

Fast Cross-validation in Harmonic Approximation

Felix Bartel Ralf Hielscher Daniel Potts

30. Rhein-Ruhr-Workshop

January 31, 2020



Mathematik!
TU Chemnitz

General problem

Setting:

- domain X
- function $f : X \rightarrow \mathbb{C}$
- node set $\mathcal{X} \subset X$
- function values $\tilde{f}_x = f(x) + \varepsilon_x$, $x \in \mathcal{X}$
- functions $\varphi_n : X \rightarrow \mathbb{C}$, $n \in \mathcal{I}$

Task:

- from \tilde{f} find $\hat{g} = (\hat{g}_n)_{n \in \mathcal{I}}$ such that

$$\sum_{n \in \mathcal{I}} \hat{g}_n \varphi_n \approx f$$

Approach

We minimize the **Tikhonov functional**

$$J_\lambda(\hat{\mathbf{g}}) = \left\| \mathbf{F}\hat{\mathbf{g}} - \tilde{\mathbf{f}} \right\|_{\mathbf{W}}^2 + \lambda \|\hat{\mathbf{g}}\|_{\hat{\mathbf{W}}}^2$$

- $\mathbf{F} = (\varphi_n(x))_{x \in \mathcal{X}, n \in \mathcal{I}}$
- $\tilde{\mathbf{f}} = (\tilde{f}_x)_{x \in \mathcal{X}}$
- $\mathbf{W} = \text{diag}(w_x)_{x \in \mathcal{X}}$
- $\hat{\mathbf{W}} = \text{diag}(\hat{w}_n)_{n \in \mathcal{I}}$

Theorem

The unique minimizer of J_λ is given by

$$\hat{\mathbf{g}}_\lambda = \left(\mathbf{F}^H \mathbf{W} \mathbf{F} + \lambda \hat{\mathbf{W}} \right)^{-1} \mathbf{F}^H \mathbf{W} \tilde{\mathbf{f}}.$$

Basic concept of leave-one-out cross-validation

- 1 calculate the Tikhonov minimizer $\hat{\mathbf{g}}_{-x,\lambda}$ from the function values \tilde{f}_y for $y \in \mathcal{X} \setminus \{x\}$
- 2 evaluate the residual of $\hat{\mathbf{g}}_{-x}$ in the node x

$$\left| [\mathbf{F}\hat{\mathbf{g}}_{-x,\lambda}]_x - \tilde{f}_x \right|$$

- 3 calculate the mean value with respect to all nodes

$$CV(\lambda) := \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \left| [\mathbf{F}\hat{\mathbf{g}}_{-x,\lambda}]_x - \tilde{f}_x \right|^2$$

- 4 choose $\lambda^* = \arg \min_{\lambda} CV(\lambda)$

Calculating the residual for full data, we obtain

$$\begin{aligned} F\hat{g}_\lambda - \tilde{\mathbf{f}} &= F \left(F^H W F + \lambda \hat{W} \right)^{-1} F^H W \tilde{\mathbf{f}} - \tilde{\mathbf{f}} \\ &= H_\lambda \tilde{\mathbf{f}} - \tilde{\mathbf{f}}. \end{aligned}$$

Theorem

[Golub, Heath, Whaba, 1979]

$$CV(\lambda) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{[H_\lambda \tilde{\mathbf{f}} - \tilde{\mathbf{f}}]_x^2}{(1 - h_{x,x})^2}$$

with $h_{x,x}$ being the diagonal entries of H_λ .

Derivative of the Cross-validation Score

Theorem

$$\frac{\partial}{\partial \lambda} CV(\lambda) = 2 \sum_{x \in \mathcal{X}} \left(\frac{([\mathbf{H}_\lambda \tilde{\mathbf{f}}]_x - \tilde{f}_x) \frac{\partial}{\partial \lambda} [\mathbf{H}_\lambda \tilde{\mathbf{f}}]_x}{(1 - h_{x,x})^2} + \frac{([\mathbf{H} \tilde{\mathbf{f}}]_x - \tilde{f}_x)^2 \frac{\partial}{\partial \lambda} h_{x,x}}{(1 - h_{x,x})^3} \right)$$

with

$$\frac{\partial}{\partial \lambda} \mathbf{H}_\lambda \tilde{\mathbf{f}} = -\mathbf{F} \left(\mathbf{F}^H \mathbf{W} \mathbf{F} + \lambda \hat{\mathbf{W}} \right)^{-1} \hat{\mathbf{W}} \hat{\mathbf{g}}_\lambda.$$

