Sampling and low-rank approximation

Bertrand Gauthier

Cardiff University - School of Mathematics

MoDA13, July 9th-14th, 2023

Table of contents

Motivations

- $(\mathscr{X}, \mathcal{A})$ a general measurable space.
- μ a general measure on $(\mathscr{X}, \mathcal{A})$.
- $K : \mathscr{X} \times \mathscr{X} \to \mathbb{C}$ a positive-semidefinite (PSD) kernel.

Integral operator defined by K and μ.

$$\mathcal{L}_{K,\mu}[f](x) = \int_{\mathscr{X}} K(x,t) f(t) \mathrm{d}\mu(t),$$

with $f : \mathscr{X} \to \mathbb{C}$ and $x \in \mathscr{X}$.

Remark: This class of operators encompasses the PSD matrices (case $\mathscr{X} = [N]$, and $\mu = \sum_{j=1}^{N} \delta_j$, $N \in \mathbb{N}$).

Notation: $[N] = \{1, \cdots, N\}$.

Problem: How to design accurate low-rank approximations of operators of the form $\mathcal{L}_{K,\mu}$?

Remark: Rank-optimal approximations correspond to truncated spectral expansions; hence, such approximations can only be implemented for operators/matrices for which an SVD is available beforehand.

Two ways to tackle the problem:

- approximation of the measure μ;
- approximation of the kernel K.

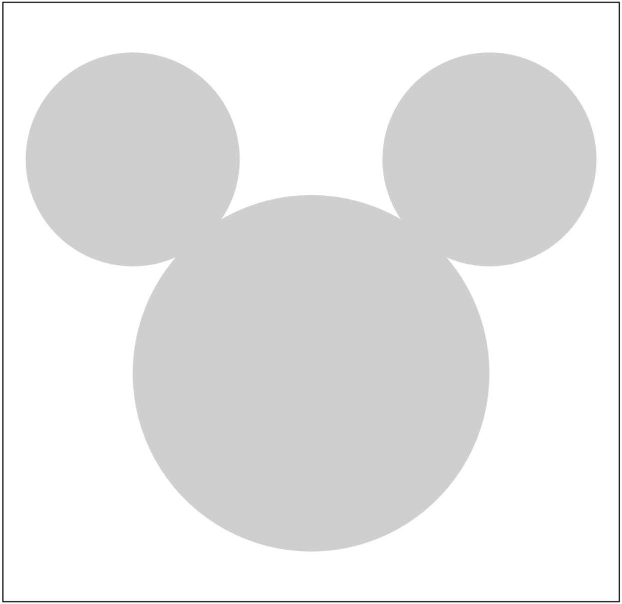
Sampling-based approximations

Notations: Let $\mathcal{H}$ be the RKHS associated with K.
For $t \in \mathscr{X}$, let $k_t \in \mathcal{H}$ be defined as $k_t(x) = K(x, t)$, $x \in \mathscr{X}$.

From a sample $\{s_1, \cdots, s_m\} \subseteq \mathscr{X}$, $m \in \mathbb{N}$, we may:

- approximate μ by $\nu = \sum_{j=1}^{m} \upsilon_j \delta_{s_j}$, $\upsilon_j \in \mathbb{C}$;
- approximate the kernel K by the reproducing kernel of the subspace $\mathrm{span}_{\mathbb{C}}\{k_{s_1}, \cdots, k_{s_m}\} \subseteq \mathcal{H}$.

New problem: How to design (sparse) samples leading to accurate approximations?



