Samplet basis pursuit

Multiresolution scattered data approximation with sparsity constraints

Michael Multerer

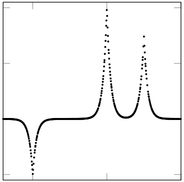
joint work with D. Baroli and H. Harbrecht

SNSF StG: “Multiresolution methods for unstructured data”

Outline

1. Samplets
2. Samplet basis pursuit
3. Numerical examples
4. Conclusion

Scattered data

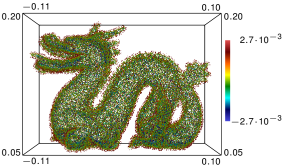


- In many applications, we are given a set of data points.
- Examples are measurements or sample values.
- We denote the set of *data sites* by

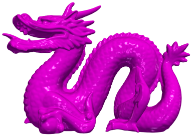
$$X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d.$$

- If the data sites have no regular structure, we refer to them as *scattered data*.

Scattered data



$$s_h(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x})$$



- Given *data values* $f(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in X$, we want to detect features or compress them.
- To make predictions, we want to interpolate or fit the data values.
- In this talk, we consider a dictionary of reproducing kernels to sparsily approximate data.
- To this end, we consider the RKHS embeddings of samplet bases.

Functional analytic setting

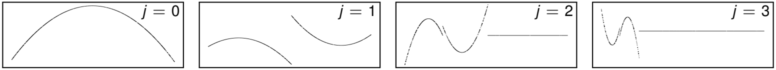
- Let $X := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega \subset \mathbb{R}^d$ denote a set of data sites.
- We introduce the Dirac-δ-distributions

$$\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N} \in [C(\Omega)]'.$$

- The Dirac-δ-distributions satisfy $(f, \delta_{\boldsymbol{x}_i})_\Omega := \delta_{\boldsymbol{x}_i}(f) = f(\boldsymbol{x}_i)$ for all $f \in C(\Omega)$.
- The *data values* $f_i := (f, \delta_{\boldsymbol{x}_i})_\Omega$, $i = 1, \ldots, N$, amount to the available information of the function $f \in C(\Omega)$.
- We define the Hilbert space $\mathcal{X} := \operatorname{span}\{\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}\} \subset [C(\Omega)]'$ with inner product given by

$$\langle u, v\rangle_{\mathcal{X}} := \sum_{i=1}^N u_i v_i, \quad \text{where } u = \sum_{i=1}^N u_i \delta_{\boldsymbol{x}_i},\ v = \sum_{i=1}^N v_i \delta_{\boldsymbol{x}_i}.$$

Samplets



- Let $\mathcal{X}_0 \subset \mathcal{X}_1 \subset \cdots \subset \mathcal{X}_J := \mathcal{X}$ be a multiresolution analysis with

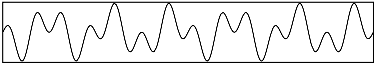
$$\mathcal{X}_j := \operatorname{span} \mathbf{\Phi}_j,\ \mathbf{\Phi}_j := \{\varphi_{j,k}\}_k.$$

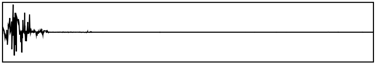
- Each distribution $\varphi_{j,k}$ is a linear combination of Dirac-δ-distributions.
- Since $\mathcal{X}_j \subset \mathcal{X}_{j+1}$, we can (orthogonally) decompose $\mathcal{X}_{j+1} = \mathcal{X}_j \oplus \mathcal{S}_j$.
- For $\mathcal{S}_j$, we introduce the orthonormal bases $\mathbf{\Sigma}_j := \{\sigma_{j,k}\}_k$.
- Recursively applying the decomposition, the set $\mathbf{\Sigma}_J = \mathbf{\Phi}_0 \cup \bigcup_{j=0}^{J-1} \mathbf{\Sigma}_j$ is a basis of $\mathcal{X}_J$, which we call *samplet basis*.
- For data compression, we may construct samplets with vanishing moments

$$(p, \sigma_{j,k})_\Omega = 0 \quad \text{for all } p \in \mathcal{P}_q(\Omega).$$

Properties of samplets

[Harbrecht-M. '22]





- The samplet basis can be constructed with cost $\mathcal{O}(N)$.
- The number of all samplets on level j scales like 2^j.
- They are orthonormal and have vanishing moments of order $q+1$.
- For $f \in C^{q+1}(O)$, $\tau \subset O$, where τ is the support of a samplet $\sigma_{j,k}$, there holds

$$|(f, \sigma_{j,k})_\Omega| \leq \left(\frac{d}{2}\right)^{q+1} \frac{\operatorname{diam}(\tau)^{q+1}}{(q+1)!} \|f\|_{C^{q+1}(O)} \|\boldsymbol{\omega}_{j,k}\|_1.$$

- Herein, $\boldsymbol{\omega}_{j,k}$ is the coefficient vector of the measure $\sigma_{j,k}$.
- For quasi-uniform data sets X, we have $\operatorname{diam}(\tau) \sim 2^{-j/d}$.