- computing $\frac{\partial}{\partial \lambda} CV(\lambda)$ has the same complexity as $CV(\lambda)$

For $X = \mathbb{T}^d$, exponential functions

$$\varphi_{\mathbf{n}} = e^{2\pi i \mathbf{n} \cdot \mathbf{x}},$$

nodes

$$\mathcal{X} = \left\{ \frac{\mathbf{m}}{N} : \mathbf{m} \in \mathbb{Z}^d \cap \prod_{t=1}^d [0, N) \right\}$$

and weights $w_{\mathbf{x}} = 1/N^d$ the diagonal entries can be computed via

$$h_{\mathbf{x}, \mathbf{x}} = w_{\mathbf{x}} \sum_{\mathbf{n} \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}}.$$

Algorithm: Computation of the cross-validation score

Input:

- \mathcal{X}
- $\mathbf{W} = \text{diag}(w_x)_{x \in \mathcal{X}}$
- $\hat{\mathbf{W}}$
- $\tilde{\mathbf{f}} = (\tilde{f}_x)_{x \in \mathcal{X}}$
- λ

Output:

- $CV(\lambda)$

① Compute $\hat{\mathbf{g}} := \left(\mathbf{F}^H \mathbf{W} \mathbf{F} + \lambda \hat{\mathbf{W}} \right)^{-1} \mathbf{F}^H \mathbf{W} \tilde{\mathbf{f}}$

② Compute $\mathbf{g} := \mathbf{H}_\lambda \tilde{\mathbf{f}} = \mathbf{F} \hat{\mathbf{g}}$

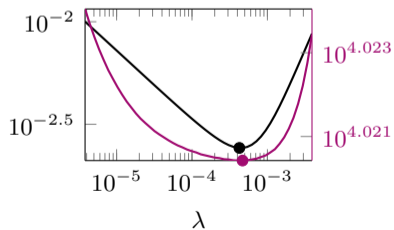
③ Compute $h_{x,x} := w_x \sum_{n \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_n}$ for $x \in \mathcal{X}$

④ Evaluate $CV(\lambda) := \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{|g_x - \tilde{f}_x|^2}{(1 - h_{x,x})^2}$

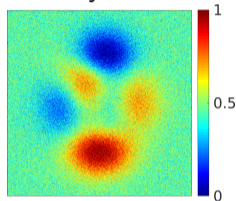
Numerical example

- 1024×1024 nodes
- 10% Gaussian noise

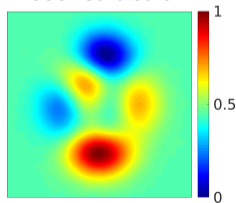
L_2 -error $\|\hat{g}_\lambda - \hat{f}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data



reconstruction



Theorem (Diagonal entries $h_{x,x}$ of the hat matrix) [B, Hielscher, Potts, 2019]

Let \mathcal{I} be an index set and \mathcal{X}, \mathbf{W} form an exact quadrature rule for $e^{2\pi i \mathbf{n} \cdot \mathbf{x}}$ with $\mathbf{n} \in \mathcal{D}(\mathcal{I}) = \{\mathbf{n}_1 - \mathbf{n}_2 : \mathbf{n}_1, \mathbf{n}_2 \in \mathcal{I}\}$. Then

$$h_{\mathbf{x}, \mathbf{x}} = w_{\mathbf{x}} \sum_{\mathbf{n} \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}}.$$

Theorem (Diagonal entries $h_{\mathbf{x},\mathbf{x}}$ of the hat matrix) [B, Hielscher, Potts, 2019]

Let \mathcal{I} be an index set and \mathcal{X} , \mathbf{W} form an exact quadrature rule for $e^{2\pi i \mathbf{n} \cdot \mathbf{x}}$ with $\mathbf{n} \in \mathcal{D}(\mathcal{I}) = \{\mathbf{n}_1 - \mathbf{n}_2 : \mathbf{n}_1, \mathbf{n}_2 \in \mathcal{I}\}$. Then

$$h_{\mathbf{x},\mathbf{x}} = w_{\mathbf{x}} \sum_{\mathbf{n} \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}}.$$

Proof:

Using quadrature, we have $\mathbf{F}^H \mathbf{W} \mathbf{F} = \mathbf{I}$ and thus

