Mathematisches Forschungsinstitut Oberwolfach

Report No. 22/2025

DOI: 10.4171/OWR/2025/22

Computational Multiscale Methods

Organized by
Björn Engquist, Austin
Daniel Peterseim, Augsburg
Barbara Verfürth, Bonn
Yunan Yang, Ithaca

27 April – 2 May 2025

Abstract. Many scientific and engineering problems exhibit complex interactions over a wide range of inseparable scales in space and time. Direct numerical simulations to solve such multiscale problems are often beyond current computational capabilities. The difficulties are exacerbated by the presence of uncertainty, randomness, and disorder and are hardly manageable for multiscale inverse problems. Therefore, the simulation of novel phenomena using multiscale models requires a new generation of multiscale computational methods. These must account for under-resolved scales, cross-scale couplings, and stochasticity in a hierarchical and adaptive manner and be able to integrate probabilistic, data-driven, and machine learning approaches. The workshop enhanced the development of a new generation of efficient multiscale computational methods and their rigorous mathematical and numerical analysis so that reliable and fast simulations of challenging multiscale problems from applications eventually become a reality.

Mathematics Subject Classification (2020): 65-XX, 35Bxx, 74Qxx, 70Fxx, 76Axx, 76Mxx, 78Mxx.

Introduction by the Organizers

This workshop concerned the algorithms and mathematics that underlie the computer-aided simulation of complex multiscale processes. Interaction of effects across multiple scales occurs in many applications such as the mechanical analysis of composite and multifunctional materials, porous media flow, wave propagation in heterogeneous media, or the simulation of condensed matter in the presence of

disorder. The main characteristic of such problems is that the inherently complex interplay of non-linear effects on various non-separable length and time scales essentially determines the overall properties that may go far beyond the usual regime in mono-scale problems, as discussed, e.g., in the talks of A. Målqvist, Y. Efendiev, and C. Döding. Although mathematical physics supplies models of partial differential equations that implicitly describe the physical phenomena arising from these microstructures, the problems are intractable for an analytical solution such that their understanding and control depend on numerical simulation. From a computational point of view, however, a direct numerical treatment of such problems is often not feasible due to the fact that the resolution of the full range of relevant scales may result in an almost insurmountable number of degrees of freedom.

The observation and prediction of physical phenomena using multiscale models, hence, require advanced numerical techniques that adaptively select the most relevant scales based on a priori and a posteriori knowledge of the problem, effectively represent unresolved scales, and quantify errors and uncertainty. We refer to such algorithms as computational multiscale methods, and this Oberwolfach Workshop *Computational Multiscale Methods* aimed at the understanding and advancement of computational techniques for the efficient simulation of multiscale processes and analytical or numerical techniques that can provide the effective properties of unresolved scales and utilise such upscaled information to efficiently attain an approximation of sufficient quality or even similar quality as a non-feasible fully resolved simulation. Amongst the particular trends have been probabilistic aspects in both the models and the methods, data-driven and machine learning approaches, and multiscale inverse problems.

The numerical homogenization methods beyond scale separation that have been central to previous Oberwolfach workshops have further matured and can handle now even more challenging linear and non-linear problems that seemed intractable a few years ago. Striking examples were given in the talks by F. Legoll, M. Khrais, A. Lozinski, and T. Sprekeler. Important questions in the design of the methods, like higher-order methods and appropriate localisation strategies, have been pushed forward and recent developments were presented by R. Maier and M. Hauck, respectively. The intimate relation with domain decomposition methods and optimal local approximations via spectral problems enables very powerful and robust algorithms that have been explained and rigorously analysed in talks by R. Scheichl and E. Chung. As a consequence, multiscale methods also inspire new efficient and robust multigrid-type or domain decomposition-type solves and preconditioners. Examples were shown in the talks by L. Zhao and G. Li. The talk of D. Gallistl underlined how such ideas fertilise time-stepping schemes that benefit from the finite propagation speed in wave equations. M. Lukáčová-Medvidová added an analytical view on dissipative (very-weak) solutions of the compressible Euler system, showing how structure-preserving Young measures and K-convergence yield globally existing coarse-grained solutions.

A previous workshop on *Computational Multiscale Methods*[1] in 2019 has already seen a shift in the community toward probabilistic aspects of both the mathematical multiscale models and numerical methods. This trend has ever increased then, making randomness a central feature in the models and a crucial computational tool in multiscale methods. F. Otto showed recent results in the macroscopic or effective description of random media. The talks by A. Teckentrup and J. Fischer highlighted recent results for the further development of multilevel Monte Carlo methods as well as their use in challenging application problems. Randomness as a computational tool for efficient model reduction was central to K. Smetana's talk, focusing on the challenges for nonlinear problems. The combination of randomness and multiscale structures still poses severe challenges for computational multiscale methods. A. Lang showed how to formulate and analyse a numerical homogenisation scheme for SPDEs and D. Kolombage presented an offline-online strategy for a class of random perturbations in multiscale materials.

There is an increasing interest in using the multiscale approach to solve inverse problems due to its ability to facilitate robust and efficient interpretation of complex datasets, which often hold information across a variety of scales. In that regard, M. Peter showed how to identify microstructure information in elastodynamics with the help of homogenisation. The identification of dynamical systems was at the heart of J. Botvinick-Greenhouses's talk, where he showed a data-driven approach. In general, recent years have seen an incredibly growing interest in data-driven methods, for instance, to learn effective or upscaled models. This trend was also clearly visible in the workshop. For instance, H. Owhadi and R. Tsai discussed data-driven learning strategies for solution operators and Hamiltonian flows, respectively.

Further recent hot topics are machine-learning methods and quantum computing, which were also reflected in the workshop. Using classic thoughts on adaptivity for numerical methods, M. Feischl opened a new perspective on the hierarchical training of neural networks. The talk by X. Liu showed how neural networks can be used in the design of multigrid strategies. The incredible potential but also current limitations of quantum computing strategies were highlighted in the talks of J. Hu and M. Deiml in the context of multiscale problems and multilevel preconditioning.

Ultimately, the workshop has discussed crucial algorithmic challenges and fundamental mathematical problems at the intersection of the scientific fields of multiscale modeling and simulation, scientific computing, computational (geo-)physics and material sciences and, in particular, numerical and mathematical analysis of partial differential equations. New bridges between these research communities have been identified that promise future progress on *Computational Multiscale Methods*.

There were 48 on-site participants from 10 countries, more specifically, 20 participants from Germany, 10 from the United States, 5 from Sweden, 4 from China,

[1] *Computational multiscale methods. Abstracts from the workshop held July 28–August 3, 2019.* Organized by B. Engquist and D. Peterseim. Oberwolfach Rep. 16(3):2099–2181, 2019.

3 from France, 2 from Austria, one participant each from the UK, Canada, Saudi-Arabia, and Singapore. Furthermore, there were 14 women and 18 young researchers – with less than 10 years of experience since the PhD – among the participants. In addition, 5 participants from the United States (3 of them female) joined the workshop virtually. On behalf of all participants, the organizers would like to thank the institute and in particular its staff for their great hospitality and support before and during the workshop.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, "US Junior Oberwolfach Fellows".

Workshop: Computational Multiscale Methods

Table of Contents

Abstracts

Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis

Houman Owhadi
(joint work with Yasamin Jalalian, Juan Felipe Osorio Ramirez, Alexander Hsu, Bamdad Hosseini)

We introduce a novel kernel-based framework for learning differential equations and their solution maps that is efficient in data requirements, in terms of solution examples and amount of measurements from each example, and computational cost, in terms of training procedures. Our approach is mathematically interpretable and backed by rigorous theoretical guarantees in the form of quantitative worst-case error bounds for the learned equation. Numerical benchmarks demonstrate significant improvements in computational complexity and robustness while achieving one to two orders of magnitude improvements in terms of accuracy compared to state-of-the-art algorithms. In comparison to equivalent neural net methods, our approach is significantly more robust to the choice of hyperparameters and does not require close human supervision during training.

References

[1] Y. Jalalian, J. F. Osorio Ramirez, A. Hsu, B. Hosseini, H. Owhadi, *Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis*, arXiv:2503.01036 (2025).

Smoothed circulant embedding and applications in multilevel Monte Carlo methods

Aretha L. Teckentrup
(joint work with Anastasia Istratuca)

Mathematical modelling and simulation are frequently used to inform decisions and assess risk. However, the parameters in mathematical models for physical processes are often impossible to determine fully or accurately, and are hence subject to uncertainty. In practice, it is crucial to study the influence of this uncertainty on the outcome of the simulations, in order to correctly quantify risk and make decisions. By assigning a probability distribution to the parameters, which is consistent with expert knowledge and observed measurements if they are available, it is then possible to propagate the uncertainty through the model and quantify the induced uncertainty in the model outputs.

A typical example problem is groundwater flow modelling, with potential applications in carbon capture and storage underground. In this setting, the precise make-up of the environment, such as the location and conductive properties of different layers of soil, is not fully known. Darcy's law for a steady state flow in

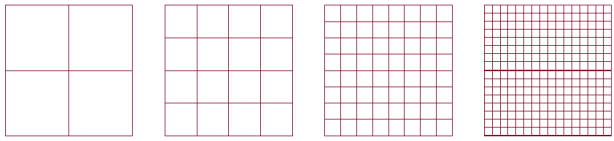


FIGURE 1. A range of meshes used in MLMC.

a porous medium, coupled with incompressibility conditions, results in a diffusion PDE for the flow of water underground:

$$-\nabla \cdot (k(\mathbf{x}, \omega)\nabla u(\mathbf{x}, \omega)) = f(\mathbf{x}), \quad \mathbf{x} \in D.$$

Here, u denotes the pressure head of the fluid, f represents the source or sink terms, and k stands for the hydraulic conductivity, that is, the ease with which a fluid can move through porous media or fractures under a given pressure gradient. Due to the sparsity of measurements available, there is considerable uncertainty in the value of k, and this is hence typically modelled as a random field.

The end goal is typically to compute quantities of interest $\mathbb{E}[Q]$ in the form of expected values of functionals Q of the PDE solution u. An easily computed example of physical interest is the pressure of the water at a given point $\mathbf{x}^* \in D$. A much more challenging example is to quantify the risk of leaked particles from a repository re-entering the human environment, which requires simulating particle flow in the velocity field $-k\nabla u$ and computing hitting times.

A standard computational method to propagate the uncertainties through the model is to use Monte Carlo sampling (also known as ensemble methods): we pick different realisations of our uncertain parameter k, simulate the model for the different scenarios, and make inference on output quantities of interest from the ensemble of simulations. This process can be notoriously time consuming, especially when applied to complex models which result in expensive simulations.

Multilevel Monte Carlo (MLMC) methods [2] alleviate the simulation cost by utilising different discretisations (or approximations) of the underlying model. Most simulations are done with a low cost/low accuracy discretisation to capture the bulk behaviour, while a few high cost/high accuracy simulations are added to increase the overall accuracy. In the context of the PDE discussed above, this corresponds to using a hierarchy of meshes ranging from coarse and computationally cheap to fine and accurate as illustrated in Figure 1.

A typical model used for $k(\mathbf{x}, \omega)$ is a log-normal random field, i.e. $k(\mathbf{x}, \omega) = \exp(Z(\mathbf{x}, \omega))$ where Z is a stationary Gaussian field. Due to the typically large physical extent of the spatial domain D (in the order of hundreds of kilometers), the length scales of variations in Z are small compared to the size of the domain and samples of Z, and hence of k, are highly oscillatory. A typical sample of Z is shown in the left plot in Figure 2.

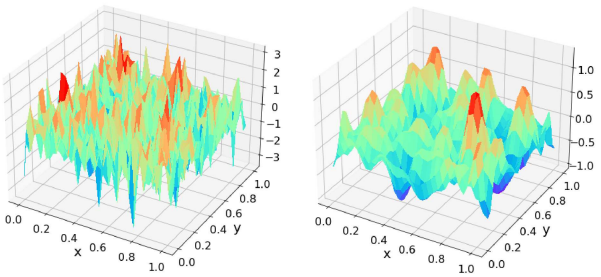


FIGURE 2. Sample from the Gaussian field Z (left) and its smoothed version $\tilde{Z}$ (right)

The highly oscillatory nature of k makes the application of MLMC methods challenging, since very coarse meshes cannot resolve the oscillations. To circumvent this issue, we propose to used smoothed versions of k on coarse meshes [3]. This smoothing is illustrated in the right plot in Figure 2.

Our smoothing technique is based on the circulant embedding method [1] for sampling from Gaussian random fields. Here, sampling from Z on the mesh points used for the numerical solution of the PDE is performed by computing

$$\mathbf{Z} = G\Lambda^{1/2}\boldsymbol{\xi},$$

where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, I)$ and $S = G\Lambda G^T$ is an eigendecomposition of a cleverly chosen extension of the covariance matrix of the random field evaluated at the mesh points. On a uniform mesh, S is circulant for $D \subset \mathbb{R}$ and block circulant with circulant blocks for $D \subset \mathbb{R}^2$, and this eigendecomposition can hence be computed efficiently in log-linear time using the Fast Fourier transform.

To obtain a smoothed random field, we drop the τ smallest eigenvalues in a given sample $\mathbf{Z} = G\Lambda^{1/2}\boldsymbol{\xi}$, which correspond to the sharpest oscillations. The choice of τ, and hence the amount of smoothing introduced, is based on the following theorem [3].

Theorem 1. *Let τ be the truncation index and $\tilde{\mathbf{Z}}$ be the resulting smoothed sample. Then, for any $p \in \mathbb{N}$:*

$$\mathbb{E}\left[\|\mathbf{Z} - \tilde{\mathbf{Z}}\|_\infty^p\right] \leq C s^{-\frac{p}{2}} \left(\max_{j=s-\tau+1,\ldots,s} \sqrt{\lambda_j}\right)^p \tau^p,$$

where $s = \dim(S)$ is the dimension of the embedding matrix S.

Choosing suitable truncation indices τ_ℓ, linked to the mesh size h_ℓ on level ℓ, the MLMC method then leads to orders of magnitude savings of MLMC over standard MC in computational cost to obtain a given accuracy.

References

[1] C.R. Dietrich, G.N. Newsam, *Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix*, SIAM J. Sci. Comput. **18** (1997), 1088–1107.

[2] M.B. Giles, *Multilevel Monte Carlo methods*, Acta Numer. **24** (2015), 259–328.

[3] A. Istratuca, A.L. Teckentrup, *Smoothed circulant embedding with applications to multilevel Monte Carlo methods for PDEs with random coefficients*, IMA J. Numer. Anal. (2025), drae102.

