A Game Theoretic Approach to Numerical Approximation and Algorithm Design

Houman Owhadi

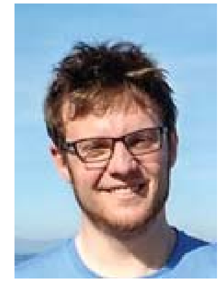
BIRS

April 26, 2018

DARPA EQUiPS / AFOSR award no FA9550-16-1-0054
(Computational Information Games)

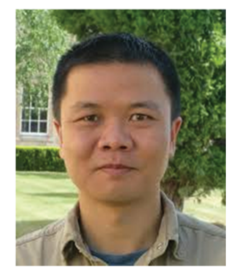
Collaborators

- Operator adapted wavelets, fast solvers, and numerical homogenization from a game theoretic approach to numerical approximation and algorithm design, 2018.
- Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis, 2017. arXiv:1703.10761. H. Owhadi and C. Scovel.
- Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity, arXiv:1706.02205, 2017. Schäfer, Sullivan, Owhadi.
- Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients, 2016. arXiv:1606.07686. H. Owhadi and L. Zhang.
- Multigrid with rough coefficients and Multiresolution operator decomposition from Hierarchical Information Games. H. Owhadi. SIAM Review, 59(1), 99149, 2017. arXiv:1503.03467
- Bayesian Numerical Homogenization. H. Owhadi. SIAM Multiscale Modeling & Simulation, 13(3), 812828, 2015. arXiv:1406.6668


Florian Schäfer


Clint Scovel


Tim Sullivan


Lei Zhang

DARPA EQUiPS / AFOSR award no FA9550-16-1-0054
(Computational Information Games)

Interplays between numerical approximation and Gaussian process regression

Pioneering work

Poincaré (1896). Sul'din (1959). Sard (1963).
Kimeldorf and Wahba (1970). Larkin (1972)

Bayesian Numerical Analysis

Diaconis (1988). Shaw (1988).
O'Hagan (1991). Skilling (1992).

Information based complexity

Woźniakowski (1986). Wasilkowski and Woźniakowski (1986).
Packel (1987). Traub, Wasilkowski and Woźniakowski (1988).
Novak and Woźniakowski (2008-2010).

Probabilistic Numerics

Briol, Chkrebtii, Campbell, Calderhead, Conrad, Duvenaud, Girolami, Griebel, Hennig, Karniadakis, Maziar, Oates, Osborne, Owhadi, Paris, Sejdinovic, Särkä, Schäfer, Schober, Scovel, Sullivan, Stuart, Venturi, Zabaras, Zhang and Zygalakis. (2014-now)

The operator

$$\Omega \subset \mathbb{R}^d$$

$\mathcal{L}$: linear, symmetric, positive, invertible, local

$$\left(H_0^s(\Omega), \|\cdot\|_{H_0^s(\Omega)}\right) \xrightarrow{\quad \mathcal{L} \quad} \left(H^{-s}(\Omega), \|\cdot\|_{H^{-s}(\Omega)}\right)$$

$\mathcal{L}$ local: $\int_\Omega u\mathcal{L}v = 0$ if u and v have disjoint supports

G : Green's function

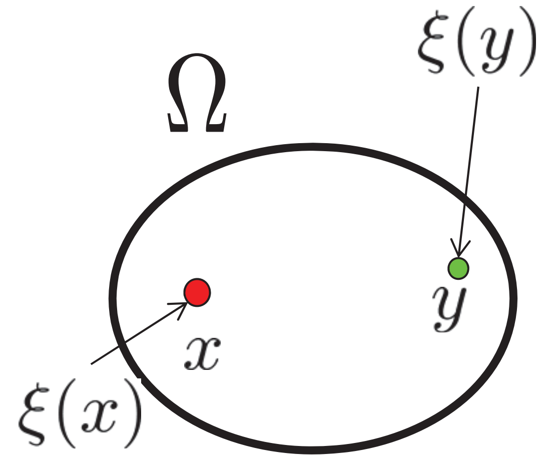
$$(1) \quad \boxed{\mathcal{L}u = f}$$

The solution of (1) is $u(x) = \int_\Omega G(x,y) f(y)\, dy$

The Gaussian field

G : symmetric positive definite kernel

$\xi \sim \mathcal{N}(0, G)$



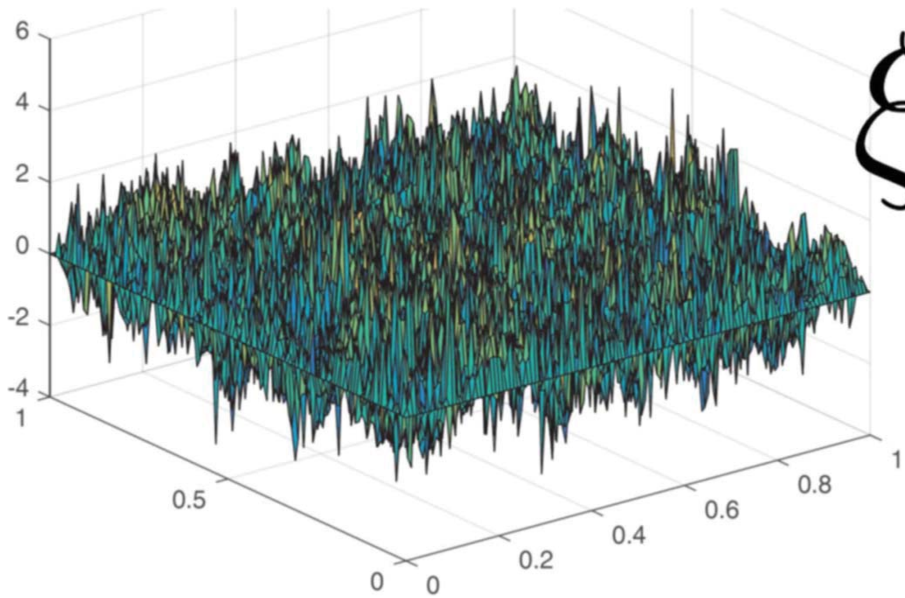
$$s > d/2$$

$\xi(x)$ is a centered Gaussian process

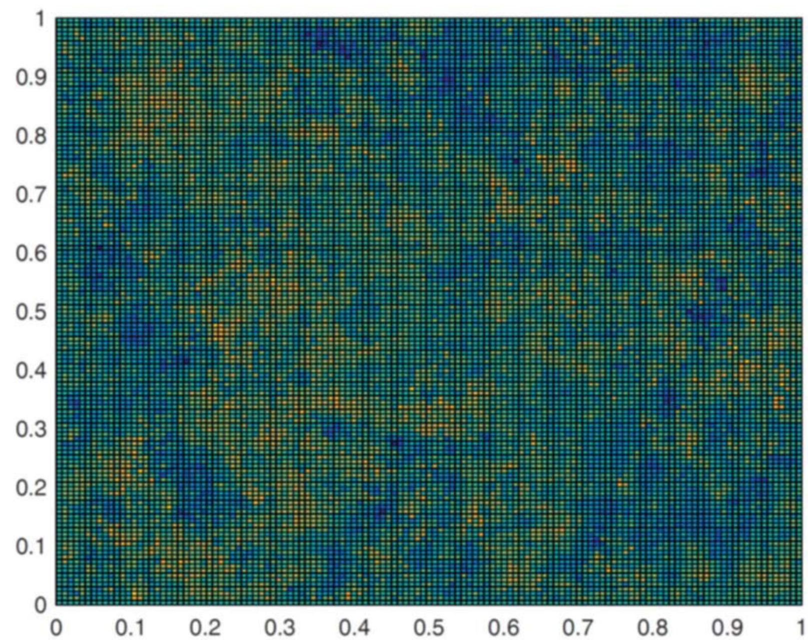
$$\mathrm{Cov}\left(\xi(x), \xi(y)\right) = G(x, y)$$

For $\phi \in H^{-s}(\Omega)$,

$$\int_\Omega \xi(x)\phi(x)\,dx \sim \mathcal{N}\left(0, \int_{\Omega^2} \phi(x)G(x,y)\phi(y)\,dx\,dy\right)$$



ξ

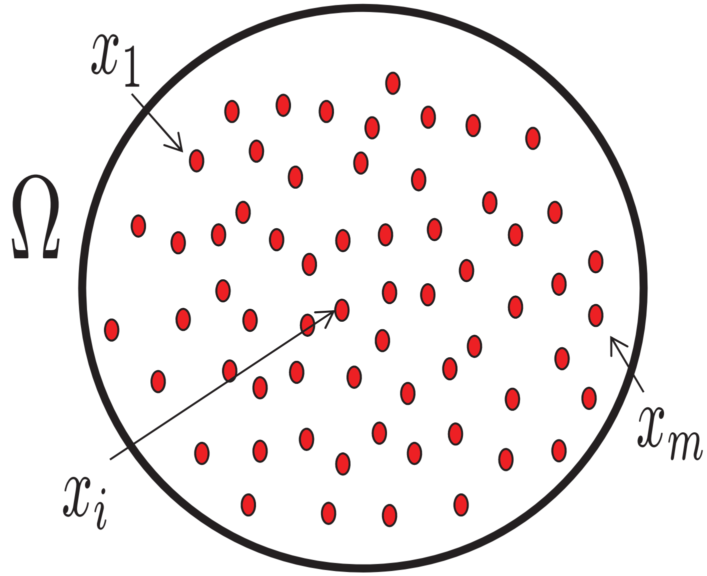


Problem $s > d/2$

$u \in H_0^s(\Omega)$ unknown

Given $u(x_1), \ldots, u(x_m)$
what is your best approximation of u?
Best: $\min_v \max_u \frac{\|u-v\|^2}{\|u\|^2}$ as small as possible

$$\|u\|^2 = \int_\Omega u \mathcal{L} u$$



Answer

$$v^\dagger(x) = \mathbb{E}\Big[\xi(x) \Big| \xi(x_i) = u(x_i) \text{ for all } i\Big]$$

Numerical approximation/Optimal recovery game

$\|\cdot\|$: Operator/Energy norm defined by $\mathcal{L}$

$\|u\|^2 = \int_\Omega u\mathcal{L}u = [\mathcal{L}u, u]$

$\phi_1, \ldots, \phi_m \in H^{-s}(\Omega)$

$[\phi, u] := \int_\Omega \phi u$

Player I

Chooses $u \in H^s_0(\Omega)$

Player II

Sees $\big([\phi_1, u], \ldots, [\phi_m, u]\big)$

Chooses $v\big([\phi_1, u], \ldots, [\phi_m, u]\big) \in H^s_0(\Omega)$

Max

Min

$$\frac{\Big\|u - v([\phi_1, u], \ldots, [\phi_m, u])\Big\|^2}{\|u\|^2}$$

Theorem

The optimal strategy of Player II is

$$v^\dagger = \mathbb{E}\Big[\xi \Big| [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \{1, \ldots, m\}\Big]$$

$v^\dagger$ is also the minimizer of

$$\inf_v \sup_{u \in H_0^s(\Omega)} \frac{\Big\| u - v([\phi_1, u], \ldots, [\phi_m, u]) \Big\|^2}{\|u\|^2}$$

C. A. Micchelli. Orthogonal projections are optimal algorithms. *Journal of Approximation Theory*, 40(2):101–110, 1984.

C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.

Representation theorem

$$v^\dagger(x) = \sum_{i=1}^m c_i \int_\Omega G(x,y)\phi_i(y)\,dy$$

$$c_i = \sum_{j=1}^m \Theta_{i,j}^{-1} \int_\Omega u\phi_j$$

$$\Theta_{i,j} = \int_{\Omega^2} \phi_i(x)G(x,y)\phi_j(y)\,dx\,dy$$

Basis functions

$$v^\dagger(x) = \sum_{i=1}^m [\phi_i, u]\psi_i(x)$$

Optimal recovery splines

$$\psi_i = \sum_{j=1}^m \Theta^{-1}_{i,j} \mathcal{L}^{-1}\phi_j$$

Elementary gambles/bets

$$\psi_i = \mathbb{E}\Big[\xi \Big| [\phi_j, \xi] = \delta_{i,j} \text{ for } j \in \{1, \dots, m\}\Big]$$

Biorthogonal system

$$[\phi_j, \psi_i] = \delta_{i,j}$$

$$[\phi, \psi] := \int_\Omega \phi\psi$$

$$\langle u, v\rangle := \int_\Omega u\mathcal{L}v$$

Theorem

The $\langle \cdot, \cdot \rangle$ orthogonal projection of $u \in H_0^s(\Omega)$ onto $\mathcal{L}^{-1}\,\mathrm{span}\{\phi_i \mid i \in \{1, \ldots, m\}\}$ is

$$\sum_{i=1}^{m} [\phi_i, u]\psi_i$$

Numerical Homogenization

$$(1) \quad \mathcal{L}u = f$$

Given m, to find $\psi_1, \ldots, \psi_m$ s.t:

1. *Accuracy.*

   $$\sup_{f \in L^2(\Omega)} \inf_{c \in \mathbb{R}^m} \frac{\|\mathcal{L}^{-1} f - \sum_{i=1}^m c_i \psi_i\|}{\|f\|_{L^2(\Omega)}}$$

   is as small as possible.

2. *Localization.* The ψ_i are as localized as possible.

Numerical Homogenization

Pioneering work in classical homogenization

Papanicolaou, Bensoussan, Lions, Murat, Tartar, Varadhan, Zhikov, Kozlov, Oleinik, Allaire, Nguetseng,... (and many others)

MsFEM Hou, Wu, Efendiev, Fish, Wagiman, Chung...

Harmonic coordinates Babuska, Caloz, Osborn, Allaire, Brizzi, Zhang, O. ...

HMM Engquist, E, Abdulle, Runborg, Schwab,...

Stochastic Homogenization

Papanicolaou, Varadhan, Zhikov, O., Bourgeat, Piatnitsky, Gloria, Otto, Lebris, Legoll, Blanc, Jing, Bal, Sougadinis, E,...

Variational Multiscale Method. Orthogonal Decomposition

Hughes, Feijóo, Mazzei, Quincy, Malqvist, Peterseim,...

Projection based methods

Nolen, Papanicolaou, Pironneau.

Variational Multiscale Method. Orthogonal Decomposition

Hughes, Feijóo, Mazzei, Quincy, Malqvist, Peterseim,...

Flux Norm. Rough Polyharmonic splines

O., Berlyand, Zhang, Symes, Bebendorf,...

Gamblets – Sparse Operator compression

O., Scovel, Schäfer, Sullivan, Hou, Zhang, Huang, Lam, Qin,...

Localization problem

Localization problem in Numerical Homogenization

[Chu-Graham-Hou-2010] (limited inclusions). [Efendiev-Galvis-Wu-2010] (limited inclusions or mask). [Babuska-Lipton 2010] (local boundary eigenvectors). [Owhadi-Zhang 2011] (localized transfer property). [Malqvist-Peterseim 2012] Local Orthogonal Decomposition. [Owhadi-Zhang-Berlyand 2013] (Rough Polyharmonic Splines). [A. Gloria, S. Neukamm, and F. Otto, 2015] (quantification of ergodicity). [Hou and Liu,DCDS-A, 2016]. [Chung-Efendiev-Hou, JCP 2016]. [Owhadi, Multiresolution operator decomposition, SIREV 2017]. [Owhadi, Zhang, gamblets for hyperbolic and parabolic PDEs, 2016]. [Hou, Qin, Zhang, 2016]. [Hou, Zhang, 2017]. [Hou and Zhang, 2017]: Higher order PDEs (localization under strong ellipticity, h sufficiently small, and higher order polynomials as measurement functions). [Hou, Huang, Lam, Zhang, 2017]. [Kornhuber, Peterseim, Yserentant, 2016]: Subspace decomposition.

Subspace decomposition/correction and Schwarz iterative methods

[J. Xu, 1992]: Iterative methods by space decomposition and subspace correction. [Griebel-Oswald, 1995]: Schwarz algorithms.

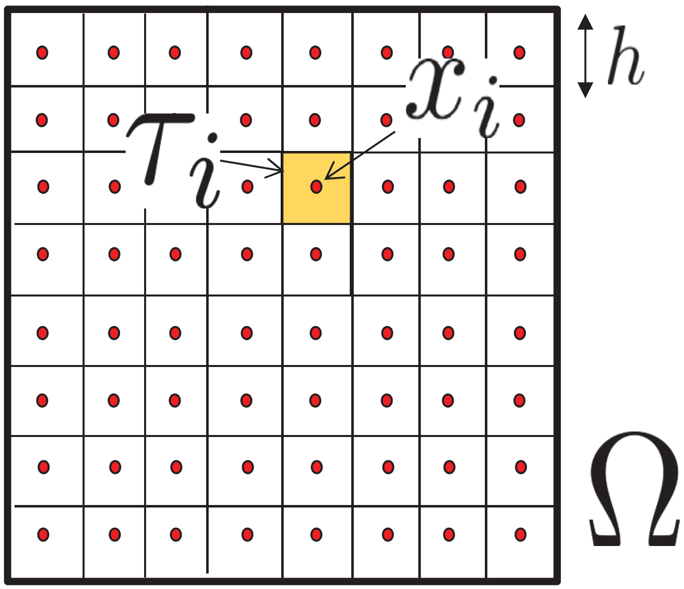
Non-conforming coarse elements. Geometric conditions.

[O. 2015]. [O. - Scovel, 2017]

Wannier basis functions

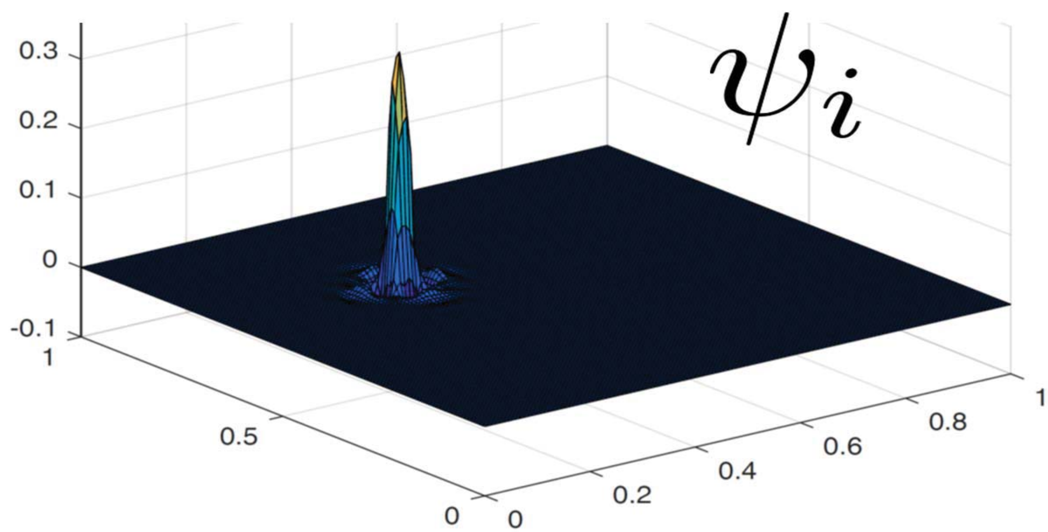
[Wannier 1962] [Kohn 1959] [Marzari, Vanderbilt, 1997]

Measurement functions



- $\phi_i = \frac{1_{\tau_i}}{\sqrt{|\tau_i|}}$.
- $\phi_i = \delta(\cdot - x_i)$, $(s > \frac{d}{2})$

Gamblets



Accuracy

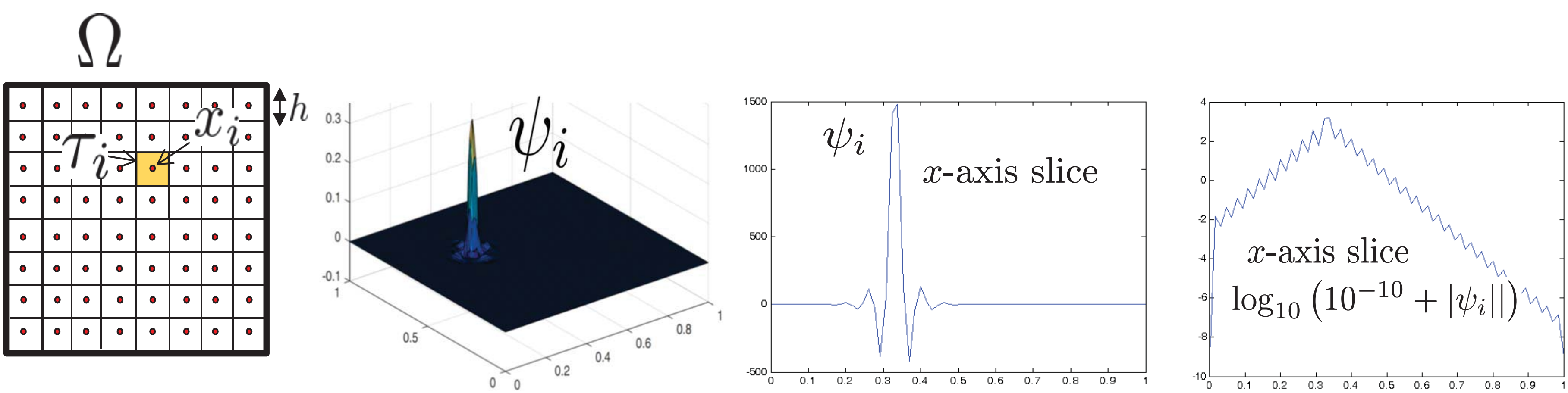
$$\inf_{v \in \text{span}\{\psi_1,\ldots,\psi_m\}} \|u - v\| \leq C h^s \|\mathcal{L}u\|_{L^2(\Omega)}$$

Achieves Kolmogorov n-width accuracy
up to multiplicative constant

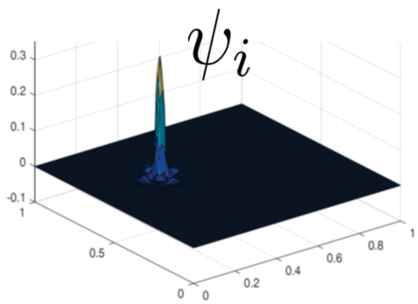
Localization

$$\|\psi_i\|_{H^s(\Omega/B(x_i,nh))} \leq C e^{-n/C}$$

4



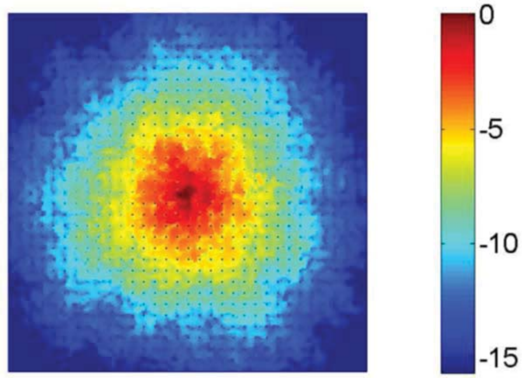
Localization of Gamblets

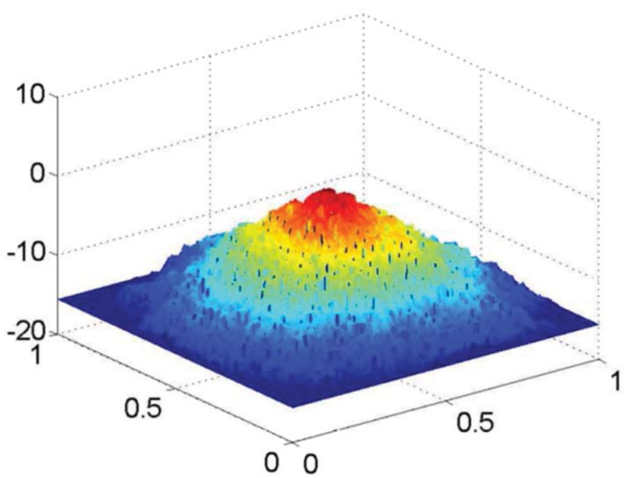


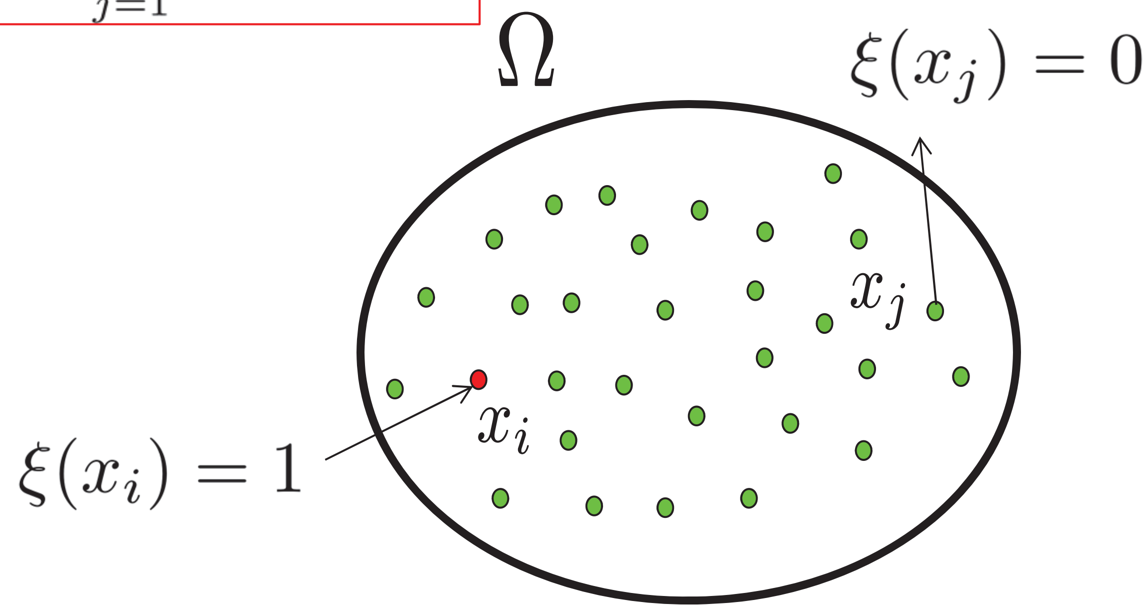
$$\xi \sim \mathcal{N}(0, \mathcal{L}^{-1})$$

$$\psi_i(x) = \mathbb{E}\big[\xi(x) \mid \xi(x_j) = \delta_{i,j} \text{ for } j \in \mathcal{I}\big]$$

$$\psi_i(x) = \sum_{j=1}^{m} \Theta^{-1}_{i,j} G(x, x_j)$$

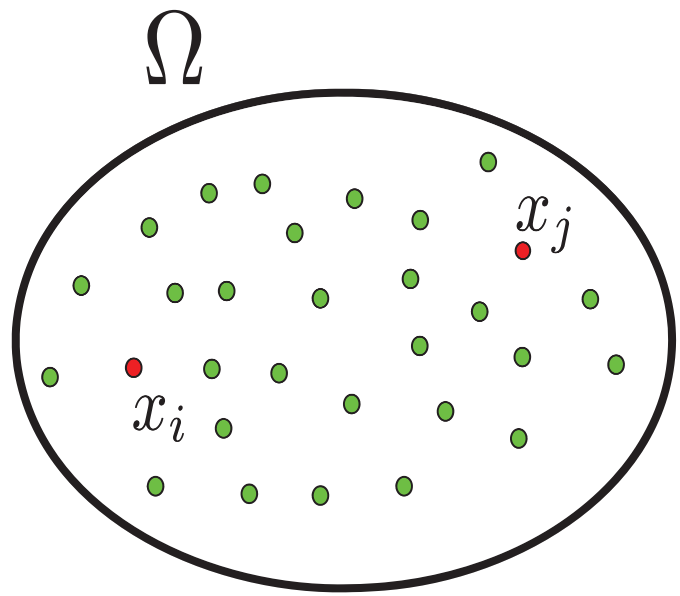






Screening effect

$$\mathrm{Cor}\left(\xi(x_i), \xi(x_j) \middle| \xi(x_l), l \neq i, j]\right) = -\frac{\langle \psi_i, \psi_j \rangle}{\|\psi_i\| \|\psi_j\|}$$



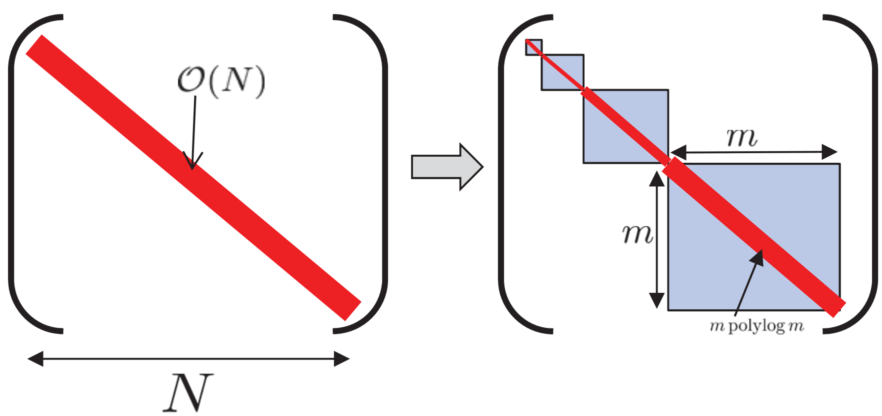
$$\mathrm{Cor}\left(\xi(x_i), \xi(x_j) \middle| \xi(x_l), l \neq i, j]\right) \leq C e^{-\,\mathrm{dist}(x_i, x_j)/h}$$

Operator adapted wavelets

$$\left(H_0^s(\Omega), \|\cdot\|_{H_0^s(\Omega)}\right) \xrightarrow{\quad\mathcal{L}\quad} \left(H^{-s}(\Omega), \|\cdot\|_{H^{-s}(\Omega)}\right)$$

How to construct operator adapted wavelets for $\mathcal{L}$?

1. Scale-orthogonal wavelets with respect to operator scalar product (leads to block-diagonalization)
2. Operator needs to be well conditioned within each subband
3. Wavelets need to be localized (compact support or exp. decay)



Operator adapted wavelets

First Generation Wavelets: Signal and imaging processing

Morlet, Grossmann, Mallat, Daubechies, Coifman, Meyer, Wickerhauser,...

First Generation Operator Adapted Wavelets (shift and scale invariant)

Cohen, Daubechies, Feauveau (Biorthogonal bases of compactly supported wavelets), Beylkin, Coifman, Rokhlin, Engquist, Osher, Zhong, Alpert, Jawerth, Sweldens, Dahlke, Weinreich, Bacry, Mallat, Papanicolaou, Bertoluzza, Maday, Ravel, Vasilyev, Paolucci, Dahmen, Kunoth, Stevenson, Candes...

Lazy wavelets (Multiresolution decomposition of solution space)

Yserentant (Multilevel splitting), Bank, Dupont, Yserentant (Hierarchical basis multigrid method),...