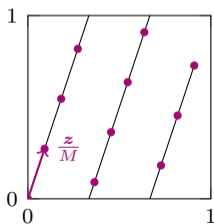
$$\begin{aligned} h_{\mathbf{x},\mathbf{x}} &= \left(\mathbf{F} \left(\mathbf{F}^H \mathbf{W} \mathbf{F} + \lambda \hat{\mathbf{W}} \right)^{-1} \mathbf{F}^H \mathbf{W} \right)_{\mathbf{x},\mathbf{x}} \\ &= \left(\mathbf{F} \operatorname{diag} \left(\frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}} \right)_{\mathbf{n} \in \mathcal{I}} \mathbf{F}^H \mathbf{W} \right)_{\mathbf{x},\mathbf{x}} = w_{\mathbf{x}} \sum_{\mathbf{n} \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}}. \end{aligned}$$

Example: Rank-1 lattices

[Kämmerer, Potts, Volkmer 2015]

- $\mathcal{X} = \Lambda(\mathbf{z}, M) = \left\{ \mathbf{x} = \frac{1}{M}(m\mathbf{z} \bmod M\mathbf{1}) \in \mathbb{T}^d : m = 0, \dots, M-1 \right\}$
- Given \mathcal{I} , there exist algorithms that find \mathbf{z} and M such that $\mathbf{F}^H \mathbf{W} \mathbf{F} = \mathbf{I}$ for $\mathbf{W} = \text{diag}(1/M)_{\mathbf{x} \in \mathcal{X}}$.
- Multiplication with \mathbf{F} can be carried out with the LFFT in $\mathcal{O}(|\mathcal{X}| \log |\mathcal{X}|)$.

$$\mathbf{z} = (1, 3)^\top, M = 11$$



Numerical example

- $f = \bigotimes_{j=1}^7 B_2$ with 5 % Gaussian noise

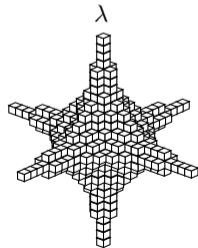
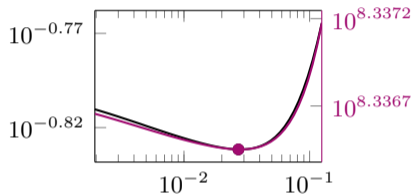
- $\hat{w}_{\mathbf{n}} = \prod_{j=1}^7 \max(|n_j|^2, 1)$

- Index set

$$\mathcal{I}_{16}^{7,\text{hc}} := \left\{ \mathbf{n} \in \mathbb{Z}^7 : \prod_{j=1}^7 \max(1, |n_j|) \leq 16 \right\}$$

- \mathcal{X} : 1 105 193 rank-1 lattice nodes

L_2 -error $\|\hat{\mathbf{g}}_{\lambda} - \mathbf{f}\|_2$ and
cross-validation score $CV(\lambda)$



Cross-validation on the Unit interval $[-1, 1]$

- domain $X = [-1, 1]$
- Chebyshev polynomials

$$\varphi_n(x) = T_n(x) = \cos(\arccos x)$$

for $n = 0, \dots, N - 1$

- scalar product

$$(f, g) = \int_{-1}^1 \frac{f(x)\overline{g(x)}}{\sqrt{1-x^2}} dx$$

Theorem

[B, Hielscher, Potts, 2019]

For exact quadrature we have $\mathbf{F}^H \mathbf{W} \mathbf{F} = \text{diag}(\pi, \pi/2, \dots, \pi/2)$ and

$$h_{x,x} = \frac{w_x}{2} \left(\frac{2}{\pi + \lambda \hat{w}_0} + \sum_{n=1}^{N-1} \frac{\cos(2n \arccos x) + 1}{\pi/2 + \lambda \hat{w}_n} \right).$$

Example: Chebyshev nodes

[B, Hielscher, Potts, 2019]

For $x_m = \cos \frac{(2m+1)\pi}{2N}$ and $w_m = \pi/N$ we have

$$h_{x_m, x_m} = \frac{w_m}{2} \left(\frac{\sqrt{N/2}}{\gamma(m)} [\mathbf{C}_{2N+1}^I \mathbf{b}]_{2m+1} + \sum_{n=1}^{N-1} \frac{1}{\pi/2 + \lambda \hat{w}_n} \right)$$

with

$$\mathbf{b} = \left(\frac{2\sqrt{2}}{\pi + \lambda \hat{w}_0}, 0, \frac{1}{\pi/2 + \lambda \hat{w}_1}, 0, \dots, \frac{1}{\pi/2 + \lambda \hat{w}_{N-1}}, 0, 0 \right)^T$$

