Fast Cross-validation in Harmonic Approximation

Felix Bartel, Ralf Hielscher, Daniel Potts

Chemnitz Symposium on Inverse Problems 2019

October 2, 2019

Mathematik!
TU Chemnitz

General problem

Setting:

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

Task:

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

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

Approach

We minimize the Tikhonov functional

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

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

Theorem

The unique minimizer of J_λ is given by

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

Basic concept of leave-one-out cross-validation

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

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

3. calculate the mean value with respect to all nodes

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

Cross-validation

Calculating the residual for full data, we obtain

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

Theorem [Golub, Heath, Whaba, 1979]

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

with $h_{x,x}$ being the diagonal entries of $\boldsymbol{H}_\lambda$.

Derivative of the Cross-validation Score

Theorem

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

with

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

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

Cross-validation on the Torus $\mathbb{T}^d$

Theorem [Tasche, Weyrich, 1996]

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

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

nodes

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

and weights $w_x = 1/N^d$ the diagonal entries can be computed via

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

Algorithm: Computation of the cross-validation score

Input:

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

Output:

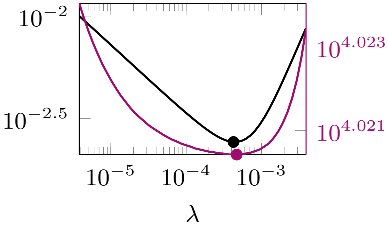
- $CV(\lambda)$

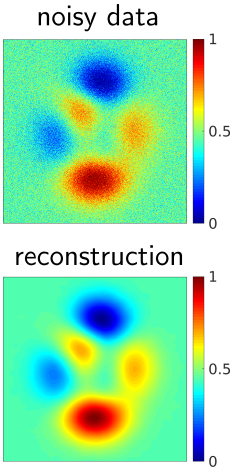
1. Compute $\hat{\boldsymbol{g}} := \left(\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{F} + \lambda\hat{\boldsymbol{W}}\right)^{-1}\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\tilde{\boldsymbol{f}}$
2. Compute $\boldsymbol{g} := \boldsymbol{H}_\lambda\tilde{\boldsymbol{f}} = \boldsymbol{F}\hat{\boldsymbol{g}}$
3. Compute $h_{x,x} := w_x \sum_{\boldsymbol{n}\in\mathcal{I}} \frac{1}{1+\lambda\hat{w}_{\boldsymbol{n}}}$ 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 $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and
cross-validation score $CV(\lambda)$





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

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

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

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

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

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

Proof:

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

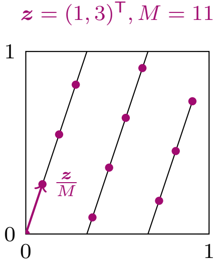
$$\begin{aligned} h_{\boldsymbol{x},\boldsymbol{x}} &= \left(\boldsymbol{F}\left(\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{F} + \lambda\hat{\boldsymbol{W}}\right)^{-1}\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\right)_{\boldsymbol{x},\boldsymbol{x}} \\ &= \left(\boldsymbol{F}\,\mathrm{diag}\left(\frac{1}{1+\lambda\hat{w}_{\boldsymbol{n}}}\right)_{\bolds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

∎

Cross-validation on the Torus $\mathbb{T}^d$

Example: Rank-1 lattices [Kämmerer, Potts, Volkmer 2015]

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



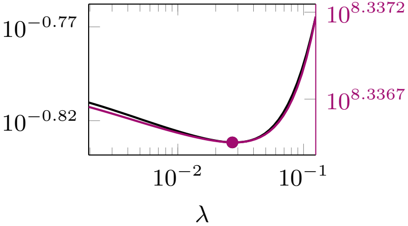
Numerical example

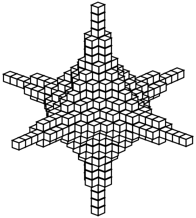
- $f = \bigotimes_{j=1}^{7} B_2$ with $5\,\%$ Gaussian noise
- $\hat{w}_{\boldsymbol{n}} = \prod_{j=1}^{7} \max\left(|n_j|^2, 1\right)$
- Index set

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

- $1\,105\,193$ rank-1 lattice nodes

L_2-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ 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 $n = 0, \ldots, N-1$
- scalar product

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

Theorem [B, Hielscher, Potts, 2019]

For exact quadrature we have $\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{F} = \mathrm{diag}(\pi, \pi/2, \ldots, \pi/2)$ and

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

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

Example: Chebyshev nodes [B, Hielscher, Potts, 2019]

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

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

with

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

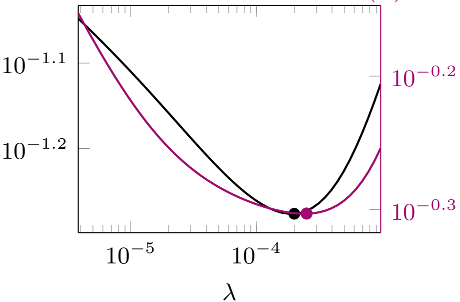
and

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

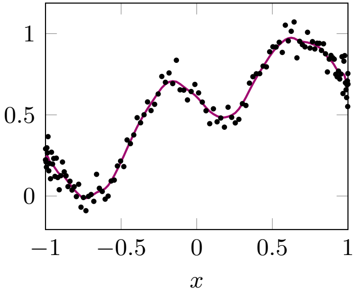
Numerical example

- 128 Chebyshev nodes
- 5 % Gaussian noise

L_2-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ 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