Second Generation Operator Adapted Wavelets

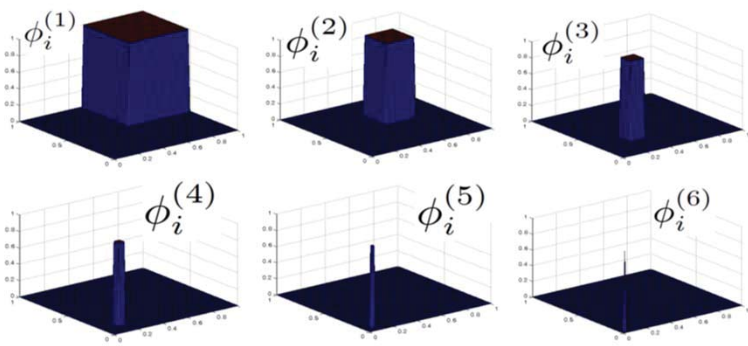
Sweldens (the lifting scheme); Dorobantu, Engquist; Vassilevski, Wang (stabilizing the hierarchical basis); Carnicer, Dahmen, Peña, Lounsbery, DeRose, Warren, Barinka, Barsch, Charton, Cohen, Dahlke, Dahmen, Urban, Cohen, Dahmen, DeVore, Chiavassa, Liandrat, Dahmen, Kunoth, Schwab, Stevenson, Sudarshan, Engquist, Runborg, Yin, Liandrat,...

$$H_0^s(\Omega) \xrightarrow{\quad \mathcal{L} \quad} H^{-s}(\Omega)$$

$$\cup$$

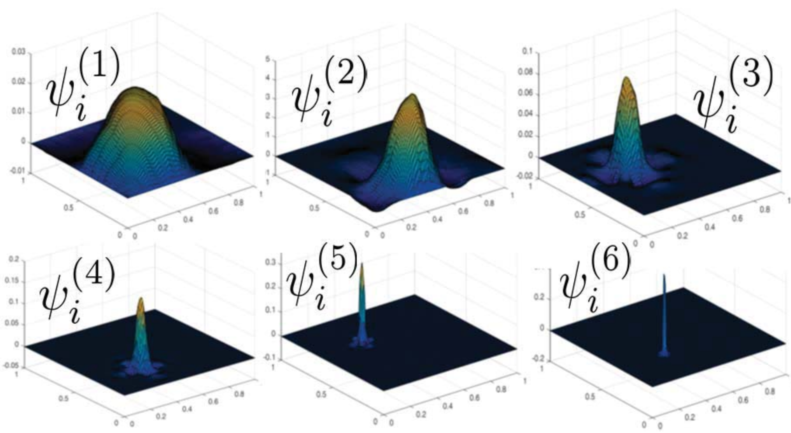
$$L^2(\Omega)$$

The method

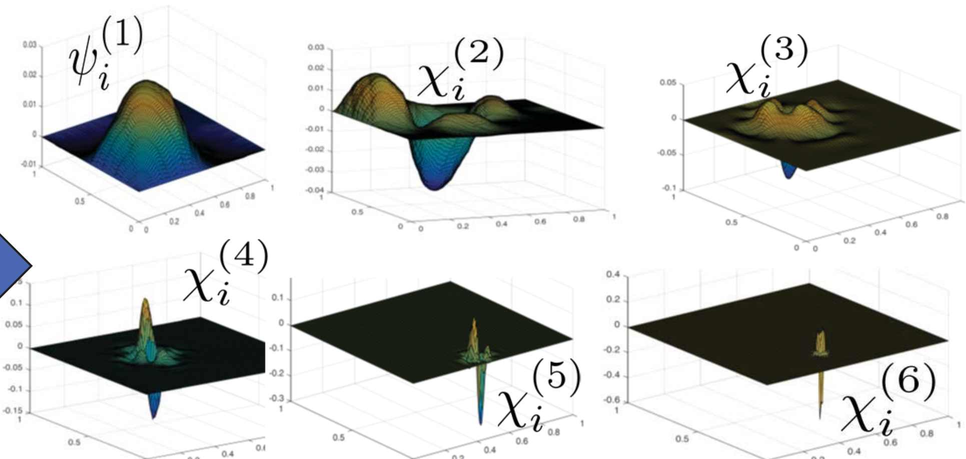


Haar pre-wavelet decomposition of $L^2(\Omega) \to H^{-s}(\Omega)$

Gamblet transform



Multi-resolution decomposition of $H_0^s(\Omega) \to H^{-s}(\Omega)$



Hierarchy of measurement functions

$$\phi_i^{(k)} \in H^{-s}(\Omega) \text{ with } k \in \{1, \ldots, q\}$$

$$\phi_i^{(k)} = \sum_j \pi_{i,j}^{(k,k+1)} \phi_j^{(k+1)}$$

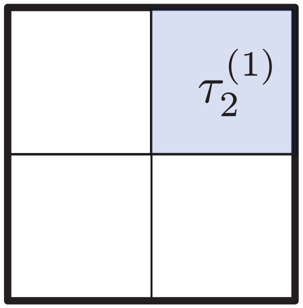
Hierarchy of gamblets

$$\psi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k)}} \Theta_{i,j}^{(k),-1} \mathcal{L}^{-1} \phi_j^{(k)}$$

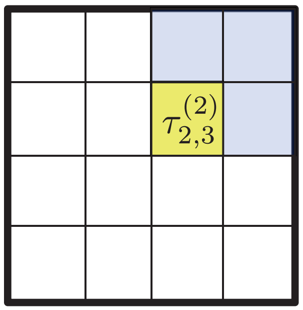
$$\Theta_{i,j}^{(k)} := \int \phi_i^{(k)} \mathcal{L}^{-1} \phi_j^{(k)}$$

Example

$\phi_i^{(k)}$: Haar (pre)-wavelets



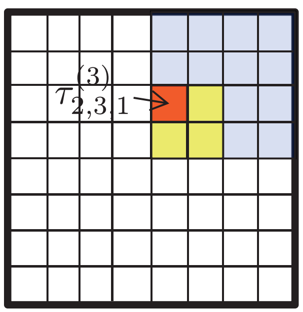
$\phi_i^{(k)}$: Weighted indicator functions of a hierarchical nested partition of Ω of resolution 2^{-k}

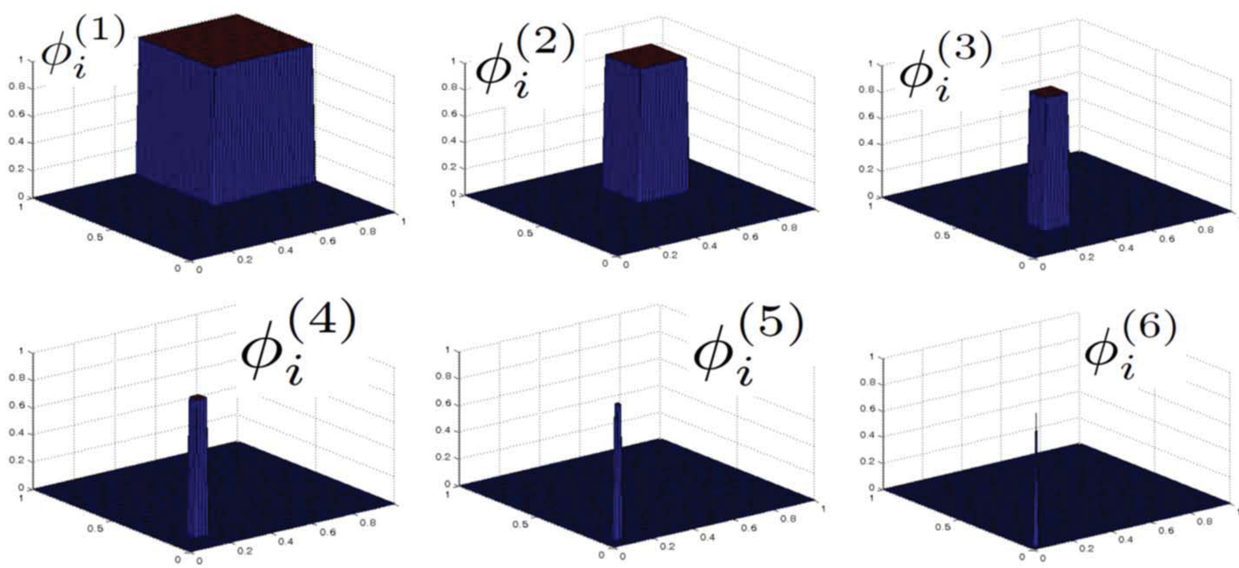


0	0	1/2	1/2
0	0	1/2	1/2
0	0	0	0
0	0	0	0

$\pi_{i,\cdot}^{(1,2)}$

$$\phi_i^{(k-1)} = \sum_j \pi_{i,j}^{(k-1,k)} \phi_j^{(k)}$$

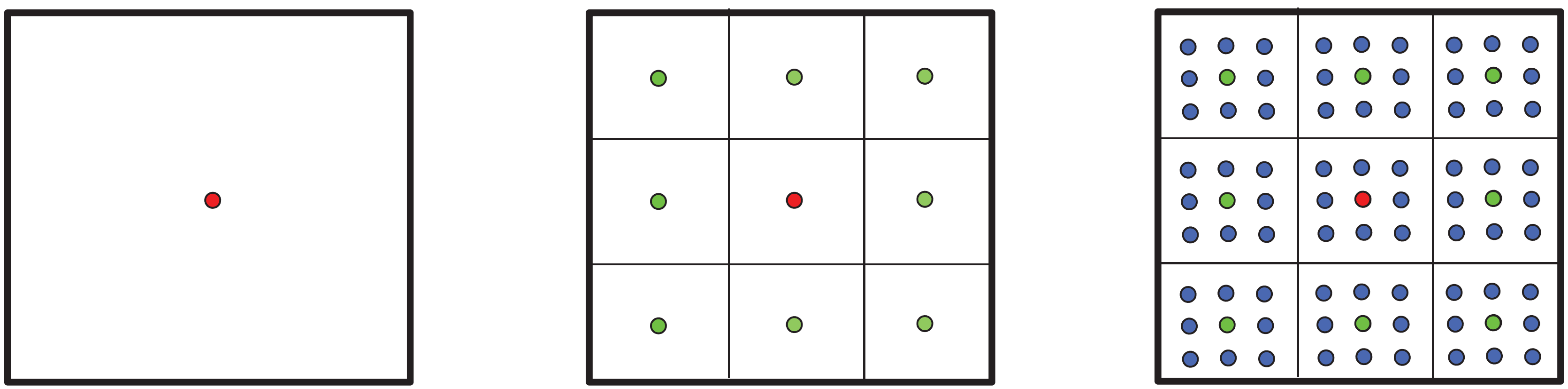




Example

$\phi_i^{(k)}$: Sub-sampled diracs

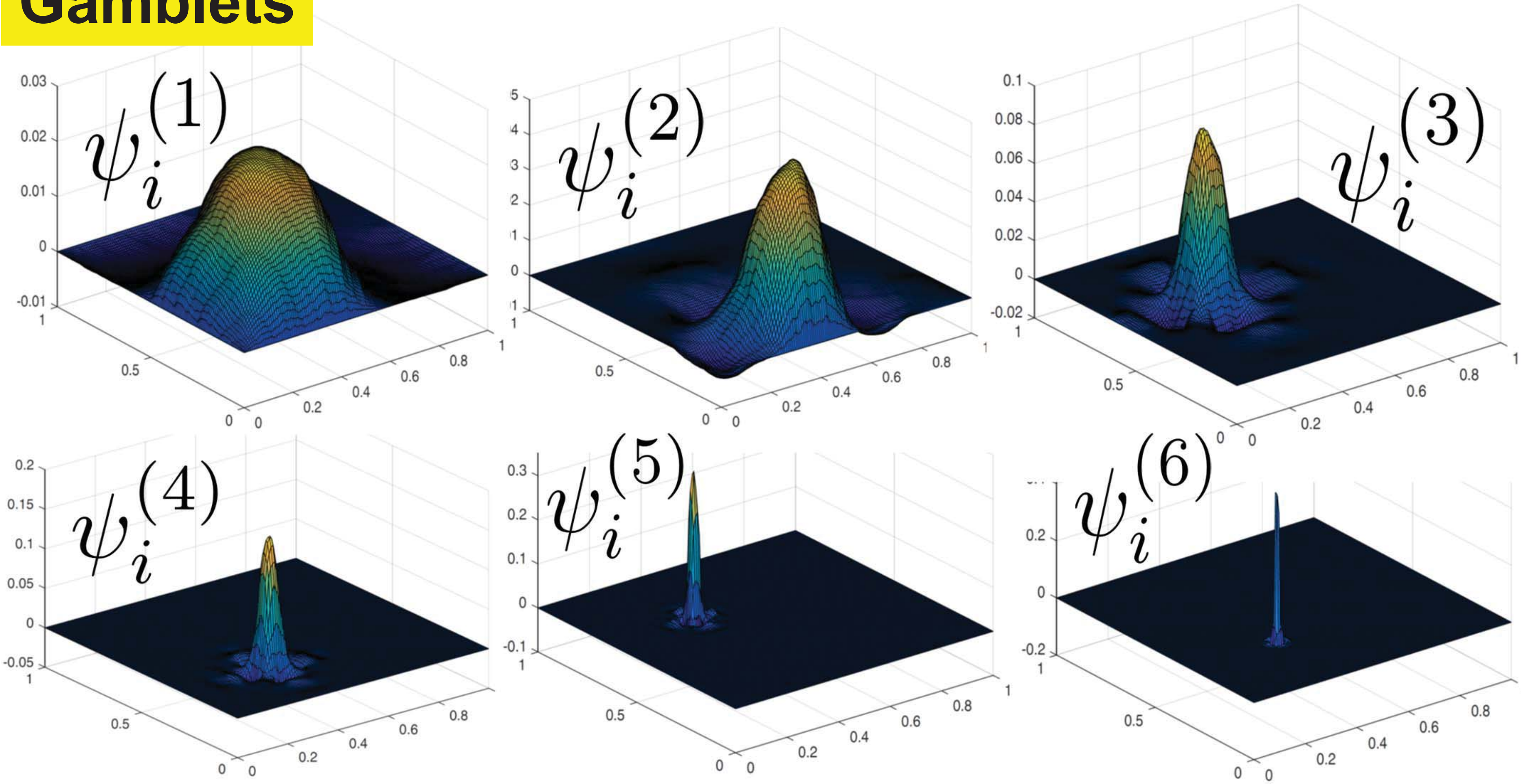
$$s > d/2$$

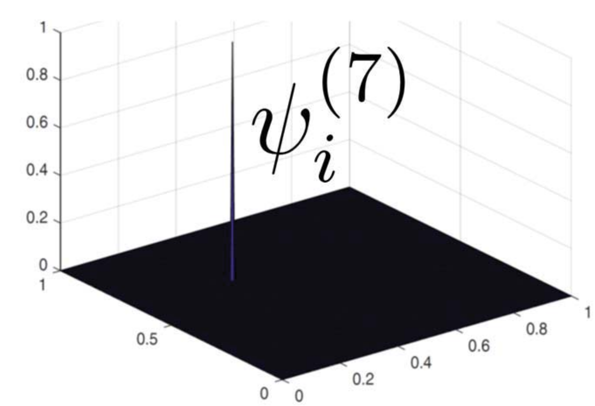


$\phi_i^{(1)} = \delta(x - x_i^{(1)})$ $\phi_j^{(2)} = \delta(x - x_j^{(2)})$ $\phi_l^{(3)} = \delta(x - x_l^{(3)})$

[Schäfer, Sullivan, O., 2017]

Gamblets





$$\psi_i^{(k)} = \sum_{j\in\mathcal{I}^{(k)}} \Theta_{i,j}^{(k),-1}\mathcal{L}^{-1}\phi_j^{(k)}$$

$$\Theta_{i,j}^{(k)} := \int \phi_i^{(k)}\mathcal{L}^{-1}\phi_j^{(k)}$$

Measurement functions are nested

$$\phi_i^{(k)} = \sum_j \pi_{i,j}^{(k,k+1)} \phi_j^{(k+1)}$$

$$\psi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k)}} \Theta_{i,j}^{(k),-1} \mathcal{L}^{-1} \phi_j^{(k)}$$

Gamblets are nested

$$\psi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k+1)}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1)}$$

$$\mathfrak{V}^{(k)} := \text{span}\{\psi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\}$$

$$\mathfrak{V}^{(k-1)} \subset \mathfrak{V}^{(k)}$$

$\mathfrak{W}^{(k)}$: Orthogonal complement of $\mathfrak{V}^{(k-1)}$ in $\mathfrak{V}^{(k)}$ w.r. to $\langle u, v \rangle := \int_\Omega u \mathcal{L} v$

$$\mathfrak{V}^{(k)} = \mathfrak{V}^{(k-1)} \oplus \mathfrak{W}^{(k)}$$

Theorem

$$H_0^s(\Omega) = \mathfrak{V}^{(1)} \oplus \mathfrak{W}^{(2)} \oplus \mathfrak{W}^{(3)} \oplus \cdots$$

Question

What are the basis functions for $\mathfrak{W}^{(k)}$?

$$\chi \perp \mathfrak{V}^{(k-1)} \Leftrightarrow [\phi_i^{(k-1)}, \chi] = 0 \text{ for all } i \in \mathcal{I}^{(k-1)}$$

$$\chi \in \mathfrak{V}^{(k)} \Leftrightarrow \chi = \sum_j c_j \psi_j^{(k)}$$

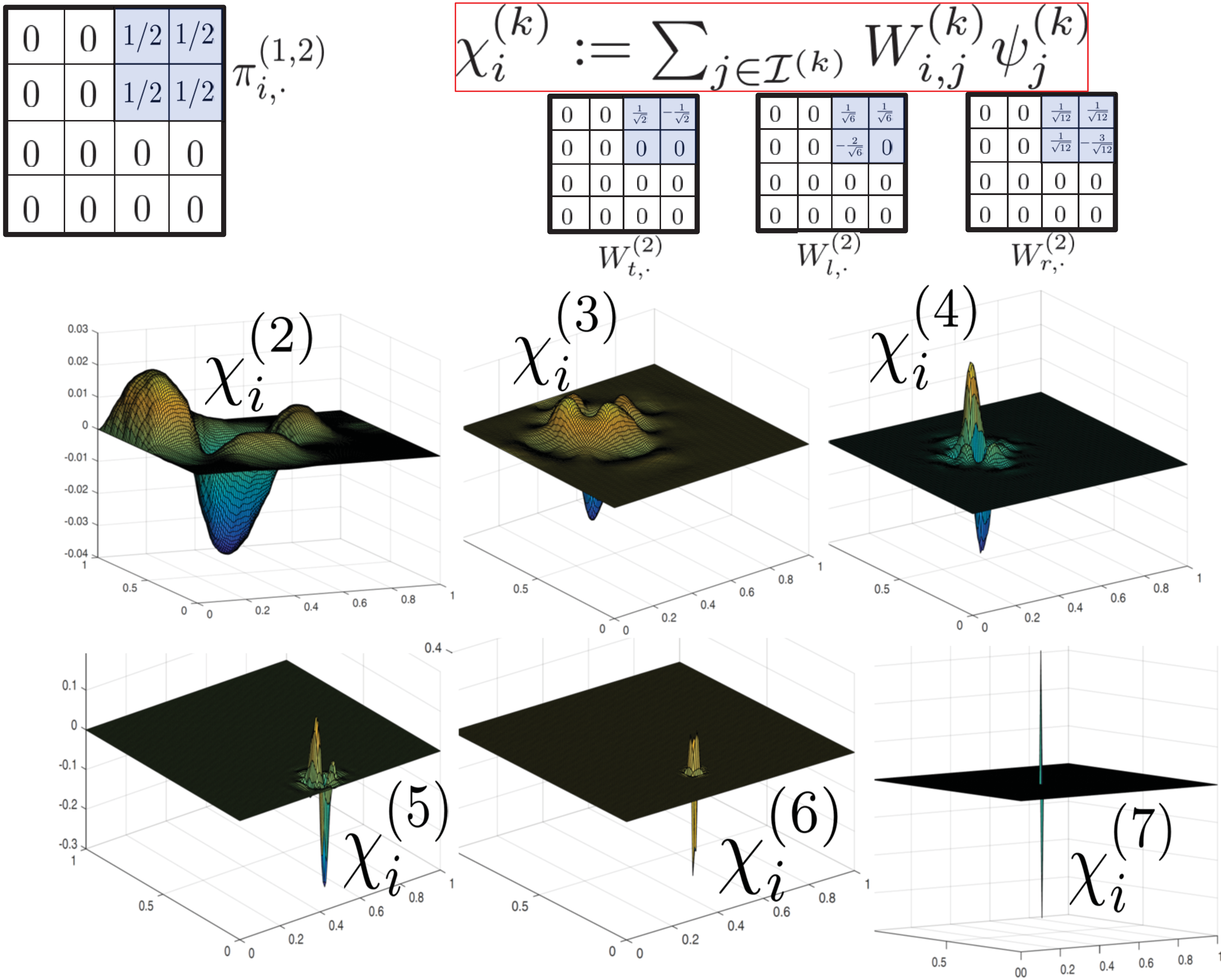
$$\chi \in \mathfrak{W}^{(k)} \Leftrightarrow \pi^{(k-1,k)} c = 0$$

Basis functions for $\mathfrak{W}^{(k)}$

$$i \in \mathcal{I}^{(k)} / \mathcal{I}^{(k-1)}$$

$$\chi_i^{(k)} := \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$$

$$\mathrm{Img}(W^{(k),T}) = \mathrm{Ker}(\pi^{(k-1,k)})$$
$$W^{(k)} (W^{(k)})^T = J^{(k)}$$



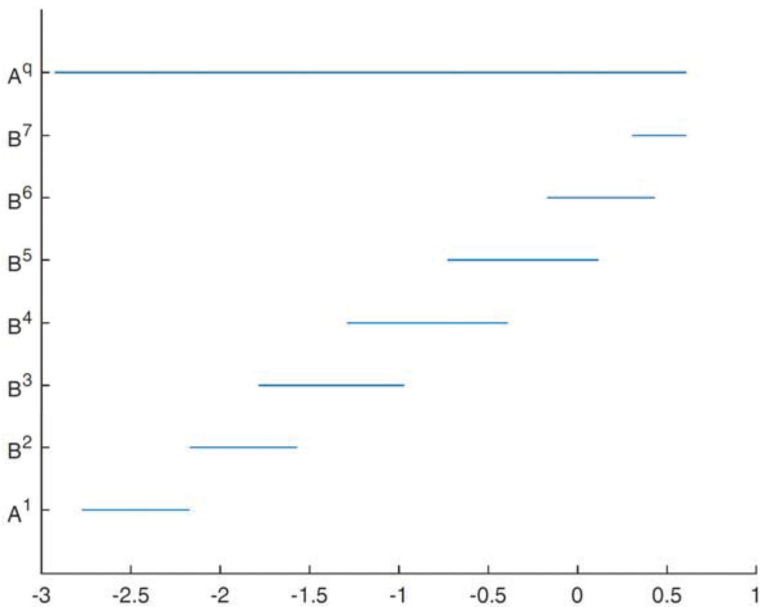
$\pi_{i,\cdot}^{(1,2)}$
$\chi_i^{(k)} := \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$
$W_{t,\cdot}^{(2)}$
$W_{l,\cdot}^{(2)}$
$W_{r,\cdot}^{(2)}$
$\chi_i^{(2)}$
$\chi_i^{(3)}$
$\chi_i^{(4)}$
$\chi_i^{(5)}$
$\chi_i^{(6)}$
$\chi_i^{(7)}$

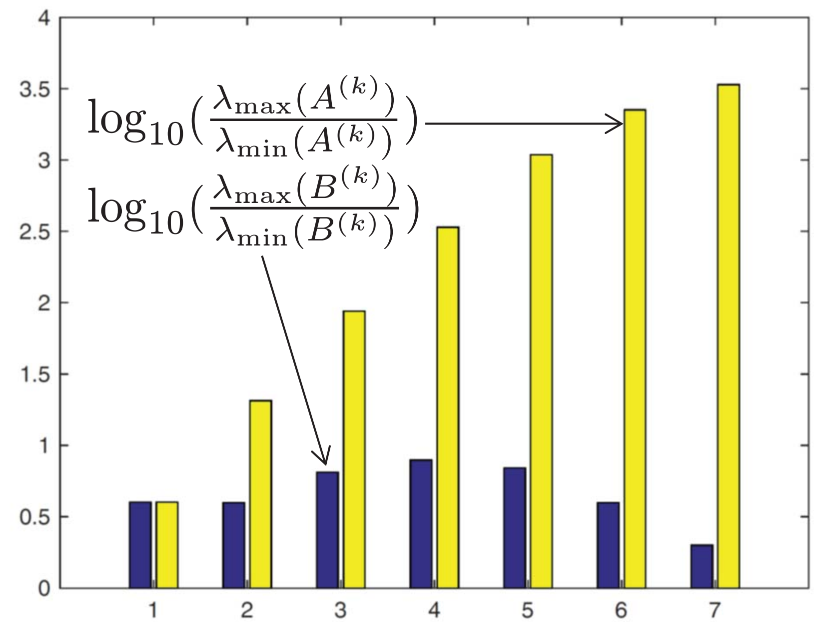
The operator is well conditioned in each subband

$$A^{(k)}_{i,j} = \langle \psi^{(k)}_i, \psi^{(k)}_j \rangle \qquad B^{(k)}_{i,j} = \langle \chi^{(k)}_i, \chi^{(k)}_j \rangle$$

Theorem

$$\mathrm{Cond}(A^{(1)}) \leq C \qquad \mathrm{Cond}(B^{(k)}) \leq C$$





[O., Scovel, 2017] [Schäfer, Sullivan, O., 2017]

Interpolation operator

$$\psi_i^{(k-1)} = \sum_{j\in\mathcal{I}^{(k)}} R_{i,j}^{(k-1,k)} \psi_j^{(k)}$$

$$[\phi_j^{(k-1)}, \psi_i^{(k-1)}] = \delta_{i,j} \qquad \psi_i^{(k-1)} \perp \mathfrak{W}^{(k)}$$

$$\psi_i^{(k-1)} = \sum_j \pi_{i,j}^{(k-1,k)} \psi_j^{(k)} + \sum_j c_j \chi_j^{(k)}$$

$$0 = (\pi^{(k-1,k)} A^{(k)} W^{(k),T})_{i,j} + (B^{(k)} c)_j$$

$$R^{(k-1,k)} = \pi^{(k-1,k)} (I^{(k)} - A^{(k)} W^{(k),T} B^{(k),-1} W^{(k)})$$

Gamblet Transform

$\psi_i^{(q)} = \varphi_i$

$A_{i,j}^{(q)} = \langle \psi_i^{(q)}, \psi_j^{(q)} \rangle$

for $k = q$ to 2 **do**

$\quad B^{(k)} = W^{(k)} A^{(k)} W^{(k),T}$

$\quad \chi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$

$\quad R^{(k-1,k)} = \pi^{(k-1,k)} (I^{(k)} - A^{(k)} W^{(k),T} B^{(k),-1} W^{(k)})$

$\quad A^{(k-1)} = R^{(k-1,k)} A^{(k)} R^{(k,k-1)}$

$\quad \psi_i^{(k-1)} = \sum_{j \in \mathcal{I}^{(k)}} R_{i,j}^{(k-1,k)} \psi_j^{(k)}$

end for

Theorem

$\mathcal{O}(N \log^{3d}(N))$ complexity

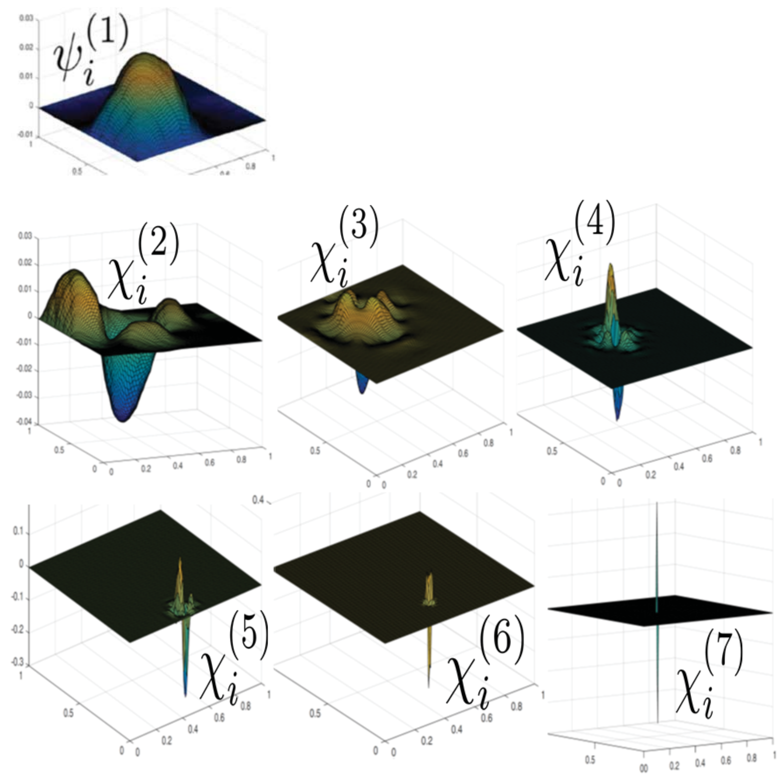
to achieve grid size accuracy in energy norm

Sparse Rank Revealing Representation of the Green's function

$$G(x,y) = \sum_{i,j} A_{i,j}^{(1),-1} \psi_i^{(1)}(x)\psi_j^{(1)}(y) + \sum_{k\geq 2}\sum_{i,j} B_{i,j}^{(k),-1} \chi_i^{(k)}(x)\chi_j^{(k)}(y)$$

$$C^{-1}h^{2s} \leq A^{(1),-1} \leq C$$

$$C^{-1}h^{2sk} \leq B^{(k),-1} \leq Ch^{2s(k-1)}$$

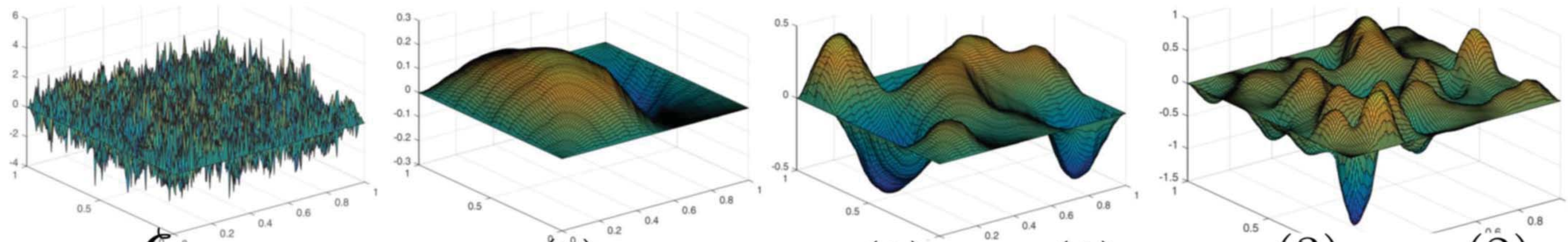


Sparse Representation Of the Gaussian Process

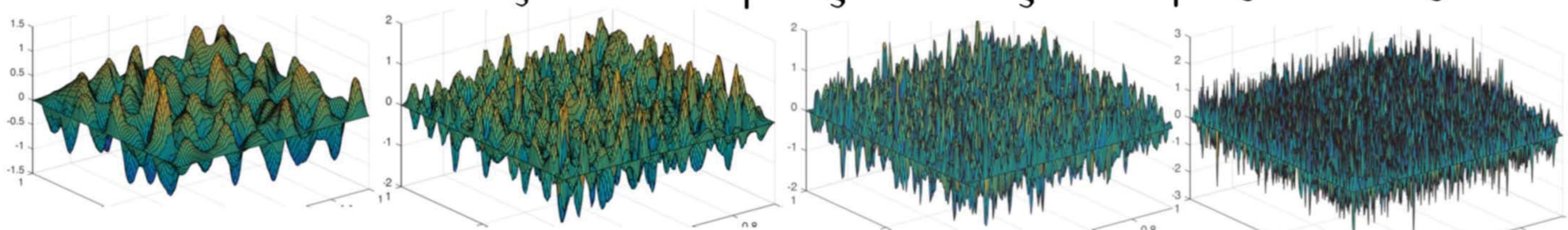
$$\xi = \xi^{(1)} + (\xi^{(2)} - \xi^{(1)}) + \cdots + (\xi^{(k)} - \xi^{(k-1)}) + \cdots$$

$$\xi^{(1)} = \sum_i Y_i^{(1)} \psi_i^{(1)} \qquad \xi^{(k)} - \xi^{(k-1)} = \sum_i Y_i^{(k)} \chi_i^{(k)}$$

$$Y^{(1)} \sim \mathcal{N}(0, A^{(1),-1}) \qquad Y^{(k)} \sim \mathcal{N}(0, B^{(k),-1})$$



$$\xi = \xi^{(1)} + \xi^{(2)} - \xi^{(1)} + \xi^{(3)} - \xi^{(2)}$$

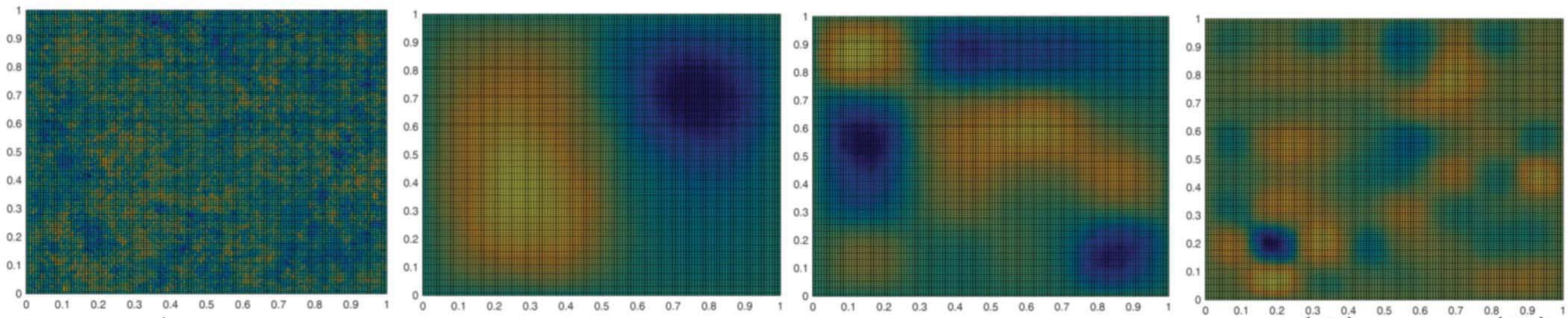


$$+\xi^{(4)} - \xi^{(3)} + \xi^{(5)} - \xi^{(4)} + \xi^{(6)} - \xi^{(5)} + \xi^{(7)} - \xi^{(6)}$$