Properties of samplets II

- The mapping $f \mapsto If := \big[(f, \delta_{\boldsymbol{x}_1})_\Omega, \ldots, (f, \delta_{\boldsymbol{x}_N})_\Omega\big]^\intercal$ defines an operator $I\colon C(\Omega) \to \mathbb{R}^N$.
- Samplets represent $\boldsymbol{f} = If$ in a multiresolution fashion, i.e.,

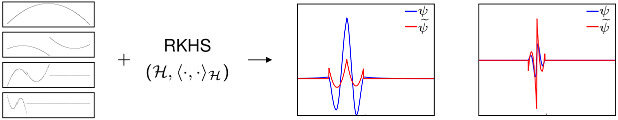
$$\boldsymbol{f}^\Sigma = \boldsymbol{T}\boldsymbol{f},$$

  where $\boldsymbol{f}^\Sigma := \big[(f, \sigma_{j,k})_\Omega\big]^\intercal$ and $\boldsymbol{T} \in \mathbb{R}^{N\times N}$ is orthogonal.
- The cost for the fast samplet transform is $\mathcal{O}(N)$.
- If $\mathcal{K} \in C(\Omega \times \Omega)$ is *asymptotically smooth* and X is quasi-uniform, we can approximate the samplet representation of $\boldsymbol{K} := (I \otimes I)\mathcal{K}$ with cost $\mathcal{O}(N \log N)$.
- Setting all entries of $\boldsymbol{K}^\Sigma$ whose corresponding samplets' supports τ, τ' satisfy $\operatorname{dist}(\tau, \tau') \geq \eta \max\{\operatorname{diam}(\tau), \operatorname{diam}(\tau')\}$ to zero yields

$$\big\|\boldsymbol{K}^\Sigma - \boldsymbol{K}^\Sigma_\varepsilon\big\|_F \lesssim m_q(\rho\eta/d)^{-2(q+1)}\big\|\boldsymbol{K}^\Sigma\big\|_F.$$

Samplets in RKHS

[Baroli-Harbrecht-M. '24]



- Let $\mathcal{H} \subset C(\Omega)$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$.
- Then, $\mathcal{H}$ is a *reproducing kernel Hilbert space* (RKHS) with reproducing kernel

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{y}) := (R\delta_{\boldsymbol{x}})(\boldsymbol{y}),$$

  by the Riesz representation theorem.
- Given a subspace $\mathcal{H}_X := \operatorname{span}\{\mathcal{K}(\boldsymbol{x}_1,\cdot), \ldots, \mathcal{K}(\boldsymbol{x}_N,\cdot)\} \subset \mathcal{H}$, samplets induce a multiresolution basis for $\mathcal{H}_X$ via the embedding

$$\sigma_{j,k} = \sum_{i=1}^{N} \omega_{j,k,i}\delta_{\boldsymbol{x}_i} \mapsto \psi_{j,k} = R\sigma_{j,k} = \sum_{i=1}^{N} \omega_{j,k,i}\mathcal{K}(\boldsymbol{x}_i,\cdot).$$

- There holds $\langle\psi_{j,k}, h\rangle_{\mathcal{H}} = 0$ if $h|_O \in \mathcal{P}_q(O)$.

Dual samplets in RKHS

- The dual basis is characterized by

$$\widetilde{\psi}_{j,k} = \sum_{i=1}^{N} \widetilde{\omega}_{j,k,i}\mathcal{K}(\boldsymbol{x}_i,\cdot),\ \widetilde{\boldsymbol{\omega}}_{j,k} := \boldsymbol{K}^{-1}\boldsymbol{\omega}_{j,k},\quad \text{where } \boldsymbol{K} = [\mathcal{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)]_{i,j=1}^{N}.$$

- For $h \in \mathcal{H}$, the orthogonal projection onto $\mathcal{H}_X$ is given by

$$s_h(\boldsymbol{x}) = \sum_{i=1}^{N}\langle\widetilde{\psi}_i, h\rangle_{\mathcal{H}}\psi_i(\boldsymbol{x}) = \sum_{i=1}^{N}\alpha_i\mathcal{K}(\boldsymbol{x}_i,\cdot),\quad \boldsymbol{\alpha} = \boldsymbol{T}^{\intercal}[\langle\widetilde{\psi}_i, h\rangle_{\mathcal{H}}]_{i=1}^{N}.$$

- It satisfies $s_h(\boldsymbol{x}_i) = h(\boldsymbol{x}_i)$ for all $\boldsymbol{x}_i \in X$.
- There holds

$$\langle\widetilde{\psi}_{j,k}, h\rangle_{\mathcal{H}} = 0 \quad \text{if } \alpha_i = p(\boldsymbol{x}_i) \text{ for some } p|_O \in \mathcal{P}_q(O).$$

- If a signal is sparse in the kernel basis, it is also sparse in the samplet basis.
- This means that sparsity in the samplet basis is the more general concept.