Towards optimal hierarchical training of neural networks

Michael Feischl

(joint work with Alexander Rieder, Fabian Zehetgruber)

This talk is based on the preprint [5]. One of the most pressing problems in deep learning is the question of how to design good networks for a given task and how to train them efficiently. The design of network architectures is often driven by intuition and/or experiment. This is particularly true for machine learning tasks in language modelling or computer vision and hence automated methods are in demand [8]. Particularly in mathematical applications of deep learning, we have the unique advantage of constructive approximation results that show that a certain architecture will, with the right weights, achieve a certain approximation quality. However, even if we know the optimal architecture, it is still an open question of how to find the optimal weights in reasonable compute time. The conventional approach of employing variants of (stochastic) gradient descend is known to sometimes converge to weights that result in networks with much worse approximation qualities than what is predicted from approximation theory. Examples of this gap can be found everywhere in the literature, even with mathematical justification [6]. This is in stark contrast to classical approximation methods, such as interpolation, spectral methods, or finite element/volume methods. There, for many problem sets, algorithms are known that design optimal approximation architectures (e.g., the mesh for a finite element method) and also equip them with the optimal parameters (e.g., the coefficients of the polynomial on each mesh element). The design of the architecture and the determination of the parameters go hand-in-hand. One prominent example of this are adaptive mesh refinement algorithms for the finite element method. The optimality results known for such algorithms are of the following form: Assume that there exists a mesh (that, in general, can not be computed practically) on which a good approximation $u_h^{\mathrm{opt}} \approx u$ exists, then the algorithm will find a similarly good mesh with a comparable amount of degrees of freedom as the optimal mesh on which it computes an approximation u_h of comparable quality, i.e.,

$$\|u - u_h\| \leq C\|u - u_h^{\mathrm{opt}}\| \quad \text{and} \quad \text{compute-time}(u_h) \leq C\,\text{compute-time}(u_h^{\mathrm{opt}}),$$

for some constant $C > 0$ that depends only on secondary characteristics of the method but is independent of the solution u, its approximation and the number of degrees of freedom. For examples of such results, we refer to the seminal works [1,

7, 2] in which the Poisson problem was tackled. The ideas in these works were generalized to many other model problems, and we refer to [3] for an overview.

The big advantage of this line of reasoning is that one does not need to know the approximability of the exact solution u and can still guarantee that the approximation is (quasi-) optimal for a given computational budget.

The present work pursues whether similar results are possible for the training of deep neural networks and gives some promising first answers. To that end, we propose an algorithm that hierarchically extends the network architecture by adding new neurons in each layer. This is done in such a way that the realization of the network is unchanged at first, i.e., the new weights are set to zero. Then, an optimal initialization of the new weights is computed by solving a small optimization problem. Under the assumption that a neural network of a certain size exists that strictly improves the loss function, we can show that this optimal initialization reduces the loss by a factor that is proportional to the optimal loss reduction if the best possible (but in general unknown) initialization would be chosen.

To rigorously establish the loss reduction, we introduce the concept of *stable networks*. The notion quantifies the amount of cancellation that is present in a network with given weights. Roughly speaking, we show that the existence of stable networks that achieve small loss guarantees that the proposed hierarchical training algorithm will find networks of similar size with similar loss. We argue that the notion of stability is necessary in the sense that if a certain loss functional can only be made small by very unstable networks (i.e., a small change in the weights results in a large change of the output), there is little hope for any training algorithm to find the correct weights for such a network. Practical experience also seems to suggest that this stability is inherent in many deep learning applications. For example, it is common practice to do inference or training of large networks in single precision in order the increase performance.

As a corollary, we obtain a computable quantity that can be used to judge the training state of a given network. The quantity measures how far a network of a given size is from optimality, i.e., the smallest possible loss for this network architecture. This is similar to the well-known Céa lemma for Galerkin methods, which states that a computed approximation is, up to a factor, as good as the best possible approximation within the given setting. The difference to the present setting is that we do not know the factor a priori, but compute a quantity that is related to this factor. In that sense, we have an a posteriori type Céa lemma for training neural networks.

Finally, we show that our hierarchical training algorithm results in improved estimates on the generalization error compared to direct algorithms. Moreover, we introduce the concept of *optimal generalization* which turns out to be necessary in order to achieve small generalization error and allows us to make statements about the generalization error without quantifying the distance of training data and possible inputs of the network. This is particularly interesting as current machine learning applications such as Large Language Models (e.g.,[4]) demonstrate remarkable generalization over inputs that are far away from the training data.

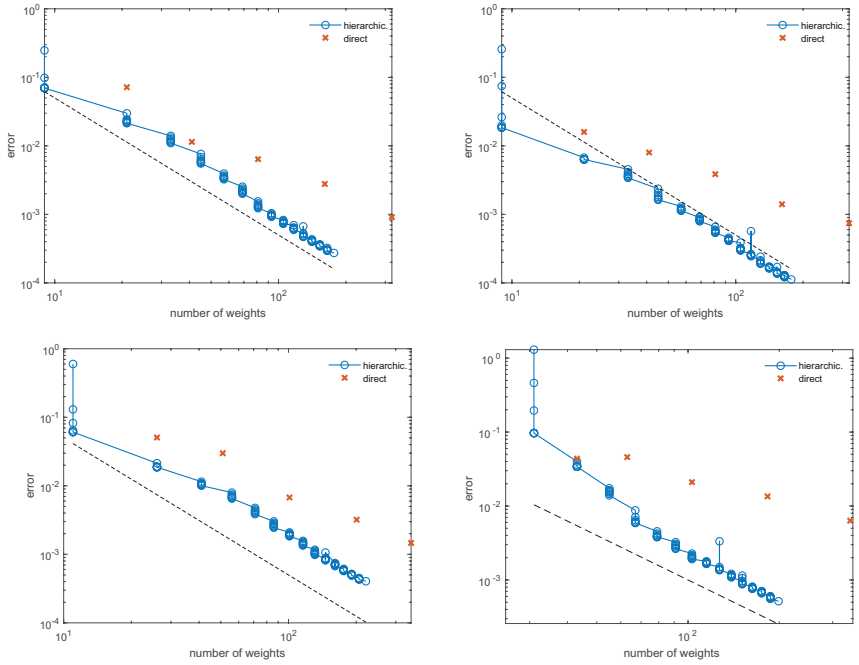


FIGURE 1. Comparison of hierarchical and direct algorithm. The dashed line represents $\mathcal{O}(n^{-2})$. Approximating of $f(x,y) = (x+y)^2/2$ (upper-left), $f(x,y,z) = (x+y+z)^2/3$ (lower-left), $f(x,y) = (x+y)^{2/3}$ (upper-right), and $f(x) := (\sum_{i=1}^{10} x_i)^2/10$ (lower-right) on the respective unit cubes.

Numerical examples. We compare the (classical) direct training with our hierarchical approach. The direct approaches receive $4 \cdot 10^5$ training epochs, where the hierarchical approach receives $2 \cdot 10^3$ epochs per iteration of the outer loop of which there are not more than 10^2 in each example.

REFERENCES

[1] P. Binev, W. Dahmen, R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math. **97** (2004), 219–268.
[2] J. M. Cascon, C. Kreuzer, R. H. Nochetto, K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. **46** (2008), 2524–2550.
[3] C. Carstensen, M. Feischl, M. Page, D. Praetorius, *Axioms of adaptivity*, Comput. Math. Appl. **67** (2014), 1195–1253.
[4] OpenAI et al., *Gpt-4 technical report*, arXiv:2303.08774 (2024).
[5] M. Feischl, A. Rieder, F. Zehetgruber, *Towards optimal hierarchical training of neural networks*, arXiv:2407.02242 (2024).
[6] P. Grohs, F. Voigtlaender, *Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces*, Found. Comput. Math., 2023.

[7] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), 245–269.
[8] B. Zoph, Q. V. Le, *Neural architecture search with reinforcement learning*, arXiv:1611.01578 (2017).

Multiscale Finite Element Methods for reaction-diffusion problems

Frederic Legoll
(joint work with Claude Le Bris, Albéric Lefort)

The Multiscale Finite Element Method (MsFEM) is a finite element approach that allows to solve partial differential equations with highly oscillatory coefficients on a coarse mesh, i.e. a mesh with elements of size much larger than the characteristic scale of the heterogeneities (see [7] and also [8, 4]). To do so, MsFEMs use pre-computed basis functions, adapted to the differential operator, thereby taking into account the small scales of the problem.

In this talk, we consider reaction-diffusion equations with oscillating coefficients. The problem is phrased in terms of an eigenvalue problem, and is motivated by applications in neutronics. It typically consists in finding the lowest eigenvalue λ^ε, along with the associated eigenvector u^ε (that we take of unit L^2 norm), to the problem

$$(1) \qquad \frac{1}{\varepsilon^2}\,\sigma^\varepsilon(x)u^\varepsilon - \operatorname{div}\left(A^\varepsilon(x)\,\nabla u^\varepsilon\right) = \frac{\lambda^\varepsilon}{\varepsilon^2}\,u^\varepsilon \ \text{ in } \Omega, \qquad u^\varepsilon = 0 \ \text{ on } \partial\Omega,$$

where Ω is a bounded domain of $\mathbb{R}^d$, σ^ε is a positive and uniformly bounded function and A^ε is a uniformly elliptic, matrix-valued field which is such that $A^\varepsilon(x)$ is a symmetric matrix for any $x \in \Omega$ (a typical case is when $\sigma^\varepsilon(x) = \sigma_{\rm per}(x/\varepsilon)$ for a periodic, positive and bounded function $\sigma_{\rm per}$, and $A^\varepsilon(x) = A_{\rm per}(x/\varepsilon)$ for a periodic, uniformly elliptic, matrix-valued field $A_{\rm per}$). The equation of interest may be scalar-valued, as in (1) (in neutronics, this corresponds to the case when neutrons all have the same energy), or vector-valued (when several energy groups are represented). In the latter case, the operator at hand is in general not self-adjoint. In both cases, the reaction term magnitude is large, which implies that this term and the diffusion term both contribute to the corrector equation.

Typical solutions of (1) are much more oscillatory than those of a purely diffusive problem. For instance, in the periodic, scalar-valued case, the lowest eigenmode is bounded in $L^2(\Omega)$ but not in $H^1(\Omega)$, and we have, at the leading order, $u^\varepsilon(x) \approx \psi(x/\varepsilon)\,v^\star(x)$ for two functions ψ and $v^\star$ independent of ε. A typical representation of u^ε is shown on Figure 1.

We make use of theoretical homogenization results in a periodic framework (see e.g. [1, 2]) to guide our intuition in order to define appropriate MsFEM basis functions. A specific attention is devoted in order to make the basis functions as insensitive as possible to the boundary conditions imposed on the local problems. To that aim, we make use of filtering approaches, initially introduced in [5, 6] in the context of molecular dynamics, and next used in [3] in the context of periodic

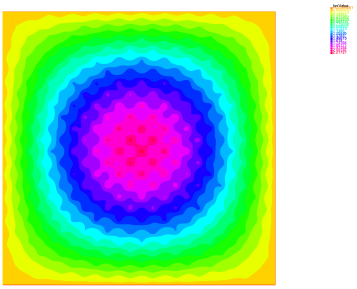


FIGURE 1. A typical solution to (1) in $\Omega = (0,1)^2$. Note the clearly visible oscillations of u^ε.

homogenization of elliptic equations. We present several numerical results, both in the scalar- and the vector-valued case, for periodic and non-periodic coefficients, which illustrate the efficiency of the approach. We also provide some rigorous error estimates, demonstrating the accuracy of the approach. We refer to [9] for more details.

REFERENCES

[1] G. Allaire, Y. Capdeboscq, *Homogenization of a spectral problem in neutronic multigroup diffusion*, Comput. Methods Appl. Mech. Engrg. **187** (2000), 91–117.

[2] G. Allaire, F. Malige, *Analyse asymptotique spectrale d'un problème de diffusion neutronique [Spectral asymptotic analysis of a neutronic diffusion problem]*, C. R. Acad. Sci. Paris, Série I, 324:939–944, 1997.

[3] X. Blanc, C. Le Bris, *Improving on computation of homogenized coefficients in the periodic and quasi-periodic settings*, Netw. Heterog. Media **5** (2010), 1–29.

[4] X. Blanc, C. Le Bris, *Homogenization theory for multiscale problems: an introduction*, MS&A (Modeling, Simulation and Applications), vol. 21, Springer, 2023.

[5] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, G. Turinici, *High-order averaging schemes with error bounds for thermodynamical properties calculations by molecular dynamics simulations*, J. Chem. Phys. **121** (2004), 10346–10355.

[6] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, G. Turinici, *Long-time averaging for integrable Hamiltonian dynamics*, Numer. Math. **100** (2005), 211–232.

[7] Y. Efendiev, T.Y. Hou, *Multiscale Finite Element Methods*, Surveys and Tutorials in the Applied Mathematical Sciences, vol. 4, Springer New York, 2009.

[8] C. Le Bris, F. Legoll, *Examples of computational approaches for elliptic, possibly multiscale PDEs with random inputs*, J. Comput. Physics **328** (2017), 455–473.

[9] A. Lefort, PhD thesis, Ecole Nationale des Ponts et Chaussées, in preparation (defense expected Fall 2025).

Approximation of multiscale stochastic partial differential equations

Annika Lang
(joint work with Per Ljung, Axel Målqvist)

Let us consider the stochastic partial differential equation

$$\partial_t X(t,x) - \nabla \cdot A(x)\nabla X(t,x) = \dot{W}(t,x), \quad X(0,x) = X_0(x), \tag{1}$$

on a sufficiently smooth domain D with a rapidly varying coefficient A and driven by a Wiener process W with covariance Q. Then the $L^2(D)$-valued noise has a basis expansion, also known as Karhunen–Loève expansion, given by

$$W(t,x) = \sum_{i=1}^{\infty} \sqrt{\gamma_i}\, \beta_i(t)\, e_i(x),$$

where $(e_i, i \in \mathbb{N})$ is an orthonormal of $L^2(D)$ and eigenbasis of Q such that $Qe_i = \gamma_i e_i$, and $(\beta_i, i \in \mathbb{N})$ denotes a sequence of independent real-valued Brownian motions. Typically, this noise is approximated by the spectral truncation

$$W^{(L)}(t,x) = \sum_{i=1}^{L} \sqrt{\gamma_i}\, \beta_i(t)\, e_i(x),$$

and the speed of convergence depends on the spatial smoothness of W which depends on the algebraic decay rate of the eigenvalues $(\gamma_i, i \in \mathbb{N})$ of Q.