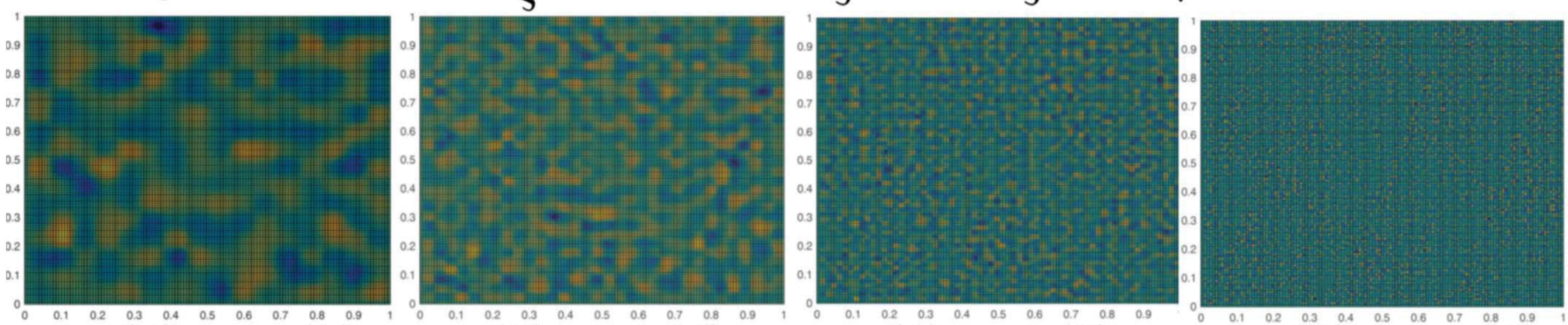
Sparse Representation Of the Gaussian Process

$$\xi = \xi^{(1)} + (\xi^{(2)} - \xi^{(1)}) + \cdots + (\xi^{(k)} - \xi^{(k-1)}) + \cdots$$

$$\xi^{(1)} = \sum_i Y_i^{(1)} \psi_i^{(1)} \qquad \xi^{(k)} - \xi^{(k-1)} = \sum_i Y_i^{(k)} \chi_i^{(k)}$$



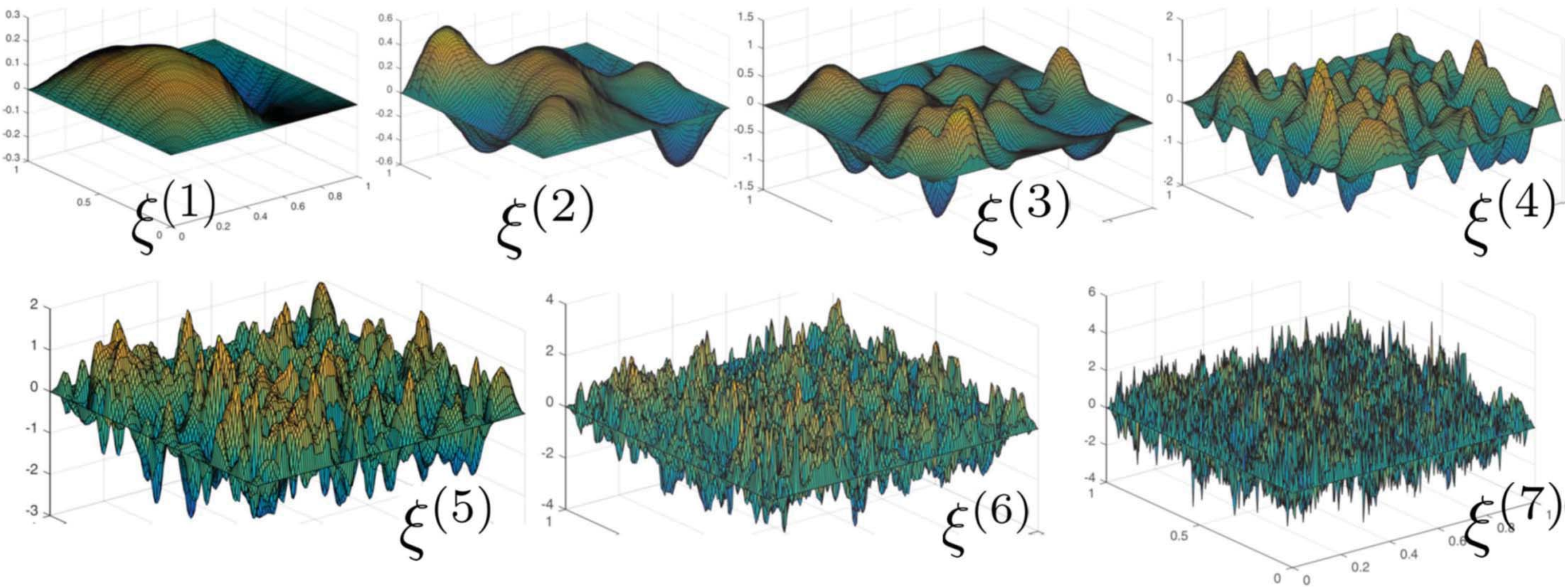
$$\xi = \xi^{(1)} + \xi^{(2)} - \xi^{(1)} + \xi^{(3)} - \xi^{(2)}$$



$$+\xi^{(4)} - \xi^{(3)} + \xi^{(5)} - \xi^{(4)} + \xi^{(6)} - \xi^{(5)} + \xi^{(7)} - \xi^{(6)}$$

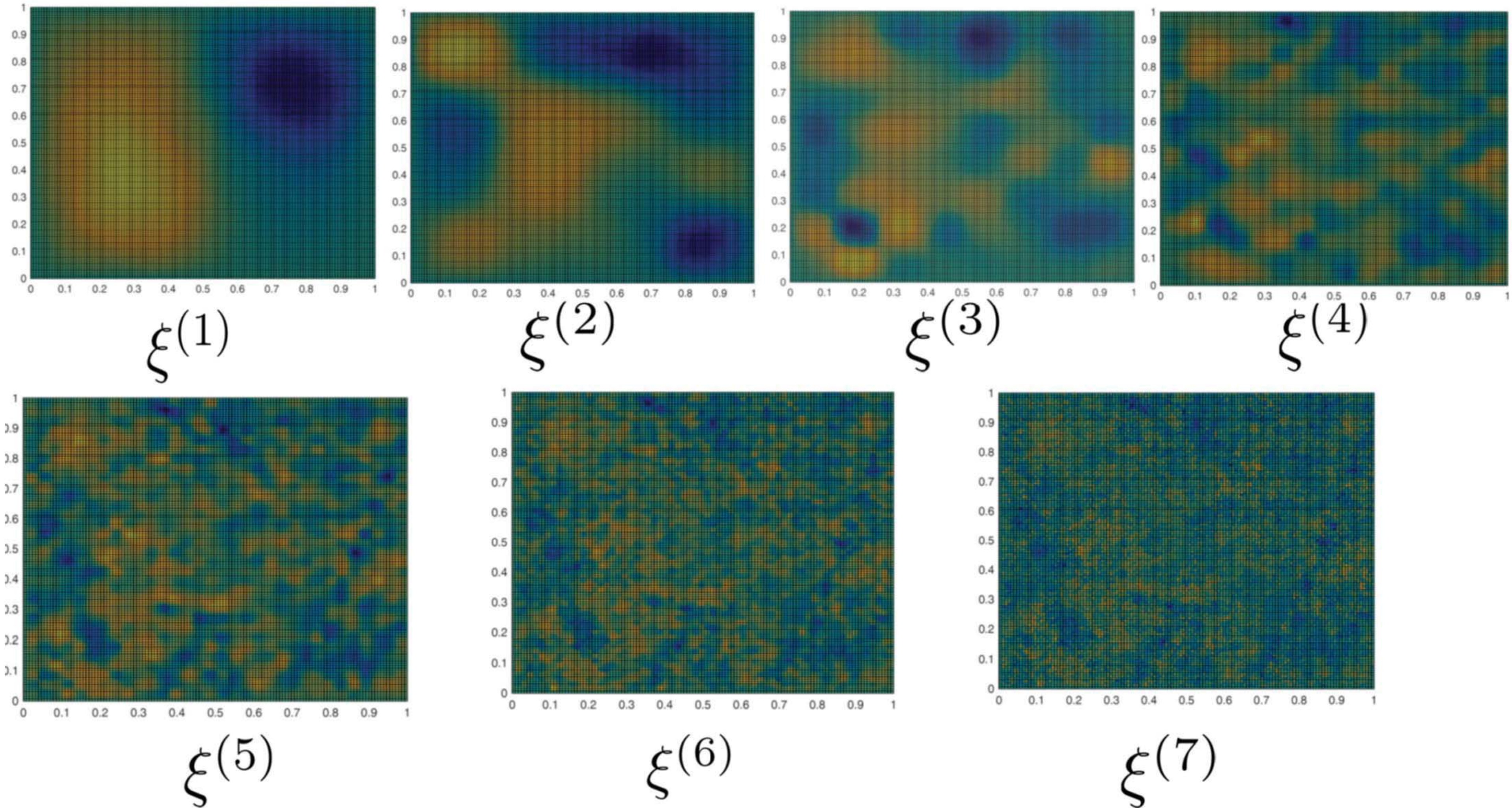
Sparse Representation Of the Gaussian Process

$$\xi^{(k)} = \sum_i Y_i^{(1)} \psi_i^{(1)} + \sum_{k'=2}^{k} \sum_i Y_i^{(k')} \chi_i^{(k')}$$



Sparse Representation Of the Gaussian Process

$$\xi^{(k)} = \sum_i Y_i^{(1)} \psi_i^{(1)} + \sum_{k'=2}^{k} \sum_i Y_i^{(k')} \chi_i^{(k')}$$



Fast solvers

Multigrid Methods [Fedorenko, 1961, Brandt, 1973, Hackbusch, 1978]

Multiresolution/Wavelet based methods

[Brewster and Beylkin, 1995, Beylkin and Coult, 1998, Averbuch et al., 1998] [Beylkin, Coifman, Rokhlin, 1992] [Engquist, Osher, Zhong, 1992] [Alpert, Beylkin, Coifman, Rokhlin, 1993] [Cohen, Daubechies, Feauveau. 1992] [Bacry, Mallat, Papanicolaou. 1993]

Robust/Algebraic multigrid

[Mandel et al., 1999], [Wan-Chan-Smith, 1999], [Xu and Zikatanov, 2004], [Xu and Zhu, 2008], [Ruge-Stüben, 1987], [Panayot - 2010]

Stabilized Hierarchical bases, Multilevel preconditioners

[Vassilevski - Wang, 1997, 1998], [Panayot - Vassilevski, 1997], [Chow - Vassilevski, 2003], [Aksoylu- Holst, 2010]

Low Rank Matrix Decomposition methods

Fast Multipole Method: [Greengard and Rokhlin, 1987]
Hierarchical Matrix Method: [Hackbusch et al., 2002], [Bebendorf, 2008]
Direct Solvers: [Martinsson, Gillman, Hao, 2009-2016] [Ho, Ying: 2016]

Graph sparsification for structured matrices

[Spielman and Teng , 2004-14], [Kelner, et al, 2013], [Koutis, Miller, Gary and Peng , 2014], [Cohen, Kyng, Miller, Pachocki, Peng, Rao, and Xu, 2014], [Kyng, Lee, Peng, Sachdeva, Spielman , 2016], [Kyng and Sachdeva, 2016]

$$(1) \quad \mathcal{L}u = f$$

Fast Solvers

How to solve (1) as fast as possible
up to a given accuracy?

How to construct a fast solver with some degree of universality?

$\mathcal{L}$: arbitrary, symmetric, positive, continuous linear bijection

$$\left(H_0^s(\Omega), \|\cdot\|_{H_0^s(\Omega)}\right) \xrightarrow{\quad\mathcal{L}\quad} \left(H^{-s}(\Omega), \|\cdot\|_{H^{-s}(\Omega)}\right)$$

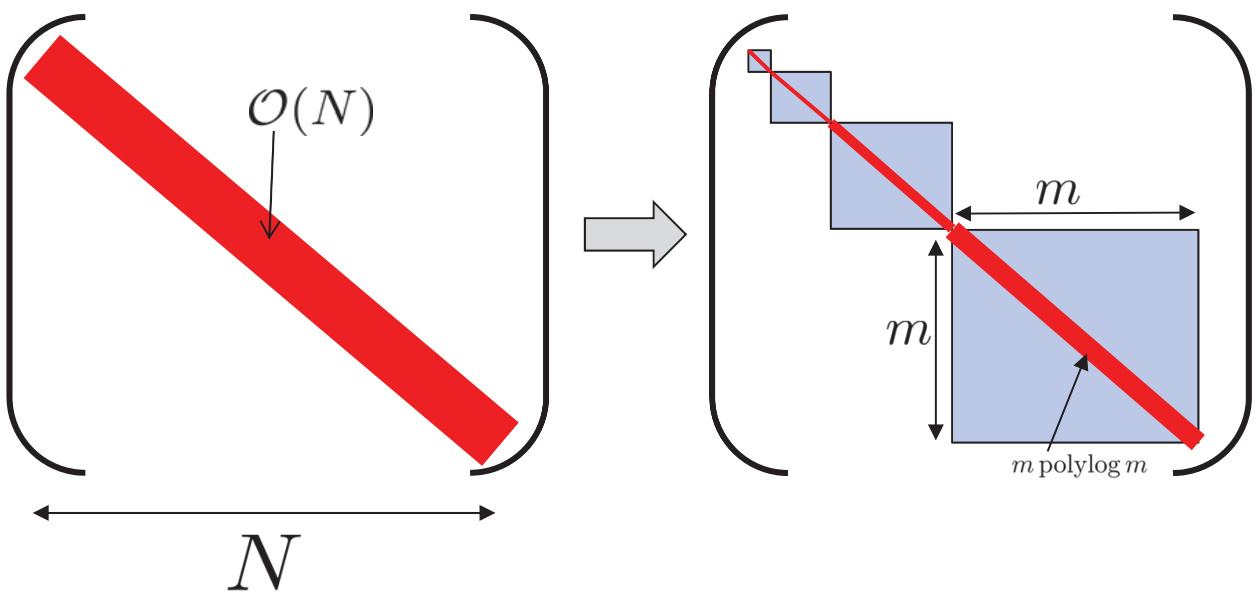
$\mathcal{L}$: Arbitrary symmetric positive continuous linear bijection

$$H_0^s(\Omega) \xrightarrow{\quad\mathcal{L}\quad} H^{-s}(\Omega)$$

Gamblet transform

$$H_0^s(\Omega) = \mathfrak{V}^{(1)} \oplus \mathfrak{W}^{(2)} \oplus \mathfrak{W}^{(3)} \oplus \cdots$$

$$\|u\|^2 := \int_\Omega u\mathcal{L}u$$



$$H_0^s(\Omega) = \mathfrak{W}^{(1)} \oplus \mathfrak{W}^{(2)} \oplus \mathfrak{W}^{(3)} \oplus \cdots$$

$$\boxed{\mathcal{L}u = f}$$

Theorem

$$u = v^{(1)} + \cdots + v^{(k)} + \cdots$$

$$v^{(k)} = \sum_{i \in \mathcal{I}^{(k)}} w_i^{(k)} \chi_i^{(k)}$$

$$\boxed{B^{(k)} w^{(k)} = f^{(k)}}$$

$$f_i^{(k)} = \int_\Omega f \chi_i^{(k)} \qquad B_{i,j}^{(k)} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle$$

Gamblet Solve

$f_i^{(q)} = \int_\Omega f \psi_i^{(q)}$

for $k = q$ to 2 **do**

$\quad w^{(k)} = B^{(k),-1} W^{(k)} f^{(k)}$

$\quad u^{(k)} - u^{(k-1)} = \sum_{i \in \mathcal{J}^{(k)}} w_i^{(k)} \chi_i^{(k)}$

$\quad f^{(k-1)} = R^{(k-1,k)} f^{(k)}$

end for

$U^{(1)} = A^{(1),-1} f^{(1)}$

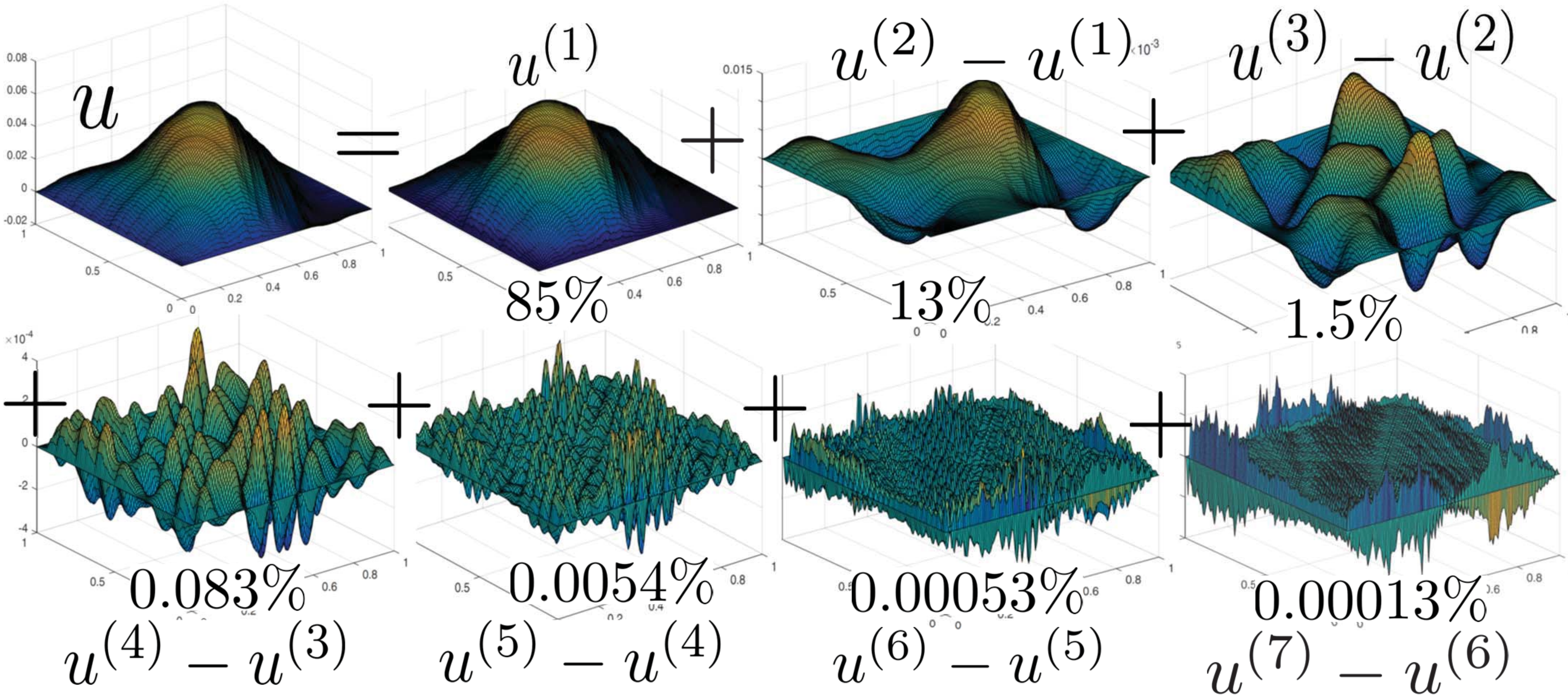
$u^{(1)} = \sum_{i \in \mathcal{I}^{(1)}} U_i^{(1)} \psi_i^{(1)}$

$u = u^{(1)} + (u^{(2)} - u^{(1)}) + \cdots + (u^{(q)} - u^{(q-1)})$

Theorem

$\mathcal{O}(N \log^{d+1}(N))$ complexity

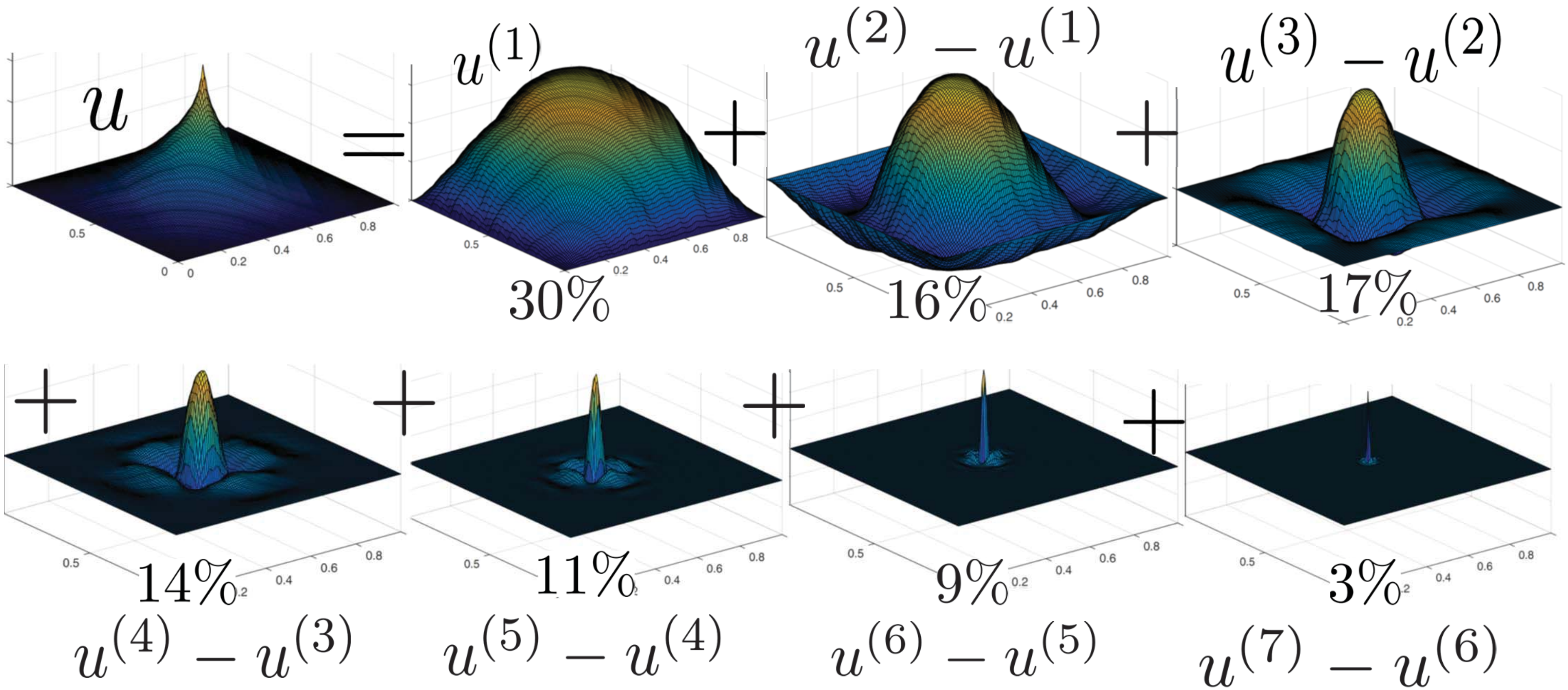
to achieve grid size accuracy in energy norm



Energy content

$$\begin{cases} -\operatorname{div}(a\nabla u) = f, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$

$$f \in C^\infty(\Omega)$$



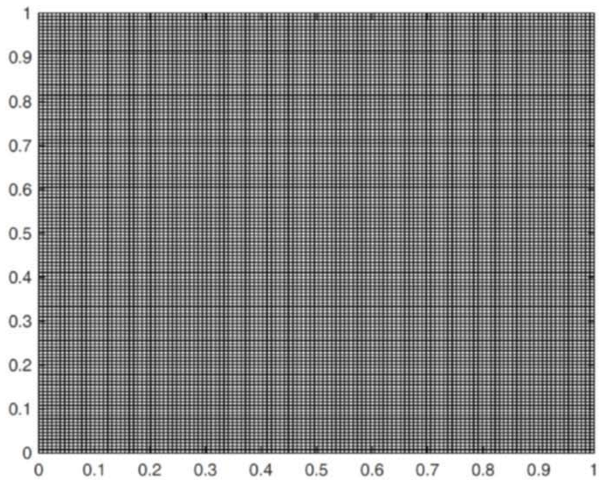
Energy content

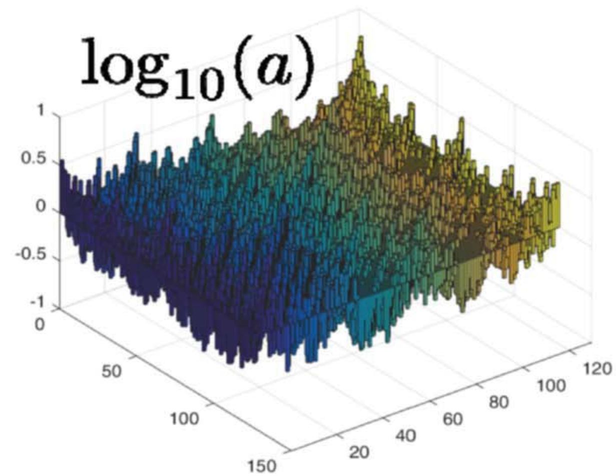
$$\begin{cases} -\operatorname{div}(a\nabla u) = f, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$

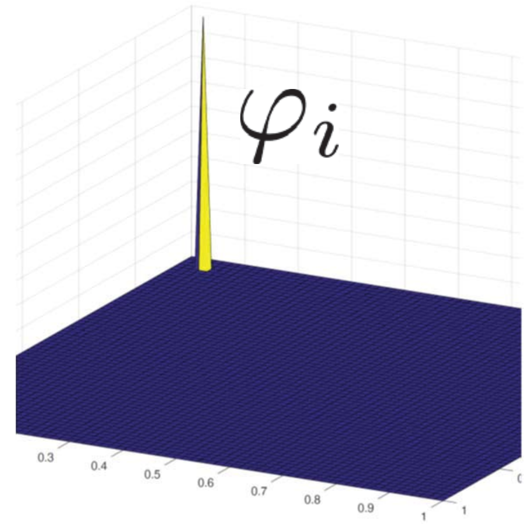
$$f = \delta(x - x_0)$$

Example

$$\begin{cases} -\operatorname{div}(a\nabla u) = f, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$





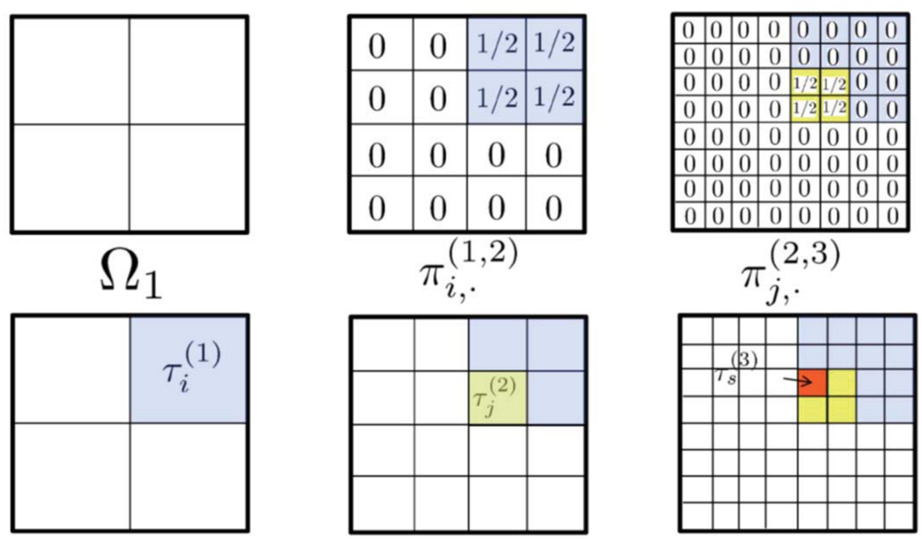


$$\mathrm{Span}\{\varphi_i | i \in \mathcal{I}\} \subset H_0^1(\Omega)$$

$$\|u\|^2 := \int_\Omega (\nabla u)^T a \nabla u$$

Inputs of the algorithm

$$A_{i,j} = \int_\Omega (\nabla \varphi_i)^T a \nabla \varphi_j$$



$$\pi^{(k-1,k)}$$

$$\pi^{(k-1,k)}(\pi^{(k-1,k)})^T = I^{(k-1)}$$

$$t^{(1)} = l^{(1)} = r^{(1)} = i$$

0	0	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$
0	0	0	0
0	0	0	0
0	0	0	0

$W^{(2)}_{t,\cdot}$

0	0	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$
0	0	$-\frac{2}{\sqrt{6}}$	0
0	0	0	0
0	0	0	0

$W^{(2)}_{l,\cdot}$

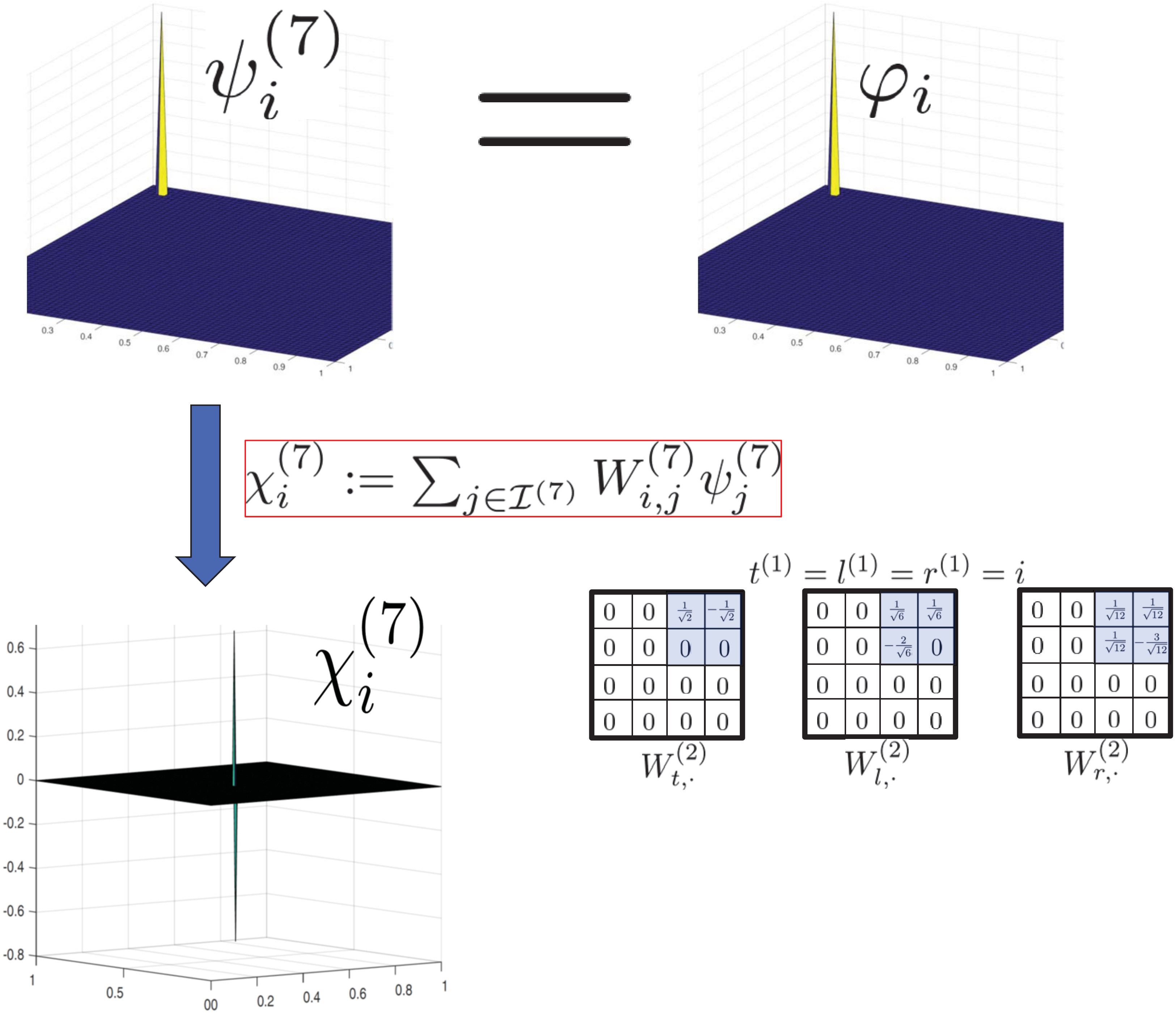
0	0	$\frac{1}{\sqrt{12}}$	$\frac{1}{\sqrt{12}}$
0	0	$\frac{1}{\sqrt{12}}$	$-\frac{3}{\sqrt{12}}$
0	0	0	0
0	0	0	0