Hilbert-Schmidt operators on RKHSs

- $\mathcal{H}$, separable RKHS with reproducing kernel K.
- $\mathrm{HS}(\mathcal{H})$, Hilbert space of all Hilbert-Schmidt (HS) op. on $\mathcal{H}$.
- $\mathcal{G}$, RKHS associated with $|K|^2$ (squared-modulus kernel);

$$|K|^2(x,t) = \left|K(x,t)\right|^2 = K(x,t)\overline{K(x,t)}, \ x \text{ and } t \in \mathscr{X}.$$

- For $a, b \in \mathcal{H}$, let $T_{a,b} \in \mathrm{HS}(\mathcal{H})$ be the rank-1 linear operator given by $T_{a,b}[h] = a\langle b \,|\, h\rangle_{\mathcal{H}}$, $h \in \mathcal{H}$. Set $S_b = T_{b,b}$.

Singular value decomposition

An operator $T \in \mathrm{HS}(\mathcal{H})$ always admits a decomposition of the form $T = \sum_{i \in \mathbb{I}} \sigma_i T_{u_i, v_i}$, $\mathbb{I} \subseteq \mathbb{N}$, where $\{\sigma_i\}_{i \in \mathbb{I}} \in \ell^2(\mathbb{I})$ is the set of all strictly-positive singular values of T, and where $\{u_i\}_{i \in \mathbb{I}}$ and $\{v_i\}_{i \in \mathbb{I}}$ are two orthonormal systems in $\mathcal{H}$.

The map Gamma...

For $T = \sum_{i \in \mathbb{I}} \sigma_i T_{u_i, v_i} \in \mathrm{HS}(\mathcal{H})$, define the $\mathbb{C}$-valued function

$$\Gamma[T](x) = \sum_{i \in \mathbb{I}} \sigma_i u_i(x) \overline{v_i(x)}, x \in \mathscr{X}.$$

. . . is a natural coisometry from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{G}$

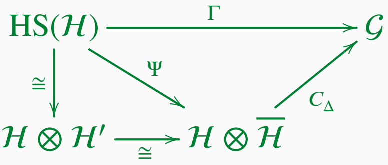
The map Γ is a natural coisometry from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{G}$, with initial space $\mathcal{I}(\Gamma) = \overline{\mathrm{span}_{\mathbb{C}}\{S_{k_x} | x \in \mathscr{X}\}}^{\mathrm{HS}(\mathcal{H})} \subseteq \mathrm{HS}(\mathcal{H})$.
For all $T \in \mathrm{HS}(\mathcal{H})$ and $x \in \mathscr{X}$, we have

$$\Gamma[T](x) = \langle S_{k_x} \,|\, T \rangle_{\mathrm{HS}(\mathcal{H})} = \langle k_x \,|\, T[k_x] \rangle_{\mathcal{H}} = T[k_x](x).$$

Through Γ, operators in $\mathcal{I}(\Gamma)$ are isometrically (and bijectively) represented as functions in the RKHS $\mathcal{G}$.

Remark:

- $\overline{\mathcal{H}}$, the RKHS associated with $\overline{K}$ (conjugate RKHS).
- $\overline{\mathcal{H}}$ is isometric to $\mathcal{H}'$, the continuous dual of $\mathcal{H}$.
- $\mathrm{HS}(\mathcal{H})$ is isometric to $\mathcal{H} \otimes \overline{\mathcal{H}}$.
- $\mathcal{G}$ is the product of $\mathcal{H}$ and $\overline{\mathcal{H}}$.
- $C_\Delta : \mathcal{H} \otimes \overline{\mathcal{H}} \to \mathcal{G}$, the pullback along the diagonal.



Basic properties:

- if $T \in \mathrm{HS}(\mathcal{H})$ is self-adjoint, then $\Gamma[T]$ is real-valued;
- if $T \in \mathrm{HS}(\mathcal{H})$ is PSD, then $\Gamma[T]$ is nonnegative;
- if $T \in \mathrm{HS}(\mathcal{H})$ is PSD and $\Gamma[T] = 0$, then $T = 0$; and
- if $T \in \mathcal{I}(\Gamma)$, then $T^* \in \mathcal{I}(\Gamma)$.

