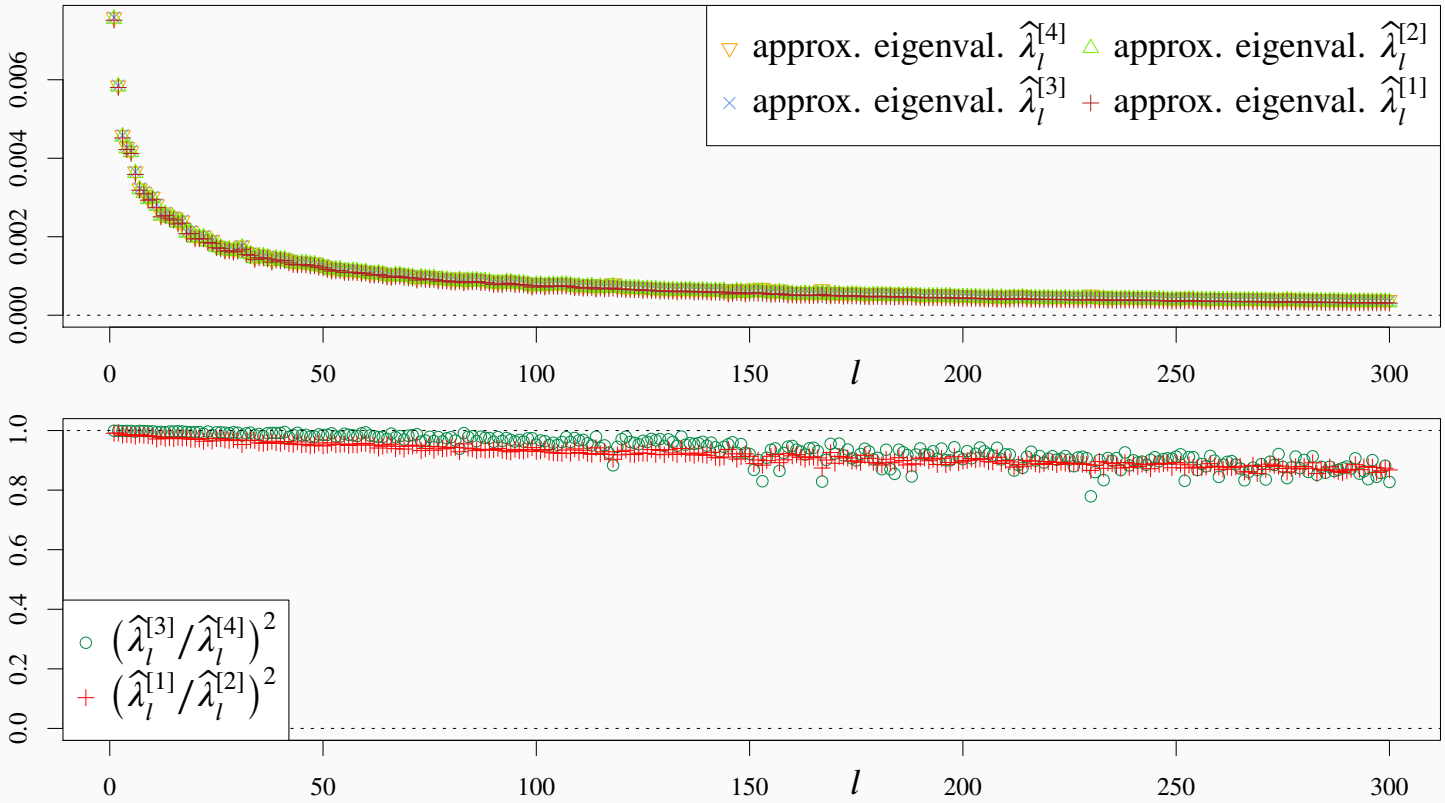


Figure 8: For the test subsample of the SUSY dataset, graphical representation of the 300 first approximate eigenvalues $\widehat{\lambda}_l^{[\cdot]}$ induced by the merged solution $\boldsymbol{v}_{[13674]}$ obtained from the approximate solution $\widehat{\boldsymbol{v}}^*$ to problem (2) with $\varkappa = 0.3$ (top); ratios $\left(\widehat{\lambda}_l^{[1]}/\widehat{\lambda}_l^{[2]}\right)^2$ and $\left(\widehat{\lambda}_l^{[3]}/\widehat{\lambda}_l^{[4]}\right)^2$ measuring the accuracy of the underlying approximate eigendirections (bottom). We only use 7000 points among 500 000.

Conclusion

Conclusion

Contribution

- QP-based strategy for "quadrature-sparsification".
- Analogy with kernel-LASSO and one-class SVM models.
- Spectral approximation with controlled error.

Numerical thought

- Main bottleneck of the approach: computation of the dual distortion term $\mathbf{S}\boldsymbol{\omega}$; this can nevertheless be massively parallelised, and GPU could be used.
- Once the dual distortion term is known, sparse solutions can be obtained readily.
- Assessing the accuracy of an approximate eigendirection trough the computation of the four associated geometric approximate eigenvalues can also prove challenging (same complexity as the distortion term); this operation is nevertheless optional, and the more affordable orthogonality test might be performed to detect poorly approximated eigendirections.

Thank you.