$W^{(2)}_{r,\cdot}$

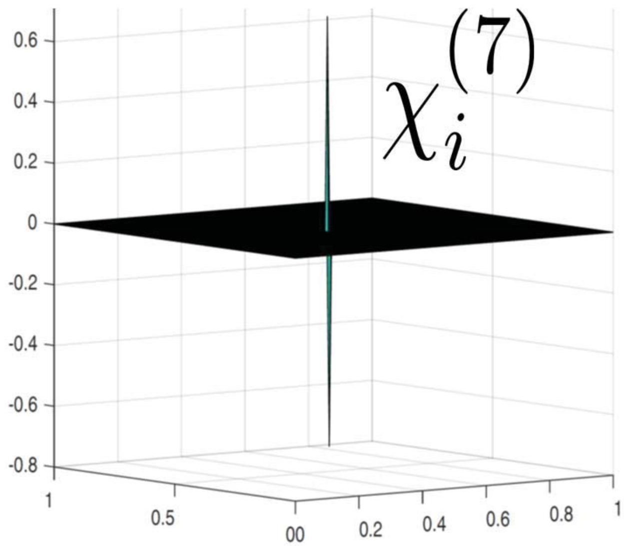
$$W^{(k)}:$$

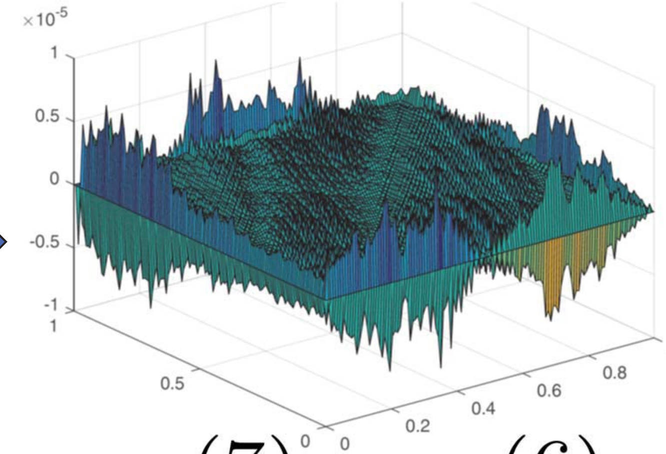
$$\text{Img}(W^{(k),T}) = \text{Ker}(\pi^{(k-1,k)})$$

$$W^{(k),T} W^{(k)} = J^{(k)}$$



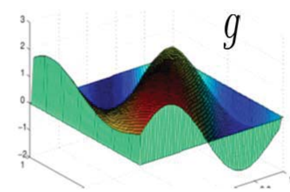
$\psi_i^{(7)}$
=
φ_i
$\chi_i^{(7)} := \sum_{j \in \mathcal{I}^{(7)}} W_{i,j}^{(7)} \psi_j^{(7)}$
$\chi_i^{(7)}$
$t^{(1)} = l^{(1)} = r^{(1)} = i$
$W_{t,\cdot}^{(2)}$
$W_{l,\cdot}^{(2)}$
$W_{r,\cdot}^{(2)}$





$$u^{(7)} - u^{(6)}$$

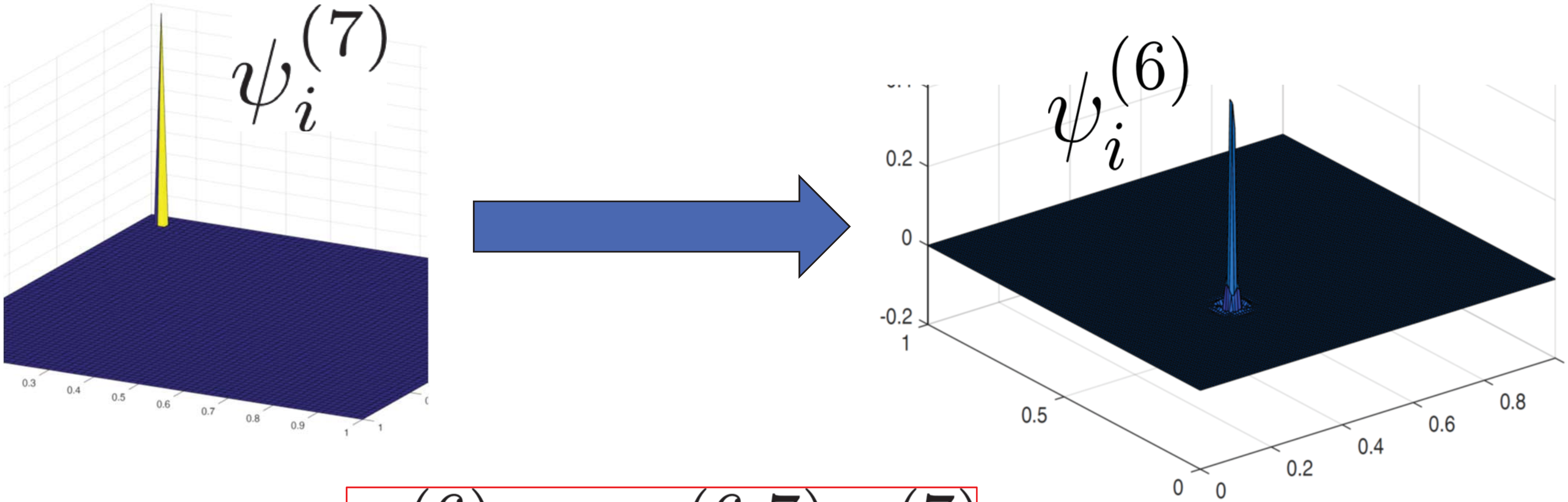
$$u^{(7)} - u^{(6)} = \sum_i w_i^{(7)} \chi_i^{(7)}$$



$$B^{(7)} w^{(7)} = W^{(7),T} f^{(7)}$$

$$f_i^{(7)} = \int_\Omega f \psi_i^{(7)}$$

$$B_{i,j}^{(7)} = \int_\Omega (\nabla \chi_i^{(7)})^T a \nabla \chi_j^{(7)} \qquad B^{(7)} = W^{(7)} A W^{(7),T}$$

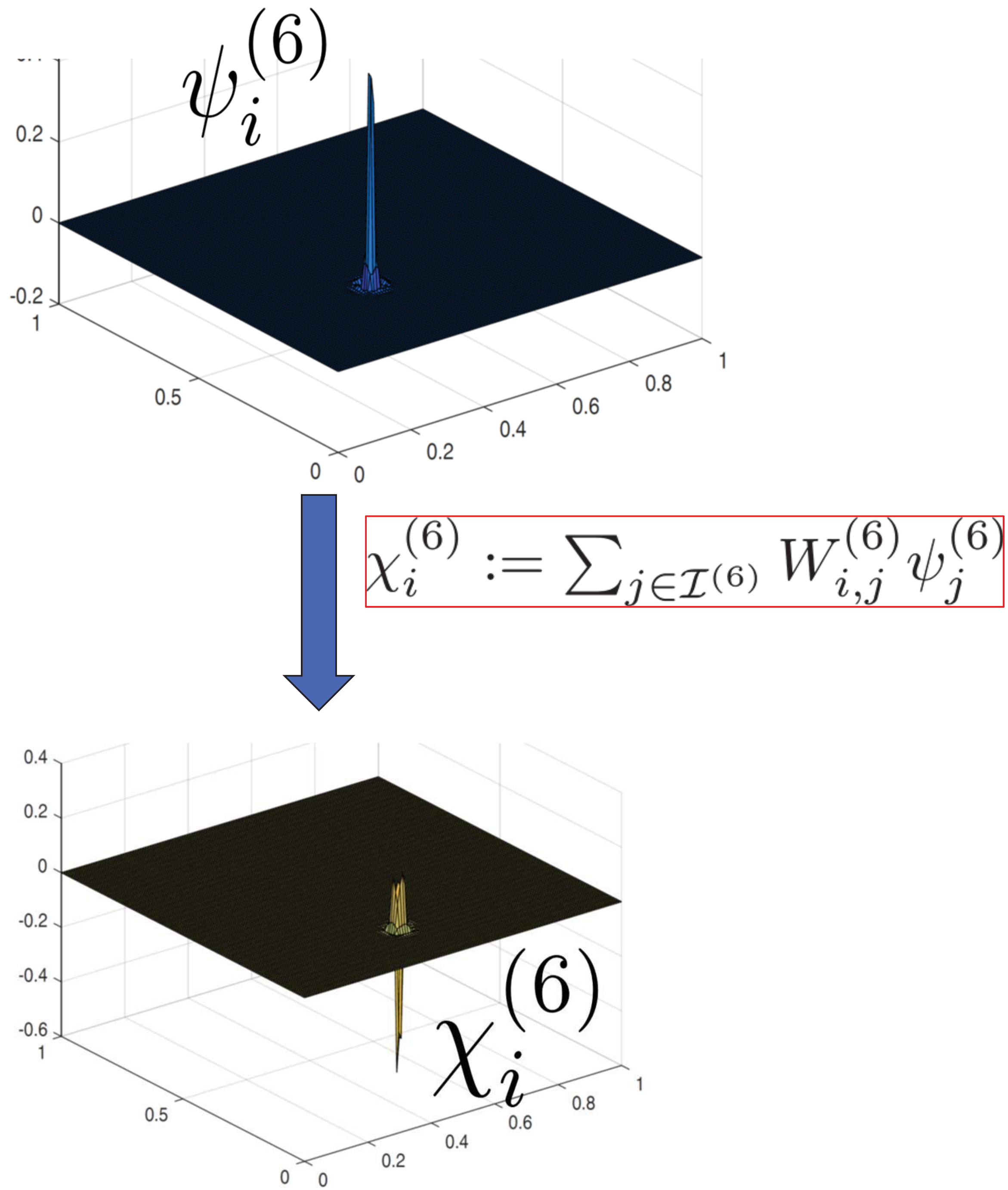


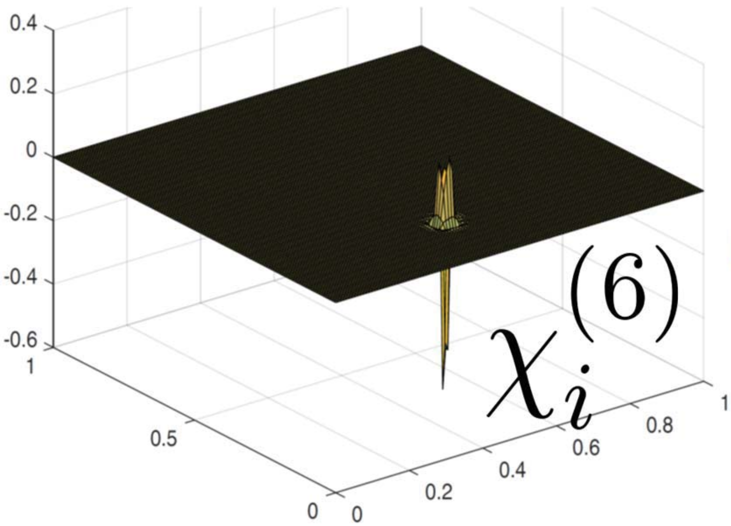
$$\psi_i^{(6)} = R_{i,j}^{(6,7)} \psi_j^{(7)}$$

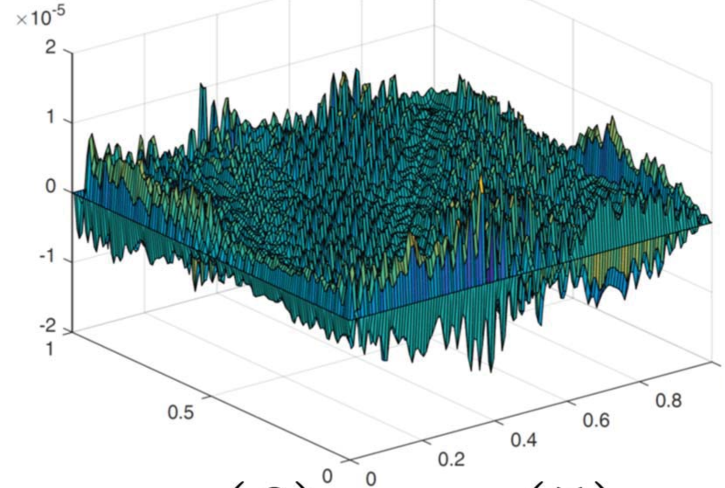
$$A^{(7)} = A$$

$$R^{(6,7)} = \pi^{(6,7)}(I^{(7)} - A^{(7)} W^{(7),T} B^{(7),-1} W^{(7)})$$

$$A^{(6)} = R^{(6,7)} A^{(7)} (R^{(6,7)})^T$$


$\psi_i^{(6)}$
0.2
0
-0.2
1
0.5
0
0
0.2
0.4
0.6
0.8
1
$\chi_i^{(6)} := \sum_{j \in \mathcal{I}^{(6)}} W_{i,j}^{(6)} \psi_j^{(6)}$
0.4
0.2
0
-0.2
-0.4
-0.6
1
0.5
0
0
0.2
0.4
0.6
0.8
1
$\chi_i^{(6)}$



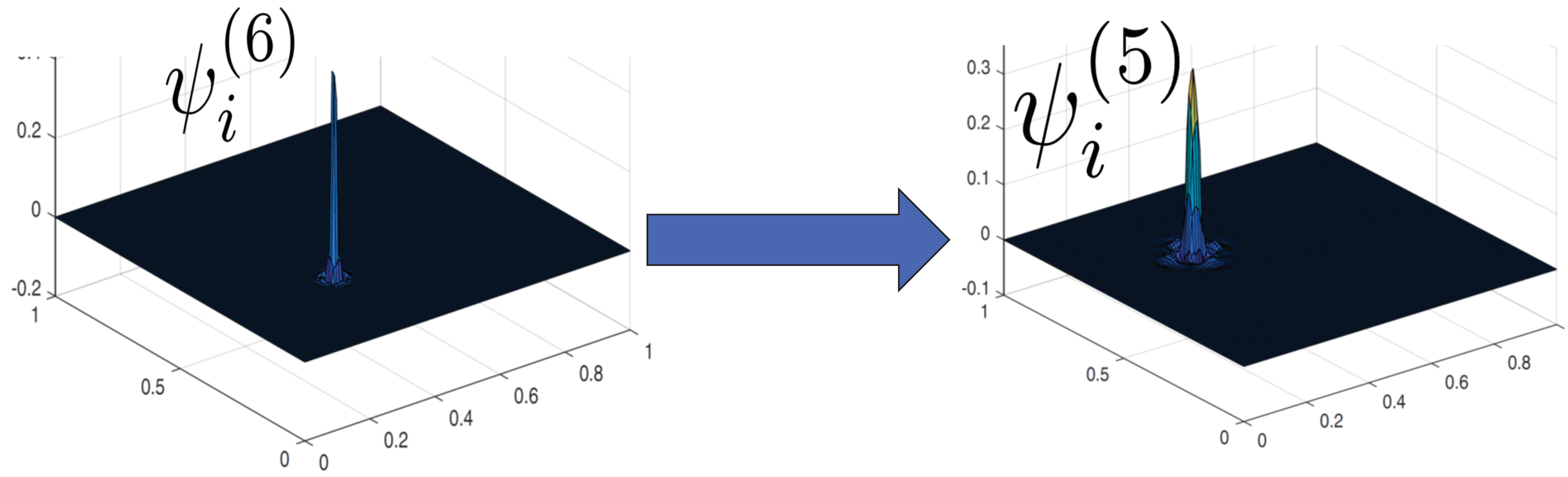


$u^{(6)} - u^{(5)}$

$$u^{(6)} - u^{(5)} = \sum_i w_i^{(6)} \chi_i^{(6)}$$

$$B^{(6)} w^{(6)} = W^{(6),T} f^{(6)} \qquad f^{(6)} = R^{(6,7)} f^{(7)}$$

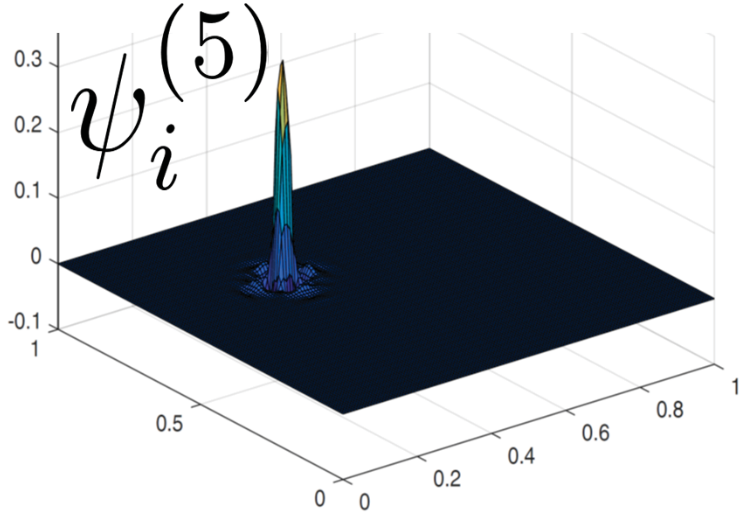
$$B^{(6)} = W^{(6)} A^{(6)} W^{(6),T}$$



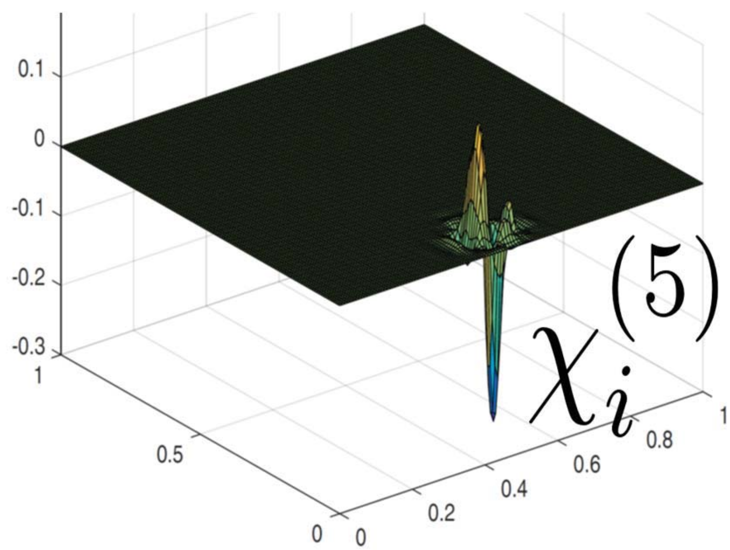
$$\psi_i^{(5)} = \sum_j R_{i,j}^{(5,6)} \psi_j^{(6)}$$

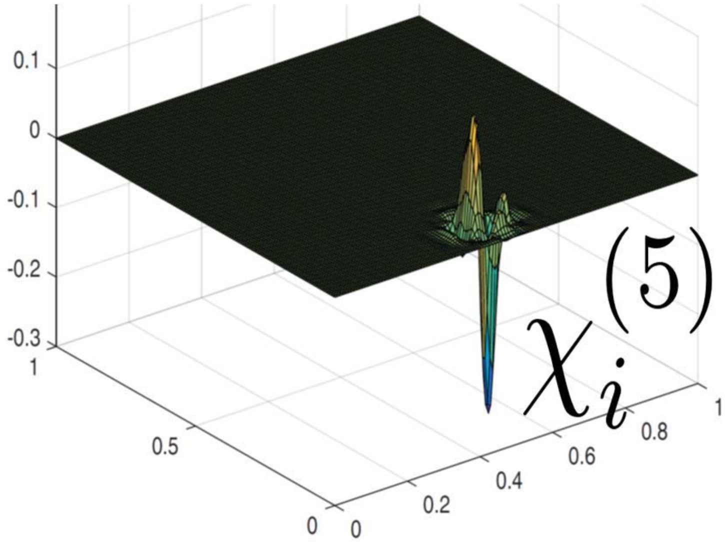
$$R^{(5,6)} = \pi^{(5,6)}(I^{(6)} - A^{(6)} W^{(6),T} B^{(6),-1} W^{(6)})$$

$$A^{(5)} = R^{(5,6)} A^{(6)} (R^{(5,6)})^T$$

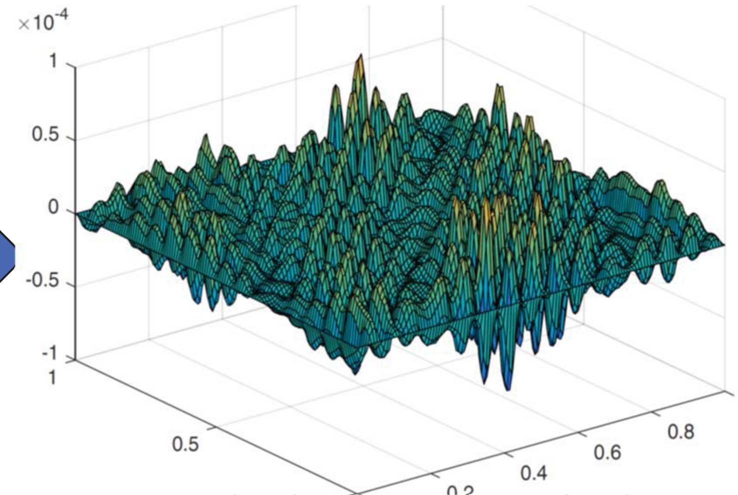


$$\chi_i^{(5)} := \sum_j W_{i,j}^{(5)} \psi_j^{(5)}$$





$\chi_i^{(5)}$



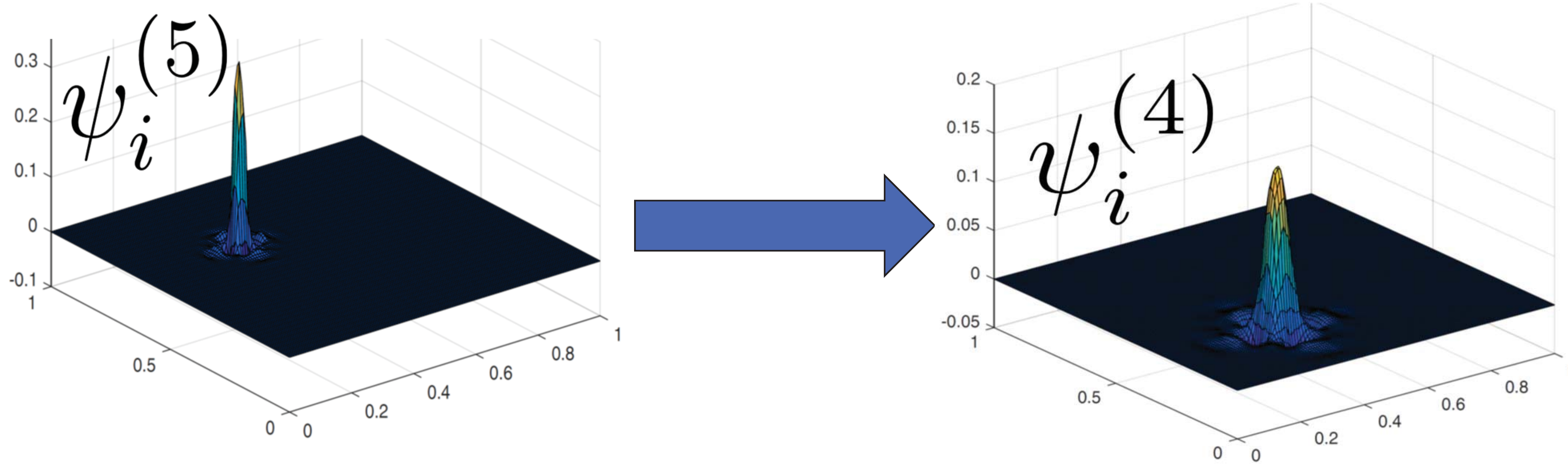
$u^{(5)} - u^{(4)}$

$$u^{(5)} - u^{(4)} = \sum_i w_i^{(5)} \chi_i^{(5)}$$

$$B^{(5)} w^{(5)} = W^{(5),T} f^{(5)}$$

$$f^{(5)} = R^{(5,6)} f^{(6)}$$

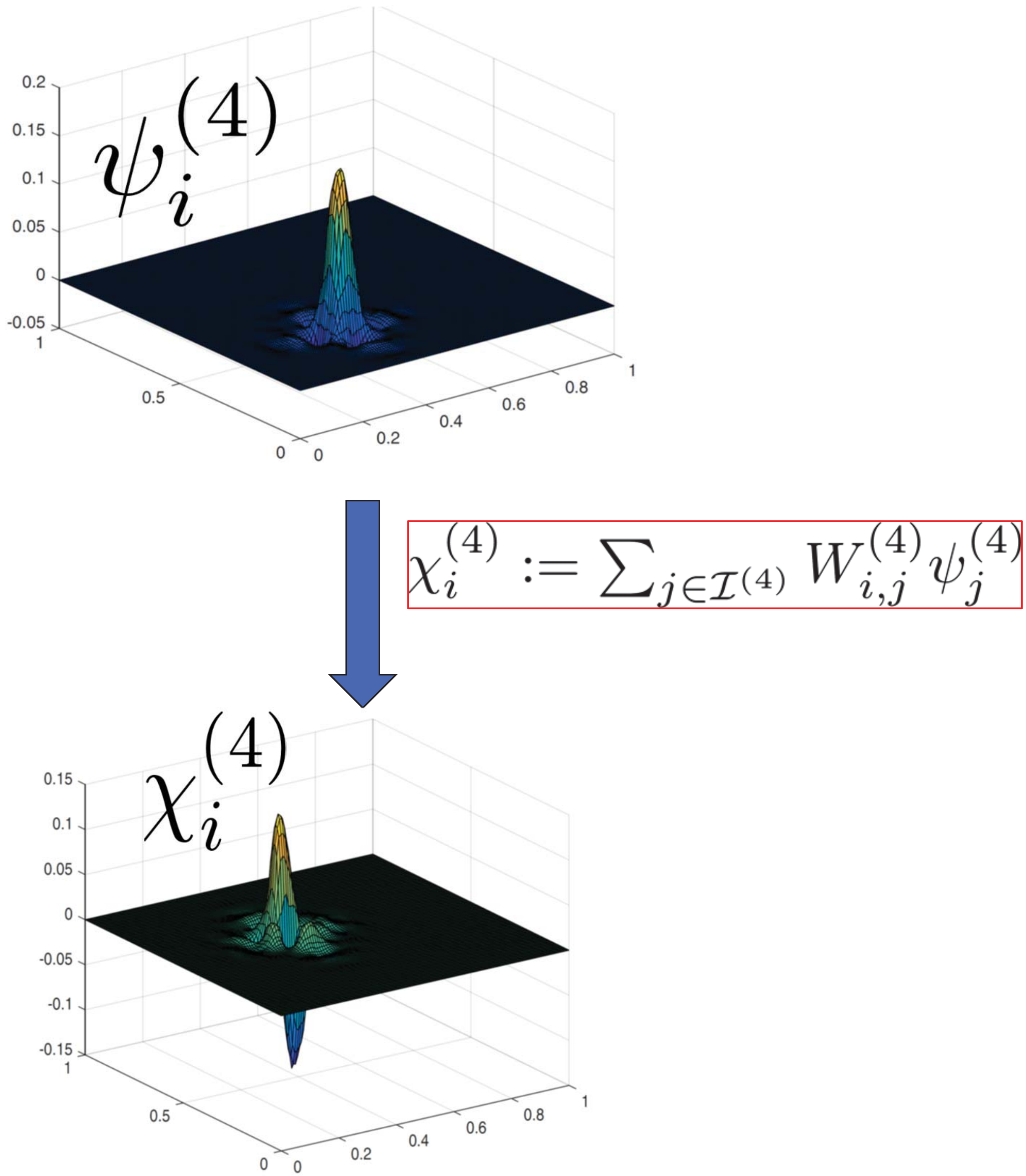
$$B^{(5)} = W^{(5)} A^{(5)} W^{(5),T}$$



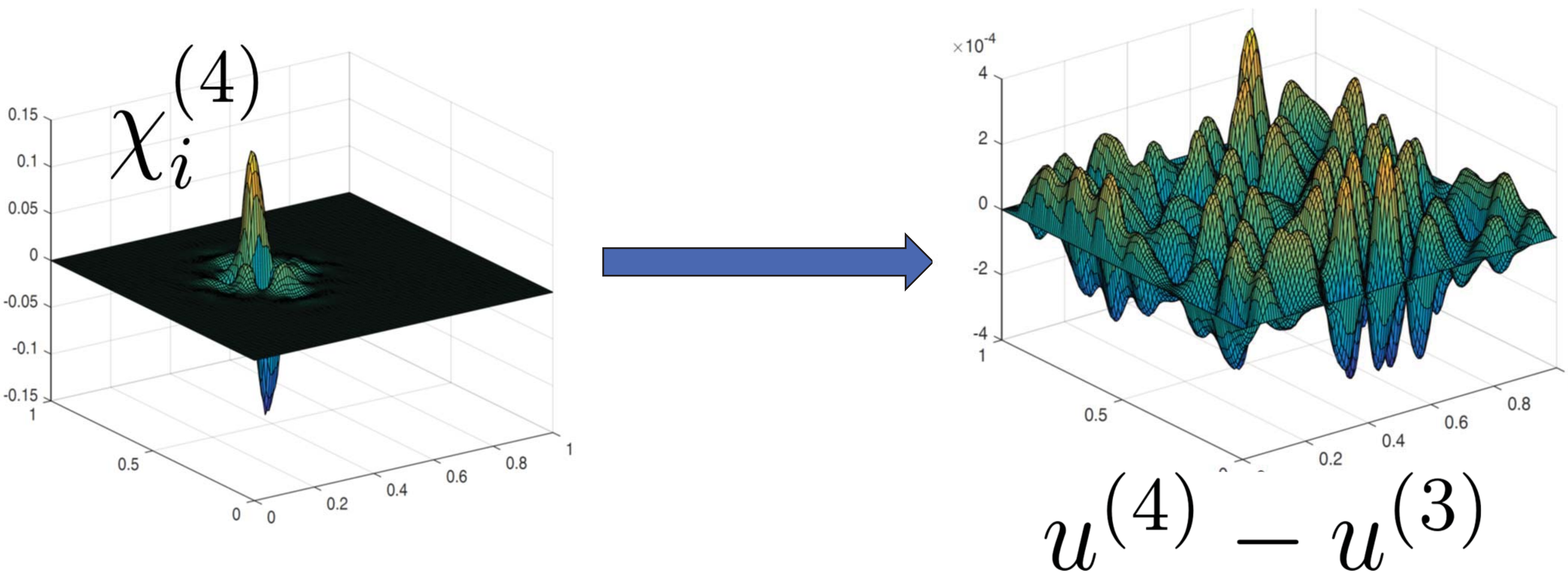
$$\psi_i^{(4)} = R_{i,j}^{(4,5)} \psi_j^{(5)}$$

$$R^{(4,5)} = \pi^{(4,5)} (I^{(5)} - A^{(5)} W^{(5),T} B^{(5),-1} W^{(5)})$$

$$A^{(4)} = R^{(4,5)} A^{(5)} (R^{(4,5)})^T$$



$\psi_i^{(4)}$
$\chi_i^{(4)} := \sum_{j \in \mathcal{I}^{(4)}} W_{i,j}^{(4)} \psi_j^{(4)}$
$\chi_i^{(4)}$