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

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

Theorem [B, Hielscher, Potts, 2019]

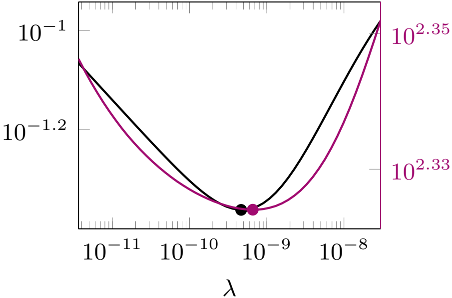
For exact quadrature we have

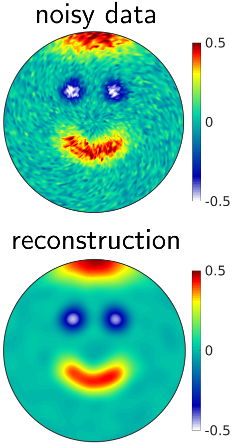
$$h_{\boldsymbol{x},\boldsymbol{x}} = \frac{w_{\boldsymbol{x}}}{4\pi}\sum_{n=0}^{N}\frac{2n+1}{1+\lambda\hat{w}_n}.$$

Numerical example

- $21\,000$ nearly equidistributed nodes
- $5\,\%$ Gaussian noise

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





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

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

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

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

Theorem [B, Hielscher, Potts, 2019]

For exact quadrature we have

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

Scattered nodes

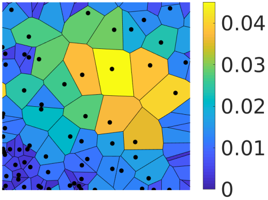
Idea: use approximative quadrature weights

Definition

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

$$V_{\boldsymbol{x}} := \left\{ \boldsymbol{y} \in \mathcal{M} : \operatorname{dist}(\boldsymbol{x}, \boldsymbol{y}) \leq \operatorname{dist}(\boldsymbol{x'}, \boldsymbol{y}),\ \forall \boldsymbol{x'} \in \mathcal{X} \right\}.$$

The Voronoi weight $w_{\boldsymbol{x}}$ is the area of the Voronoi cell $V_{\boldsymbol{x}}$.



Nodes without Exact Quadrature

By this approximative quadrature we obtain

$$\boldsymbol{F}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{F} \approx \boldsymbol{I}$$

which motivates

Definition

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

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

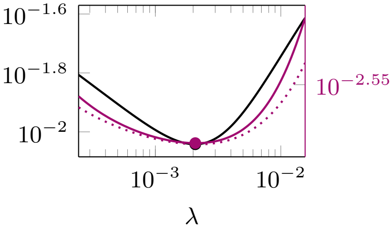
for

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

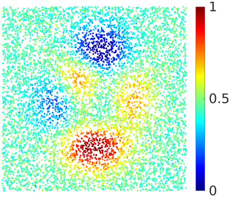
Numerical example

- 8 192 uniformly scattered nodes
- 5 % Gaussian noise

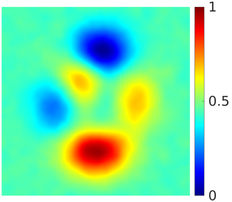
L_2-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data



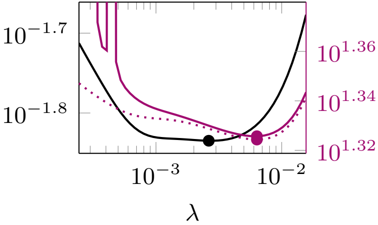
reconstruction



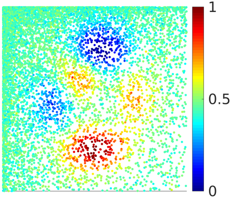
Numerical example

- 8 192 scattered nodes
- 5 % Gaussian noise

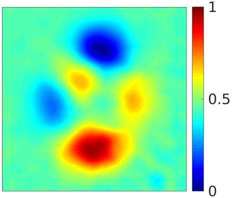
L_2-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and
cross-validation score $CV(\lambda)$



noisy data



reconstruction



Nodes without Exact Quadrature

Lemma [B, Hielscher, Potts, 2019]

The diagonal entries $h_{x,x}$ of the hat matrix $\boldsymbol{H}_\lambda$ satisfy

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

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

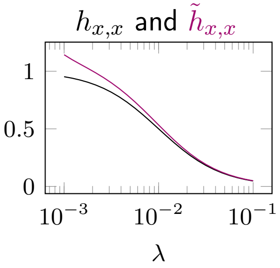
Nodes without Exact Quadrature

Lemma [B, Hielscher, Potts, 2019]

The diagonal entries $h_{x,x}$ of the hat matrix $\boldsymbol{H}_\lambda$ satisfy

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

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



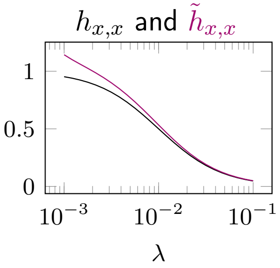
Nodes without Exact Quadrature

Lemma [B, Hielscher, Potts, 2019]

The diagonal entries $h_{x,x}$ of the hat matrix $\boldsymbol{H}_\lambda$ satisfy

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

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



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