Let us set $\Lambda = -\nabla \cdot A\nabla$ for abbreviation, then (1) is interpreted as the integral equation

$$X(t) = X_0 - \int_0^t \Lambda X(s)\,\mathrm{d}s + W(t)$$

on $L^2(D)$, which we approximate by a Galerkin finite element (FEM) approximation with mesh width h in space and a backward Euler scheme with step size k in time given by

$$(1 + k\Lambda_h)X_h^n = X_h^{n-1} + P_h(W(t_n) - W(t_{n-1})). \tag{2}$$

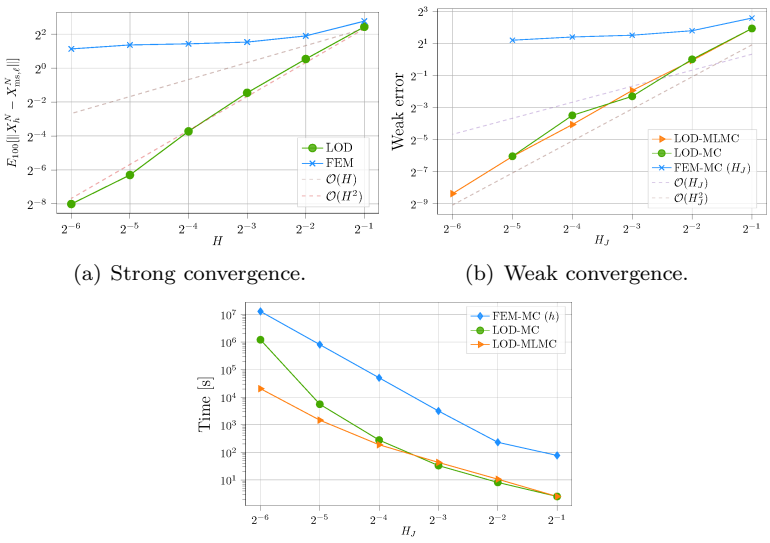
We denote the finite element space of first order polynomials by V_h and the $L^2(D)$-orthogonal projection on V_h by P_h. It is known, see, e.g., [5, 10, 11], that the approximation converges for $\mu \in [1,2]$ asymptotically with strong error bound

$$\mathbb{E}\big[\|X_h^n - X(t_n)\|_{L^2(D)}^2\big]^{1/2} \lesssim (k^{\mu/2} + h^{\mu})\big(\|X_0\|_{\dot{H}^{\mu}(D)} + \|\Lambda^{(\mu-1)/2}Q^{1/2}\|_{\mathrm{HS}}\big), \tag{3}$$

where "HS" denotes the Hilbert–Schmidt norm and $\dot{H}^{\mu}(D)$ is defined by fractional powers of Λ. For all Lipschitz test functions, this yields weak convergence, since

$$\big|\mathbb{E}[\varphi(X_h^n) - \varphi(X(t_n))]\big| \lesssim \mathbb{E}\big[\|X_h^n - X(t_n)\|_{L^2(D)}^2\big]^{1/2}.$$

In simulations, these convergence rates can only be observed when the chosen mesh width h resolves the fine oscillations of A but fails to be visible on coarse grids as can be seen in Figures 1(a) and 1(b). To overcome this bottleneck, we modify our finite element space such that the resulting space is suitable for highly oscillating coefficients even on coarse grids. We use the localized orthogonal decomposition



(a) Strong convergence.

(b) Weak convergence.

(c) Running time.

FIGURE 1. FEM and LOD error simulations and running times.

(LOD) method as introduced in [9] and obtain the multiscale space V_ℓ^{ms} with patch size N_ℓ by modification of the finite element space V_H on the coarse grid with mesh size H. Denoting by $\Lambda_\ell^{\mathrm{ms}}$ the Galerkin approximation of Λ and by $P_{\ell,h}^{\mathrm{ms}}$ the orthogonal projection, the fully discrete approximation scheme reads

$$(1 + k\Lambda_\ell^{\mathrm{ms}})X_{\mathrm{ms},\ell}^n = X_{\mathrm{ms},\ell}^{n-1} + P_{\ell,h}^{\mathrm{ms}}(W(t_n) - W(t_{n-1})). \tag{4}$$

In [6], we showed that the same convergence rates on coarse grids can be obtained as asymptotically with standard FEM approximations. This is confirmed in simulations, see Figures 1(a) and 1(b). Furthermore, we get with suitably chosen sample sizes the same rate of weak convergence with a multilevel Monte Carlo method as pioneered in [2], which is computationally cheaper as observed in Figure 1(c).

In the suggested scheme (4), we need to compute the increments of W and project them onto V_ℓ^{ms}, i.e., we generate

$$P_{\ell,h}^{\mathrm{ms}}(W(t_n) - W(t_{n-1})) = \sqrt{k}P_{\ell,h}^{\mathrm{ms}}U,$$

where $U \sim \mathcal{N}(0, Q)$. In practice, we used the truncated process $W^{(L)}$ for a sufficiently large L and computed the projections. This is not very convenient. As a possible idea for future research, I introduce an idea for FEM, where we generate U for special choices of Q directly on V_h using discrete white noise. More specifically, we look at generalized Whittle–Matérn covariances that have become popular in spatial statistics in the stationary setting, see [8]. These are given as

solutions to special stochastic partial differential equations of the form

$$U = \gamma(\mathcal{L})\mathcal{W},$$

where $\mathcal{L} = V - \nabla\mathcal{D}\nabla$, γ is a sufficiently smooth power spectral density, $\mathcal{W}$ spatial white noise, and V and $\mathcal{D}$ are a sufficiently smooth potential and diffusion matrix, respectively. These solutions can be approximated by finite element approximations and either a quadrature approximation of the Dunford–Taylor representation in case γ is of the form $x^{-\beta}$ or a Chebyshev approximation, else. Convergence results on spheres and hypersurfaces with surface FEM approximations were presented from [3, 4].

If we now combine such results with (2), i.e., using the finite element solution to $\mathcal{L}^{\beta}W(t) = \mathcal{W}(t)$ instead of P_hW, we obtain similar convergence results as in (3) as shown in [1].

Finally, we looked at the transformation used in [7] to use SPDE solutions to generate stochastic manifolds to give an outlook into future research.

References

[1] Ø. S. Auestad, G.-A. Fuglstad, E. R. Jakobsen, A. Lang, *Finite element approximation of parabolic SPDEs with Whittle–Matérn noise*, arXiv:2406.11041, 2025.

[2] S. Heinrich, *Multilevel Monte Carlo methods*, in *Large-Scale Scientific Computing*, volume 2179 in *Lecture Notes in Computer Science*, Berlin:Springer, 58–67, 2001.

[3] E. Jansson, M. Kovács, A. Lang, *Surface finite element approximation of spherical Whittle–Matérn Gaussian random fields*, SIAM J. on Sci. Comput. **44** (2022), A825–A842.

[4] E. Jansson, A. Lang, M. Pereira, *Non-stationary Gaussian random fields on hypersurfaces: sampling and strong error analysis*, arXiv:2406.08185, 2024.

[5] R. Kruse, *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, volume 2093 of *Lecture Notes in Mathematics*, Springer Cham, 2014.

[6] A. Lang, P. Ljung, A. Målqvist, *Localized orthogonal decomposition for a multiscale parabolic stochastic partial differential equation*, Multiscale Model. Simul. **22** (2024), 204–229.

[7] A. Lang, Ch. Schwab, *Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations*, Ann. Appl. Probab. **25** (2015), 3047–3094.

[8] F. Lindgren, H. Rue, J. Lindström, *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach*, J. R. Stat. Soc., Ser. B, Stat. Methodol. **73** (2011), 423–498.

[9] A. Målqvist, D. Peterseim. *Localization of elliptic multiscale problems*, Math. Comp. **83** (2014), 2583–2603.

[10] Y. Yan, *Semidiscrete Galerkin approximation for a linear stochastic parabolic partial differential equation driven by an additive noise*, BIT Numer. Math. **44** (2004), 829–847.

[11] Y. Yan, *Galerkin finite element methods for stochastic parabolic partial differential equations*, SIAM J. Numer. Anal. **43** (2005), 1363–1384.

Linearized Localized Orthogonal Decomposition for Quasilinear Nonmonotone Elliptic PDE

Maher Khrais
(joint work with Barbara Verfürth)

In this work we present a multiscale method in the framework of the Local Orthogonal Decomposition (LOD) for solving nonlinear nonmonotone elliptic PDEs. Consider the following prototypical nonlinear nonmonotone model, find $u \in H_0^1(\Omega)$ that solves

$$\mathcal{A}(u; u, v) := (\alpha(x, u)\nabla u, \nabla v) = (f, v) = F(v), \quad \forall v \in H_0^1(\Omega).$$

Where $f \in L^2(\Omega)$ and $\Omega \subset \mathbb{R}^d$ is a bounded domain with $d \leq 3$. The coefficient $\alpha \in L^\infty(\Omega; \mathbb{R}^d)$ satisfies certain assumptions, including Lipschitz continuity, ellipticity, and boundedness. The presented material is based on the detailed manuscript [3]. The goal of the LOD method is to incorporate the (spatial) fine-scale behavior of the coefficient α into the basis of the coarse FE space resulting in a new modified low-dimensional function space with good approximation properties. However, these arguments cannot easily be transferred to the nonlinear case. For quasilinear problems of monotone type, a multiscale method was proposed and analyzed in [1] which uses a linearization of the problem to compute the correction operators.

1. Multiscale method and linearization techniques

We focus on the construction of the multiscale space and, in particular, on the linearization techniques involved for our nonmonotone PDE. Consider a decomposition of the domain of interest Ω into a partition $\mathcal{T}_H$. Assume $H := \max_{K \in \mathcal{T}_H} H_K$, and let V_H be the standard lowest-order conforming finite element subspace of $H_0^1(\Omega)$. Let $I_H : H_0^1(\Omega) \to V_H$ be a bounded local linear projective operator that satisfies H^1 stability and the approximation property. The fine-scale incorporation into the coarse FE space is obtained by using correction operators. Given $W = \ker I_H$, we define the corrector operator

$$Q : V_H \to W$$

that satisfies the following linear elliptic problem

$$\mathcal{A}_L(p^*, u_H - Qu_H, w) = (A_L(x, p^*, \nabla(u_H - Qu_H)), \nabla w) = 0, \ \forall w \in W.$$

For a suitable choice of linearization point $p^* \in H_0^1(\Omega)$, two linearization techniques are considered to convert the corrector problem into a linear elliptic problem:

- The first technique is known as Kačanov-type linearization, where we freeze the nonlinearity of $\alpha(x, v)\nabla v$ to get the following linear form $A_L(x, p^*, \nabla v) = \alpha(x, p^*)\nabla v$.
- Second, Fréchet derivative linearization in the direction of $(u - p^*)$ to obtain the linear approximation form $\alpha(x, v)\nabla v \approx A_L(x, p^*, \nabla v) = \alpha(x, p^*)\nabla v + (v - p^*)\alpha_u(x, p^*)\nabla p^*$.

The correction operator is then computed in a localized fashion on small patches of coarse mesh elements with oversampling parameter k. The multiscale space is defined as

$$V_{H,k} = (id - Q_k)V_H = \{v_H - Q_k v_H : v_H \in V_H\}.$$

This space is constructed only once and it is then employed to solve the nonlinear multiscale problem with a Galerkin approach, which in practice of course uses an iterative method.

2. Error Analysis and numerical example

The nonmonotone nature requires several changes in the error analysis in comparison to [1]. Our main contribution is thus to extend the ideas from [1] to the problem of interest above, which is a non-trivial task.

Theorem 1. *Let u be the analytical solution, and let $u_{H,k}$ be the multiscale solution. Then for sufficiently large patch, it holds that*

$$|u - u_{H,k}|_1 \lesssim H\|f\|_0 + G(u - p^*, u_{H,k} - p^*),$$

where $G(u - p^*, u_{H,k} - p^*)$ satisfies the following inequalities:

(1) *Kačanov-type linearization*

$$G(u - p^*, u_{H,k} - p^*) \lesssim \|u - p^*\|_\infty \|\nabla u\|_0 + \|\nabla u_{H,k}\|_0 \|u_{H,k} - p^*\|_\infty.$$

(2) *Fréchet derivative linearization*

$$G(u - p^*, u_{H,k} - p^*) \lesssim \|u - p^*\|_\infty \|\nabla u\|_0 + \|\nabla u_{H,k}\|_0 \|u_{H,k} - p^*\|_\infty + \|(\alpha_{ss}(x, q^*)\nabla q^*)\|_\infty \|u - p^*\|_0^2,$$

where $q^* \in H_0^1(\Omega)$ is an intermediate function, i.e., $q^* = p^* + t(u - p^*)$, for some $t \in [0, 1]$.

Our a priori error estimate is of optimal order with respect to the mesh size without any dependence on the spatial variations of α or (higher) regularity of the exact solution. The estimates for the linearization error gives guidance for the choice of the linearization point for the correction problems.

Practically, we consider the so-called exponential model $\alpha(x, u) = c(x)\exp(2u)$ in $\Omega = [0, 1]^2$ with heterogeneous spatial term $c(x)$. The right-hand side $f(x_1, x_2)$ is defined as 0.1 for $x_2 \leq 0.1$ and 1 otherwise. For further numerical experiments we refer to [3]. To study the convergence performance, we use the following relative errors $e_{\text{LOD}} := \frac{|u_h - u_{H,k}|_1}{|u_h|_1}$, u_h is a reference solution solved on $h = 2^{-7}$.

We compare the behavior of the convergence for different linearization techniques and several linearization points:

$$\begin{cases} u_H, u_h & \text{FEM solutions, } H = \frac{1}{32}, \text{ and } h = \frac{1}{128}, \\ \text{ulod} & \text{LOD solution } H = \frac{1}{16} \text{ and } k = 4, \\ g(x) = 10xy(1-x)y(1-y), & \\ g_1(x) = 0.5xy(1-x)y(1-y)e^{(5(x+y))}. & \end{cases}$$

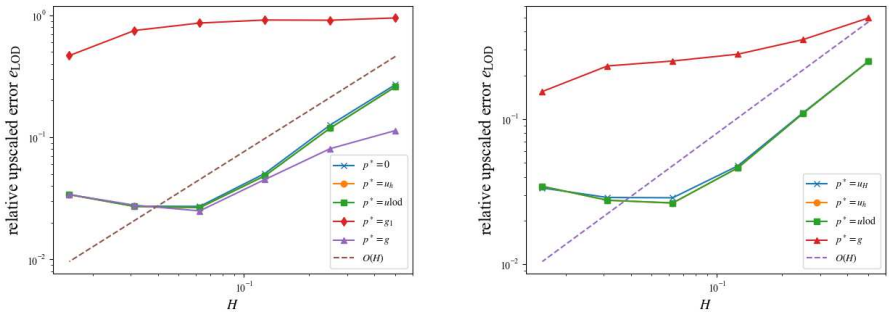


FIGURE 1. Convergence history of e_{LOD} for Kačanov method (left) and Fréchet method (right) with different linearization points and fixed oversampling parameter $k = 3$.

The numerical experiments illustrate the impact of the choice of the linearization points on the performance of the method. Some linearization points perform better than others, depending on how close they are to the analytical solution. For example, the linearization points $p^* \in \{0, u_H, u_h, \mathrm{ulod}\}$ performs almost the same as they are closed to u and $u_{H,k}$. However, the linearization points that are not close to both u and $u_{H,k}$, we notice that the first order convergence is lost due to the dominance of linearization error. The two linearization techniques especially differ when the linearization point is far from the solution. The linearization technique that depends on Fréchet derivative exhibits greater sensitivity to the chosen linearization point. An iteration of our multiscale method by updating the linearization point improves the performance considerably. This indicates that iterative multiscale methods may also be promising for nonmonotone quasilinear problems, which is left for future research.

REFERENCES

[1] B. Verfürth, *Numerical homogenization for nonlinear strongly monotone problems*, IMA J. Numer. Anal., **42** (2022), 1313–1338.

[2] A. Målqvist, D. Peterseim, *Numerical homogenization by localized orthogonal decomposition*, Vol. 5. SIAM Spotlights. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2021.

[3] M. Khrais, B. Verfürth, *Linearized Localized Orthogonal Decomposition for Quasilinear Nonmonotone Elliptic PDE*, arXiv:2502.13831 (2025).