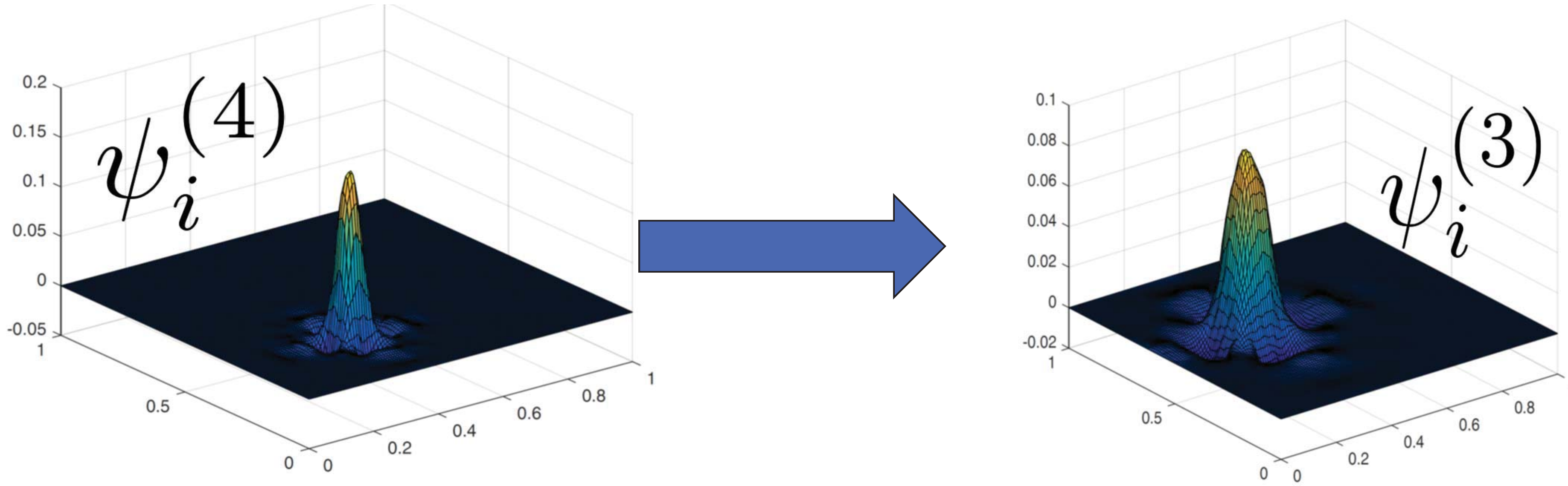


$$u^{(4)} - u^{(3)} = \sum_i w_i^{(4)} \chi_i^{(4)}$$

$$B^{(4)} w^{(4)} = W^{(4),T} f^{(4)}$$

$$f^{(4)} = R^{(4,5)} f^{(5)}$$

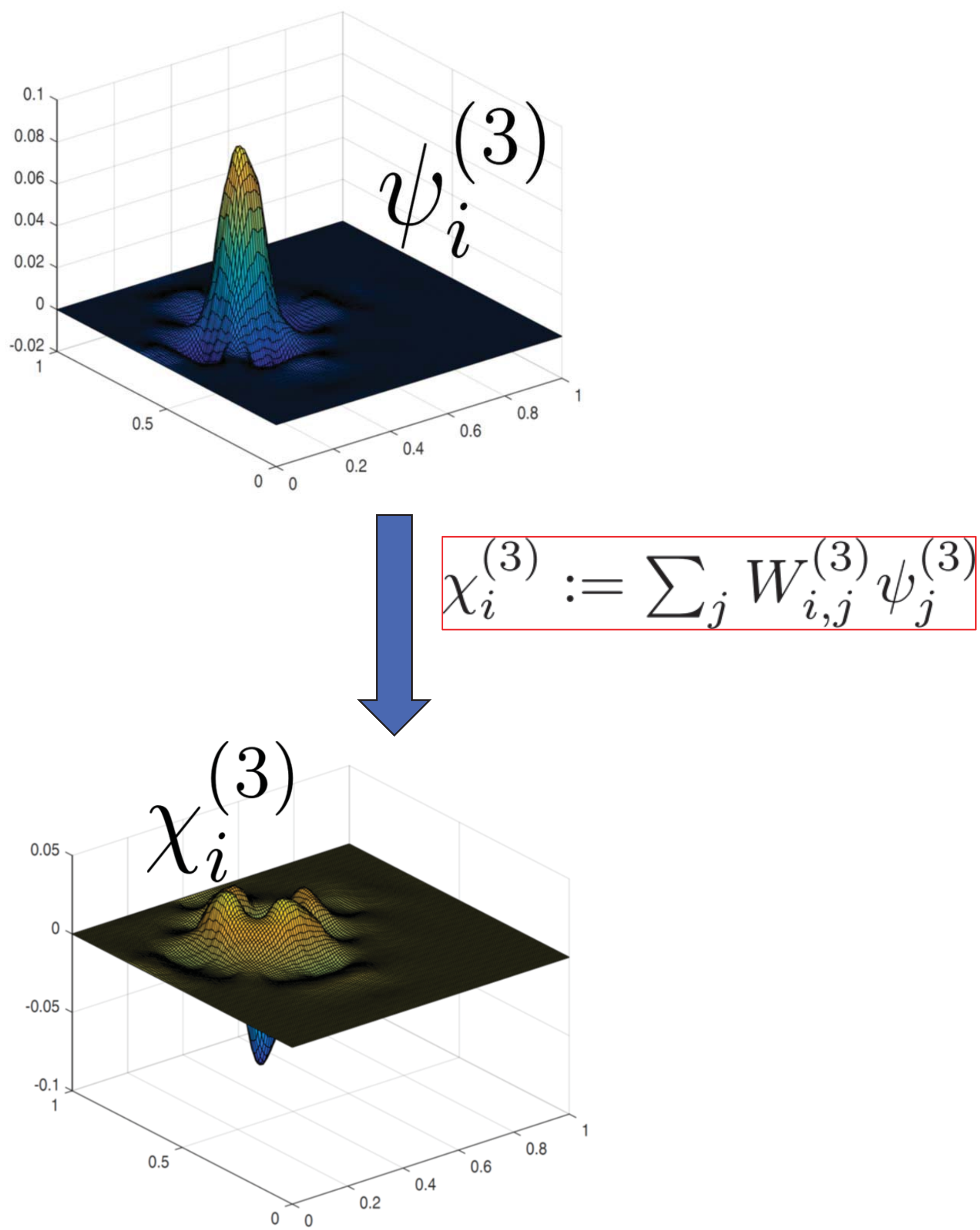
$$B^{(4)} = W^{(4)} A^{(4)} W^{(4),T}$$

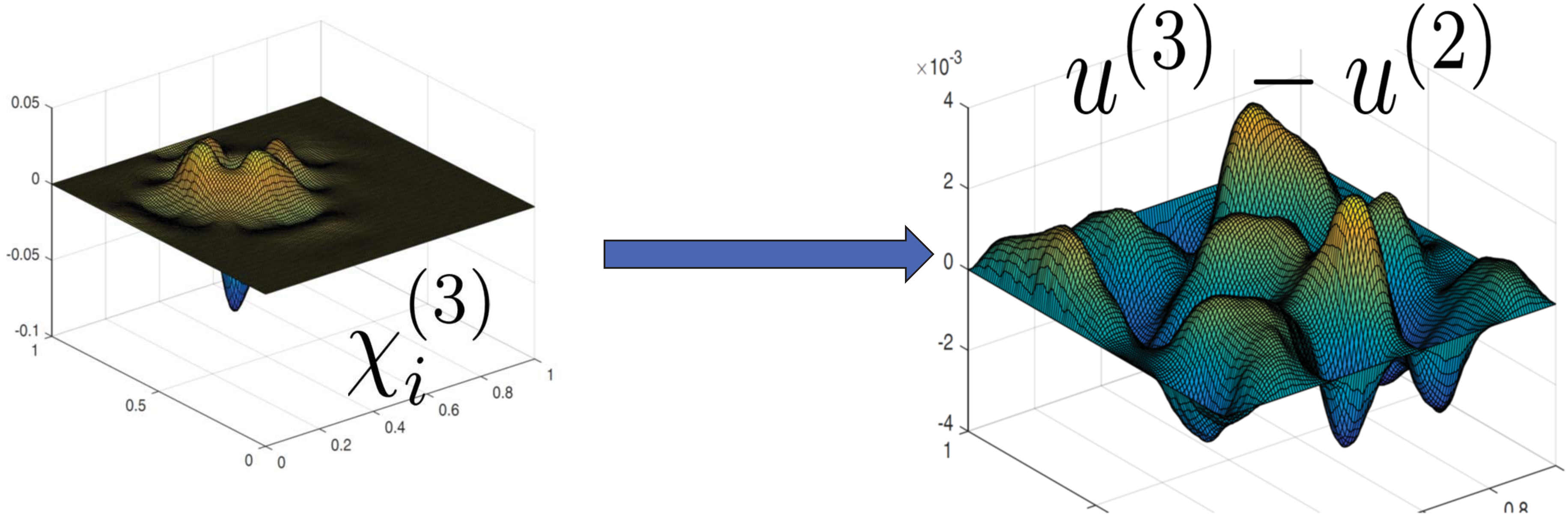


$$\psi_i^{(3)} = \sum_j R_{i,j}^{(3,4)} \psi_j^{(4)}$$

$$R^{(3,4)} = \pi^{(3,4)}(I^{(4)} - A^{(4)} W^{(4),T} B^{(4),-1} W^{(4)})$$

$$A^{(3)} = R^{(3,4)} A^{(4)} (R^{(3,4)})^T$$

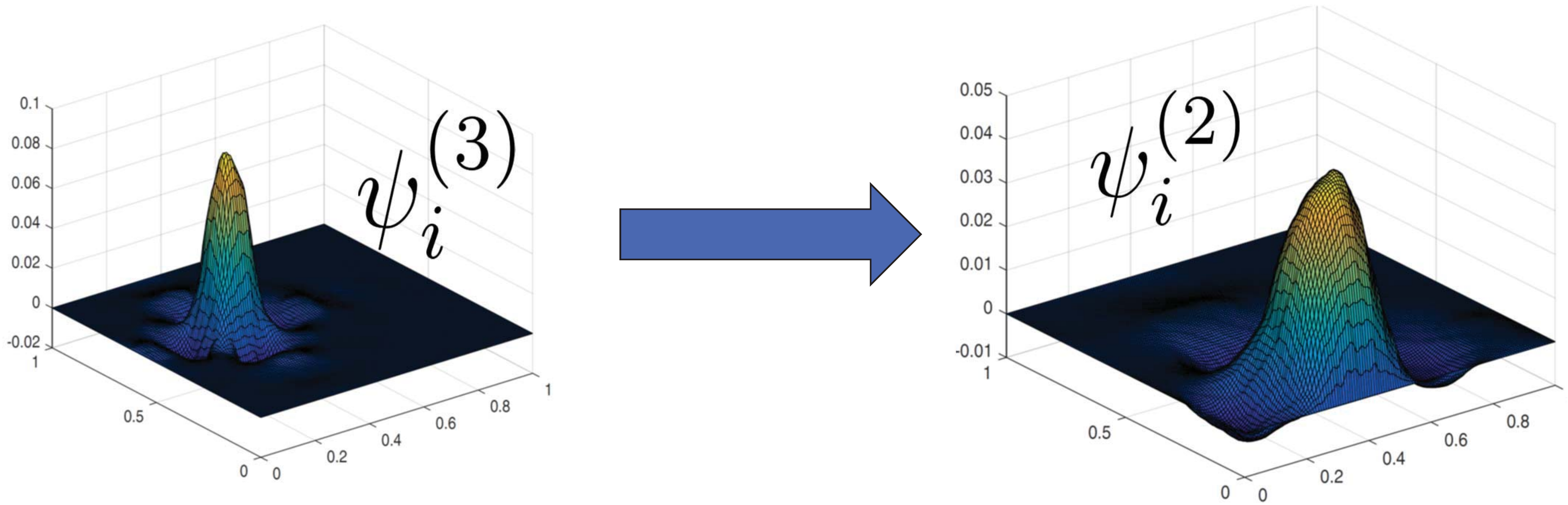

$\psi_i^{(3)}$
0.1
0.08
0.06
0.04
0.02
0
-0.02
1
0.5
0
0
0.2
0.4
0.6
0.8
1
$\chi_i^{(3)} := \sum_j W_{i,j}^{(3)} \psi_j^{(3)}$
$\chi_i^{(3)}$
0.05
0
-0.05
-0.1
1
0.5
0
0
0.2
0.4
0.6
0.8
1



$$u^{(3)} - u^{(2)} = \sum_i w_i^{(3)} \chi_i^{(3)}$$

$$B^{(3)} w^{(3)} = W^{(3),T} f^{(3)} \qquad f^{(3)} = R^{(3,4)} f^{(4)}$$

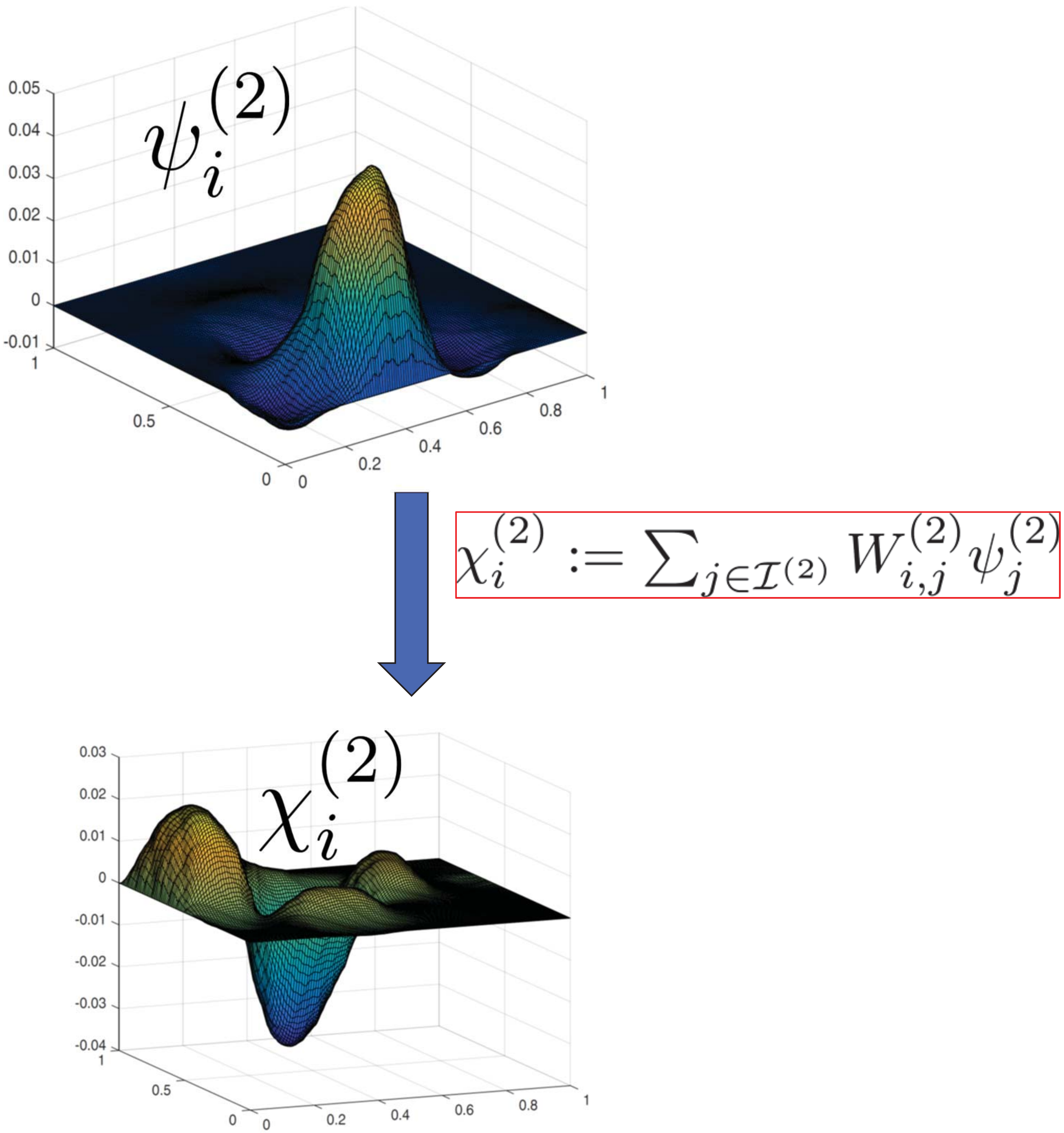
$$B^{(3)} = W^{(3)} A^{(3)} W^{(3),T}$$

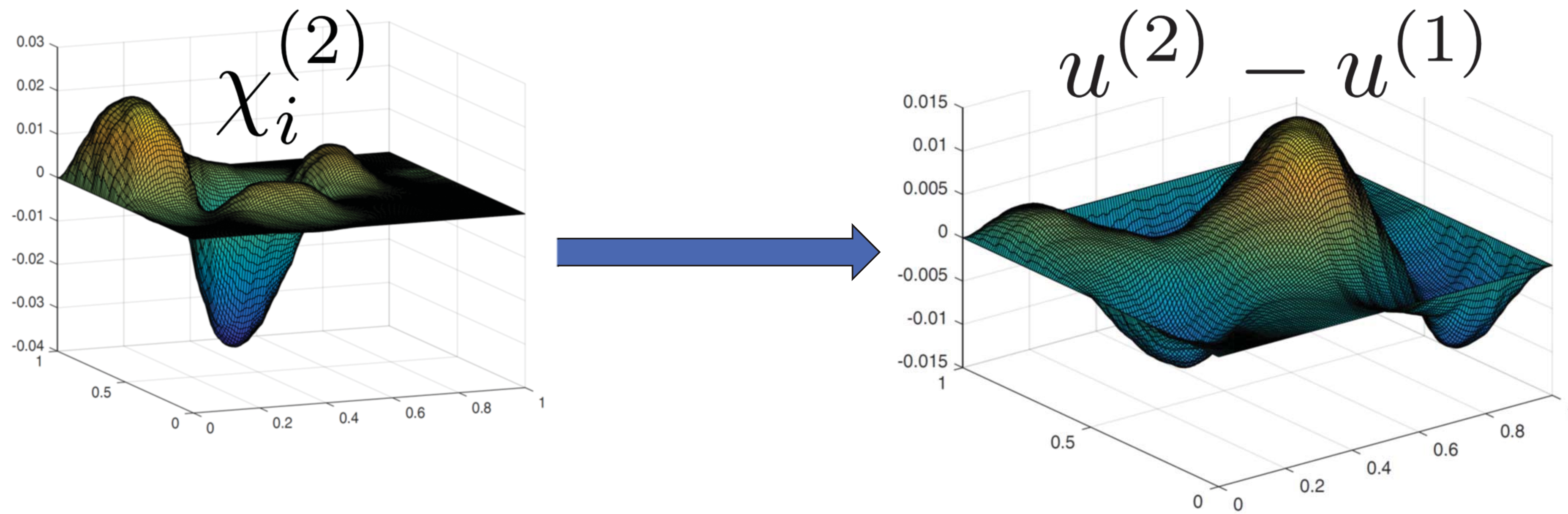


$$\psi_i^{(2)} = \sum_j R_{i,j}^{(2,3)} \psi_j^{(3)}$$

$$R^{(2,3)} = \pi^{(2,3)}(I^{(3)} - A^{(3)} W^{(3),T} B^{(3),-1} W^{(3)})$$

$$A^{(2)} = R^{(2,3)} A^{(3)} (R^{(2,3)})^T$$


$\psi_i^{(2)}$
0.05
0.04
0.03
0.02
0.01
0
-0.01
1
0.5
0
0
0.2
0.4
0.6
0.8
1
$\chi_i^{(2)} := \sum_{j \in \mathcal{I}^{(2)}} W_{i,j}^{(2)} \psi_j^{(2)}$
$\chi_i^{(2)}$
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
1
0.5
0
0
0.2
0.4
0.6
0.8
1

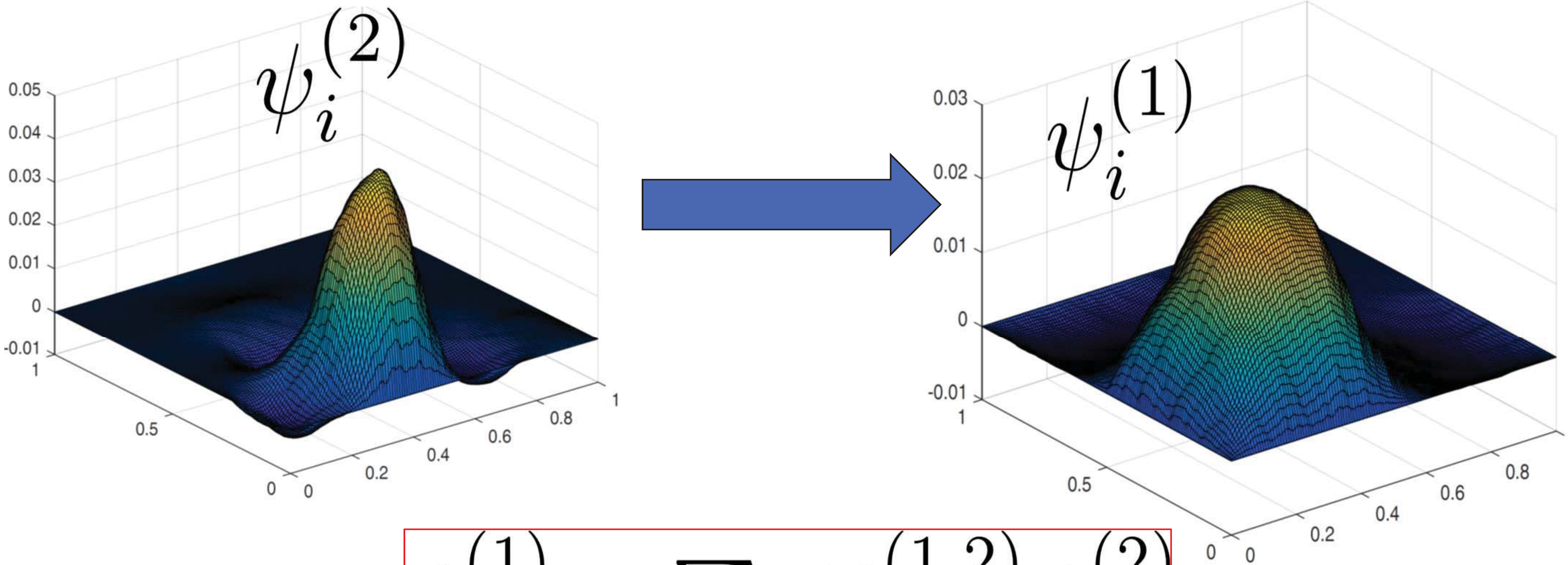


$$u^{(2)} - u^{(1)} = \sum_i w_i^{(2)} \chi_i^{(2)}$$

$$B^{(2)} w^{(2)} = W^{(2),T} f^{(2)}$$

$$f^{(2)} = R^{(2,3)} f^{(3)}$$

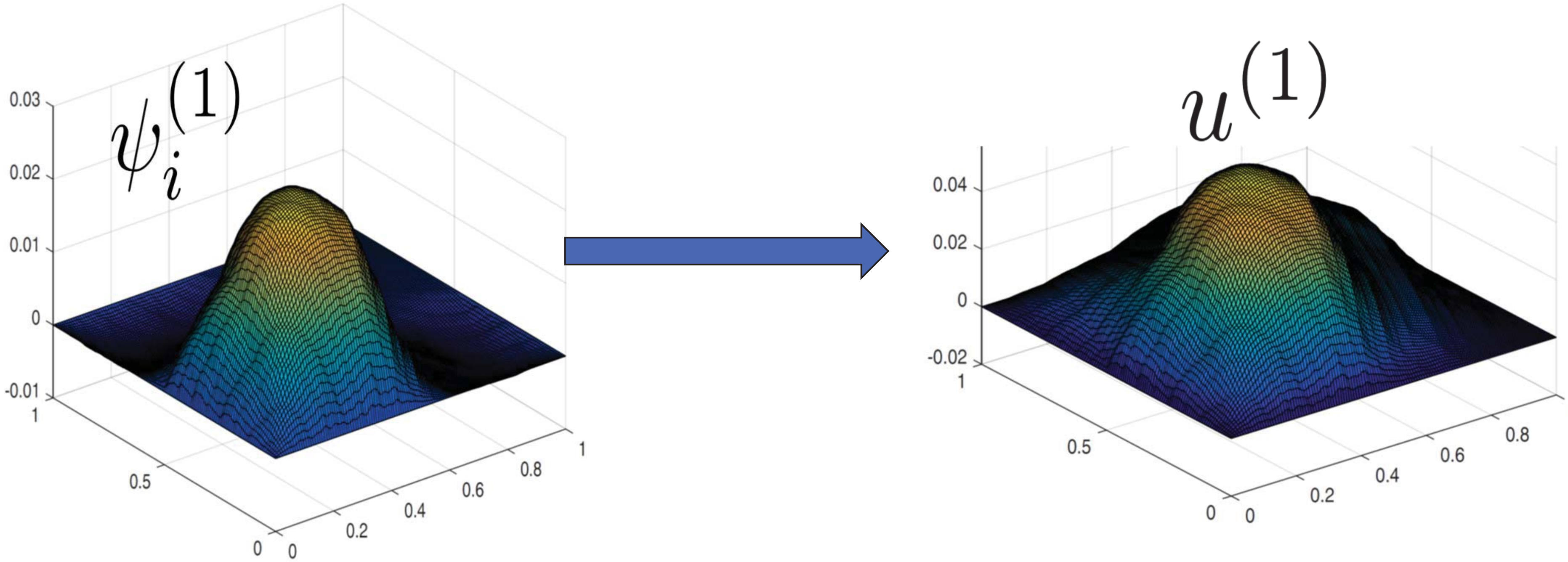
$$B^{(2)} = W^{(2)} A^{(2)} W^{(2),T}$$



$$\psi_i^{(1)} = \sum_j R_{i,j}^{(1,2)} \psi_j^{(2)}$$

$$R^{(1,2)} = \pi^{(1,2)} \left(I^{(2)} - A^{(2)} W^{(2),T} B^{(2),-1} W^{(2)} \right)$$

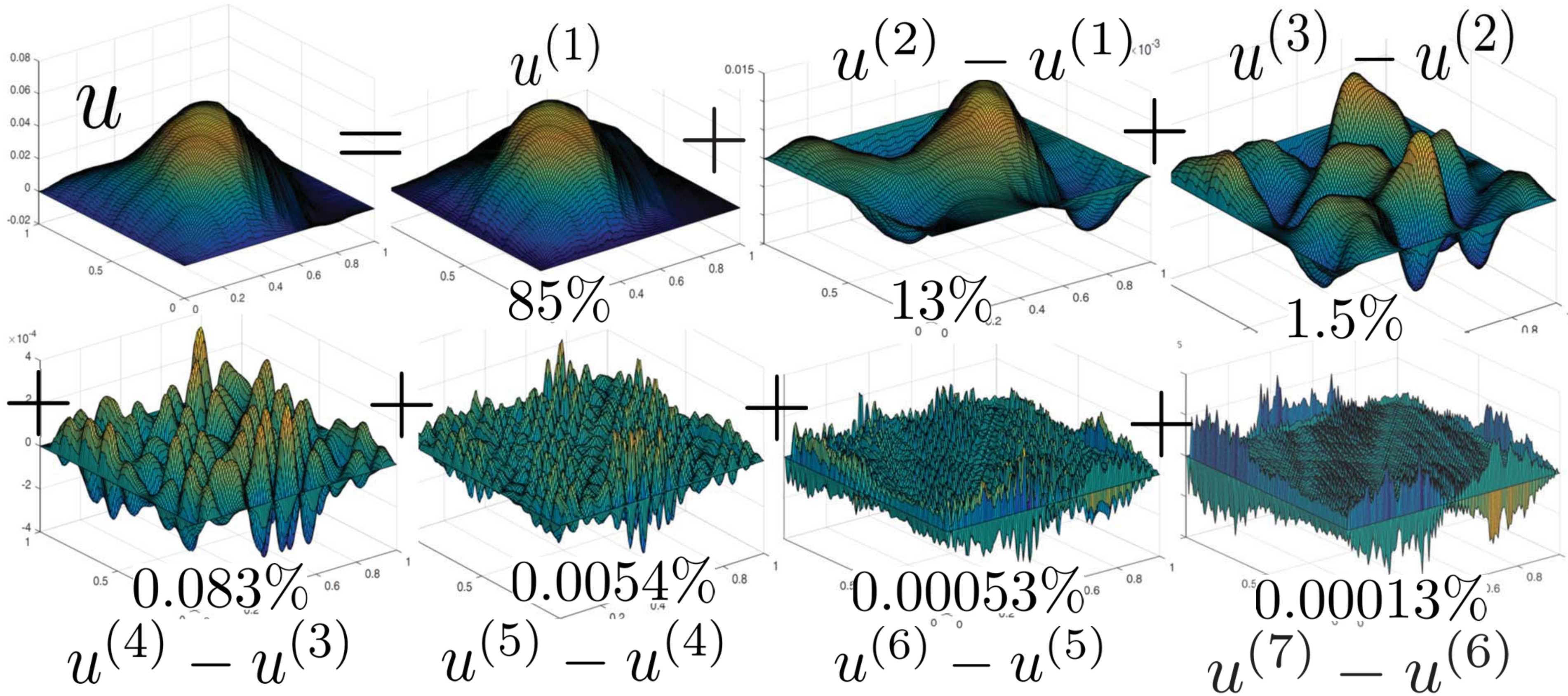
$$A^{(1)} = R^{(1,2)} A^{(2)} \left(R^{(1,2)} \right)^T$$



$$u^{(1)} = \sum_i v_i^{(1)} \psi_i^{(1)}$$

$$A^{(1)} v^{(1)} = f^{(1)}$$

$$f^{(1)} = R^{(1,2)} f^{(2)}$$



$$
\begin{cases} -\operatorname{div}(a\nabla u) = f, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}
$$

Dense kernel matrices

- Schäfer, Sullivan, O. 2017.
- arXiv:1706.02205

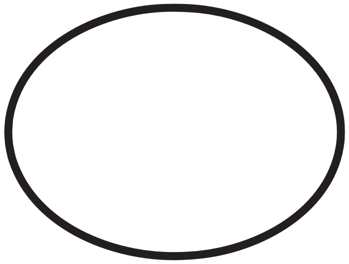
Florian Schäfer

Tim Sullivan

Software

https://github.com/f-t-s/nearLinKernel.git

Covariance function/kernel



$\Omega \subset \mathbb{R}^d$: Bounded domain

$\mathcal{L}$: arbitrary continuous positive symmetric linear bijection

$$(H_0^s(\Omega), \|\cdot\|_{H_0^s(\Omega)}) \xrightarrow{\quad \mathcal{L} \quad} (H^{-s}(\Omega), \|\cdot\|_{H^{-s}(\Omega)})$$

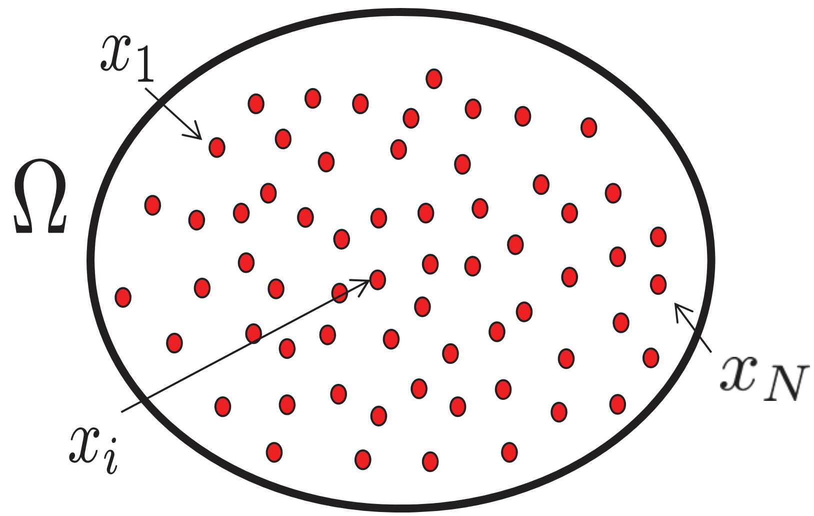
$\mathcal{L}$: is local $\quad \int_\Omega u\mathcal{L}v = 0$ if u and v have disjoint supports

Covariance function = its Green functions

$$G = \mathcal{L}^{-1}$$

The kernel/covariance matrix

$s > \frac{d}{2}$ ➡ $G(x, y)$ is continuous



$x_1, \ldots, x_N$: Approximately homogeneous

Θ: $N \times N$ symmetric positive definite matrix

$$\Theta_{i,j} := G(x_i, x_j)$$

The kernel/covariance matrix

Important in

- Computational Physics
- (Gaussian process) statistics
- Kernel methods for machine learning (e.g. Support Vector Machines)

Computational bottleneck

Θ is dense, naively we have

- Storage, $\mathcal{O}(N^2)$
- Θv, $\mathcal{O}(N^2)$
- $\Theta^{-1} v$, $\mathcal{O}(N^3)$
- $\det(\Theta)$, $\mathcal{O}(N^3)$
- $\mathrm{PCA}(\Theta)$, $\mathcal{O}(N^3)$

Algorithm

For $\epsilon > 0$ knowing only Ω and $\{x_i\}_{1 \le i \le N}$, we will

- Select $\mathcal{O}(N \operatorname{polylog}(N) \operatorname{polylog}(\frac{1}{\epsilon}))$ entries of Θ and an ordering P of $\{x_i\}_{1 \le i \le N}$.
- From these entries compute a lower triangular matrix L such that $nnz(L) = \mathcal{O}(N \operatorname{polylog}(N) \operatorname{polylog}(\frac{1}{\epsilon}))$.