Remark: The map Γ is also well-defined when all the involved Hilbert spaces are real. We in this case have $\mathcal{I}(\Gamma) = \overline{\mathrm{span}_{\mathbb{R}}\{S_{k_x} | x \in \mathscr{X}\}}^{\mathrm{HS}(\mathcal{H})}$, and the operators in $\mathcal{I}(\Gamma)$ are self-adjoint; also, if $T^* = -T$, then $\Gamma[T] = 0$. By comparison, in the complex case, if $T^* = -T$, then the function $\Gamma[T]$ is pure-imaginary.

Isometric representation of integral operators

Measurability conditions

- For all $t \in \mathscr{X}$, $k_t \in \mathcal{H}$ is measurable.
- The diagonal of K is measurable.

For a measure μ on $(\mathscr{X}, \mathcal{A})$, define

$$\tau_\mu = \int_{\mathscr{X}} K(t,t)\mathrm{d}|\mu|(t) \in \mathbb{R}_{\geqslant 0} \cup \{+\infty\};$$

and let $\mathcal{T}_+(K)$, $\mathcal{T}(K)$ and $\mathcal{T}_{\mathbb{C}}(K)$ be the sets of all nonnegative, signed and complex measures such that τ_μ is finite.
Set $\mathcal{T}_{\mathbb{F}}(K) = \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$.

For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, set

$$L_\mu = \int_{\mathscr{X}} S_{k_t} \mathrm{d}\mu(t) \in \mathrm{HS}(\mathcal{H}) \quad \text{and} \quad g_\mu = \int_{\mathscr{X}} |k_t|^2 \mathrm{d}\mu(t) \in \mathcal{G};$$

in particular, for $h \in \mathcal{H}$ and $x \in \mathscr{X}$,

$$L_\mu[h](x) = \int_{\mathscr{X}} K(x,t)h(t)\mathrm{d}\mu(t) \text{ and } g_\mu(x) = \int_{\mathscr{X}} \left|K(x,t)\right|^2 \mathrm{d}\mu(t);$$

also, L_μ is trace-class, with $\mathrm{trace}(|L_\mu|) \leqslant \tau_\mu$.

Remark: For all $g \in \mathcal{G}$, $\langle g_\mu \,|\, g \rangle_{\mathcal{G}} = \int_{\mathscr{X}} g(t)\mathrm{d}\overline{\mu}(t)$.

Isometric representation

For all $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $L_\mu \in \mathcal{I}(\Gamma)$ and $\Gamma[L_\mu] = g_\mu$.

Proof: $\Gamma^*[g_\mu] = L_\mu$. □

Generalised integral probability metric (IPM)

Set $B_{\mathcal{G}} = \left\{ g \in \mathcal{G} \middle| \|g\|_{\mathcal{G}} \leqslant 1 \right\}$, and introduce

$$\mathfrak{M}_{\mathcal{G}}(\mu, \nu) = \sup_{g \in \mathcal{B}_{\mathcal{G}}} \left| \int_{\mathcal{X}} g(t) \mathrm{d}\mu(t) - \int_{\mathcal{X}} g(t) \mathrm{d}\nu(t) \right|, \ \mu \text{ and } \nu \in \mathcal{T}_{\mathbb{F}}(K).$$

Quadrature approximation as generalised IPM

$$\|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})} = \|g_\mu - g_\nu\|_{\mathcal{G}} = \mathfrak{M}_{\mathcal{G}}(\mu, \nu).$$

Remark: $\mathfrak{M}_{\mathcal{G}}(\mu, \nu) = \mathfrak{M}_{\mathcal{G}}(\overline{\mu}, \overline{\nu})$ (as $|K|^2$ is $\mathbb{R}$-valued).

Measures and projections

Projections defined by measures

For $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, set $\mathcal{H}_\nu = \overline{\mathrm{range}(L_{|\nu|})}^{\mathcal{H}}$ and let P_ν be the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_\nu$. Also, let K_ν be the reproducing kernel of $\mathcal{H}_\nu$.