and

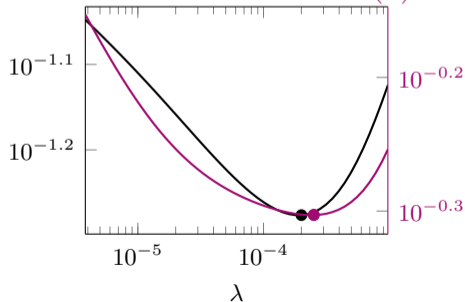
$$\gamma(n) = \begin{cases} 1/\sqrt{2} & n = 0 \\ 1 & n = 1, \dots, N-1. \end{cases}$$

Cross-validation on the Unit interval $[-1, 1]$

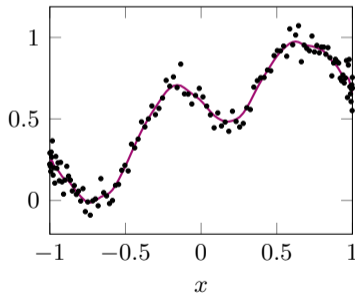
Numerical example

- 128 Chebyshev nodes
- 5 % Gaussian noise

L_2 -error $\|\hat{g}_\lambda - \hat{f}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data and reconstruction



Cross-validation on the two-dimensional sphere \mathbb{S}^2

- domain $X = \mathbb{S}^2$
- spherical harmonics

$$\varphi_{n,k}(\theta, \varphi) = Y_{n,k}(\theta, \varphi) = \sqrt{\frac{2n+1}{4\pi}} P_{|k|}^n(\cos \theta) e^{ik\varphi}$$

for $n = 0, \dots, N$, $k = -n, \dots, n$

Theorem

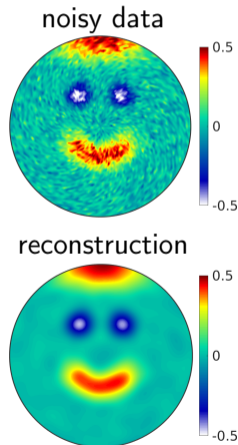
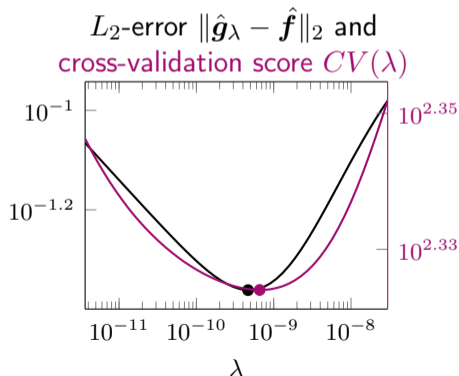
[B, Hielscher, Potts, 2019]

For exact quadrature we have

$$h_{\mathbf{x}, \mathbf{x}} = \frac{w_{\mathbf{x}}}{4\pi} \sum_{n=0}^N \frac{2n+1}{1 + \lambda \hat{w}_n}.$$

Numerical example

- 21 000 nearly equidistributed nodes
- 5 % Gaussian noise



Cross-validation on the rotation group $SO(3)$

- domain $X = SO(3) = \{\mathbf{R} \in \mathbb{R}^{3 \times 3} : \mathbf{R}^\top \mathbf{R} = \mathbf{I}, \det \mathbf{R} = 1\}$
- Wigner-D functions

$$\varphi_n^{kk'}(\mathbf{R}) = D_n^{kk'}(\mathbf{R}) = \int_{\mathbb{S}^2} Y_n^k(\mathbf{R}\boldsymbol{\xi}) \overline{Y_n^{k'}(\boldsymbol{\xi})} d\boldsymbol{\xi}$$

for $n = 0, \dots, N$ and $k, k' = -n, \dots, n$

Theorem

[B, Hielscher, Potts, 2019]

For exact quadrature we have

$$h_{\mathbf{x}, \mathbf{x}} = w_{\mathbf{x}} \sum_{n=0}^N \frac{2n+1}{\frac{8\pi^2}{2n+1} + \lambda \hat{w}_n}.$$

Scattered nodes

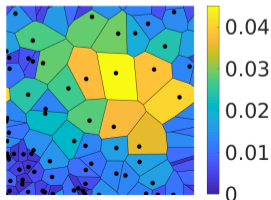
Idea: use approximative quadrature weights

Definition

Voronoi cell V_x corresponding to $x \in \mathcal{X} \subset \mathbb{T}^d$ is

$$V_x := \{ \mathbf{y} \in \mathcal{M} : \text{dist}(\mathbf{x}, \mathbf{y}) \leq \text{dist}(\mathbf{x}', \mathbf{y}), \forall \mathbf{x}' \in \mathcal{X} \}.$$

The Voronoi weight w_x is the area of the Voronoi cell V_x .