Theorem

The above can be done in complexity $N \operatorname{polylog}(N) \operatorname{polylog}(\frac{1}{\epsilon})$, in time and space, such that

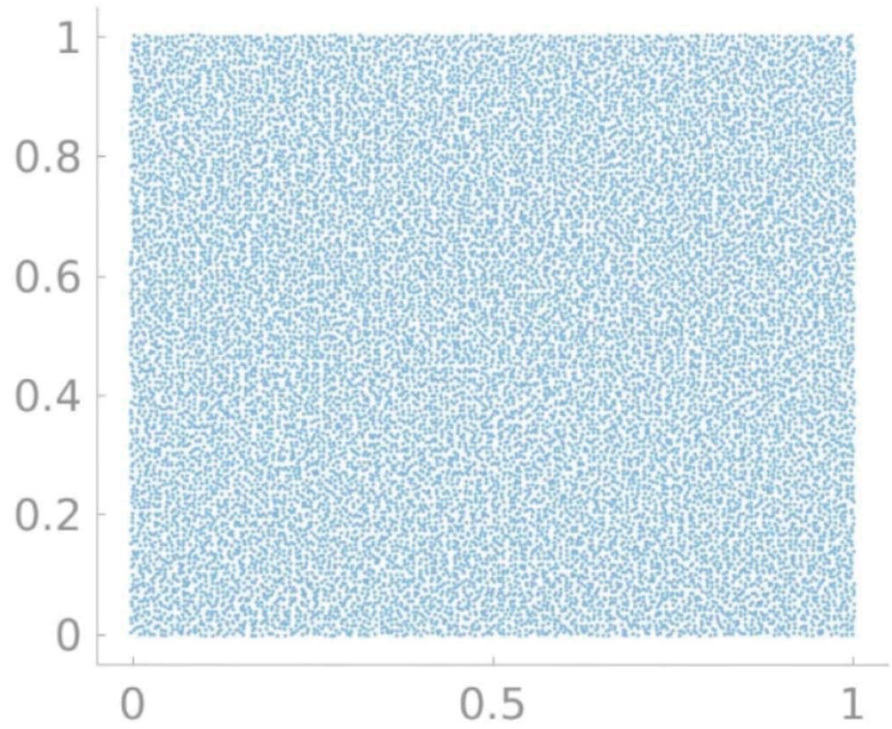
$$\|\Theta - PLL^T P^T\| \le \epsilon$$

Allows to approximate Θv, $\Theta^{-1} v$, $\det(\Theta)$, in $\mathcal{O}(N \operatorname{polylogN} \operatorname{polylog} \frac{1}{\epsilon})$ complexity

A simple algorithm

- Decompose $\{x_i\}_{i \in \mathcal{I}}$ into a nested hierarchy:

$$\{x_i\}_{i \in \mathcal{I}^{(1)}} \subset \{x_i\}_{i \in \mathcal{I}^{(2)}} \subset \{x_i\}_{i \in \mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i \in \mathcal{I}^{(q)}}$$



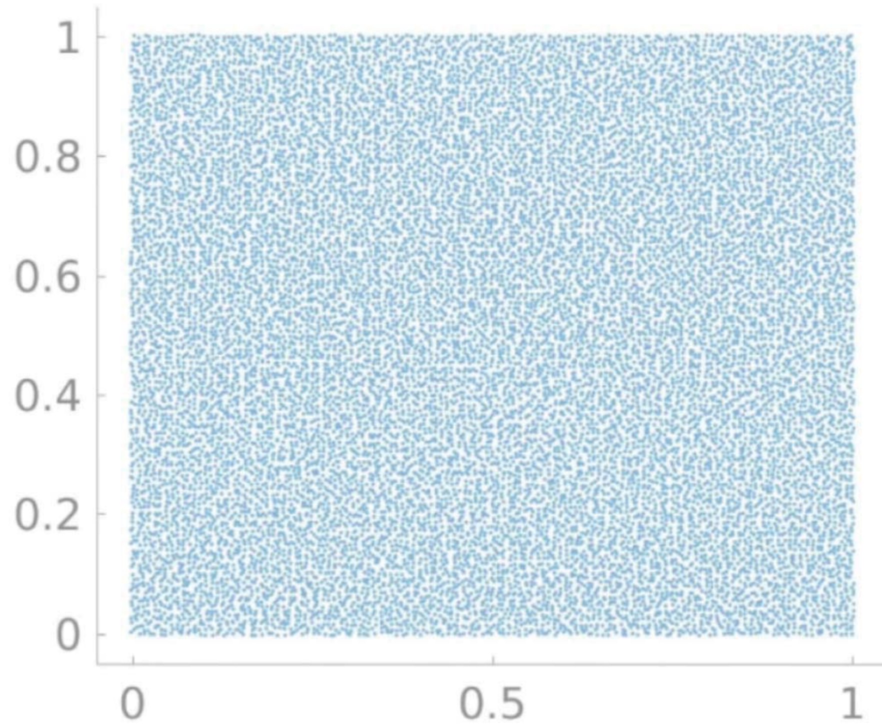
A simple algorithm

- Decompose $\{x_i\}_{i\in\mathcal{I}}$ into a nested hierarchy:

$$\{x_i\}_{i\in\mathcal{I}^{(1)}} \subset \{x_i\}_{i\in\mathcal{I}^{(2)}} \subset \{x_i\}_{i\in\mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i\in\mathcal{I}^{(q)}}$$

- Define

$$\mathcal{J}^{(k)} := \mathcal{I}^{(k)} / I^{(k-1)}$$



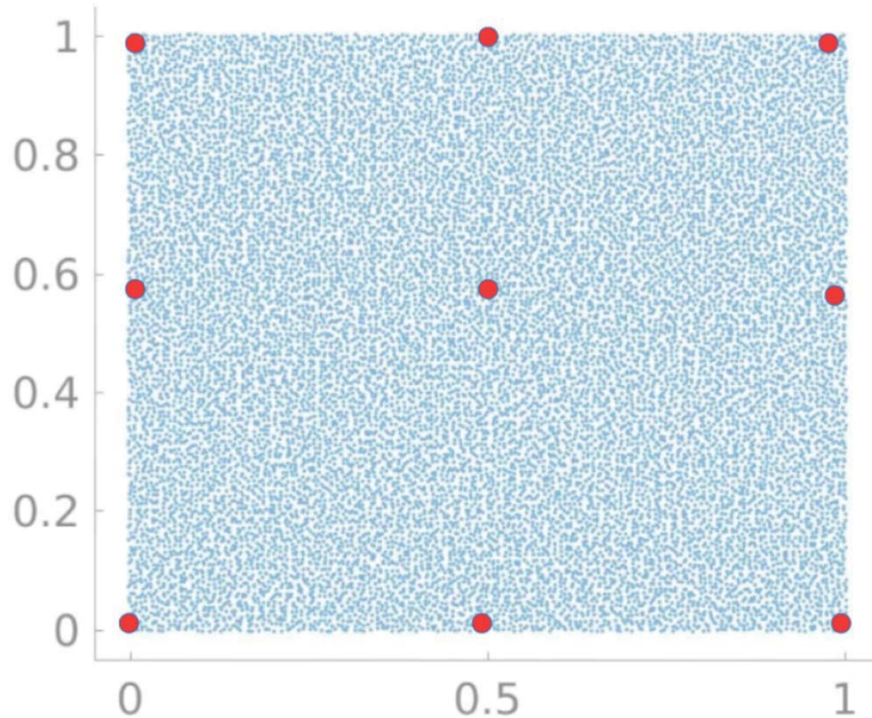
A simple algorithm

- Decompose $\{x_i\}_{i\in\mathcal{I}}$ into a nested hierarchy:

$\{x_i\}_{i\in\mathcal{I}^{(1)}} \subset \{x_i\}_{i\in\mathcal{I}^{(2)}} \subset \{x_i\}_{i\in\mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i\in\mathcal{I}^{(q)}}$

- Define

$$\mathcal{I}^{(1)} = \mathcal{J}^{(1)}$$



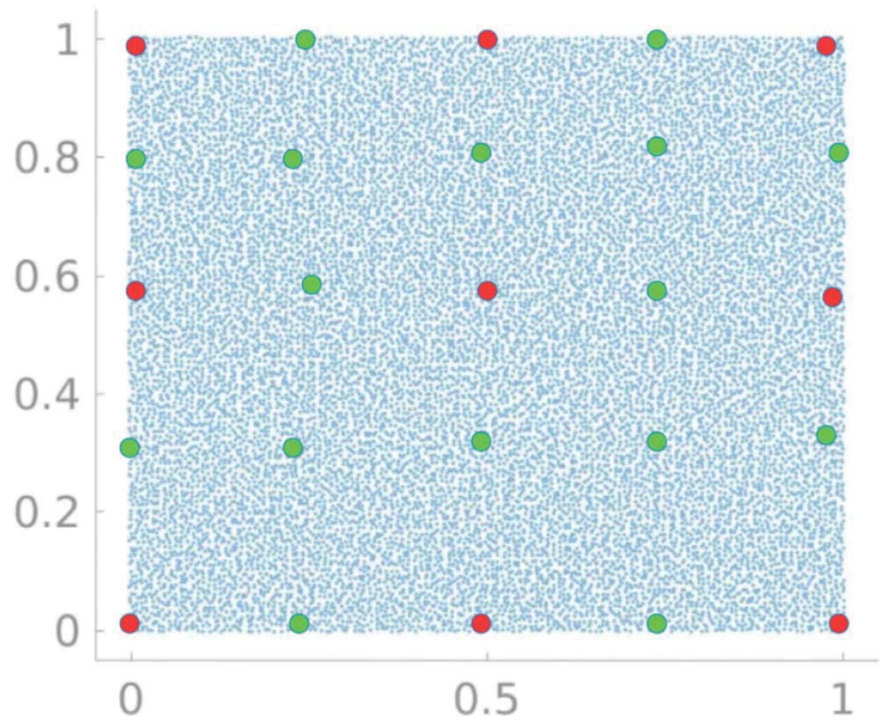
A simple algorithm

- Decompose $\{x_i\}_{i \in \mathcal{I}}$ into a nested hierarchy:

$$\{x_i\}_{i \in \mathcal{I}^{(1)}} \subset \{x_i\}_{i \in \mathcal{I}^{(2)}} \subset \{x_i\}_{i \in \mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i \in \mathcal{I}^{(q)}}$$

- Define

$$\mathcal{I}^{(2)} = \mathcal{J}^{(1)} \cup \mathcal{J}^{(2)}$$



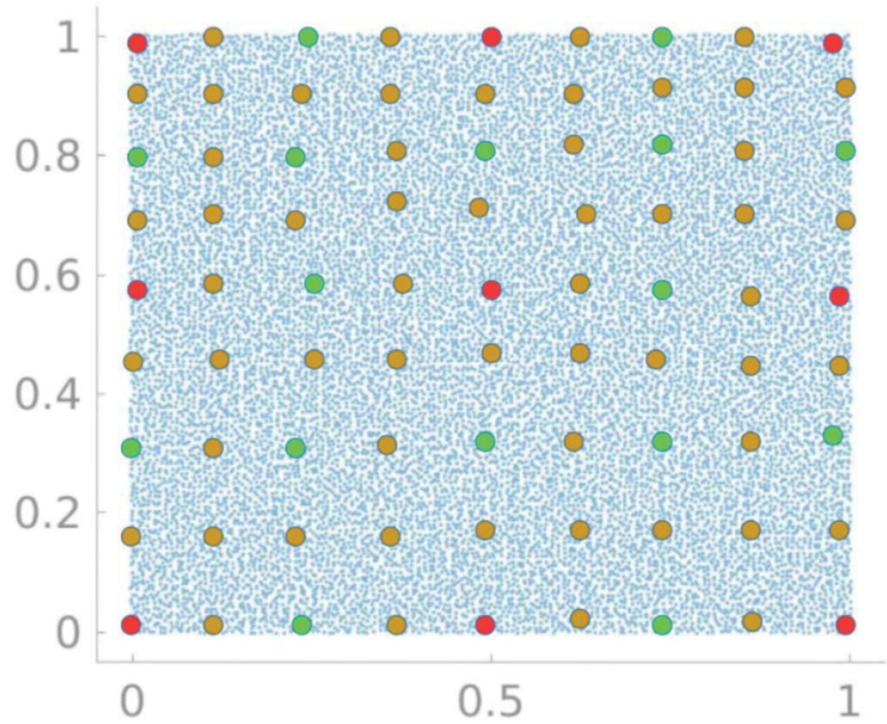
A simple algorithm

- Decompose $\{x_i\}_{i\in\mathcal{I}}$ into a nested hierarchy:

$$\{x_i\}_{i\in\mathcal{I}^{(1)}} \subset \{x_i\}_{i\in\mathcal{I}^{(2)}} \subset \{x_i\}_{i\in\mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i\in\mathcal{I}^{(q)}}$$

- Define

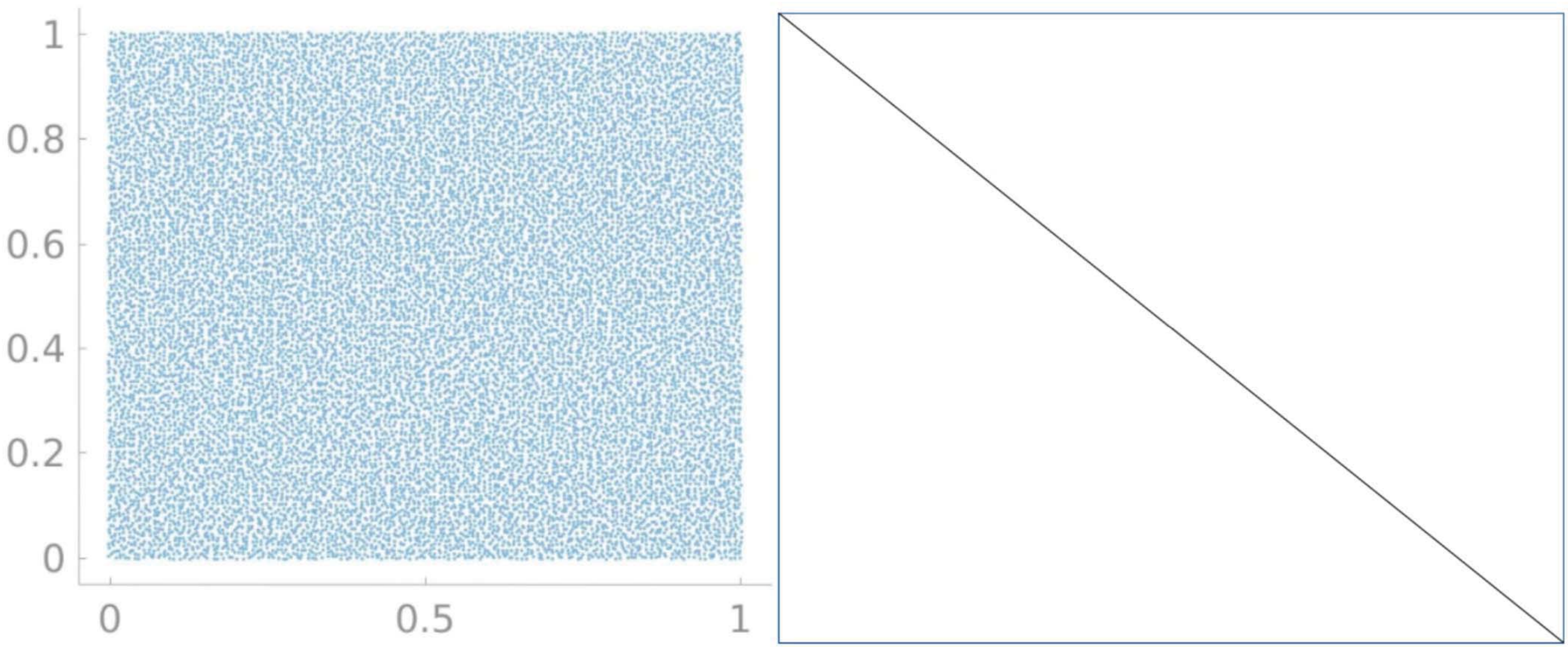
$$\mathcal{I}^{(3)} = \mathcal{J}^{(1)} \cup \mathcal{J}^{(2)} \cup \mathcal{J}^{(3)}$$



A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

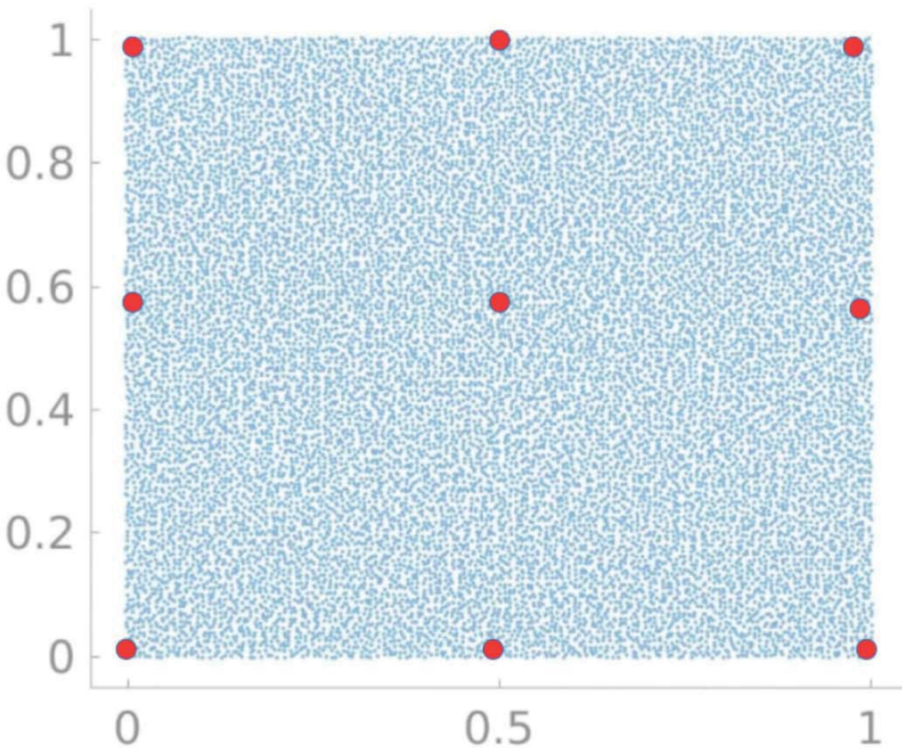
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\ j \in \mathcal{J}^{(l)},\ \mathrm{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

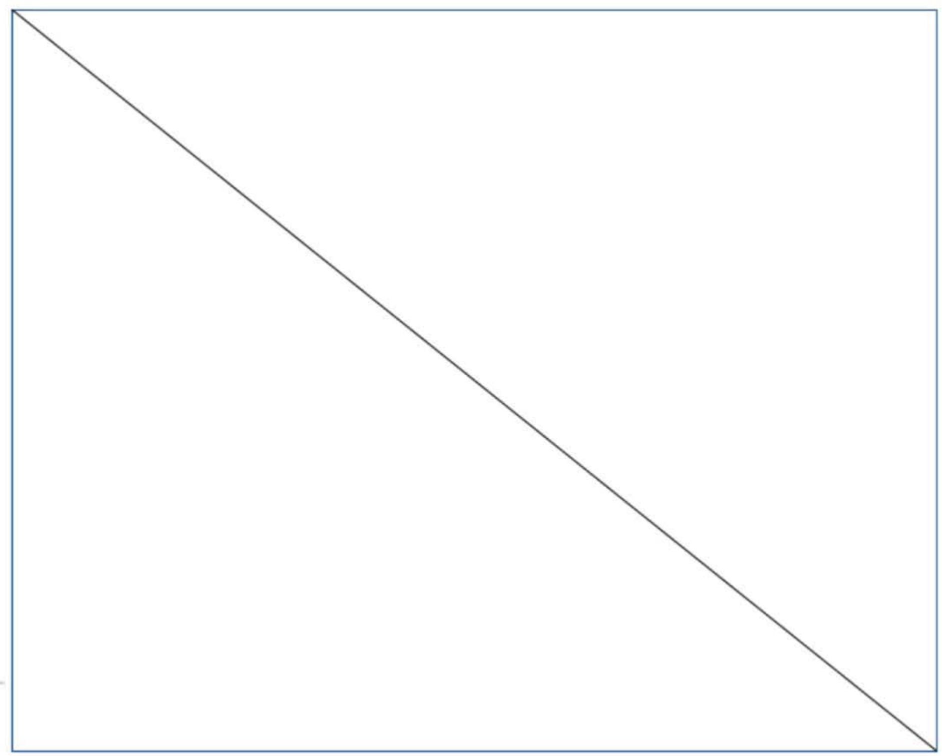


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

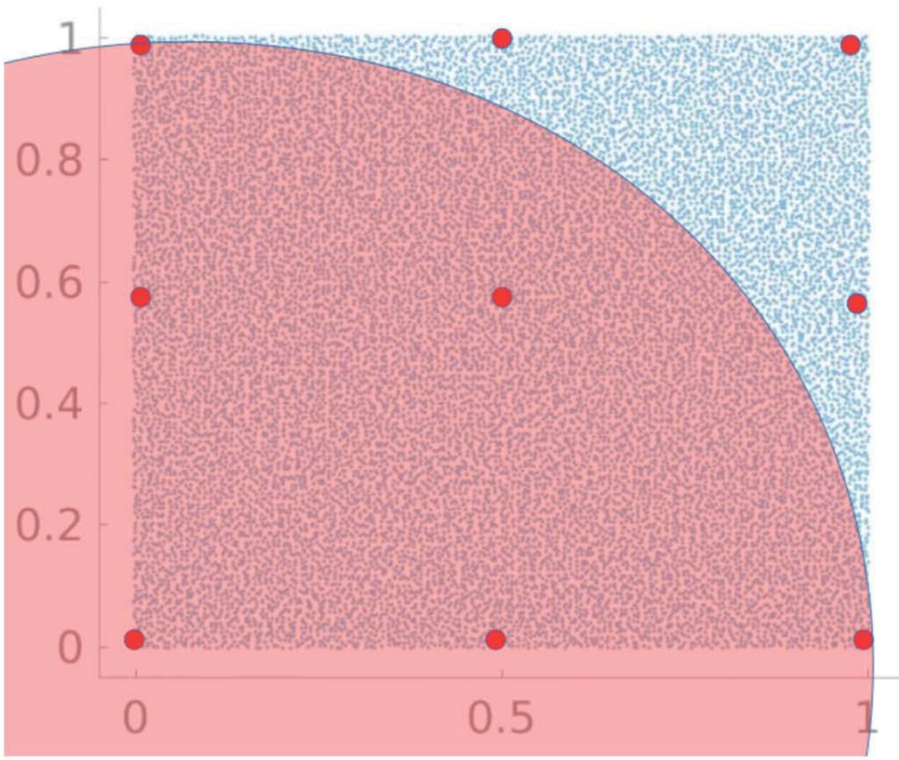


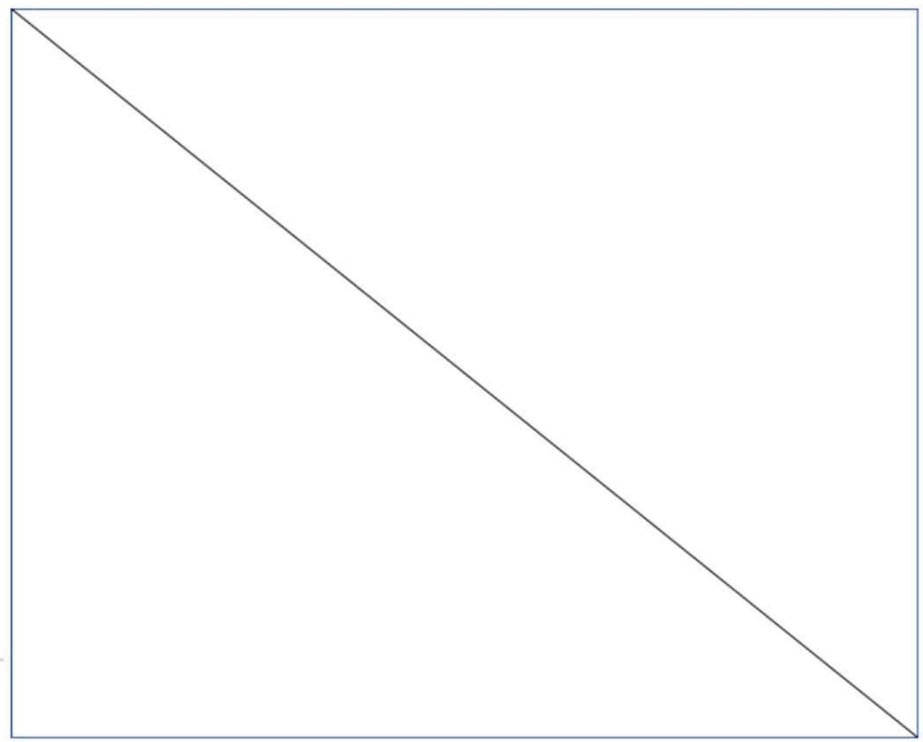


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\ j \in \mathcal{J}^{(l)},\ \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

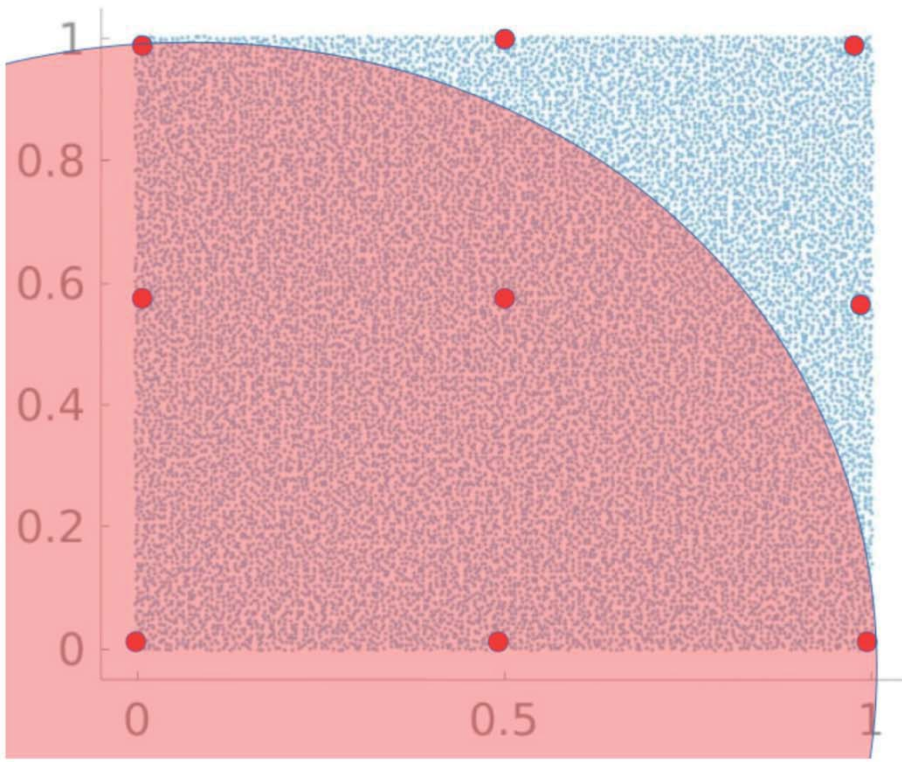


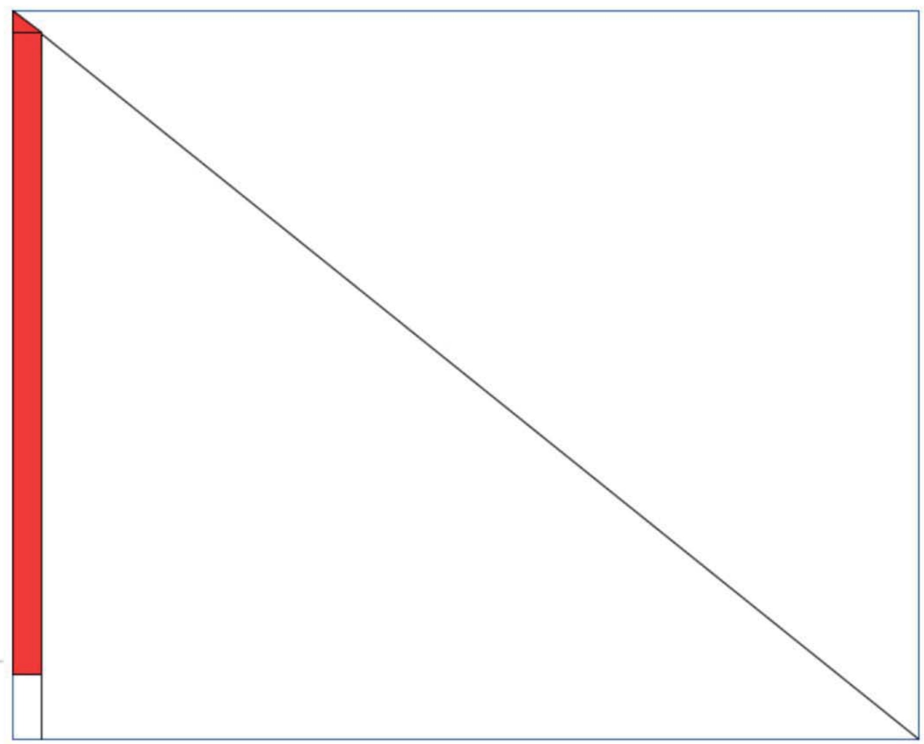


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{ (i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)} \right\}$$