Through P_ν, and in addition to L_ν, a measure $\nu \in \mathcal{T}_{\mathbb{F}}(K)$ also defines the approximations $P_\nu L_\mu$, $L_\mu P_\nu$ or $P_\nu L_\mu P_\nu$ of L_μ.

Remark: We have $\overline{\mathrm{range}(L_\nu)}^{\mathcal{H}} \subseteq \overline{\mathrm{range}(L_{|\nu|})}^{\mathcal{H}}$ and $L_\nu = P_\nu L_\nu = L_\nu P_\nu = P_\nu L_\nu P_\nu$; also, if $\nu = \sum_{i=1}^{m} \upsilon_i \delta_{s_i}$, with $\upsilon_i \in \mathbb{C}$, $\upsilon_i \neq 0$, then $\mathcal{H}_\nu = \mathrm{span}_{\mathbb{C}}\{k_{s_1}, \cdots, k_{s_n}\}$.

For $h \in \mathcal{H}$ and $x \in \mathscr{X}$, we have

$$P_\nu L_\mu[h](x) = \int_{\mathscr{X}} K_\nu(x,t)h(t)\mathrm{d}\mu(t),$$

so that $P_\nu L_\mu \in \mathrm{HS}(\mathcal{H})$ can be regarded as an integral operator defined by K_ν and μ. The following inequalities hold:

$$\|L_\mu - P_\nu L_\mu\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|L_\mu - P_\nu L_\mu P_\nu\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})}.$$

Remark:

$$\|L_\mu - P_\nu L_\mu\|^2_{\mathrm{HS}(\mathcal{H})} = \iint_{\mathscr{X}} [K(x,t) - K_\nu(x,t)]K(t,x)\mathrm{d}\mu(t)\mathrm{d}\overline{\mu}(x);$$

$$\|L_\mu - P_\nu L_\mu P_\nu\|^2_{\mathrm{HS}(\mathcal{H})} = \iint_{\mathscr{X}} |K(x,t)|^2 - |K_\nu(x,t)|^2\mathrm{d}\mu(t)\mathrm{d}\overline{\mu}(x).$$

Nonnegative measures and partial L^2-embeddings

For $\mu \in \mathcal{T}_+(K)$, let $L^2(\mu)$ be the Hilbert space of all ($\mathbb{C}$-valued) square-integrable functions with respect to μ.

From the Cauchy-Schwartz inequality in $\mathcal{H}$, we have

$$\int_{\mathscr{X}} |h(t)|^2 \mathrm{d}\mu(t) = \int_{\mathscr{X}} \left|\langle k_t \mid h\rangle_{\mathcal{H}}\right|^2 \mathrm{d}\mu(t) \leqslant \|h\|_{\mathcal{H}}^2 \tau_\mu, h \in \mathcal{H}.$$

Embedding of $\mathcal{H}$ in $L^2(\mu)$

For $\mu \in \mathcal{T}_+(K)$, the linear map $\iota_\mu : \mathcal{H} \to L^2(\mu)$, with $\iota_\mu[h]$ the equiv. class of all meas. fcts μ-a.e. equal to $h \in \mathcal{H}$, is HS.
For $f \in L^2(\mu)$ and $x \in \mathscr{X}$, we have

$$\iota_\mu^*[f](x) = \langle \iota_\mu[k_x] \mid f \rangle_{L^2(\mu)} = \int_{\mathscr{X}} K(x,t) f(t) \mathrm{d}\mu(t),$$

so that $\iota_\mu^* : L^2(\mu) \to \mathcal{H}$ is a *natural interpretation* of $\mathcal{L}_{K,\mu}$.