Optimal artificial boundary conditions for random media

Felix Otto
(joint work with Jianfeng Lu, Lihan Wang)

Given a uniformly elliptic coefficient field $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$

$$\xi.a\xi \geq \lambda|\xi|^2 \quad \text{and} \quad \xi.a\xi \geq |a\xi|^2 \qquad \text{for all } \xi \in \mathbb{R}^d,$$

we will consider the solution u of the equation

$$\nabla.a\nabla u = \nabla.g \qquad \text{in } \mathbb{R}^d, \tag{1}$$

where the right-hand side g is a vector field that is localized at a scale l, i.e.

$$g(x) = \hat{g}(\frac{x}{l}) \qquad \text{with } \hat{g} \in C_0^2(B_1) \text{ and } |\int \hat{g}| = 1.$$

We are interested in observables of the form $\int \omega\nabla u$ where the average is taken on a scale R, namely

$$\omega(x) = \frac{1}{R^d}\hat{\omega}(\frac{x}{R}) \qquad \text{with } \hat{\omega} \in C_0^2(B_1) \text{ and } \int \hat{\omega} = 1.$$

The topic is on how well this observable can be predicted if the medium a is only known on a large ball B_L. Without any further structure, $\int \omega\nabla u$ is determined by $a|_{B_L}$ only up to $O((\frac{l}{L})^d)$ in the regime $R \ll l \ll L$, where this scaling arises from the dipolar decay of u from the system size L and the source size l. This scaling is saturated by imposing homogeneous Dirichlet boundary conditions:

$$\int \omega\nabla u = \int \omega\nabla\tilde{u} + O\big((\frac{l}{L})^d\big)$$

where $\tilde{u}$ solves the homogeneous Dirichlet problem

$$\nabla.a\nabla\tilde{u} = \nabla.g \quad \text{in } B_L \qquad \text{with } \tilde{u} = 0 \text{ on } \partial B_L.$$

In certain situations one can do better, namely if one has the a priori information that a is drawn from an ensemble (=probability measure) that

- is stationary, i.e. $a(\cdot + y) \overset{\text{law}}{=} a$ for all $y \in \mathbb{R}^d$,
- has unit range, i.e $a|_D$ and $a|_{D'}$ are independent for $\text{dist}(D, D') \geq 1$.

Loosely speaking, [2, 3] establishes that for $1 \ll R \ll l \ll L$ and with overwhelming probability

$$\int \omega\nabla u \quad \text{is determined by } a|_{B_L} \quad \text{up to } O\big((\frac{l}{L})^d(\frac{1}{L})^{\frac{d}{2}}\big),$$

yielding an improvement by the central limit theorem-type factor $(\frac{1}{L})^{\frac{d}{2}}$, involving the system size L and the range 1. More precisely, we have that

$$\int \omega\nabla u = \int \omega\nabla\tilde{u} + O\big((\frac{l}{L})^d(\frac{1}{L})^{\frac{d}{2}-}\big)$$

where $\tilde{u}$ solves the inhomogeneous Dirichlet problem

$$\nabla.a\nabla\tilde{u} = \nabla.g \quad \text{in } B_{\frac{L}{2}} \qquad \text{with } \tilde{u} = \tilde{u}^D \text{ on } \partial B_{\frac{L}{2}} \tag{2}$$

for some computable Dirichlet boundary data $\tilde{u}^D$, see below.

This is the best that we can do information-theoretically:

Theorem 1. [2, Theorem 2] *For "most" ensembles as above there exists $R \sim 1$ such that*

$$\mathbb{E}^{\frac{1}{2}}\Big|\int \omega\nabla u - \mathbb{E}\big[\int \omega\nabla u \,|\, a_{|B_L}\big]\Big|^2 \gtrsim (\frac{l}{L})^d(\frac{1}{L})^{\frac{d}{2}} \qquad \textit{for all } 1 \ll l \ll L.$$

Here $\mathbb{E}[\cdot|a_{|B_L}]$ denotes the expectation conditioned on the restriction of a onto B_L. For a generic class of ensembles, Theorem 1 is a consequence of the following sensitivity result, which bounds from below how much the observable depends on the medium in some ball $B_R(y)$.

Theorem 1'. [2, Lemma 7] *For "most" ensembles as above there exists $R \sim 1$ such that*

$$\mathbb{E}^{\frac{1}{2}}\big|\mathbb{E}\big[\int \omega\nabla u \,|\, a_{|B_R(y)}\big] - \mathbb{E}\big[\int \omega\nabla u\big]\big|^2 \gtrsim (\frac{l}{|y|})^d(\frac{R}{|y|})^d \qquad \textit{for all } 1 \ll l \ll |y|.$$

In the sequel, we sketch the two ingredients for Theorem 1'.

General homogenization theory (cf. [4]). To a coefficient field $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, one associates a constant coefficient matrix $\bar{a} \in \mathbb{R}^{d\times d}$ and fields

$$\phi^i : \mathbb{R}^d \to \mathbb{R}, \qquad \sigma^i : \mathbb{R}^d \to \mathbb{R}^{d\times d} \text{ skew symmetric}$$

for $i = 1, \dots, d$, such that the following Helmholtz-type decomposition holds

$$a(e^i + \nabla\phi^i) = \bar{a}e^i + \nabla.\sigma^i.$$

The correctors ϕ^i and flux correctors σ^i yield an intertwining relation between the heterogeneous operator $\nabla.a\nabla$ and the homogeneous $\nabla.\bar{a}\nabla$ one:

$$\nabla.a\nabla(1 + \phi^i\partial_i)\bar{u} = \nabla.\bar{a}\nabla\bar{u} + \nabla.(\phi^i a + \sigma^i)\nabla\partial_i\bar{u},$$

via what is known as the two scale-expansion[1] $(1 + \phi^i\partial_i)\bar{u}$, and with the error in divergence form. In order for the error $(\phi^i a + \sigma^i)\nabla\partial_i\bar{u}$ to be of higher order, the potentials ϕ^i and σ^i should grow only sublinearly in x.

Fixing an exponent β with $0 < 1 - \beta \ll 1$, one monitors this sublinearity around a point z in terms of the minimal radius $r_*(z) \in (0, \infty]$ satisfying

$$\frac{1}{r}\Big(\fint_{B_r(z)} \sum_{i=1}^{d} (\phi^i - \fint_{B_r(z)} \phi^i)^2 + |\sigma^i - \fint_{B_r(z)} \sigma^i|^2\Big)^{\frac{1}{2}} \le (\frac{r_*(z)}{r})^\beta \quad \text{for all } r \ge r_*(z).$$

For the class of ensembles described above, it is known that $r_*(z) \lesssim 1$ with overwhelming probability, [2, Lemma 4].

[1] where we use Einstein's summation convention of summation over repeated indices.

Homogenization and defects (cf. [1]). Recall the elliptic problem in (1), $\nabla.a\nabla u = \nabla.g$ and consider its homogenized variant $\nabla.\bar{a}\nabla\bar{u} = \nabla.g$. Given a uniformly elliptic coefficient field a_0, we insert a_0 as defect in $B_R(y)$:

$$a' = a_0 \text{ in } B_R(y) \qquad \text{and} \qquad a' = a \text{ else},$$
$$\bar{a}' = a_0 \text{ in } B_R(y) \qquad \text{and} \qquad \bar{a}' = \bar{a} \text{ else}$$

and consider the corresponding elliptic equations

$$\nabla.a'\nabla u' = \nabla.g \quad \text{and} \quad \nabla.\bar{a}'\nabla\bar{u}' = \nabla.g.$$

The following proposition, which shows that the insertion of defects and homogenization commute, is the main ingredient in the proof of Theorem 1'.

Proposition 1. [2, Proposition 2] *There exist constants $\alpha(d,\lambda,\beta) > 0, C(d,\lambda,\beta) < \infty$ such that for all l, R, y with $r_*(0) \le l$ and $max\{r_*(0), r_*(y)\} \le R \le \frac{|y|}{2}$, we have*

$$\Big(\fint_{B_R} \big|\nabla\big((u'-u) - (1+\phi^i\partial_i)(\bar{u}'-\bar{u})\big)\big|^2\Big)^{\frac{1}{2}} \le C\big(\frac{l}{|y|}\big)^d\big(\frac{R}{|y|}\big)^d\Big(\big(\frac{R}{|y|}\big)^\beta + \big(\frac{r_*(y)}{R}\big)^\alpha\Big).$$

Algorithm for artificial boundary data $\tilde{u}^D$. Recall the initial problem in (1) and its homogenized approximation

$$\nabla.a\nabla u = \nabla.g, \qquad \nabla.a\nabla\bar{u} = \nabla.g \qquad \text{in } \mathbb{R}^d,$$

together with the local problem in (2)

$$\nabla.a\nabla\tilde{u} = \nabla.g \quad \text{in } B_{\frac{L}{2}} \qquad \text{with } \tilde{u} = \tilde{u}^D \text{ on } \partial B_{\frac{L}{2}}.$$

For $d = 2$, we introduce the two-scale expansion of $\bar{u}$, corrected by a dipolar term

$$u^D = (1+\phi^i\partial_i)\bar{u} + \Big(\int g.\nabla\phi^i\Big)\partial_i\bar{G} \quad \text{on } \partial B_{\frac{L}{2}}, \tag{3}$$

where $\bar{G}$ is the fundamental solution of $\nabla.\bar{a}\nabla$. Since we have no access to ϕ^i and $\bar{a}$, we use as proxy the solution $\tilde{\phi}^i$ of

$$\nabla.a(e^i + \nabla\tilde{\phi}^i) = 0 \text{ in } B_L \quad \text{and} \quad \tilde{\phi}^i = 0 \text{ on } \partial B_L$$

and the constant matrix $\widetilde{\bar{a}}e^i = \int \omega_L a(e^i + \nabla\tilde{\phi}^i)$. With the proxies $\tilde{\phi}^i$ and $\widetilde{\bar{a}}$ at hand, we define $\tilde{u}^D$ as in (3).

Theorem 2. [2, Theorem 1] *For $d = 2$ there exist constants $C(\lambda,\beta) < \infty, \gamma(\lambda,\beta) > 0$ such that for all $\tilde{r}_*(0) \le R \le L$*

$$\Big|\int \omega\nabla u - \int \omega\nabla\tilde{u}\Big| \lesssim \Big(\fint_{B_R} |\nabla(u-\tilde{u})|^2\Big)^{\frac{1}{2}} \le C\big(\frac{l}{L}\big)^d\big(\frac{\tilde{r}_*(0)}{L}\big)^{\beta=\frac{d}{2}-}$$

with probability $1 - \exp(-\frac{L^\gamma}{C})$, and where $\tilde{r}_*(0)$ is determined by $(\tilde{\phi}^i, \tilde{\sigma}^i)$ as $r_*(0)$ by (ϕ^i, σ^i).

In [3, Theorem 1.2], the construction of $\tilde{u}^D$ has been extended to $d = 3$, also achieving the near-optimal rate:

Corollary 2. *For $d = 2, 3$ and $\beta < \frac{d}{2}$, there exists a finite constant $C(d, \lambda, \beta)$ such that for ensembles as above and for all $1 \sim R \ll l \ll L$*

$$\mathbb{E}^{\frac{1}{2}} \Big| \int \omega \nabla u - \int \omega \nabla \tilde{u} \Big|^2 \leq C(\frac{l}{L})^d (\frac{1}{L})^\beta.$$

References

[1] X. Blanc, C. Le Bris, P.L. Lions, *Local profiles for elliptic problems at different scales: defects in, and interfaces between periodic structures*, Comm. Partial Differential Equations **40** (2015), 2173–2236.

[2] J. Lu, F. Otto, *Optimal artificial boundary condition for random elliptic media*, Found. Comput. Math. **21** (2021), 1643–1702.

[3] J. Lu, F. Otto, L. Wang, *Optimal artificial boundary conditions based on second-order correctors for three dimensional random elliptic media*, Comm. Partial Differential Equations **49** (2024), 609–670.

[4] L. Tartar, *The general theory of homogenization: a personalized introduction*, Springer Science & Business Media (2009).

LOD for heterogeneous Stokes equations

Alexei Lozinski

(joint work with Moritz Hauck)

We present the work from [1] on a multiscale method for a heterogeneous Stokes problem in a bounded polytope $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$

$$\begin{cases} -\operatorname{div}(\nu \nabla u) + \sigma u + \nabla p = f & \text{in } \Omega, \\ \operatorname{div} u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

where the coefficients ν and σ are non-constant and can be highly oscilating. The method is based on the Localized Orthogonal Decomposition (LOD) methodology and has approximation properties independent of the regularity of these coefficients.

The basic idea of the LOD is to decompose the solution space into a fine-scale space and its orthogonal complement with respect to the energy inner product induced by the considered problem. By choosing the fine-scale space to consist of functions that average out on coarse scales, one obtains a finite-dimensional mesh-based complement space that is adapted to the problem at hand and has uniform approximation properties under minimal structural assumptions on the coefficients. It possesses exponentially decaying basis functions whose computation can thus be localized to subdomains, resulting in a practically feasible method, provided the basis functions are pre-computed on sufficiently fine meshes on these subdomains.

For Stokes problems, the divergence-free constraint poses a major challenge to LOD-type methods, since directly incorporating the constraint can lead to ill-posed problems or slowly decaying basis functions. To overcome this problem, we

reformulate the Stokes problem relaxing the divergence-free constraint, and working rather in the space of $H_0^1(\Omega)^n$-functions with piecewise constant divergence with respect to the coarse mesh. The divergence-free velocity is then recovered using a piecewise constant Lagrange multiplier defined on the same mesh. We choose the fine-scale space for the velocity as the functions whose averages vanish on all faces of the coarse mesh, and consider its finite-dimensional orthogonal complement as the prototypical LOD approximation space. This choice assures that the natural projection on the prototypical LOD space preserves the divergence, so that the approximation given by the resulting LOD method (with piecewise constant Lagrange multipliers) is exactly divergence-free and thus pressure robust. Our construction of the fine-scale space is inspired by the concept of "quantities of interest" in [2] and by Operator-adapted wavelets in [3].

The *a priori* error analysis of the proposed method reveals optimal orders of convergence. More specifically, for L^2-right-hand sides we prove first- and second-order convergence for the L^2- and H^1-errors of the velocity approximation. If the right-hand side f is H^1-regular, we can squeeze out an additional order of convergence for the velocity. We also prove the exponential decay of the basis functions, thus allowing for a feasible LOD method with localized basis functions that should be computed on the subdomains whose size should increase logarithmically with the desired accuracy. The error of this method is fully analyzed theoretically and its convergence is demonstrated by numerical experiments. A first-order approximation for the pressure can be also easily recovered by post-processing.

We have also briefly outlined the ongoing work in two directions: (i) a better localization strategy removing the $1/H$ prefactor from the localization error estimate, at the expense of more expensive calculation of the basis functions; (ii) extension to higher convergence orders by enriching the approximation spaces using the polynomials weights on edges and elements of the mesh.

References

[1] M. Hauck, A. Lozinski, *A Localized Orthogonal Decomposition Method for Heterogeneous Stokes Problems*, arXiv:2410.14514 (2024).

[2] R. Altmann, P. Henning, D. Peterseim, *Numerical homogenization beyond scale separation*, Acta Numer. **30** (2021), 1–86.

[3] M. Budninskiy, H. Owhadi, M. Desbrun, *Operator-adapted wavelets for finite-element differential forms*, J. Comp. Phys. **388** (2019), 144–177.