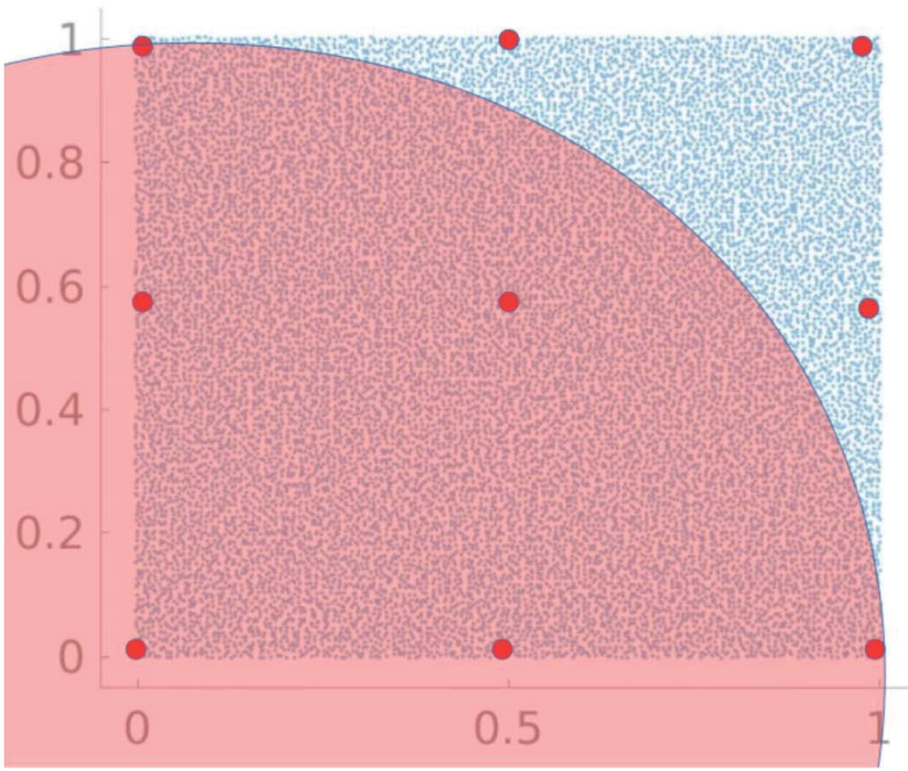


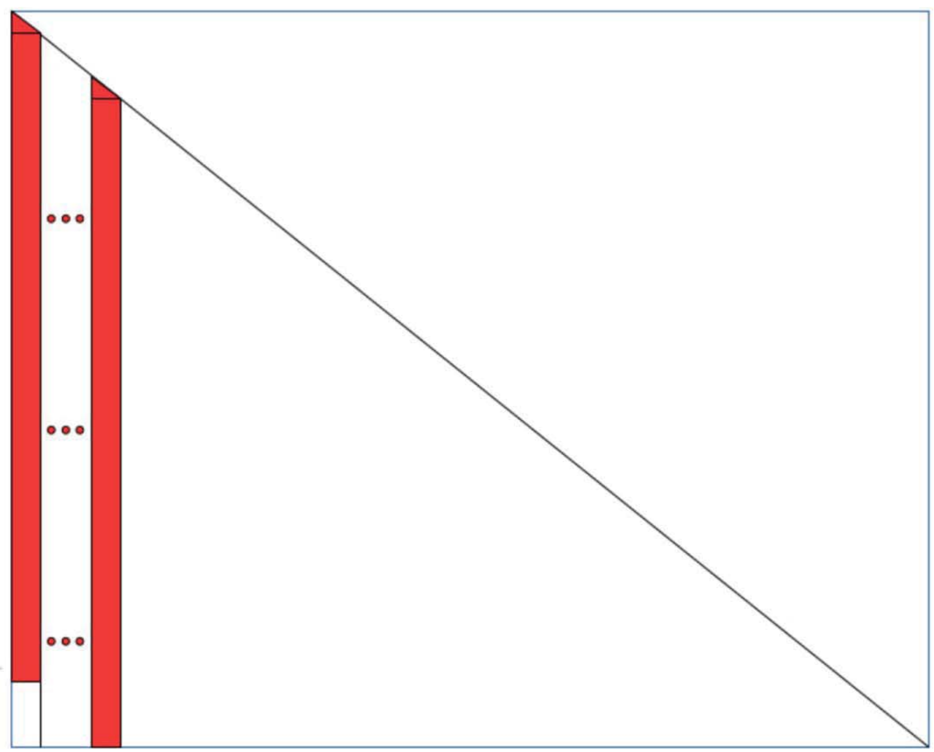


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

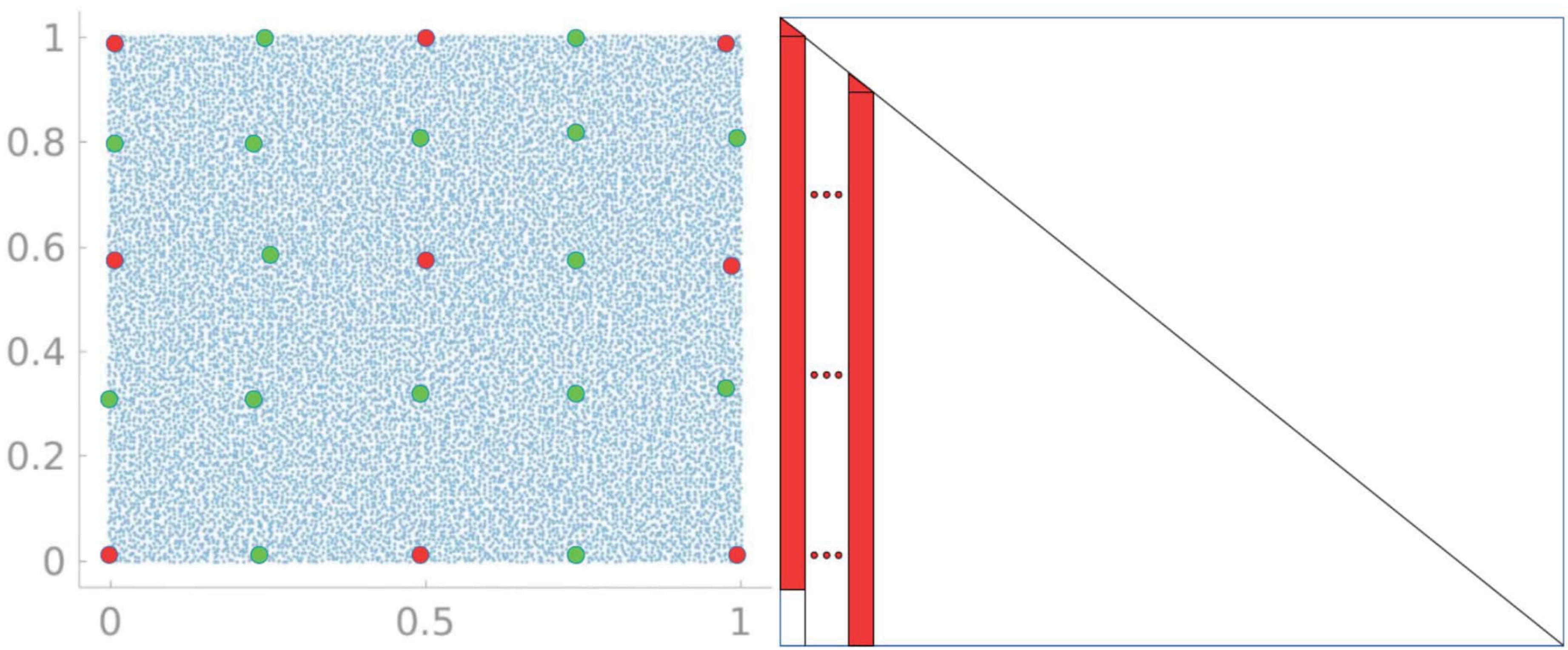




A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

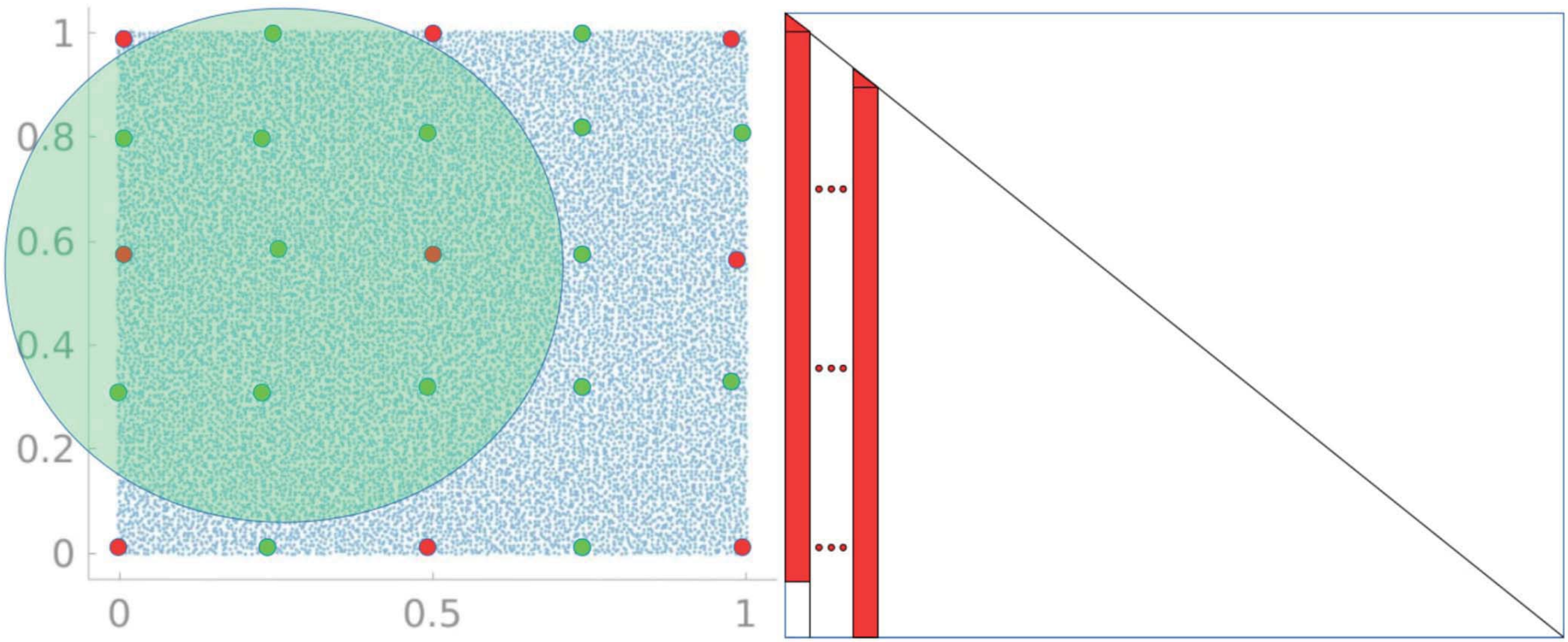
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$



A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

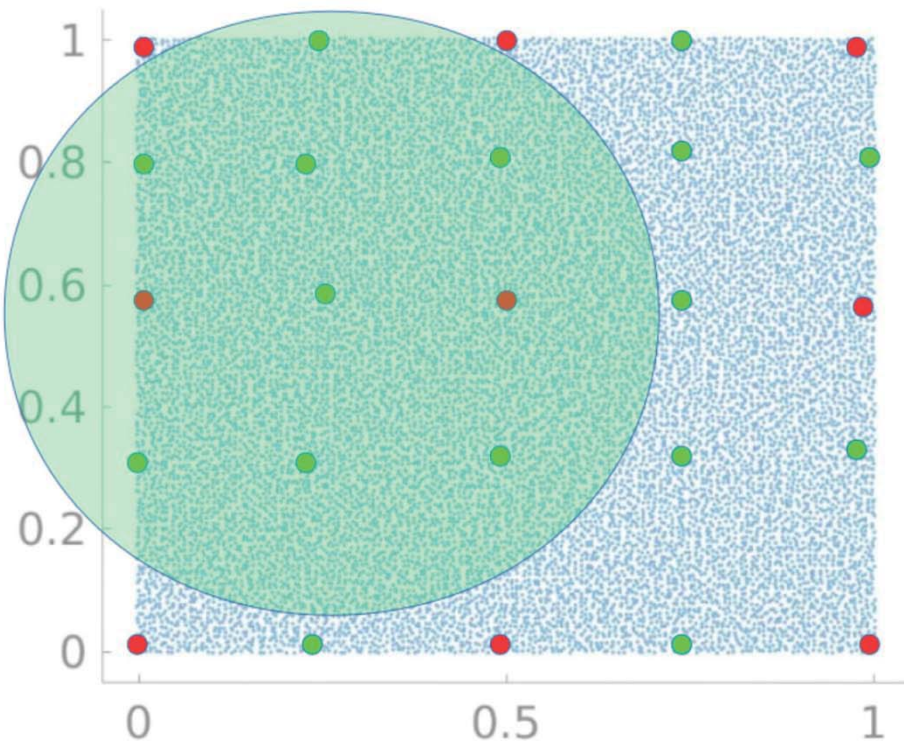
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \mathrm{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

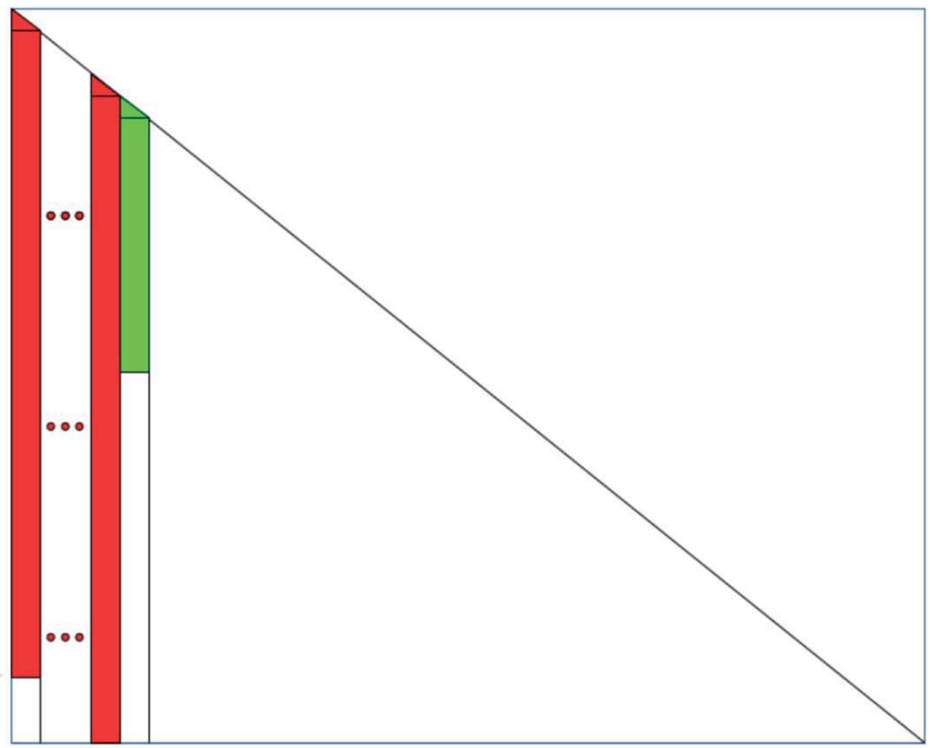


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\ j \in \mathcal{J}^{(l)},\ \mathrm{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

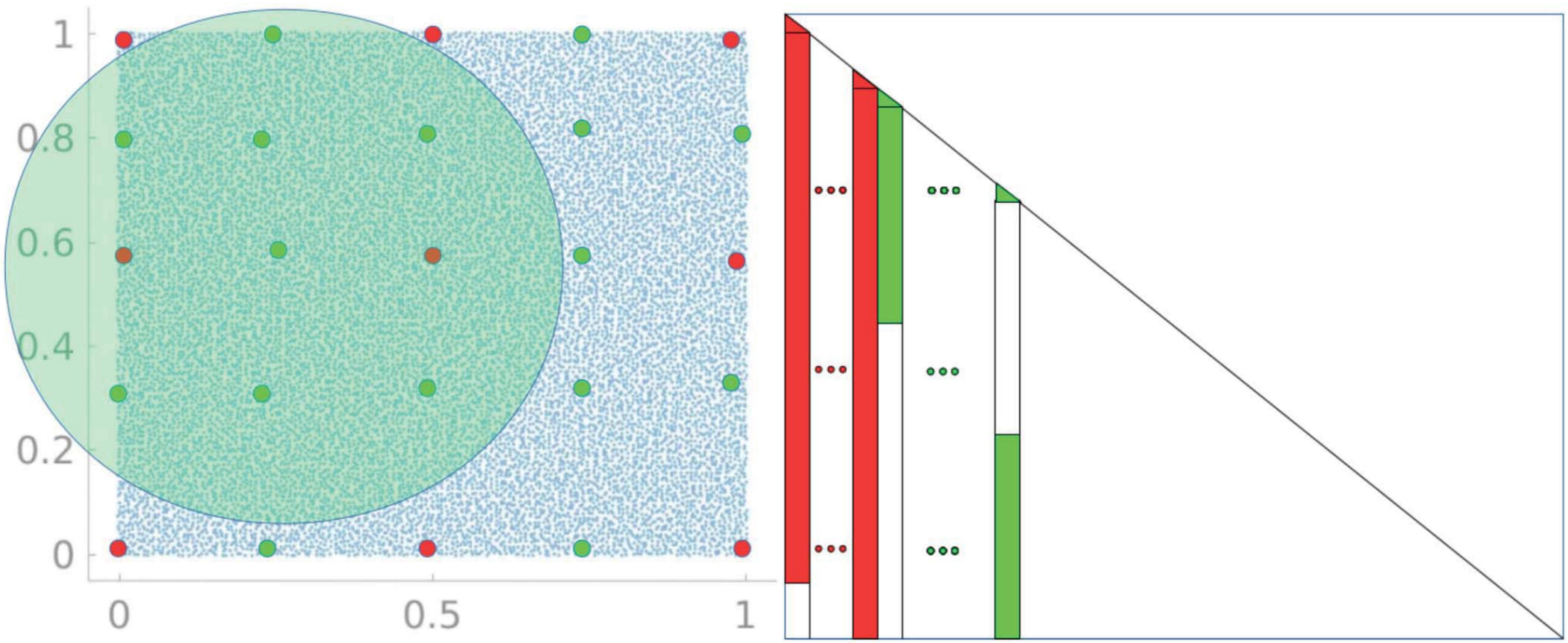




A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

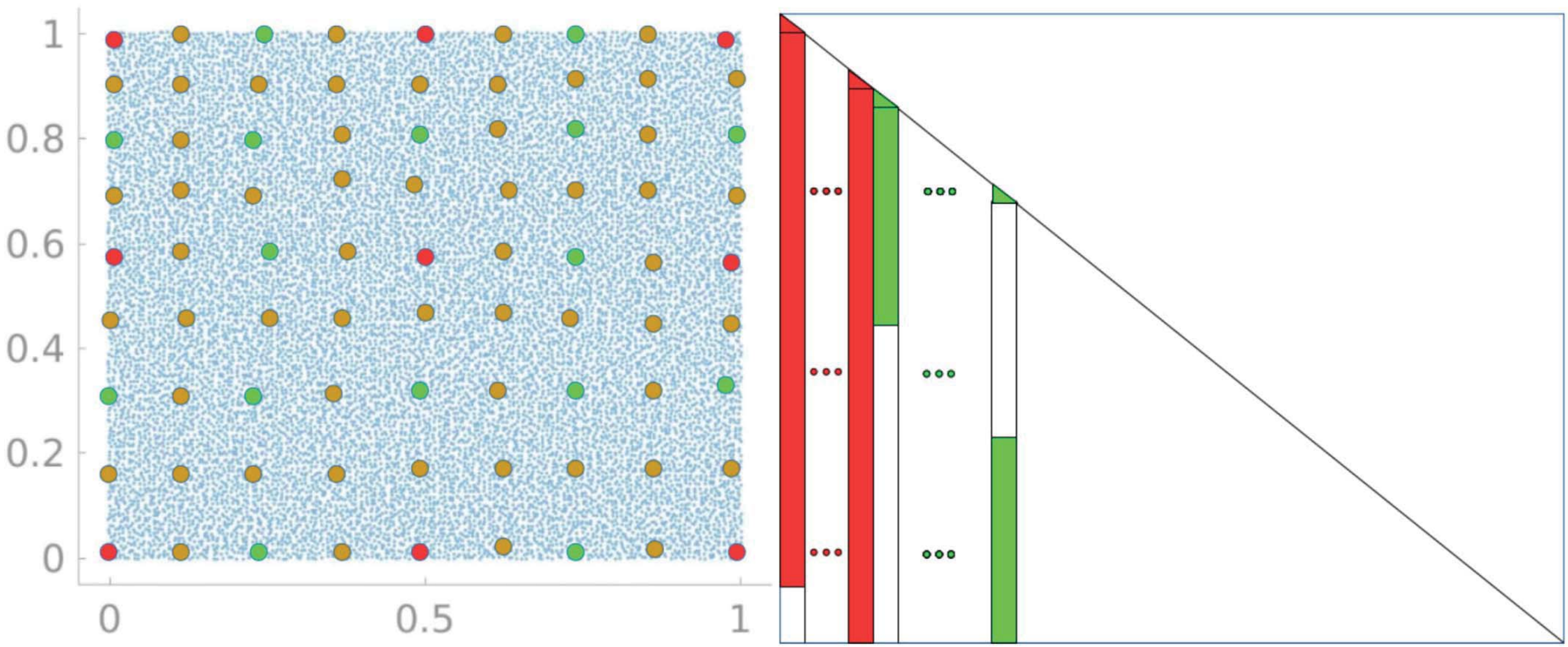
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$



A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

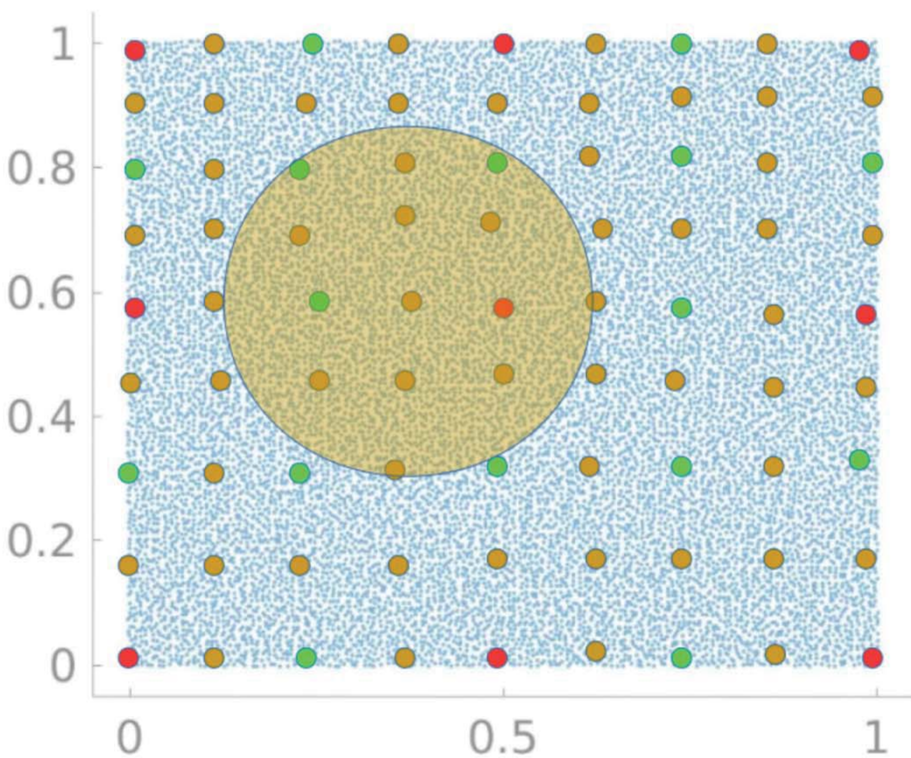
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \mathrm{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

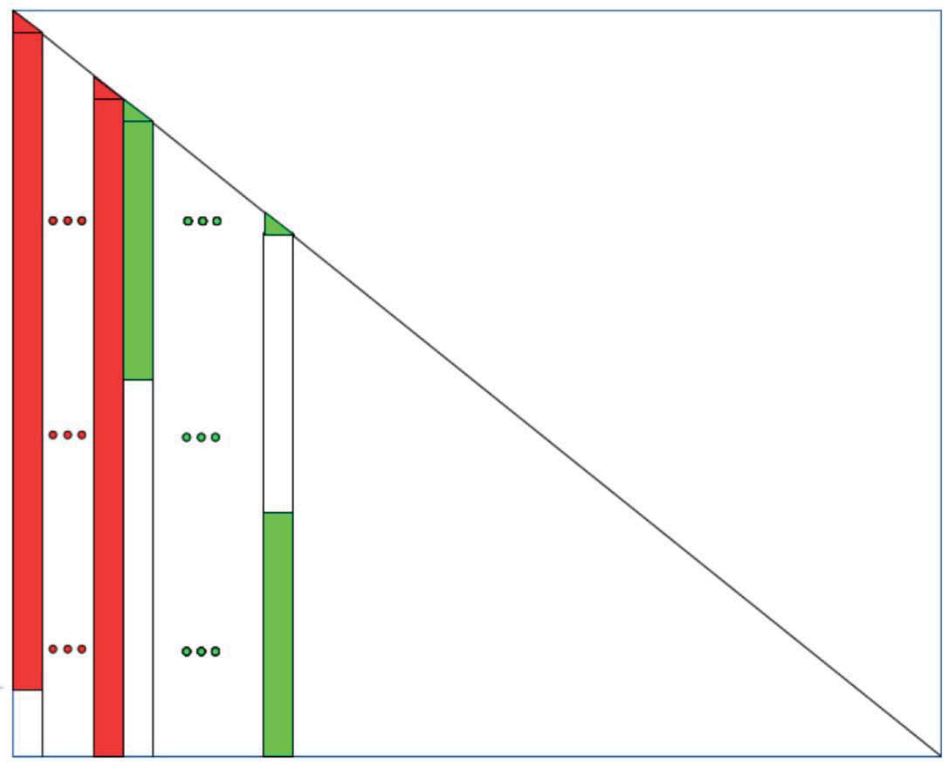


A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

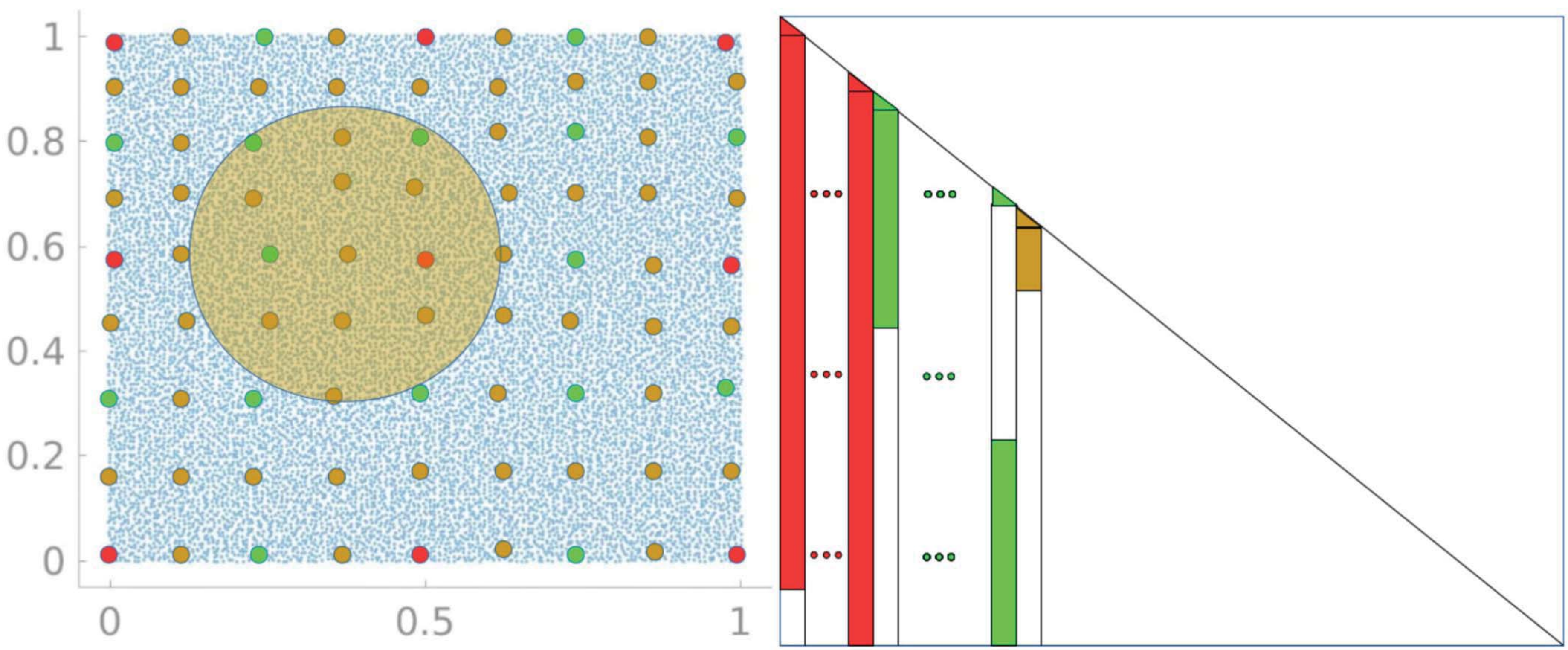




A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

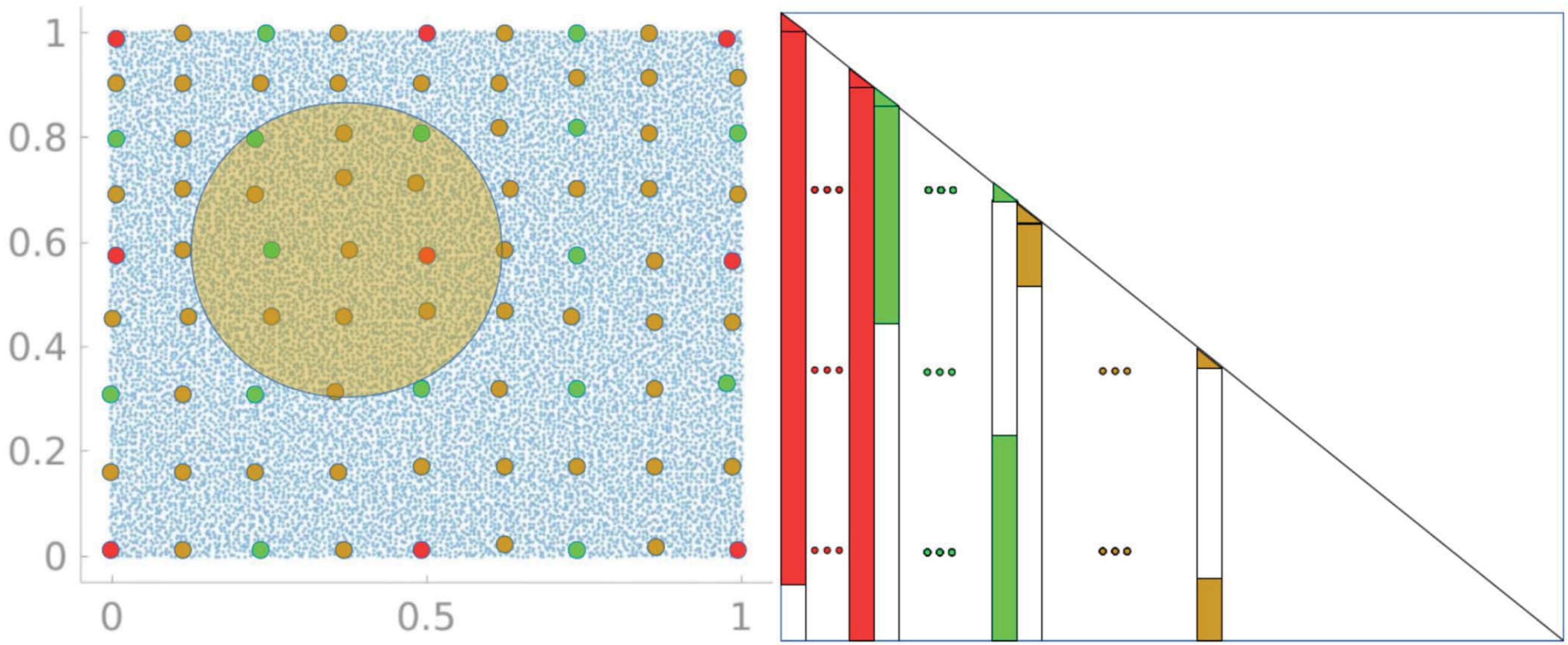
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$