Samplet basis pursuit

- For data $\boldsymbol{h} := [h(\boldsymbol{x}_i)]_{i=1}^N$, and a dictionary of kernels $\mathcal{K}_1, \ldots, \mathcal{K}_L$ we are looking for a sparse approximation

$$s_h(\boldsymbol{x}) = \sum_{j=1}^{L} \sum_{i=1}^{N} \beta_i^{(j)} \psi_i^{(j)}(\boldsymbol{x}).$$

- To this end, we solve the underdetermined least squares problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{L\cdot N}} \frac{1}{2}\|\boldsymbol{h} - \boldsymbol{K}\boldsymbol{\alpha}\|_2^2 + \sum_{i=1}^{L\cdot N} w_i|\beta_i|,$$

where $\boldsymbol{K} := [\boldsymbol{K}_1, \ldots, \boldsymbol{K}_L]$, $\boldsymbol{\alpha}^\intercal := [\boldsymbol{\alpha}_1^\intercal, \ldots, \boldsymbol{\alpha}_L^\intercal]$ and $\boldsymbol{\beta}^\intercal := [(\boldsymbol{T}\boldsymbol{\alpha}_1)^\intercal, \ldots, (\boldsymbol{T}\boldsymbol{\alpha}_L)^\intercal]$ (we set $w_i = \lambda$).

- Note that the Euclidean norm is invariant under the samplet transform.
- Computations can efficiently be carried out in samplet coordinates.

Numerical solution

[Griesse-Lorenz '08]

- Numerical algorithms to solve the ℓ^1-regularized minimization problem are based on soft-thresholding.
- The optimal coefficient can be retrieved from the root finding problem

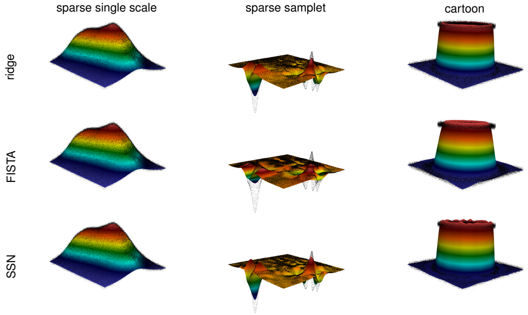
$$\mathbf{0} = \boldsymbol{\beta}^\star - \mathsf{SS}_{\gamma \boldsymbol{w}}\left(\boldsymbol{\beta}^\star + \gamma(\boldsymbol{K}^\Sigma)^\intercal(\boldsymbol{h}^\Sigma - \boldsymbol{K}^\Sigma\boldsymbol{\beta}^\star)\right) \quad \text{for any } \gamma > 0.$$

- Herein, the soft-shrinkage operator is given by

$$\mathsf{SS}_{\boldsymbol{w}}(\boldsymbol{v}) := \operatorname{sign}(\boldsymbol{v}) \max\{\mathbf{0}, |\boldsymbol{v}| - \boldsymbol{w}\}.$$

- The root finding problem is solved by the semi-smooth Newton method (with iterative regularization).

Sparse reconstruction



sparse single scale

	ridge	FISTA	MRSSN
iterations	103	10 000	411
comp. time	4.51s	1 777.42s	58.48s
final $\|\mathcal{A}\|$	973 731	32 963	101
residual	$8.3 \cdot 10^{-8}$	$1.2 \cdot 10^{-2}$	$2.0 \cdot 10^{-9}$

sparse samplet

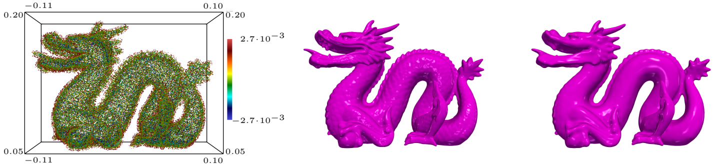
	ridge	FISTA	MRSSN
iterations	170	10 000	422
comp. time	11.3s	1 857.19s	57.86s
final $\|\mathcal{A}\|$	827 864	35 194	163
residual	$1.2 \cdot 10^{-8}$	$2.6 \cdot 10^{-2}$	$2.3 \cdot 10^{-7}$

cartoon

	ridge	FISTA	MRSSN
iterations	118	10 000	646
comp. time	5.17s	1970.76s	80.46s
final $\|\mathcal{A}\|$	978 740	218 570	261
residual	$9.6 \cdot 10^{-8}$	$7.1 \cdot 10^{-2}$	$3.4 \cdot 10^{-7}$

- The kernel is a Matérn-3/2 kernel.
- We employ samplets with $q + 1 = 4$ vanishing moments.
- Examples 1 and 2 have 10 non vanishing coefficients each.