Multiscale Hamiltonian simulations via deep learning and parallel-in-time methods

Richard Tsai
(joint work with Rui Fang)

Hamiltonian systems with multiple timescales arise in molecular dynamics, classical mechanics, theoretical physics, etc. These systems are governed by

$$\dot{u}(t) = f(u(t)) := J^{-1}\nabla H(u(t)), \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad u = (p, q) \in \mathbb{R}^{2d},$$

and conserve energy and symplectic structure. Stable and accurate long-time numerical integration of these systems demands very small time steps to resolve the fastest dynamics, resulting in substantial computational costs. Moreover, many practical applications involve simulating ensembles of trajectories for uncertainty quantification, sensitivity analyses, or varied initial conditions, further intensifying the computational demands. This work introduces two complementary strategies to accelerate simulations of multiscale Hamiltonian systems:

(i) Neural network-based flow maps $\Phi_\theta(u_0, t) \approx \phi_t(u_0)$ that approximate the solution operator directly, bypassing classical time step constraints.
(ii) A Procrustes parareal algorithm that stabilizes the standard parareal algorithm, a famous parallel-in-time method introduced in [1], for oscillatory Hamiltonian systems through data-driven phase corrections.

1. Neural network-based flow maps

We first develop a learning framework to approximate the continuous-time flow map $\phi_t : \mathbb{R}^{2d} \mapsto \mathbb{R}^{2d}$ of Hamiltonian systems using deep neural networks. We construct simple neural networks $\Phi_\theta(u_0, t)$

$$\Phi_\theta(u_0, t) \approx \phi_t(u_0), \quad u_0 \in \Omega \subset \mathbb{R}^{2d},\ t \in [0, T],$$

by minimizing a residual-based loss functional. The neural network is trained by minimizing the squared residual defined by a chosen convergent numerical scheme:

$$\mathcal{J}[\Phi_\theta] = \frac{1}{2}\int_\Omega \int_0^T \left\| R_h[\Phi_\theta](u, t) \right\|^2 d\nu(t)\, d\rho(u),$$

where ν and ρ are probability measures on $[0, T]$ and Ω respectively. For example, the forward Euler scheme with step size h:

$$R_h[\Phi_\theta](u, t) := \frac{\Phi_\theta(u, t + h) - \Phi_\theta(u, t)}{h} - f(\Phi_\theta(u, t)).$$

Compared with the exact residual approach employed in many physics-informed neural networks (PINNs), the scheme-based residual approach offers increased computational and data efficiency and enhanced robustness.

To obtain a network model that is sufficiently accurate in short time intervals and stable when used recurrently for long-time simulations, we address the following design aspects:

• **Network architecture**: Since there is no dissipation to damp out approximation errors in Hamiltonian dynamics, it is important to approximate the flow map with a smaller error at the initial time. We propose using neural networks that are consistent with the flow and its first few time derivatives at $t = 0$. For example,

$$\Psi_\theta(u,t) := \frac{\sigma(w_1 t)}{w_1}\left(f(u) + \frac{\sigma(w_2 t)}{w_2}\left(\frac{1}{2}f'(u)f(u) + \sigma(w_3 t)R_{\theta_2}(u, f(u), t)\right)\right),$$

with $\theta = \{w_1, w_2, w_3\} \cup \theta_2$ denoting the trainable parameter set and $\sigma(z) = \tanh(z)$, matches the Taylor expansion of ϕ_t up to the second order term, and R_{θ_2} is a simple multilayer perceptron approximating the remainder term. This approach provides inductive bias aligned with the system's known dynamics for small t while retaining expressivity for complex dynamical behaviors.

• **Phase space data and sampling**: Given the energy-conserving nature of Hamiltonian dynamics, training data must respect the invariant measure on energy level sets corresponding to test trajectory energies. To this end, we propose HMC-H_0, a Monte Carlo algorithm based on Hamiltonian Monte Carlo that samples phase space states from the microcanonical ensemble corresponding to a fixed energy H_0. This approach significantly outperforms other ad hoc sampling approaches, both in data efficiency and long-term stability of the learned flow maps.

• **Energy-based loss function**: Multiscale Hamiltonian systems often feature both stiff and nonstiff modes. A loss function measured with the l_2-norm may underweight fast dynamics. To address this, we construct energy-balanced norms that properly weight different contributions to the loss, improving the accuracy and robustness of the learned flow maps.

We apply our approach to several benchmark problems, including a nearly-periodic coupled oscillators problem, the Fermi-Pasta-Ulam-Tsingou problem from chaos theory, the non-integrable gravitational three-body problem, and the α-particle problem, governed by a non-separable non-canonical Hamiltonian, relevant to magnetic confinement fusion. For all problems other than the three-body problem, the neural network flow maps can achieve exceptional short-term and long-term accuracy. Remarkably, in cases with extremely large scale separations, the network maintains high accuracy and demonstrates runtime performance superior to classical symplectic solvers such as the Störmer-Verlet method. For the α-particle problem, we additionally demonstrate the flexibility of our framework in learning the dependence on the problem parameter ϵ, thereby facilitating further studies on parameter sensitivity. The non-integrable three-body problem remains a challenging test case, due to the extremely high stiffness in the dynamics during close encounters of two bodies. The learned flow maps as trained without adaptivity in the step size h used in R_h or batch samples during training, exhibited good numerical performance in terms of accuracy as long as the force is bounded below some constant (as long as the particles do not get too close to each other).

2. Stabilized parareal algorithm

As a second contribution, we develop a stabilized variant of the parareal algorithm for multiscale Hamiltonian systems. The standard parareal algorithm splits $[0, T]$ into N subintervals, $\Delta t = T/N$, and computes the solution $u_n^{(k)}$ at time $t_n = n\Delta t$ iteratively by combining a coarse solver $C_{\Delta t}$ and a fine solver $F_{\Delta t}$ via:

$$u_{n+1}^{(k+1)} = C_{\Delta t} u_n^{(k+1)} + \left(F_{\Delta t} u_n^{(k)} - C_{\Delta t} u_n^{(k)} \right),$$

with $u_{n+1}^{(0)} = C_{\Delta t} u_n^{(0)}$ and $u_0^{(k)} = u_0$. While effective for dissipative problems, the standard parareal algorithm suffers from instability and slow convergence for oscillatory systems due to phase mismatches between coarse and fine propagators (see, e.g., [2, 3, 4, 5]). Following previous works [4, 5], we stabilize the parareal algorithm for Hamiltonian systems by constructing correction operators using online-generated data during parareal iterations. Introducing the data-driven correction operator $\theta_n^{(k+1)}$ such that $\theta_n^{(k+1)} C_{\Delta t} \approx F_{\Delta t}$, we obtain the θ-parareal update

$$u_{n+1}^{(k+1)} = \theta_n^{(k+1)} C_{\Delta t} u_n^{(k+1)} + \left(F_{\Delta t} u_n^{(k)} - \theta_n^{(k+1)} C_{\Delta t} u_n^{(k)} \right).$$

To correct for phase mismatches, we restrict $\theta_n^{(k+1)}$ to phase-aligned transformations. Specifically, we define a coordinate transform Λ that isolates oscillatory modes, and determine $\theta_n^{(k+1)}$ by minimizing the discrepancy between coarse and fine solvers across a collection of previous states:

$$\theta_n^{(k)} = \Lambda^\dagger \Omega_n^{(k)} \Lambda, \quad \Omega_n^{(k)} := \operatorname{argmin}_{\Omega \in O(d)} \sum_{u \in \mathcal{D}_n^{(k)}} \|\Lambda F_{\Delta t} u - \Omega \Lambda C_{\Delta t} u\|.$$

This problem is known as the orthogonal Procrustes problem, and the solution can be efficiently computed via singular value decomposition of the correlation matrix between fine and coarse solutions.

Numerical experiments demonstrate that the proposed phase correction can successfully stabilize the parareal iterations, thereby accelerating the convergence and extending the viable time window for accurate long-term predictions. This underscores the importance of correcting phase errors in oscillatory systems.

Finally, we demonstrate that the neural network-based flow map Φ_θ can act as a surrogate coarse solver $C_{\Delta t}$ within the parareal framework, leading to further acceleration and improved performance.

References

[1] J.L. Lions, Y. Maday, G. Turinici, *A "parareal" in time discretization of PDE's*, Comptes Rendus de l'Acadmie des Sciences-Series I-Mathematics **332** (2001), 661–668.

[2] M.J. Gander, E. Hairer, *Analysis for parareal algorithms applied to Hamiltonian differential equations*, J. Comput. Appl. Math. **259** (2014), 2–13.

[3] G. Ariel, S.J. Kim, R. Tsai, *Parareal multiscale methods for highly oscillatory dynamical systems*, SIAM J. Sci. Comput. **38** (2016), A3540–A3564.

[4] G. Ariel, H. Nguyen, R. Tsai, *θ-parareal schemes*, arXiv:1704.06882 (2017).

[5] H. Nguyen, R. Tsai, *A stable parareal-like method for the second order wave equation*, J. Comput. Phys. **405** (2020), 109156.

Invariant Measures for Data-Driven Dynamical System Identification

Jonah Botvinick-Greenhouse
(joint work with Yunan Yang)

Constructing accurate models of dynamical systems in the face of data-sparsity, measurement errors, and uncertainty is a critically important task across a wide range of scientific disciplines. In this talk, we introduce a variety of techniques, based upon the comparison of simulated and observed invariant measures, which are designed to be robust to such data imperfections. More specifically, we consider the problem of reconstructing the unknown parameters $\theta^* \in \Theta \subseteq \mathbb{R}^p$ of a dynamical system, e.g., a parameterized differential equation $\dot{x} = v(x;\theta)$, based upon observed trajectory data $\{x^*(t_k)\}_{k=1}^N \subseteq \mathbb{R}^d$. One notable method for this task is given by the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm [1], which computes divided difference approximations along the trajectory $\{x^*(t_k)\}_{k=1}^N$ to obtain a local estimate of the velocity and subsequently performs sparse regression over the set Θ. Another class of approaches, given by Neural ODEs and shooting methods [2], formulate the system identification task as an optimization problem

$$\min_{\theta\in\Theta} \mathcal{J}(\theta), \qquad \mathcal{J}(\theta) := \frac{1}{N}\sum_{k=1}^{N} \|x^*(t_k) - x(t_k;\theta)\|_2^2, \tag{1}$$

where $\{x(t_k;\theta)\}_{k=1}^N$ denotes a simulated trajectory according to the vector field $v(\theta)$. While these techniques have proven to be successful at recovering the unknown parameters $\theta^* \in \Theta$ when the available data $\{x^*(t_k)\}_{k=1}^N$ is densely sampled and noise-free, their effectiveness is known to diminish when the available data is sparse, noisy, and chaotic.

Rather than seeking a pointwise matching with the observed trajectory data, or its time-derivatives, we follow [3, 4] and instead consider the occupation measure

$$\rho^* := \frac{1}{N}\sum_{k=1}^{N} \delta_{x(t_k)}, \tag{2}$$

where as $N \to \infty$ mild assumptions of ergodicity ensure that ρ^* weakly converges to an invariant measure of the underlying system. The occupation measure (2) quantifies the global time-invariant statistics of the underlying system and can be well-approximated even in the presence of noise, chaos, and data uncertainty. Thus, we recast the system identification task as the optimization

$$\min_{\theta\in\Theta} \mathcal{J}(\theta), \qquad \mathcal{J}(\theta) := \mathcal{D}(\rho^*, \rho(\theta)), \tag{3}$$

where $\mathcal{D}$ denotes a metric or divergence on the space of probability measures, e.g., the KL divergence or Wasserstein distance, and $\rho(\theta)$ is an invariant measure of the system $\dot{x} = v(x;\theta)$. While the objective (1) can be sensitive to chaotic dynamics, noisy observations, and slow sampling of the dynamics, we observe clear robustness when using (3) to reconstruct the parameters $\theta^* \in \Theta$.

In cases when the number of unknown parameters is large, e.g., for a neural network parameterization of the velocity, one must solve the optimization (3) using gradient-based methods. Rather than approximating $\rho(\theta)$ as an occupation measure from a long trajectory simulation, which can make gradient computations challenging, we follow [4] in constructing a time-independent partial-differential equation surrogate model. That is, we approximate the invariant measure as a time-invariant solution of the stationary Fokker–Planck equation

$$\nabla \cdot (\rho v) = D\Delta\rho, \tag{4}$$

which we discretize using a first order upwind finite-volume scheme. After discretization, computing the stationary solution (4) amounts to solving the eigenvector problem $M\rho = \rho$, where $M \in \mathbb{R}^{n\times n}$ is a Markov matrix approximating the differential operator of the Fokker–Planck equation. To ensure uniqueness and improve conditioning, we apply the teleportation regularization scheme from Google's PageRank algorithm to the Markov matrix [5]. When the computation of $\theta \mapsto \rho(\theta)$ is achieved by solving (4), the gradient of (3) can be seamlessly computed using the adjoint-state method, making the proposed approach of invariant-measure based system identification compatible with large-scale parameterizations of the velocity [7].

While we observe many benefits when using the invariant measure to model the underlying dynamics, an outstanding challenge related to the uniqueness of the parameter recovery still remains. In particular, infinitely many distinct dynamical systems can all share the same invariant measure, i.e., the map $\theta \mapsto \rho(\theta)$ may not be injective. To address this difficulty, we instead propose studying the invariant measure in a new coordinate frame that introduces crucial time-dependent data back into the modeling problem. Motivated by Takens' seminal embedding theory [6], we consider the delay-coordinate invariant measure

$$\hat{\mu}^{(m)}_{(y,T)} := \Psi^{(m)}_{(y,T)}\#\mu, \qquad \Psi^{(m)}_{(y,T)}(x) := (y(x), y(T(x)), \ldots, y(T^{m-1}(x))), \tag{5}$$

where μ is an invariant measure of the discrete-time dynamical system $T : \mathbb{R}^d \to \mathbb{R}^d$, which one can regard as the time-Δt flow of some vector field. In (5), m is the so-called embedding dimension, $y : \mathbb{R}^d \to \mathbb{R}$ is a smooth scalar observation function, and $\#$ denotes the measure-theoretic pushforward operation.

While there can be no theoretical guarantee for unique recovery of the underlying dynamical system when using the invariant measure, we show that the delay-coordinate invariant measure (5) can in fact distinguish between large classes of dynamical systems. We present two key results along this direction. First, we show that equality of two delay-coordinate invariant measures implies topological conjugacy of the underlying dynamical systems. Moreover, we go on to show that under mild assumptions finitely many delay-coordinate invariant measures, corresponding to different observation functions, can be used to uniquely identify the dynamical system. These two theoretical results resolve a significant challenge of non-uniqueness arising in the invariant measure-based system identification approach (3) and pave the way for novel computational approaches for performing data-driven system identification.

We showcase the effectiveness of our methodologies on a variety of dynamical systems subject to challenging data assumptions, including the Lorenz-63 system, the Van der Pol oscillator, Hall-effect thruster dynamics, and the Kuramoto–Sivashinsky equation. Across these examples, we demonstrate that the invariant measure is a valuable modeling tool for obtaining robustness against chaos, noise, data-sparsity, and slow sampling, when performing data-driven system identification.

References

[1] S. Brunton, J. Proctor, J. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings Of The National Academy Of Sciences **113** (2016), 3932–3937.

[2] R. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, *Neural ordinary differential equations*, Advances In Neural Information Processing Systems **31** (2018).

[3] C. Greve, K. Hara, R. Martin, D. Eckhardt, J. Koo, *A data-driven approach to model calibration for nonlinear dynamical systems*, J. Appl. Phys. **125** (2019).