Four natural interpretations for $\mathcal{L}_{K,\mu}$

$$\iota_\mu^* \in \mathrm{HS}(\mu, \mathcal{H}), \qquad \iota_\mu \iota_\mu^* \in \mathrm{HS}(\mu),$$
$$L_\mu = \iota_\mu^* \iota_\mu \in \mathrm{HS}(\mathcal{H}), \qquad \iota_\mu \iota_\mu^* \iota_\mu \in \mathrm{HS}(\mathcal{H}, \mu).$$

For $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, ι_μ can be approximated by $\iota_\mu P_\nu$.

For $f \in L^2(\mu)$ and $x \in \mathscr{X}$, we have

$$P_\nu \iota_\mu^*[f](x) = \langle \iota_\mu P_\nu[k_x] \,|\, f \rangle_{L^2(\mu)} = \int_{\mathscr{X}} K_\nu(x,t) f(t) \mathrm{d}\mu(t).$$

Approximations induced by $\iota_\mu P_\nu$

$$P_\nu \iota_\mu^* \in \mathrm{HS}(\mu, \mathcal{H}), \qquad \iota_\mu P_\nu \iota_\mu^* \in \mathrm{HS}(\mu),$$

$$P_\nu \iota_\mu^* \iota_\mu P_\nu \in \mathrm{HS}(\mathcal{H}), \qquad \iota_\mu P_\nu \iota_\mu^* \iota_\mu P_\nu \in \mathrm{HS}(\mathcal{H}, \mu).$$

Remark:

$$\|\iota_\mu^* - P_\nu \iota_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} = \int_{\mathscr{X}} K(t,t) - K_\nu(t,t) \mathrm{d}\mu(t);$$

$$\|\iota_\mu \iota_\mu^* - \iota_\mu P_\nu \iota_\mu^*\|^2_{\mathrm{HS}(\mu)} = \iint_{\mathscr{X}} \left|K(x,t) - K_\nu(x,t)\right|^2 \mathrm{d}\mu(t) \mathrm{d}\mu(x).$$

We also have $\|\iota_\mu \iota_\mu^* - \iota_\mu P_\nu \iota_\mu^*\|_{\mathrm{HS}(\mu)} \leqslant \|\iota_\mu^* \iota_\mu - P_\nu \iota_\mu^* \iota_\mu\|_{\mathrm{HS}(\mathcal{H})}$.

Error maps

- For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$:

$$D_\mu(\nu) = \|L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})}, \quad C^{\mathrm{P}}_\mu(\nu) = \|L_\mu - P_\nu L_\mu\|^2_{\mathrm{HS}(\mathcal{H})} \text{ and}$$

$$C^{\mathrm{PP}}_\mu(\nu) = \|L_\mu - P_\nu L_\mu P_\nu\|^2_{\mathrm{HS}(\mathcal{H})}, \nu \in \mathcal{T}_{\mathbb{F}}(K).$$

- For $\mu \in \mathcal{T}_+(K)$:

$$C^{\mathrm{tr}}_\mu(\nu) = \|\iota^*_\mu - P_\nu \iota^*_\mu\|^2_{\mathrm{HS}(\mu,\mathcal{H})} \quad \text{and} \quad C^{\mathrm{F}}_\mu(\nu) = \|\iota_\mu \iota^*_\mu - \iota_\mu P_\nu \iota^*_\mu\|^2_{\mathrm{HS}(\mu)}.$$

- ***Bonus:*** Introduction of an invariance under rescaling in D_μ;

$$R_\mu(\nu) = \min_{c \geqslant 0} D_\mu(c\nu)$$

$$= \begin{cases} \|g_\mu\|^2_{\mathcal{G}} - \Re\left(\langle g_\mu \mid g_\nu \rangle_{\mathcal{G}}\right)^2 / \|g_\nu\|^2 \text{ if } \Re\left(\langle g_\mu \mid g_\nu \rangle_{\mathcal{G}}\right) > 0, \\ \|g_\mu\|^2_{\mathcal{G}} \text{ otherwise.} \end{cases}$$