A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

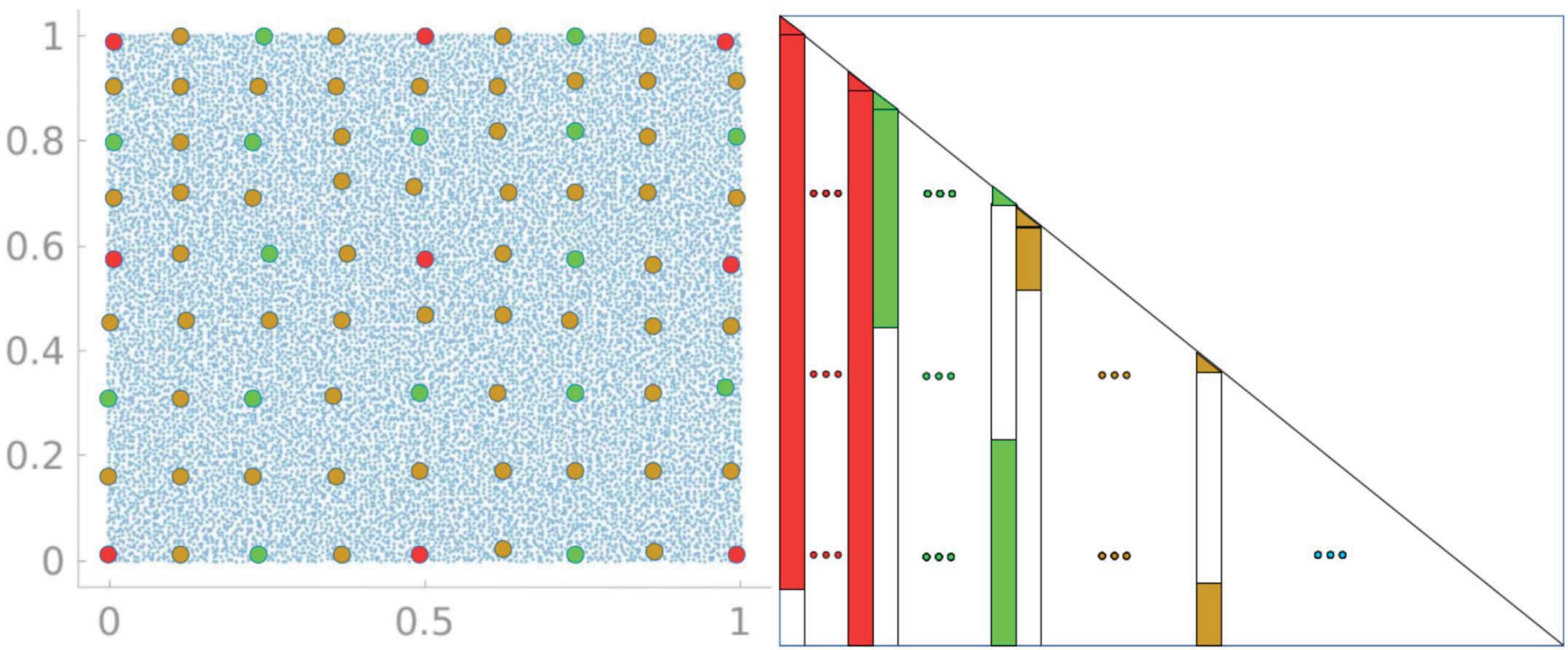
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$



A simple algorithm

- We order the degrees of freedom from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ and define the sparsity pattern:

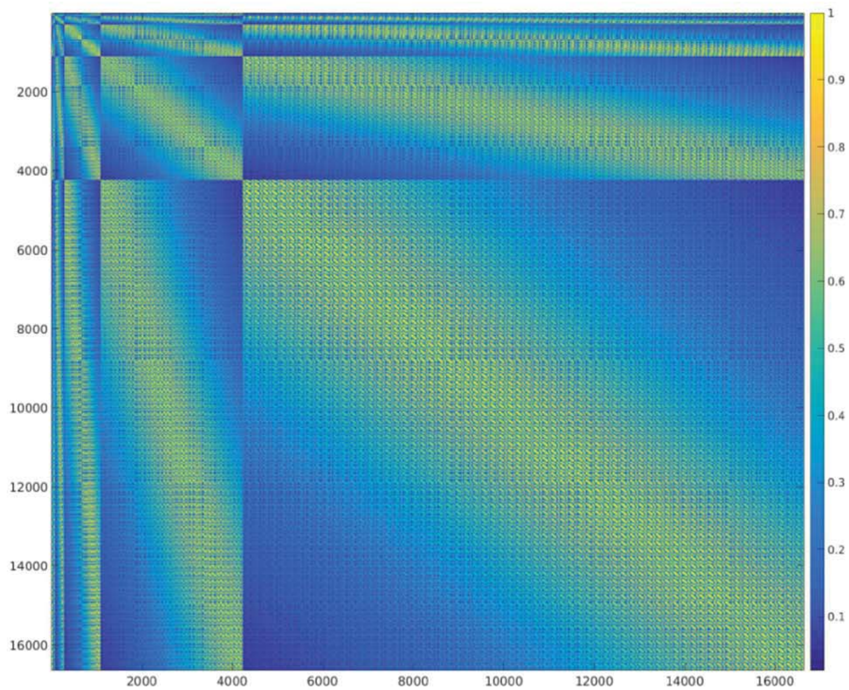
$$S := \left\{(i,j) \in \mathcal{I} \times \mathcal{I} \middle| i \in \mathcal{J}^{(k)},\, j \in \mathcal{J}^{(l)},\, \operatorname{dist}(x_i, x_j) \leq 2 \times 2^{-\min(k,l)}\right\}$$

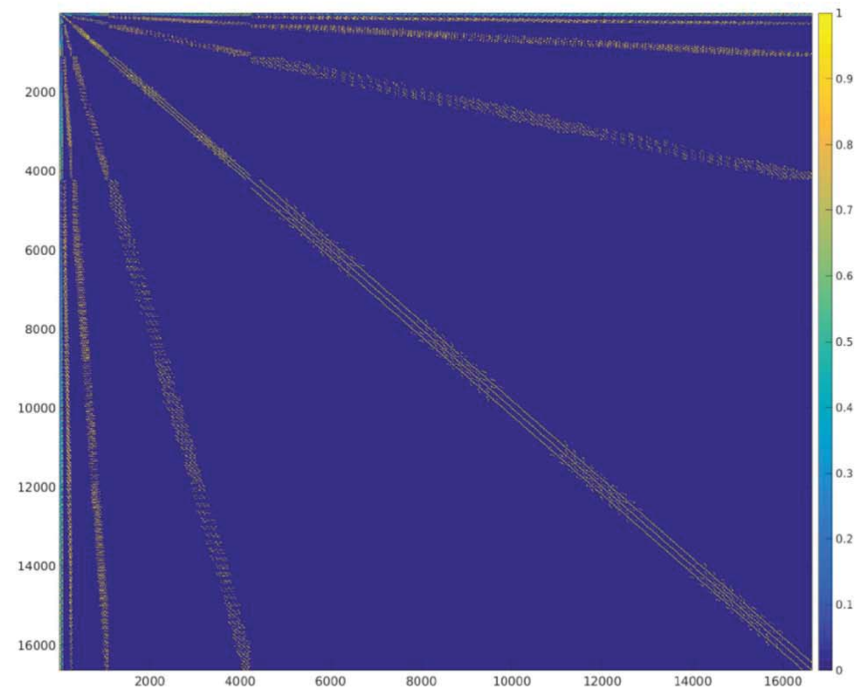


A simple algorithm

- Write the entries in matrix:

$$\Gamma_{i,j} := \begin{cases} \Theta_{i,j}, & \text{for } (i,j) \in S \\ 0 & \text{else} \end{cases}$$





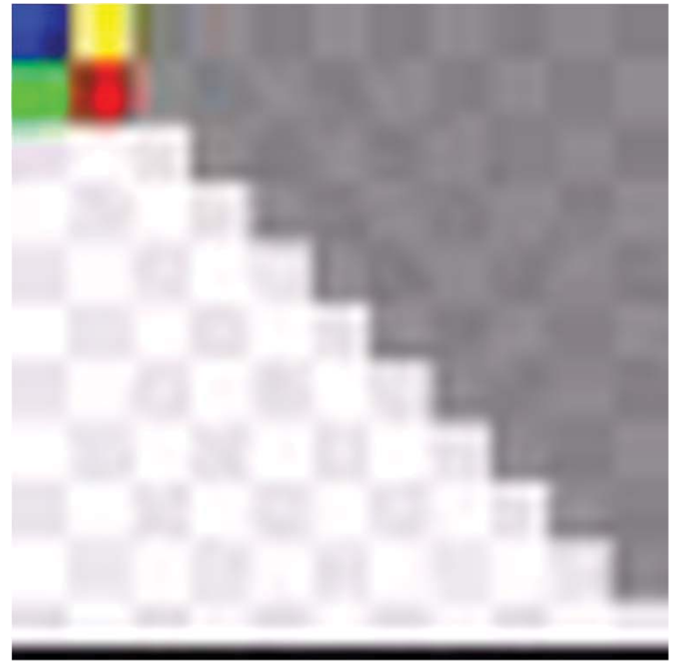
A simple algorithm

Cholesky factorization $A = LL^T$ can be computed as

Algorithm 1: Cholesky factorisation

for $i \leftarrow 1$ *to* N **do**
- $A_{i,i} \leftarrow \sqrt{A_{i,i}}$;
- **for** $j \leftarrow i+1$ *to* N **do**
  - **for** $k \leftarrow j$ *to* N **do**
    - $A_{k,j} \leftarrow A_{k,j} - A_{k,i}A_{j,i}/A_{i,i}$;
- $A_{:,i} \leftarrow A_{:,i}/\sqrt{A_{i,i}}$;

return LowerTriang (A)

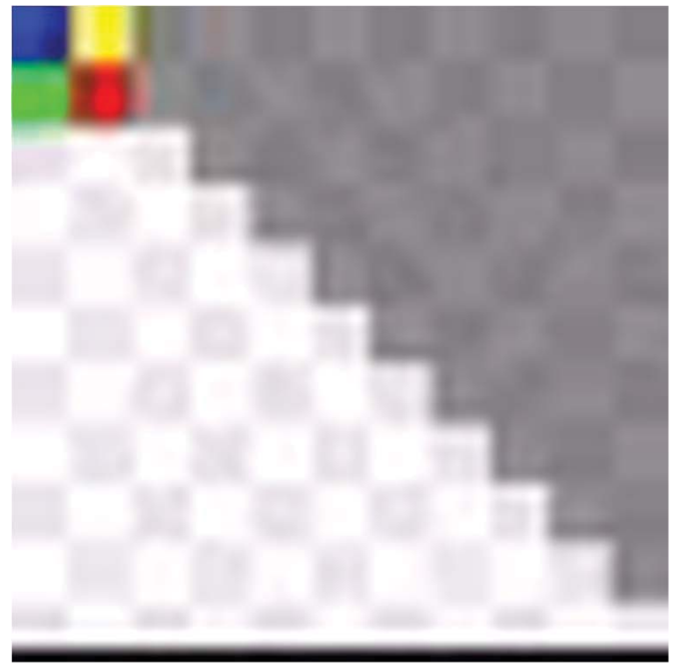


A simple algorithm

Cholesky factorization $A = LL^T$ can be computed as

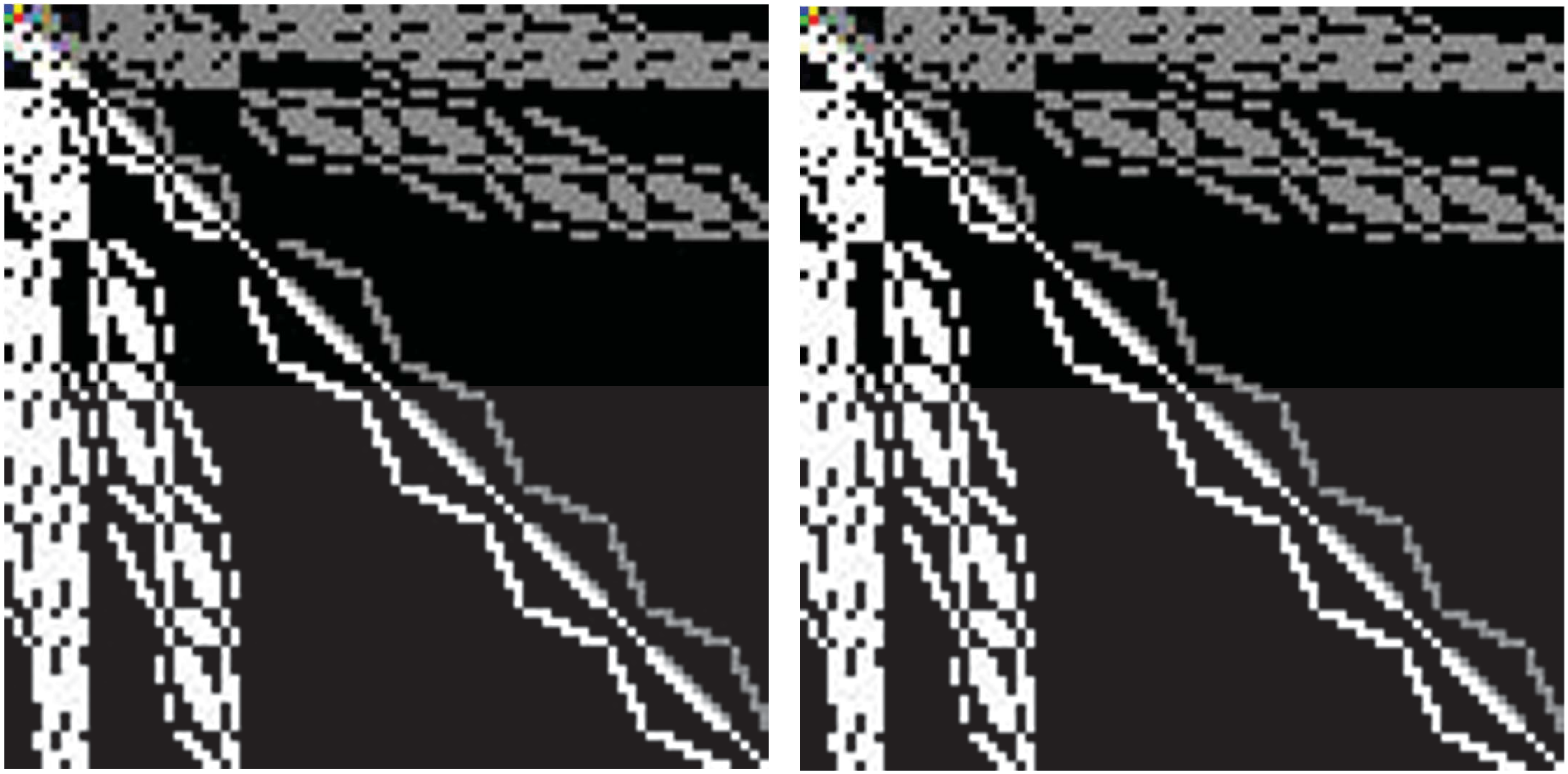
Algorithm 1: Cholesky factorisation

```
for i ← 1 to N do
    A_{i,i} ← √A_{i,i};
    for j ← i + 1 to N do
        for k ← j to N do
            A_{k,j} ← A_{k,j} − A_{k,i} A_{j,i} / A_{i,i};
    A_{:,i} ← A_{:,i} / √A_{i,i};
return LowerTriang(A)
```



- One small Tweak: Skip all operations, for which (k, j), (k, i), or (j, i) are outside of the sparsity pattern.

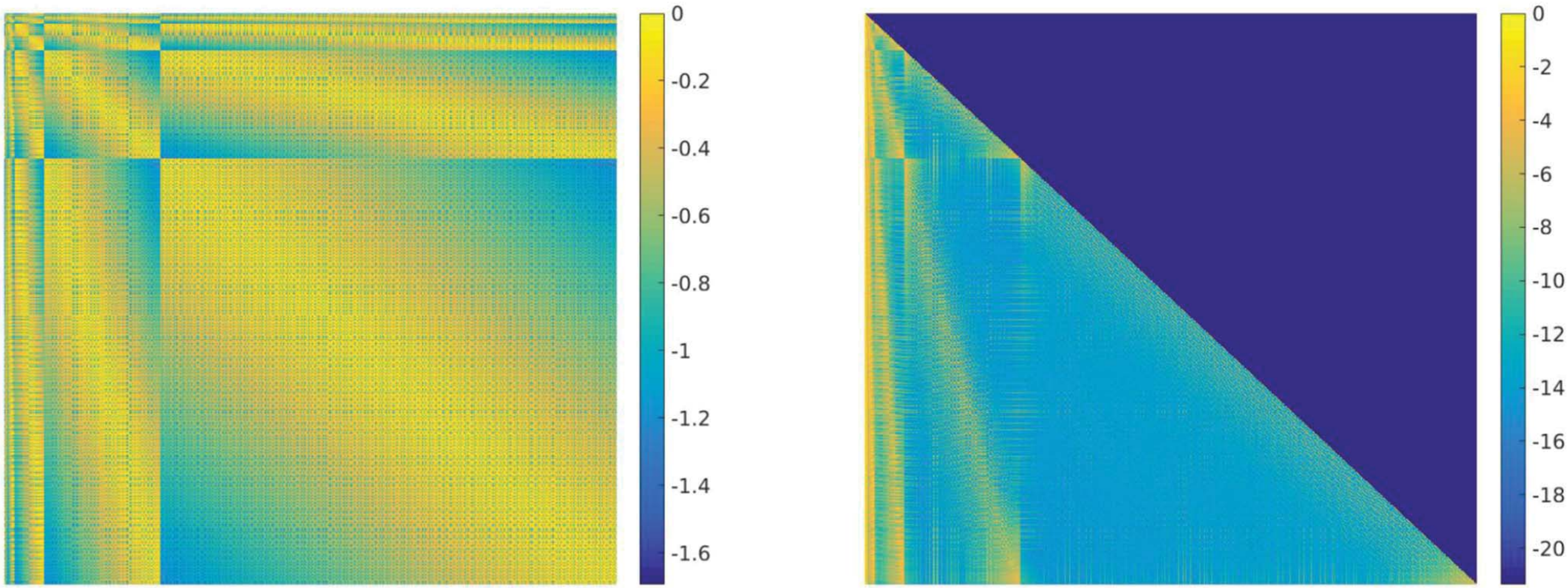
A simple algorithm



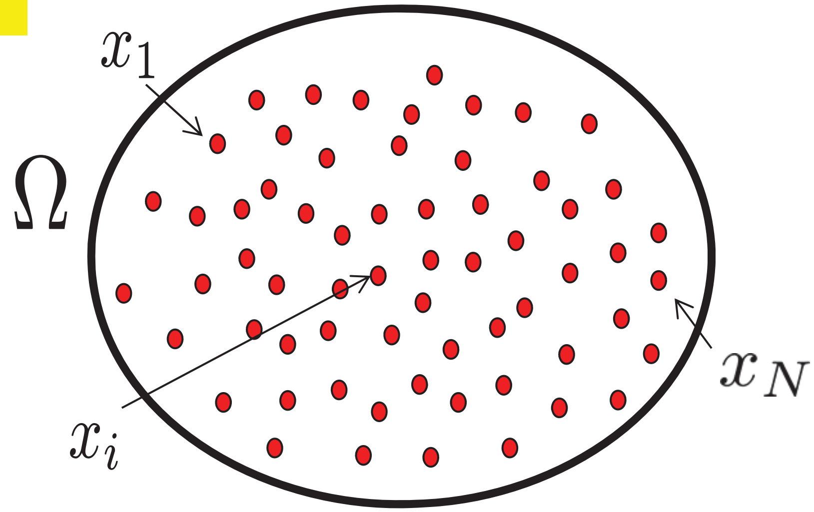
The algorithm is oblivious to exact knowledge of the PDE and uses only the geometry of the discretisation.

Why does it work?

- Θ has *almost* sparse Cholesky factors!



Analysis with gamblets



$\phi_i^{(q)} = \delta(x - x_i)$: Delta Dirac functions

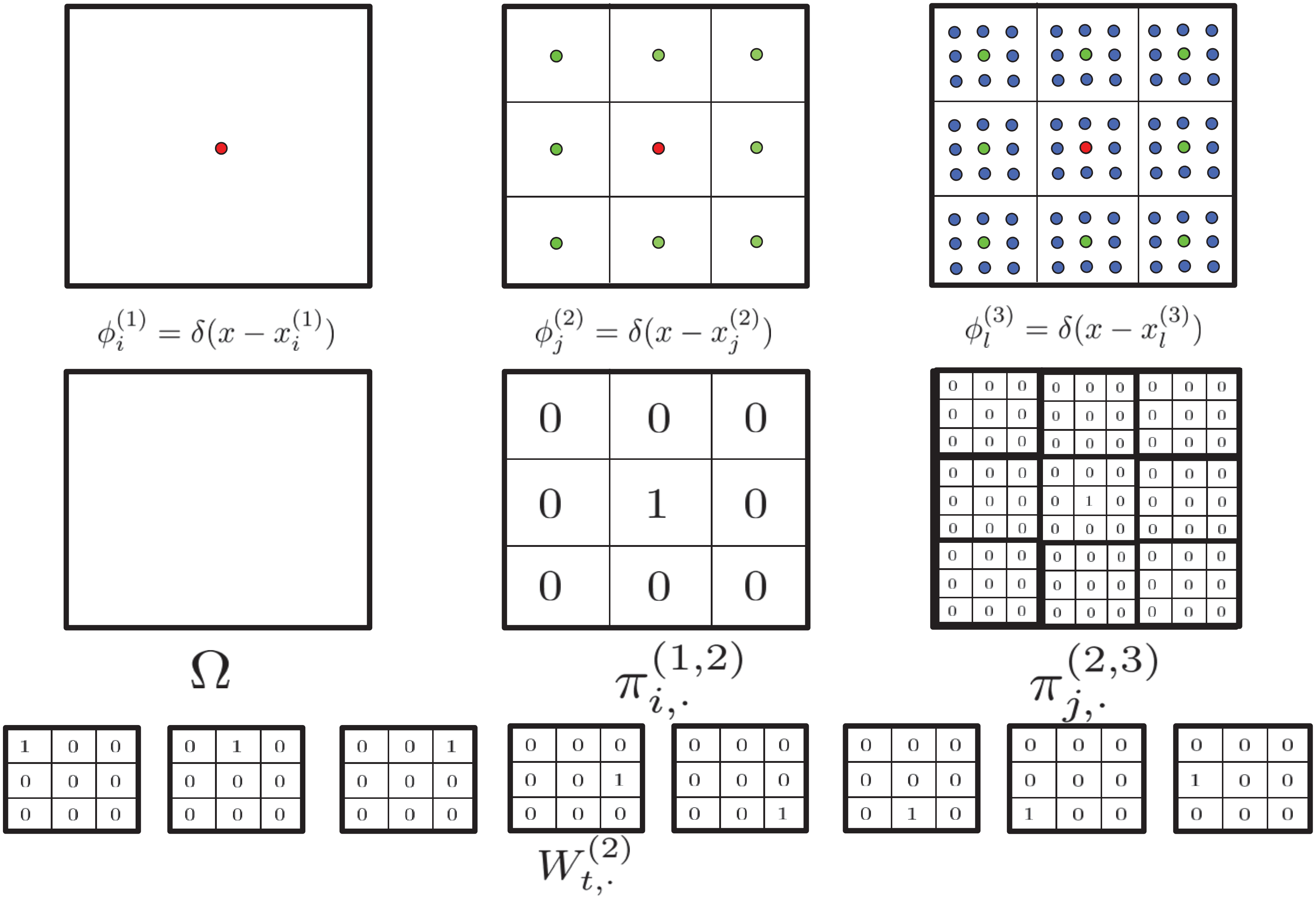
Gamblet Transform

$\psi_i^{(q)}(x) = \sum_j \Theta_{i,j}^{-1} G(x, x_j)$

$$\Theta_{i,j}^{-1} = \langle \psi_i^{(q)}, \psi_j^{(q)} \rangle$$

$\psi_i^{(q)}$ localized $\Rightarrow \Theta^{-1}$ sparse

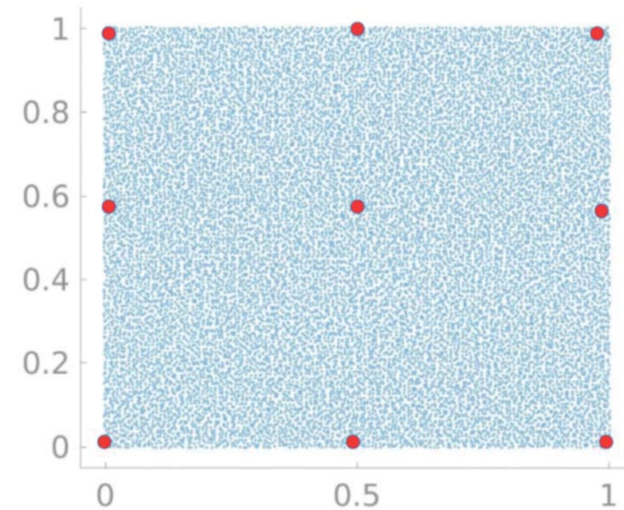
$\phi_i^{(k)}$: Sub-sampled diracs

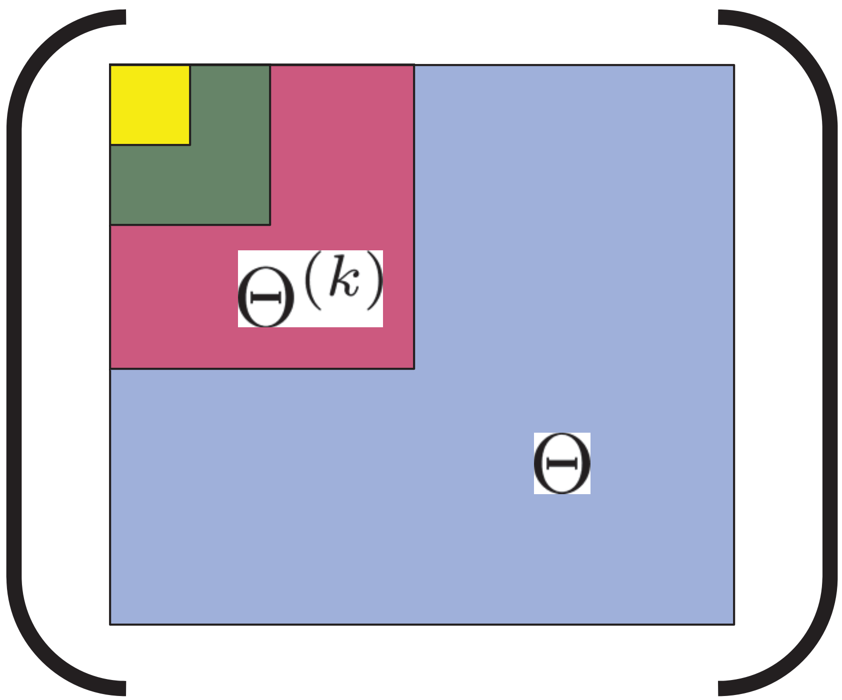


$$\Theta^{(k)}_{i,j} = \int_\Omega \phi^{(k)}_i \mathcal{L}^{-1} \phi^{(k)}_j$$

$$\Theta^{(k)}_{i,j} = G(x^{(k)}_i, x^{(k)}_j)$$

$\Theta^{(k)}_{i,j}$: Sub-matrix of Θ





Single step of (Block-) Cholesky decomposition

$$\Theta^{(k)} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad A = \Theta^{(k-1)}$$

$$= \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}$$

$A^{-1}B$ ← $[\phi_j^{(k)}, \psi_i^{(k-1)}]$

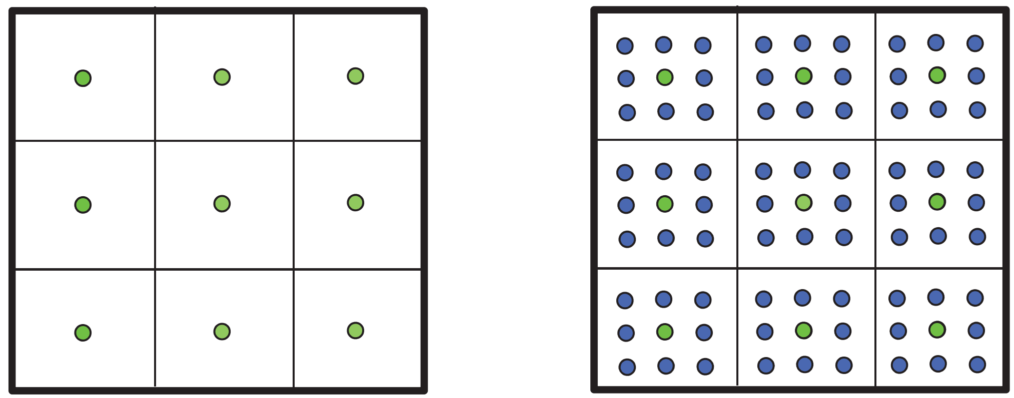
$D - CA^{-1}B$ ← $\left(B^{(k)}\right)^{-1}$

$$B_{i,j}^{(k)} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle \qquad \left(B^{(k)}\right)^{-1}$$

$\left(B^{(k)}\right)^{-1}$ = Schur complement of the block $\Theta^{(k-1)}$

of the matrix $\Theta^{(k)}$

$$B_{i,j}^{(k)} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle$$



$\phi_l^{(k-1)}$ $\phi_i^{(k)},\ \phi_j^{(k)}$

$$(B^{(k)})_{i,j}^{-1} = \mathrm{Cov}\left(\xi(x_i^{(k)}), \xi(x_j^{(k)}) \middle| \xi(x_l^{(k-1)}),\ \forall l\right)$$

$$\psi_i^{(k-1)}(x_j^{(k)}) = \mathbb{E}\left[\xi(x_j^{(k)}) \middle| \xi(x_l^{(k-1)}) = \delta_{i,l}\right]$$

Thank you

- Operator adapted wavelets, fast solvers, and numerical homogenization from a game theoretic approach to numerical approximation and algorithm design, 2018.
- Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis, 2017. arXiv:1703.10761. H. Owhadi and C. Scovel.
- Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity, arXiv:1706.02205, 2017. Schäfer, Sullivan, Owhadi.
- Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients, 2016. arXiv:1606.07686. H. Owhadi and L. Zhang.
- Multigrid with rough coefficients and Multiresolution operator decomposition from Hierarchical Information Games. H. Owhadi. SIAM Review, 59(1), 99149, 2017. arXiv:1503.03467
- Bayesian Numerical Homogenization. H. Owhadi. SIAM Multiscale Modeling & Simulation, 13(3), 812828, 2015. arXiv:1406.6668


Florian Schäfer

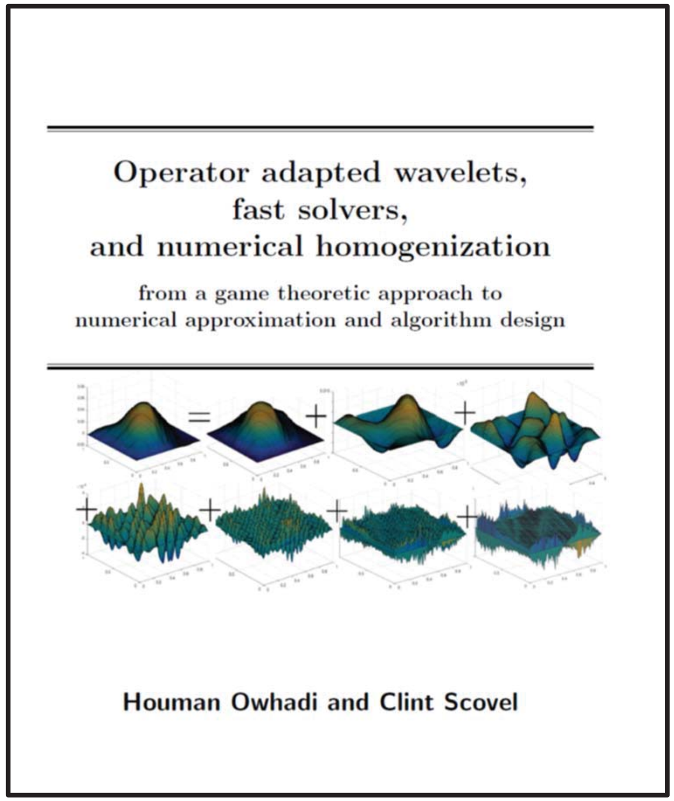

Clint Scovel


Tim Sullivan


Lei Zhang

DARPA EQUiPS / AFOSR award no FA9550-16-1-0054
(Computational Information Games)

Book


Operator adapted wavelets,
fast solvers,
and numerical homogenization
from a game theoretic approach to
numerical approximation and algorithm design
Houman Owhadi and Clint Scovel