[4] Y. Yang, L. Nurbekyan, E. Negrini, R. Martin, M. Pasha, *Optimal transport for parameter identification of chaotic dynamics via invariant measures*, SIAM J. Appl. Dyn. Syst. **22** (2023), 269–310.

[5] L. Page, S. Brin, R. Motwani, T. Winograd, *The PageRank citation ranking: Bringing order to the web*, Stanford infolab (1999).

[6] F. Takens, *Detecting strange attractors in turbulence*, Dynamical Systems And Turbulence, Warwick 1980: Proceedings Of A Symposium Held At The University Of Warwick 1979/80, 366–381, 2006.

[7] J. Botvinick-Greenhouse, R. Martin, Y. Yang, *Learning dynamics on invariant measures using PDE-constrained optimization*, Chaos: An Interdisciplinary Journal Of Nonlinear Science **33** (2023).

[8] J. Botvinick-Greenhouse, R. Martin, Y. Yang, *Invariant Measures in Time-Delay Coordinates for Unique Dynamical System Identification*, arXiv:2412.00589 (2024).

The multiscale preconditioner for highly heterogeneous flow

Lina Zhao

(joint work with Shubin Fu, Eric Chung)

Subsurface fluid simulation through porous media has many practical applications, such as reservoir simulation, nuclear water storage and underground water contamination. In reservoir simulation, the media's properties such as the permeability and porosity have significant influence on the flow behaviors. These properties usually vary several orders of magnitude and have complicated multiscale structures. To accurately capture these properties, it is quite necessary to develop robust and efficient precondiitoner.

Consider the following steady state single-phase incompressible flow in a mixed formulation:

$$\begin{aligned} \kappa^{-1}\boldsymbol{v} + \nabla p &= \mathbf{0} \quad \text{in } \Omega, \\ \nabla \cdot \boldsymbol{v} &= f \quad \text{in } \Omega, \\ \boldsymbol{v} \cdot \boldsymbol{n} &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where $\Omega \subset \mathbb{R}^d, d = 2, 3$ is the computational domain and $\boldsymbol{n}$ is the unit outward normal vector of the boundary of Ω, the source function f satisfies $\int_\Omega f = 0$, κ is a scalar permeability field of the porous media.

Let $\mathcal{T}_H$ be a usual shape regular, conforming partition of Ω into quadrilaterals or tetrahedrons K_i with diameter H_i so that $\overline{\Omega} = \cup_{i=1}^N \overline{K}_i$, where N is the number of coarse elements. We further partition each coarse element K_i into a finer mesh with mesh size h_i. Let $\mathcal{T}_h = \cup_{i=1}^N \mathcal{T}_h(K_i)$ be the union of all these partitions, which is a fine mesh partition of the domain Ω. We use $\mathcal{F}_h$ to represent the union of all the edges of $\mathcal{T}_h$ and use $\mathcal{F}_h^0$ to denote the union of all the interior edges.

Let $\boldsymbol{V}_h \subset H(\mathrm{div}, \Omega)$ and $Q_h \subset L^2(\Omega)$ represent the lowest-order Raviart-Thomas finite element spaces with respect to the prescribed triangulation $\mathcal{T}_h$ for the approximation of (1). Then, the discrete formulation for (1) reads as follows: Find $(\boldsymbol{v}_h, p_h) \in \boldsymbol{V}_h^0 \times Q_h$ such that

$$\begin{aligned} \int_\Omega \kappa^{-1} \boldsymbol{v}_h \cdot \boldsymbol{w}_h - \int_\Omega \nabla \cdot \boldsymbol{w}_h p_h = 0, \qquad & \forall \boldsymbol{w}_h \in \boldsymbol{V}_h^0, \\ - \int_\Omega \nabla \cdot \boldsymbol{v}_h q_h = - \int_\Omega f q_h, \qquad & \forall q_h \in Q_h, \end{aligned} \tag{2}$$

where $\boldsymbol{V}_h^0 := \{\boldsymbol{v} \in \boldsymbol{V}_h : \boldsymbol{v} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega\}$.

The above system can be written in terms of matrix representations as

$$\begin{bmatrix} M & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_h \\ p_h \end{bmatrix} = \begin{bmatrix} 0 \\ F \end{bmatrix}.$$

If we use the trapezoidal rule to compute $\int_\Omega \kappa^{-1} \boldsymbol{\phi}_i \boldsymbol{\phi}_j$, then M will become a diagonal matrix (cf. [1]) and thus can be inverted easily. Therefore we can eliminate $\boldsymbol{v}_h$ in (2) and obtain

$$\big(BM^{-1}B^T\big) p_h = F,$$

where $BM^{-1}B^T$ is a symmetric and semi-positive definite matrix.

Generalized multiscale space based two-grid preconditioner. The local snapshot $W_I^{i,\mathrm{snap}}$ is simply the restriction of Q_h to K_i, i.e., $W_I^i = Q_h(K_i)$. In each coarse element K_i, we solve the following spectral problem in the snapshot space $W^{i,\mathrm{snap}}$:

$$a_i(p, w) = \lambda s_i(p, w) \quad \forall w \in W^{i,\mathrm{snap}}, \tag{3}$$

where $a_i(p, w)$ and $s_i(p, w)$ are defined as

$$a_i(p, w) = \sum_{e \in \mathcal{F}_h(K_i)} \tilde{\kappa}_e [p][w] \quad \text{and} \qquad s_i(p, w) = \int_{K_i} \kappa p w,$$

where $[\cdot]$ is the jump operator. Here $\mathcal{F}_h(K_i)$ denotes the union of all the fine-grid edges that lies inside the coarse element K_i. In addition, $\tilde{\kappa}_e := 2/(\kappa_1^{-1} + \kappa_2^{-1})$ is the harmonic average of κ_1 and κ_2 with $\kappa_i, i = 1, 2$ being the permeability κ on the fine-grid elements τ_1 and τ_2, respectively. After (3) are solved numerically, we sort the eigenvalues in a increasing order, that is,

$$\lambda_1^i \le \lambda_2^i \le \cdots \le \lambda_{M^{i,\mathrm{snap}}}^i,$$

Algorithm 1 Two-grid preconditioner for solving $(BM^{-1}B^T)x = F$
Given initial guess x^0, **do**
1: $x^1 = x^0 + S\big(F - (BM^{-1}B^T)x^0\big)$ { Pre-smoothing, S is a smoother}
2: $r_1 = R\big(F - (BM^{-1}B^T)x^1\big)$ {Compute coarse grid residual}
Solve $\big(R(BM^{-1}B^T)R^T\big)x_c = r_1$ {Coarse grid correction}
$x^2 = x^1 + R^T x_c$ {Project coarse grid correction to fine grid}
3: $x^3 = x^2 + S\big(F - (BM^{-1}B^T)x^2\big)$ { Post-smoothing, S is a smoother}

TABLE 1. An illustration of the two-grid preconditioner

where λ_l^i is the l-th smallest eigenvalue for each coarse element K_i. Denote the eigenvectors corresponding to λ_l^i as $\zeta_l^i = (\zeta_{lj}^i)_{j=1}^{M^{i,\mathrm{snap}}}$ with ζ_{lj}^i being the jth component of the discrete eigenvector ζ_l^i. Then we select the first L_i eigenfunctions whose eigenvalues are less than a preset value δ to form the local generalized multiscale (GMs) space. Specifically, the generalized multiscale functions can be constructed as

$$\Psi_l^{i,\mathrm{GMs}} = \sum_{j=1}^{M^{i,\mathrm{snap}}} \zeta_{lj}^i \Psi_j^{i,\mathrm{snap}}, \quad l = 1, 2, \cdots, L_i. \tag{4}$$

Then the global GMs space is defined by

$$Q^{\mathrm{GMs}} = \mathrm{span}\{\Psi_l^{i,\mathrm{GMs}} : 1 \le l \le L_i, 1 \le i \le N\}.$$

Let R be the restriction matrix composed by all multiscale bases, i.e., $R = [\Psi_1; \Psi_2; \cdots; \Psi_{M^{\mathrm{Gms}}}]$, then R^T is the projection matrix. The proposed two-grid preconditioner consists of two components, i.e., a smoother to remove high-frequency errors and a coarse preconditioner to exchange global information. With the restriction matrix R and projection matrix R^T, we summarize the main steps of our two-grid preconditioner in Algorithm 1.

Analysis. The key is to show the smoothing property and the approximation property. We construct the interpolation operator $I_H : Q_h \to Q^{\mathrm{GMs}}$, which is defined by

$$I_H q = \sum_{i=1}^{N} \sum_{l=1}^{L_i} \Big(\int_{K_i} \kappa q \Psi_l^{i,\mathrm{GMs}}\, dx \Big) \Psi_l^{i,\mathrm{GMs}} \quad \forall q \in Q_h,$$

where $\Psi_l^{i,\mathrm{GMs}}$ is defined in (4). For an arbitrary $K \in \mathcal{T}_H$, then there exists a positive constant C_1 independent of the meshsize and κ such that

$$\int_\Omega \kappa (q - I_H q)^2\, dx \le C_1 \frac{1}{\lambda_A} a(q, q) \quad \forall q \in Q_h,$$

where $C_1 > 1$.

We define the Galerkin projection $P_0 : Q_h \to Q^{\mathrm{GMs}}$ satisfying

$$a(P_0 q, \chi) = a(q, \chi) \quad \forall q \in Q_h,\ \forall \chi \in Q^{\mathrm{GMs}}.$$

Let $a(\cdot,\cdot)$ be the bilinear form of the reduced system, then we have

$$\begin{aligned} a((I-P_0)K_1p, K_1p) &\leq (1-\frac{1}{C_1})a(K_1p,p) \\ &\leq (1-\frac{1}{C_1})a(p,p) \quad \forall p \in Q_h. \end{aligned}$$

We can define the error propagation operator for the two-grid algorithm by

$$E_1 = K_1(I - P_0 + E_0P_0)K_1.$$

Moreover, we let

$$\|E_1\|_A := \sup_{q\in Q_h} \frac{a(E_1q,q)}{a(q,q)}.$$

Then the following estimate is satisfied (cf. [2])

$$\|E_1\|_A \leq (1-\eta)\|E_0\|_A + \eta,$$

where $\eta = 1 - \frac{1}{C_1}$. Therefore, it holds $\|E_0\|_A < 1$ implies that $\|E_1\|_A < 1$.

References

[1] T. Arbogast and M. Wheeler, I. Yotov, *Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences*, SIAM J. Numer. Anal. **34** (1997), 828–852.

[2] S. Fu, C. Chung, L. Zhao, *An efficient multiscale preconditioner for large-scale highly heterogeneous flow*, SIAM J. Sci. Comput. **46** (2024), S352–S377.

Integral Representations of Sobolev Spaces via ReLU^k Activation Function and Optimal Error Estimates for Linearized Networks

Xinliang Liu

(joint work with Tong Mao, Jinchao Xu)

This paper presents two main theoretical results concerning shallow neural networks with ReLU^k activation functions. We establish a novel integral representation for Sobolev spaces, showing that every function in $\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)$ can be expressed as an $\mathcal{L}^2$-weighted integral of ReLU^k ridge functions over the unit sphere. This result mirrors the known representation of Barron spaces and highlights a fundamental connection between Sobolev regularity and neural network representations. Moreover, we prove that linearized shallow networks–constructed by fixed inner parameters and optimizing only the linear coefficients–achieve optimal approximation rates $\mathcal{O}(n^{-\frac{1}{2}-\frac{2k+1}{2d}})$ in Sobolev spaces.

1. Introduction

The following shallow neural networks function class serves as a fundamental building block of artificial intelligence technology:

$$\Sigma_n^\sigma = \left\{ \sum_{j=1}^n c_j \sigma(\theta_j \cdot \tilde{x}) : c_j \in \mathbb{R}, \theta_j \in \mathbb{R}^{d+1} \right\},$$

Among various activation functions, the Rectified Linear Unit (ReLU) has emerged as the dominant choice in modern deep learning. In this paper, we focus on ReLU activation function, $\mathrm{ReLU}(x) = \max(0, x)$, and its variants, $\sigma_k(x) = \mathrm{ReLU}^k(x)$ (k is a nonnegative integer). Due to the homogeneity of σ_k, we assume the parameters lie on the unit sphere $\mathbb{S}^d$. The corresponding function class of shallow ReLU^k neural networks, denoted by $\Sigma_n^k = \Sigma_n^{\sigma_k}$, is then

$$\Sigma_n^k = \left\{ \sum_{j=1}^n c_j \sigma_k(\theta_j \cdot \tilde{x}) : c_j \in \mathbb{R}, \theta_j \in \mathbb{S}^d \right\},$$

where $\tilde{x} = \begin{pmatrix} x \\ 1 \end{pmatrix}$ and $\theta_j = \begin{pmatrix} w_j \\ b_j \end{pmatrix}$. To ensure encodability and numerical stability, it is essential to impose restrictions on the size of the parameters in a neural network. Consequently, given a parameter M, which essentially determines the complexity of the function class, the stable shallow neural network function class is defined as

$$\Sigma_{n,M}^k = \left\{ f \in \Sigma_n^k : \sum_{j=1}^n |c_j| \le M \right\}. \tag{1}$$

In the early works, qualitative convergence has been extensively investigated since the 1990s (see, e.g., [1, 2]).

1.1. Shallow neural networks, Barron spaces and Sobolev spaces.

Given an activation function $\sigma : \mathbb{R} \to \mathbb{R}$ and $G \subset \mathbb{R}^{d+1}$, we denote the dictionary

$$\mathbb{D}_\sigma := \{ \pm\sigma(\theta \cdot \tilde{x}) : \ \theta \in G \},$$

and the $\mathcal{L}^2$-closure of its convex hull $B_1 := \overline{\mathrm{conv}(\mathbb{D}_\sigma)}$. We denote the Barron space $\mathcal{B}^\sigma(\Omega)$ as follows:

$$\mathcal{B}^\sigma(\Omega) := \{ f \in \mathcal{L}^2(\Omega) : \ \|f\|_{\mathcal{B}^\sigma(\Omega)} < \infty \},$$

where

$$\|f\|_{\mathcal{B}^\sigma(\Omega)} = \inf \{ t > 0 : \ f \in tB_1 \}.$$

In particular, when $\sigma = \sigma_k$, we take $G = \mathbb{S}^d$ as σ_k is homogeneous. In this case, we denote

$$\mathcal{B}^k(\Omega) = \mathcal{B}^{\sigma_k}(\Omega).$$

The Barron spaces can be characterized by some qualitative approximation property of the following compact subset of the (ReLU^k) neural networks (1) in [4, 3].

In this paper, we establish an approximation rate of $\mathcal{O}(n^{-\frac{r}{d}})$ for functions in the Sobolev space $\mathcal{H}^r(\Omega)$ under the condition that these parameters are well-distributed on $\mathbb{S}^d$.

Definition 1 (Well-distributed and quasi-uniform)**.** *Let $d \in \mathbb{N}$, a set of points $\{\theta_j^*\}_{j=1}^n \subset \mathbb{S}^d$ is said to be well-distributed if*

$$\max_{\theta\in\mathbb{S}^d} \min_{1\le j\le n} \rho(\theta, \theta_j^*) \lesssim n^{-\frac{1}{d}}, \tag{2}$$

Moreover, a well-distributed collection $\{\theta_j^*\}_{j=1}^n$ is said to be quasi-uniform if

$$\max_{\theta\in\mathbb{S}^d} \min_{1\le j\le n} \rho(\theta, \theta_j^*) \lesssim \min_{i\neq j} \rho(\theta_i^*, \theta_j^*).$$

The corresponding constants are independent of n.