Nodes without Exact Quadrature

By this approximative quadrature we obtain

$$\mathbf{F}^H \mathbf{W} \mathbf{F} \approx \mathbf{I}$$

which motivates

Definition

The approximated cross-validation score $\tilde{C}V(\lambda)$ is defined by

$$\tilde{C}V(\lambda) = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x} \in \mathcal{X}} \frac{[\mathbf{H} \tilde{\mathbf{f}} - \tilde{\mathbf{f}}]_{\mathbf{x}}^2}{(1 - \tilde{h}_{\mathbf{x}, \mathbf{x}})^2}$$

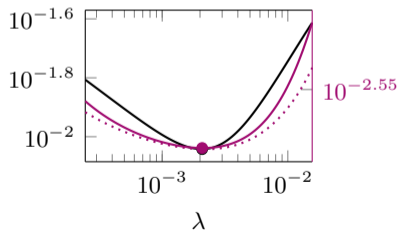
for

$$\tilde{h}_{\mathbf{x}, \mathbf{x}} = w_{\mathbf{x}} \sum_{\mathbf{n} \in \mathcal{I}} \frac{1}{1 + \lambda \hat{w}_{\mathbf{n}}} \approx h_{\mathbf{x}, \mathbf{x}}.$$

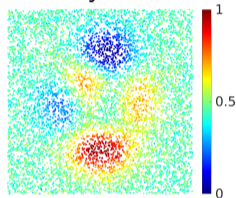
Numerical example

- 8 192 uniformly scattered nodes
- 5 % Gaussian noise

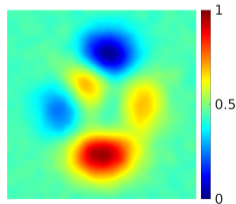
L_2 -error $\|\hat{g}_\lambda - \hat{f}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data



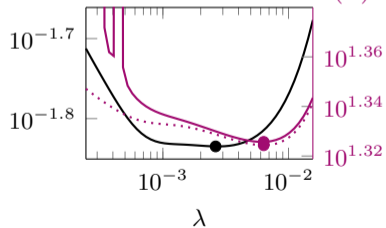
reconstruction



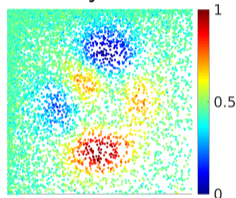
Numerical example

- 8 192 scattered nodes
- 5 % Gaussian noise

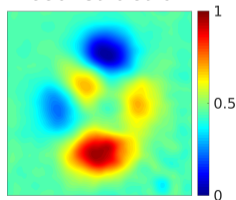
L_2 -error $\|\hat{g}_\lambda - \hat{f}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data



reconstruction



The diagonal entries $h_{x,x}$ of the hat matrix \mathbf{H}_λ satisfy

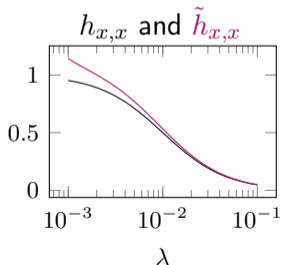
$$h_{x,x} < 1$$

for all $\lambda > 0$ and $x \in \mathcal{X}$.

The diagonal entries $h_{x,x}$ of the hat matrix \mathbf{H}_λ satisfy

$$h_{x,x} < 1$$

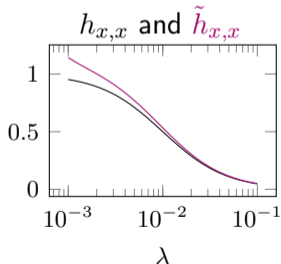
for all $\lambda > 0$ and $x \in \mathcal{X}$.



The diagonal entries $h_{x,x}$ of the hat matrix \mathbf{H}_λ satisfy

$$h_{x,x} < 1$$

for all $\lambda > 0$ and $x \in \mathcal{X}$.



$$\tilde{C}V(\lambda) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{[\mathbf{H}_\lambda \tilde{\mathbf{f}} - \tilde{\mathbf{f}}]_x^2}{(1 - \tilde{h}_{x,x})^2}$$