Surface reconstruction



- We consider $N = 10^6$ values of the signed distance function.
- The same regularization parameter for ridge- and ℓ^1-regularization is used.
- The kernel is a scaled exponential kernel.
- The approximate signed distance function is evaluated on a grid with $400 \times 400 \times 400$ points.
- For ℓ^1-regularization, we obtain $\|\boldsymbol{\beta}\|_0 = 6\,233$ non-zero coefficients.

Setup

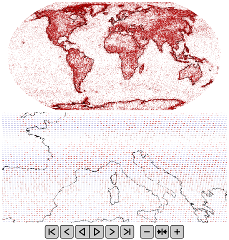
[Contains modified Copernicus Climate Change Service information 2022]

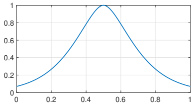
- We approximate an adaptive subsample ($N = 1.2 \cdot 10^6$) of the ERA5 monthly temperature data set of 2022.
- We set $q + 1 = 4$ and $\lambda = 10^{-6}$.
- The data are rescaled to $[0, 1]^3$. We employ the kernels

$$\mathcal{K}_1\big((\boldsymbol{x}, t), (\boldsymbol{y}, t')\big) := k_{3/2}(\|\boldsymbol{x} - \boldsymbol{y}\|_2) k_{\mathrm{per}}(|t - t'|)$$

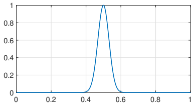
and

$$\mathcal{K}_2(\boldsymbol{z}, \boldsymbol{z}') := k_{\exp}(\|\boldsymbol{z} - \boldsymbol{z}'\|_2), \ \boldsymbol{z} := (\boldsymbol{x}, t), \boldsymbol{z}' := (\boldsymbol{y}, t').$$

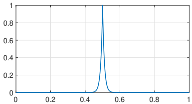




$k_{3/2}(r) := (1 + 5\sqrt{3})e^{-5\sqrt{3}r}$

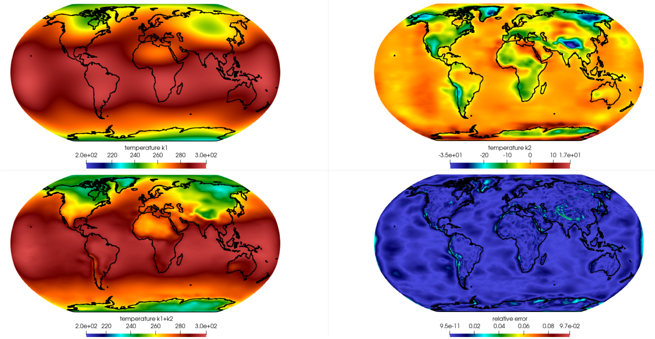


$k_{\mathrm{per}}(r) := e^{-50\sin^2(\pi r)}$



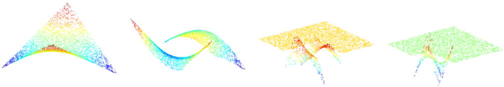
$k_{\exp}(r) := e^{-100r}$

Reconstruction



- We obtain $\|\boldsymbol{\beta}\|_0 = 6\,554$ with $\|\boldsymbol{\beta}_1\|_0 = 671$ and $\|\boldsymbol{\beta}_2\|_0 = 5\,883$.
- The relative error is computed at all ($\approx 12M$) points of the data set and smaller than 9.73% for all time steps.

Conclusion



- Samplets are a multiresolution analysis of discrete signed measures with vanishing moments.
- Samplets can be constructed on general scattered data sets.
- The embedding of samplets into RKHS is straightforward.
- Samplets compress kernel matrices of asymptotically smooth kernels to $N \log N$ relevant entries.
- Samplet basis pursuit facilitates the retrieval of sparse representations.
- Code (C++ with pybind11) is available on `github.com/muchip/fmca`.

Thank you

References

- D. Baroli, H. Harbrecht and M. Multerer. Samplet basis pursuit: Multiresolution scattered data approximation with sparsity constraints, IEEE Trans. Sig. Proc., 2024.
- R. Griesse and D.A. Lorenz. A semismooth Newton method for Tikhonov functionals with sparsity constraints. Inverse Probl., 2008.
- H. Harbrecht and M. Multerer. Samplets: Construction and scattered data compression. J. Comput. Phys., 2022.
- H. Hersbach, et al. ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.bd0915c6.