1.2. **Main results.** We now present the main results of this paper. The first theorem focuses on analyzing the approximation properties of the ReLU^k activation function with predetermined parameters $\{\theta_j^*\}_{j=1}^n$.

Theorem 1. *Let $d, n \in \mathbb{N}$, $k \in \mathbb{N}_0$, $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, and $\{\theta_j^*\}_{j=1}^n \subset \mathbb{S}^d$. Then for any $f \in \mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)$, with some*

$$M \simeq \|f\|_{\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)},$$

$$\inf_{f_n\in L^k_{n,M}} \left\|f - f_n\right\|_{\mathcal{L}^2(\Omega)} \lesssim h^{\frac{d+2k+1}{2}} \|f\|_{\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)}.$$

where

$$h = \max_{\theta\in\mathbb{S}^d} \min_{1\le j\le n} \rho(\theta, \theta_j^*).$$

More generally, let $r \le \frac{d+2k+1}{2}$ and $s \le \min\{k, r\}$. Then for any $f \in \mathcal{H}^r(\Omega)$, with some $M \simeq h^{-\frac{d+2k+1-2r}{2}} \|f\|_{\mathcal{H}^r(\Omega)}$,

$$\inf_{f_n\in L^k_{n,M}} \left\|f - f_n\right\|_{\mathcal{H}^s(\Omega)} \lesssim h^{r-s} \|f\|_{\mathcal{H}^r(\Omega)}.$$

If the collection $\{\theta_j^*\}_{j=1}^n \subset \mathbb{S}^d$ is well-distributed, then for any $f \in \mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)$, with some $M \simeq \|f\|_{\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)}$, we have

$$\inf_{f_n\in L^k_{n,M}} \left\|f - f_n\right\|_{\mathcal{L}^2(\Omega)} \lesssim n^{-\frac{1}{2}-\frac{2k+1}{2d}} \|f\|_{\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)}.$$

All the corresponding constants are independent of n, $\{\theta_j^*\}_{j=1}^n$, and f.

Interestingly, it leads to an important function space embedding result, previously established in [5]:

$$\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega) \hookrightarrow \mathcal{B}^k(\Omega),$$

Corollary 1 (Theorem 1. [5])**.** *Let k, d, Ω be as in (2), then there is the continuous embedding*

$$\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega) \hookrightarrow \mathcal{B}^k(\Omega).$$

Proof. Let $f \in \mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)$ and $f_n \in L^k_{n,M}$ be the approximants of f in Theorem 1. Then

$$f_n \in L^k_{n,M} \subset \Sigma^k_{n,M}.$$

Notice that $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{L}^2(\Omega)} = 0$ implies

$$f = \lim_{n\to\infty} f_n \subset \overline{\bigcup_{n=1}^{\infty} \Sigma^k_{n,M}} = M\overline{\mathrm{conv}(\mathbb{D}_{\sigma_k})}.$$

It means

$$\|f\|_{\mathcal{B}^k(\Omega)} \leq M \lesssim \|f\|_{\mathcal{H}^{\frac{d+2k+1}{2}}(\Omega)}.$$

□

References

[1] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Systems **2** (1989), 303–314.

[2] K. Hornik, M. Stinchcombe, H. White, *Multilayer feedforward networks are universal approximators*, Neural Networks **2** (1989), 359–366.

[3] J. W. Siegel, J. Xu, *Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks*, Found. Comput. Math. (2022), 1–57.

[4] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39** (1993), 930–945.

[5] T. Mao, J. W. Siegel, J. Xu, *Approximation Rates for Shallow ReLU^k Neural Networks on Sobolev Spaces via the Radon Transform*, arXiv:2408.10996 (2024).

A quantum gradient descent algorithm for optimizing Gaussian Process models using hierarchical matrices

Junpeng Hu

(joint work with Shi Jin, Jinglai Li, Lei Zhang)

Gaussian Process Regression (GPR) is a nonparametric supervised learning method, widely valued for its abilit to quantify uncertainty. Despite its advantages and broad applications, classical GPR implementations face significant scalability challenges, as they involve matrix operations with a *cubic* complexity in relation to the dataset size. This computational challenge is further compounded by the demand of optimizing the Gaussian Process model over its hyperparameters, rendering the total computational cost prohibitive for data intensive problems. To address this issue, we propose a quantum gradient descent algorithm to optimize the Gaussian Process model.

The hyperparameters play an important role in determining the flexibility, expressiveness of the kernel (covariance) function, and overall performance of the GP model. As such an essential step in GPR is to optimize the covariance kernel with respect to these hyperparameters, and this is often done by a maximum likelihood estimation approach to ensure the kernel accurately models the underlying data. We assume a general covariance function $k_{\boldsymbol{\theta}}$, parameterized by a vector of hyperparameters $\boldsymbol{\theta}$. Given training points $(\mathbf{X}, \mathbf{y})$, We shall find the optimal

values for $\boldsymbol{\theta}$ by maximizing the marginal likelihood $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ – it is called "marginal" likelihood because it is considered to be obtained by marginalizing the joint distribution $p(\mathbf{f}_*, \mathbf{y})$ over $\mathbf{f}_*$. It can be found that the logarithm of the marginal likelihood (LML), a more convenient form for optimization, is given by:

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top K^{-1}\mathbf{y} - \frac{1}{2}\log|K| - \frac{N}{2}\log 2\pi, \tag{1}$$

where $K = K_f + \sigma_n^2\mathbf{I}$ is the covariance matrix of the noisy observations $\mathbf{y}$, and $K_f = K(\mathbf{X}, \mathbf{X})$. The gradient of LML can be formally derived as,

$$\frac{\partial}{\partial\theta_j}\log p(\mathbf{y}|X, \boldsymbol{\theta}) = \frac{1}{2}\mathbf{y}^\top K^{-1}\frac{\partial K}{\partial\theta_j}K^{-1}\mathbf{y} - \frac{1}{2}\mathrm{tr}\left(K^{-1}\frac{\partial K}{\partial\theta_j}\right), \tag{2}$$

for $j = 1, ...d$ (assuming $\boldsymbol{\theta}$ is d-dimensional). One can see from Eq (2), the computational complexity for evaluating the LML derivatives is the same (i.e. $\mathcal{O}(N^3)$) as evaluating LML itself (Eq. (1)), which is primarily determined by inverting the $N \times N$ matrix K.

To work on a quantum computer, we must first construct a block-encoding of the matrix K, which must be unitary, as quantum computers operate only through unitary transformations (i.e., quantum gates). The general complexity of block-encoding a matrix A is $\mathcal{O}(s\,\|A\|_{\max})$, where s represents the sparsity and $\|\cdot\|_{\max}$ represents the max norm. To reduce this cost, we adopt the method proposed in [2]. Specifically, we decompose K into a hierarchical structure of *admissible* blocks, where each block corresponds to a pair of clusters whose separation exceeds the sum of their radii:

$$K = \sum_{l=2}^{L} K^{(l)} + K_{\mathrm{ad}}.$$

Assuming the kernel function exhibits polynomial decay, each admissible block– being a dense matrix with elementwise decay–is block-encoded via a two-step approach: first using the standard procedure for dense matrices, followed by the procedure for block-sparse matrices. Summing across all hierarchical levels yields a complete block-encoding of K. The total complexity of the encoding is given by

$$\alpha = 3 + 3\sum_{l=2}^{L} 2^{(L-l)(1-p)},$$

where p denotes the decay rate. When $p \geq 1$, the overall complexity becomes polylog(N).

Provided with the oracles $O_{\mathbf{y}}$, O_K and $O_{\mathrm{d}K}$, several quantum linear system solvers can be employed to obtain a block encoding of K^{-1}, such as QSVT and HHL. In this context, we define the procedure as an oracle $O_{K^{-1}}$. To calculate the derivative of LML, we first reformulate it into:

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\log p(\mathbf{y}|X, \theta) &= \frac{1}{2}\mathbf{y}^\top K^{-1}\frac{\partial K}{\partial\theta_j}K^{-1}\mathbf{y} - \frac{1}{2}\mathrm{tr}\left(K^{-1}\frac{\partial K}{\partial\theta_j}\right) \\ &= \frac{1}{2}\mathbf{y}^\top K^{-1}\frac{\partial K}{\partial\theta_j}K^{-1}\mathbf{y} - \frac{1}{2}\sum_{j\in[2^n]}\mathbf{e}_j^\top K^{-1}\frac{\partial K}{\partial\theta_j}\mathbf{e}_j\end{aligned}$$

The quantum states are defined as

$$|\tilde{\mathbf{y}}_l\rangle = \frac{1}{\mathcal{N}_l}\tilde{\mathbf{y}}_l = \sum_{j\in[2^n]} \frac{1}{\mathcal{N}_l} |0\rangle_{i_1} |j\rangle_{i_2} |j\rangle_w + \frac{\|\mathbf{y}\|}{\mathcal{N}_l} |1\rangle_{i_1} |0^n\rangle_{i_2} |\mathbf{y}\rangle_w ,$$

$$|\tilde{\mathbf{y}}_r\rangle = \frac{1}{\mathcal{N}_r}\tilde{\mathbf{y}}_r = \sum_{j\in[2^n]} \frac{-C_0}{\mathcal{N}_r} |0\rangle_{i_1} |j\rangle_{i_2} |j\rangle_w + \frac{\|\mathbf{y}\|}{\mathcal{N}_r} |1\rangle_{i_1} |0^n\rangle_{i_2} |\mathbf{y}\rangle_w ,$$

where the subscripts i_1, i_2 indicate 'index register', w indicates 'work register', and $\mathcal{N}_l$, $\mathcal{N}_r$ are the normalization factors. The initial states are prepared as

$$|\varphi_l\rangle = |\tilde{\mathbf{y}}_l\rangle \left|0^{a+1}\right\rangle_{a_1} \left|0^{a+1}\right\rangle_{a_2} \left|0^b\right\rangle_b |\theta\rangle , \quad |\varphi_r\rangle = |\tilde{\mathbf{y}}_r\rangle \left|0^{a+1}\right\rangle_{a_1} \left|0^{a+1}\right\rangle_{a_2} \left|0^b\right\rangle_b |\theta\rangle ,$$

where a_1, a_2 and b indicate 'ancilla register' used for the oracles. The unitary operator is defined as

$$U = [O_{K^{-1}}]_{w,a_2,\theta} [O_{dK}]_{w,b,\theta} [O_{K^{-1}}]^{i_1}_{w,a_1,\theta} ,$$

where the subscript denotes the target register and the superscript denotes the control register. Finally it can be verified that

$$\langle\varphi_l| U |\varphi_r\rangle = \frac{2C_0^2}{\mathcal{N}_l\mathcal{N}_r} \frac{\partial}{\partial\theta} \log p(\mathbf{y}|X,\theta),$$

which implies that the derivative of the log marginal likelihood (LML) can be computed through the inner product $\langle\varphi_l| U |\varphi_r\rangle$. This quantity is subsequently estimated using the Hadamard test, enhanced by quantum phase estimation (QPE) for improved efficiency. In a standard gradient descent iteration, the design parameter θ is updated as,

$$\begin{aligned}\theta_{t+1} &= \theta_t - \eta_t \frac{\partial}{\partial\theta} \log p(\mathbf{y}|X,\theta) \\ &= \theta_t - \frac{\eta_t\mathcal{N}_l\mathcal{N}_r}{2C_0^2} \langle\varphi_l| U |\varphi_r\rangle \\ &= \theta_t - \frac{\mu_t}{2C_0^2} \langle\varphi_l| U |\varphi_r\rangle ,\end{aligned}$$

where η_t is the step size and $\mu_t = \eta_t\mathcal{N}_l\mathcal{N}_r$. Since $\theta = \mathcal{O}(1)$, η_t should be chosen small enough to satisfy $\mu_t = \mathcal{O}(1)$.

The overall complexity of the method across multiple iteration steps, as stated in the following corollary.

Corollary 1. *Given sparse accesses to the s-sparse kernel matrix $K = K_f + \sigma_N^2\mathbf{I}$ and its derivative $\frac{dK}{d\theta}$, the $(s, \log_2(s)+1)$-block-encodings O_K and O_{dK} can be constructed. The proposed quantum gradient descent method for $T \geq 0$ steps prepares a solution $|\theta_T\rangle$ to final accuracy ε' in cost*

$$\tilde{\mathcal{O}}\left(sT(1+C_2)^T \mu polylog(N)/\delta\sigma_N^6\varepsilon'\right) ,$$

with some logarithmic terms ignored. When K is a dense kernel matrix with polynomial decay, the cost is

$$\tilde{\mathcal{O}}\left(T(1+C_2)^T \mu\kappa polylog(N)/\delta\sigma_N^6\varepsilon'\right) ,$$

The entire gradient descent procedure is embedded within the quantum circuit. By incorporating recent advancements in quantum linear algebra algorithms, and through rigorous runtime and error analyses, our method achieves an exponential speedup in computing the gradients of the log marginal likelihood. This approach markedly enhances the scalability of Gaussian Process Regression (GPR) model optimization, rendering it computationally tractable for large-scale problems.

References

[1] J. Hu, J. Li, Z. Lei, S. Jin, *A quantum gradient descent algorithm for optimizing Gaussian Process models*. arXiv:2503.17780 (2025).

[2] Q.T. Nguyen, B.T. Kiani, S. Lloyd, *Block-encoding dense and full-rank kernels using hierarchical matrices: applications in quantum numerical linear algebra*. Quantum **6** (2022), 876.

Quantum Multilevel Preconditioning

Matthias Deiml

(joint work with Daniel Peterseim)

Quantum computers can provide exponential speed-ups in many applications of computational mathematics. One exciting area where this might be possible is the numerical solution of partial differential equations (PDEs). In particular, the analysis of [1] concluded that certain elliptic PDEs could be solved up to an error of $\varepsilon > 0$ in time almost linear in ε^{-1} independent of the spatial dimension $d \in \mathbb{N}$, whereas classical computers require ε^{-d} operations in general. This, however, requires preconditioning of the linear system of equations resulting from discretization of the PDE, which is addressed in our work [2].

As a prototypical example, we consider the steady-state diffusion equation

$$-\operatorname{div}(A\nabla u) = f, \tag{1}$$

in a bounded domain $D \subset \mathbb{R}^d$, along with homogeneous Dirichlet boundary conditions $u = 0$ on ∂D. This equation has been the focus of many previous efforts. The high dimensionality of the domain D or the multiscale features encoded in the diffusion coefficient A can make numerical methods on classical computers extremely expensive, but quantum algorithms have the potential to solve this problem. Existing quantum algorithms for this class of problems include *Schrödingerization* [4, 5] where (1) is transformed into a Schrödinger equation, which can be solved efficiently on a quantum computer. Another popular strategy, as discussed, for example, in [6, 7], is finite element or finite difference discretization, followed by the use of a quantum linear system solver. However, either approach suffers from the condition number of the system, which is of order h^{-2} where $h > 0$ is the mesh width. In the absence of strong regularity assumptions, the mesh size is at best proportional to the error tolerance ε, leading to a conditioning-induced multiplicative factor in the computational complexity of order ε^{-2} and a total runtime of order ε^{-3}. Note that this can be even worse for very rough coefficients in multiscale problems.