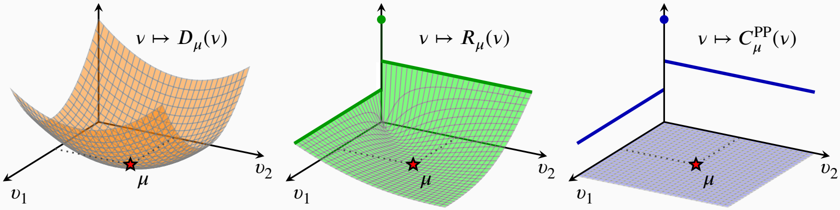


Figure 1: Representation of the maps D_μ, R_μ and C_μ^{PP} on $\mathcal{T}_+(K)$; the considered measures are of the form $\nu = v_1\delta_{x_1} + v_2\delta_{x_2}$.

Remark: For all $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have:

$$C_\mu^{\mathrm{F}}(\nu) \leqslant C_\mu^{\mathrm{P}}(\nu) \leqslant C_\mu^{\mathrm{PP}}(\nu) \leqslant R_\mu(\nu) \leqslant D_\mu(\nu).$$

A quick word about matrices

Consider a PSD matrix $\mathbf{K} \in \mathbb{C}^{N\times N}$, $N \in \mathbb{N}$.

- Let $\mathcal{E}$ be the Euclidean space $\mathbb{C}^N$.
- Consider the measure $\mu = \sum_{j=1}^{N} \delta_j$ on $\mathscr{X} = [N]$; then $L^2(\mu)$ can be identified with $\mathcal{E}$.
- The entries of $\mathbf{K}$ are the values of the kernel of a RKHS of $\mathbb{C}$-valued functions on $[N]$.
- This RKHS can be identified with $\mathcal{H} = \mathrm{span}_{\mathbb{C}}\{\mathbf{K}\} \subseteq \mathbb{C}^N$, with $\langle \boldsymbol{h} \mid \boldsymbol{f} \rangle_{\mathcal{H}} = \boldsymbol{h}^* \mathbf{K}^{\dagger} \boldsymbol{f}$, $\boldsymbol{h}$ and $\boldsymbol{f} \in \mathcal{H}$.
- The support $I \subseteq [N]$ of a measure ν on $[N]$ defines a sample of columns of $\mathbf{K}$.
- Set $\mathcal{H}_I = \mathrm{span}_{\mathbb{C}}\{\mathbf{K}_{\bullet,I}\}$, and denote by P_I the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_I$.
- We have $P_I \mathbf{K} = \mathbf{K}_{\bullet,I}(\mathbf{K}_{I,I})^{\dagger}\mathbf{K}_{I,\bullet} = \hat{\mathbf{K}}(I)$, the low-rank approximation of $\mathbf{K}$ induced by the sample of columns $\mathbf{K}_{\bullet,I}$.

Conclusion

- Equivalence between the quadrature approximation of trace-class integral operators with PSD kernels and the approximation of integral functionals on RKHSs with squared-modulus kernels.
- (In combination with sparsity-inducing mechanisms), quadrature approximation may be used as a differentiable and numerically efficient surrogate for the characterisation of projection-based approximations.

Thank you

⋄ Bertrand Gauthier and Johan Suykens. "Optimal quadrature-sparsification for integral operator approximation". In: *SIAM Journal on Scientific Computing* (2018)

⋄ Matthew Hutchings and Bertrand Gauthier. "Local optimisation of Nyström samples through stochastic gradient descent". In: *Machine Learning, Optimization, and Data Science - LOD 2022*. 2023

⋄ Matthew Hutchings and Bertrand Gauthier. "Energy-based sequential sampling for low-rank PSD-matrix approximation". In: *Preprint* (2023)

⋄ Bertrand Gauthier. "Isometric representation of integral operators with positive-semidefinite kernels". In: *Technical report* (2023)

⋄ Bertrand Gauthier. "Kernel embedding of measures and low-rank approximation of integral operators". In: *In preparation* (2023)