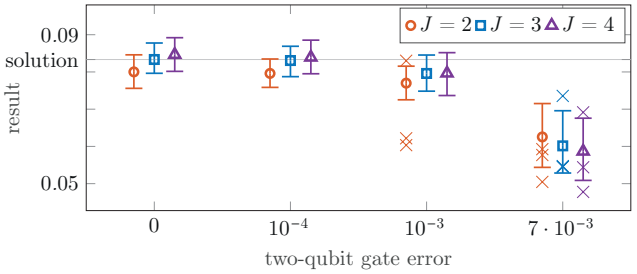


FIGURE 1. Results of a model problem with varying numbers of solver steps J with simulated noise, as well as real hardware ($\times$).

We follow the approach of a typical finite element discretization, specifically using d-linear finite elements, followed by the application of a quantum linear system solver. This allows us to use the vast amount of existing literature on preconditioning of finite element systems. Within the context of linear computations on quantum hardware, we find that *block encodings* [10] are a useful tool of abstraction: the state of a quantum computer can be represented by a $\mathbb{C}$-linear combination of the classical states $|0\rangle, \dots, |2^n - 1\rangle$ where n is the number of qubits in the machine. Any operation or program of the computer is then determined by a unitary matrix acting on this state space, and we aim to encode the stiffness matrix of our problem within a block of this unitary, thus the name *block encoding*.

The resulting block encoding can then be passed to a quantum linear system solver. These solvers have an asymptotically optimal runtime of $\mathcal{O}(\tilde{\gamma}\kappa \log \varepsilon^{-1})$, which matches the characteristics of classical Krylov solvers up to the factor $\tilde{\gamma}$ called *subnormalization*. Here we run into a problem. The multiplication operation of a stiffness matrix S and preconditioner P leads to an encoding with a subnormalization of $\tilde{\gamma}_{P \cdot S} \geq \min\{\kappa(S), \kappa(P)\}$. Since the condition of S can only be improved if the condition number of P is large, specifically $\kappa(PS) \geq \kappa(S)/\kappa(P)$, this proves that the runtime of the linear solver is not improved by this multiplicative preconditioning.

We instead resort to assembling S in an already preconditioned basis. For this we use the well-known BPX preconditioner [8] along with the machinery of frames [9]. The condition of the resulting matrix is then independent of h; however, it might scale exponentially with the dimension d. Still, this shows asymptotic quantum advantage for solving PDEs even for dimension $d = 2$, whereas previous methods are faster only for $d \geq 4$.

For details we refer to the [2], in which the full preconditioned quantum FEM algorithm and its analysis are presented. We have also tested our method on a model problem with 15 degrees. Figure 1 shows computational results on simulated and real hardware building on and extending the ones in [2, Sec. 7.1], specifically the IBM quantum computers `ibm_fez` and `ibm_marrakesh` (with two-qubit gate errors of $7 \cdot 10^{-3}$ and $3 \cdot 10^{-3}$ respectively). These results suggest that the next generation of machines will be able to accurately solve our model problem. More

significant speedups are expected for high-dimensional problems, particularly in settings where A includes randomness with short correlation length, which is the subject of our current work in progress. Additionally, we are addressing nonlinear problems in [3].

Acknowledgment. This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 865751 – RandomMultiScales).

References

[1] A. Montanaro, S. Pallister, *Quantum algorithms and the finite element method*, Phys. Rev. A **93** (2016), 032324.
[2] M. Deiml, D. Peterseim, *Quantum Realization of the Finite Element Method*, arXiv:2403.19512 (2024).
[3] M. Deiml, D. Peterseim, *Nonlinear quantum computing by amplified encodings*, arXiv:2411.16435 (2024).
[4] J. Hu, S. Jin, N. Liu, L. Zhang, *Quantum circuits for partial differential equations via schrödingerisation*, arXiv:2403.10032 (2024).
[5] J. Hu, S. Jin, L. Zhang, *Quantum algorithms for multiscale partial differential equations*, Multiscale Model. Simul. **22** (2024), 1030–1067.
[6] Y. Cao, A. Papageorgiou, I. Petras, J. Traub, S. Kais, *Quantum algorithm and circuit design solving the Poisson equation*, New J. Phys. **15** (2013), 013021.
[7] A. M. Childs, J. Liu, *Quantum spectral methods for differential equations*, Comm. Math. Phys. **375** (2020), 1427–1457.
[8] J. H. Bramble, J. E. Pasciak, J. Xu, *Parallel multilevel preconditioners*, Math. Comp. **55** (1990), 1–22.
[9] P. Oswald, *Multilevel Finite Element Approximation*, Teubner Skripten Zur Numerik, Vieweg+Teubner Verlag, Wiesbaden (1994).
[10] A. Gilyén, Y. Su, G. H. Low, N. Wiebe, *Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics*, In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pages 193–204 (2019).

Dissipative solutions of compressible flows: multiscale viewpoint

Mária Lukáčová-Medviďová
(joint work with Eduard Feireisl)

In this talk we have reviewed recent results on generalized solutions of compressible flows, the so-called dissipative solutions, putting the focus on their multiscale features. The dissipative solutions are obtained as a limit of suitable structure-preserving, consistent and stable numerical schemes [1, 2, 3]. Roughly speaking, dissipative solutions satisfy the weak formulation of the underlying PDE system, say the Euler equations of gas dynamics, modulo defect measures. Dissipative solutions are the observable quantities (the density, momentum and the entropy) and can be seen as coarse-grained quantities. In fact, they are the expected values with the respect to the underlying Young measure, the space-time parameterized probability measure. Fine scale oscillations are governed by the defect measures arising in the momentum equation and the energy balance.

In the case that the strong solution to the above equations exists, the dissipative solutions coincide with the strong solution on its life span [1].

If the consistent approximations only converge weakly and not strongly, the limit is not a weak solution. This means the defects are nonzero. We can apply a newly developed concept of $\mathcal{K}$-convergence and prove the strong convergence of the empirical means of numerical solutions to the observable quantities, the dissipative solution [4], [5].

We have also discussed a relation between numerical approximations of oscillatory solutions and the Dafermos criterion on maximization of the entropy production. Indeed, if we select a dissipative solution that is maximal with respect to the entropy production rate, then the defects vanish. Consequently, the selected solution is a weak solution. On the other hand, if solutions are oscillatory numerical simulations with standard structure-preserving numerical methods yield truly dissipative solutions with nonzero defects. In our recent work [6] we have formulated the following conjecture:

Conjecture: *If a numerical scheme converges to a weak solution of the Euler system, the limit is a maximal computable dissipative solution. If the same numerical scheme is oscillatory, the maximal (computable) dissipative solution does not exist.*

Theoretical results were illustrated by a series of numerical simulations.

References

[1] E. Feireisl, M. Lukáčová-Medvid'ová, H. Mizerová, B. She, *Numerical analysis of compressible fluid flows* (2021), Springer.

[2] E. Feireisl, M. Lukáčová-Medvid'ová, H. Mizerová, *Convergence of finite volume schemes for the Euler equations via dissipative-measure valued solutions*. Found. Comput. Math. **20** (2020), 923–966.

[3] E. Feireisl, M. Lukáčová-Medvid'ová, H. Mizerová, B. She, *Convergence of a finite volume scheme for the compressible Navier-Stokes system*. ESAIM: Math. Model. Num. **53** (2019), 1957–1979.

[4] E. Feireisl, M. Lukáčová-Medvid'ová, H. Mizerová, *$\mathcal{K}$-convergence as a new tool in numerical analysis*, IMA J. Numer. Anal. **40** (2020), 2227–2255.

[5] E. Feireisl, M. Lukáčová-Medvid'ová, B. She, Y. Wang, *Computing oscillatory solutions of the Euler system via $\mathcal{K}$-convergence*, Math. Math. Models Methods Appl. Sci. **31** (2021), 537–576.

[6] E. Feireisl, M. Lukáčová-Medvid'ová, C. Yu, *Oscillatory approximations and maximum entropy principle for the Euler system of gas dynamics*, arXiv:2505.02070 (2025).

A multilevel Monte Carlo method for the Dean-Kawasaki equation from fluctuating hydrodynamics

Julian Fischer
(joint work with Federico Cornalba)

Many PDE models in continuum mechanics arise as the continuum limit of an interacting particle system. On microscopic scales, thermal fluctuations due to the finite number of particles may become important, leading to a breakdown of the deterministic continuum description. Fluctuating hydrodynamics [12, 13] provides a framework to incorporate thermal fluctuations in continuum models via an SPDE approach, substantially enhancing the regime of validity of continuum models.

The Dean-Kawasaki equation [5, 10]

$$\partial_t \rho = \tfrac{1}{2}\Delta\rho + N^{-1/2}\nabla\cdot\big(\sqrt{\rho}\xi\big) \tag{1}$$

is one of the most basic equations of fluctuating hydrodynamics, describing density fluctuations in a system of $N \gg 1$ independent diffusing particles. Here, $\rho(x,t) \geq 0$ denotes the particle density and ξ denotes vector-valued space-time white noise. Like many SPDEs from fluctuating hydrodynamics, it has suffered from a lack of a rigorous mathematical justification: In its formulation (1), it is a highly singular SPDE that is not renormalizable by regularity structures or paracontrolled calculus. For smooth initial data ρ_0, it has even been shown to not admit any martingale solutions [11]. In recent years, it has been understood that SPDEs of fluctuating hydrodynamics typically require a regularization to become meaningful; we refer to [1, 3, 4, 7] for the Dean-Kawasaki equation and to [6] and [8] for analogues in the case of the symmetric simple exclusion process respectively the zero range process.

In the recent works [1, 3], the Dean-Kawasaki equation (1) and its version for weakly interacting particles $\partial_t\rho = \frac{1}{2}\Delta\rho + \nabla\cdot(\rho(\nabla V * \rho)) + N^{-1/2}\nabla\cdot(\sqrt{\rho}\xi)$ have found a justification as a "recipe for coarse-graining" of the underlying particle dynamics: Formal discretizations of the Dean-Kawasaki equation have been shown to accurately predict statistical properties of density fluctuations in the underlying interacting particle system, up to a numerical error. These results are valid in the regime of a large number of particles per grid cell, i. e. as long as the numerical discretization is indeed an attempt at coarse-graining of the particle system.

In particular, under the aforementioned assumption statistical quantities like $\mathbb{E}[(N^{1/2}\int_{\mathbb{T}^d}(\rho_h(\cdot,T) - \mathbb{E}[\rho_h(\cdot,T)])\eta\,dx)^p]$ (where ρ_h denotes the solution to a suitable discretization of (1), $p \geq 2$ is an integer, and η is a sufficiently smooth test function) approximate the corresponding quantity for the particle system $\mathbb{E}[(N^{1/2}\int_{\mathbb{T}^d}\eta\,d(\mu(\cdot,T) - \mathbb{E}[\mu(\cdot,T)]))^p]$ (where μ is the empirical measure), up to a numerical error.

Building on ideas from the analysis in [1, 3], in [2] we have proposed and analyzed a multilevel Monte Carlo method [9] for the Dean-Kawsaki equation (1). Employing a central finite difference discretization on a family of uniformly spaced grids $G_\ell := \{0, h_\ell, 2h_\ell, \dots, 1-h_\ell\}^d \subset \mathbb{T}^d$ with $h_\ell := 2^{-\ell}$ and an Euler-Maruyama

scheme for discretization in time with $\Delta t_\ell := h_\ell^2$, the scheme for a single level ℓ reads

$$\frac{\rho_\ell^{t+\Delta t_\ell} - \rho_\ell^t}{\Delta t_\ell} = \Delta_{h_\ell} \rho_\ell^t + \nabla_{h_\ell} \cdot \big(\sqrt{\rho_\ell^t} \xi_\ell^t\big).$$

Here, the discrete white noise $\xi_\ell^t : G_\ell \to \mathbb{R}$ corresponds to the rescaled increments $\xi_\ell^t(x) := h_\ell^{-d/2}(\beta_x^\ell(t+\Delta t) - \beta_x^\ell(t))$ of independent Brownian motions $(\beta_x^\ell)_{x \in G_\ell}$.

To achieve a variance reduction in the multilevel Monte Carlo scheme, an effective coupling between the noises ξ_ℓ^t and $\xi_{\ell+1}^t$ is needed. In [2] we propose and analyze two coupling schemes, one being the nearest-neighbor coupling in space

$$\xi_\ell^t(x) := \sum_{y \in \{0,1\}^d} \sum_{k \in \{0,1,2,3\}} \xi_{\ell+1}^{t+k\Delta t_{\ell+1}}(x + h_{\ell+1} y)$$

and another coupling scheme defined in terms of spatial Fourier modes.

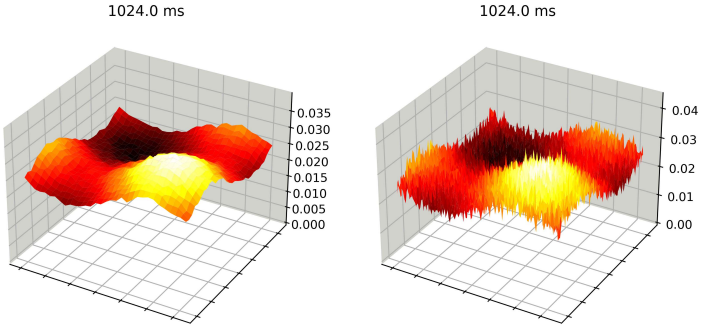


FIGURE 1. A solution to the discretized Dean-Kawasaki equation (1) for a coarse spatial discretization (left) and for a fine spatial discretization (right). One clearly observes an increase in strength of the fluctuations on finer grids, leading to an eventual breakdown of positivity of solutions on even finer grids. This illustrates the origin of the N-dependent lower bound on the discretization parameter h required in [1, 2, 3].

We then prove the following result for our multilevel Monte Carlo method [2]: Suppose that we aim to approximate a statistical quantity of the particle system like $\mathbb{E}[(N^{1/2} \int_{\mathbb{T}^d} \eta \, d(\mu(\cdot, T) - \mathbb{E}[\mu(\cdot, T)]))^p]$ for an integer $p \geq 2$ and a smooth test function η (where again μ denotes the empirical measure), up to a mean-square error $\varepsilon^2 > 0$. In the regime of large particle numbers $N\varepsilon^{d+2} \geq 1$, the multilevel Monte Carlo method for the Dean-Kawasaki equation achieves this approximation

at a computational cost of

$$C_{MLMC} \leq \begin{cases} \varepsilon^{-2-d/2} & \text{for the nearest-neighbor coupling,} \\ \varepsilon^{-2} & \text{for the Fourier coupling in } d = 1, \\ \varepsilon^{-2}|\log \varepsilon|^2 & \text{for the Fourier coupling in } d = 2, \\ \varepsilon^{-1-d/2} & \text{for the Fourier coupling in } d \geq 3. \end{cases}$$

Note that a standard Monte Carlo method based on the same discretization has a cost of order $\varepsilon^{-3-d/2}$.

References

[1] F. Cornalba, J. Fischer, *The Dean-Kawasaki equation and the structure of density fluctuations in systems of diffusing particles*, Arch. Ration. Mech. Anal. **247** (2023), 76.

[2] F. Cornalba, J. Fischer, *Multilevel Monte Carlo methods for the Dean-Kawasaki equation from fluctuating hydrodynamics*, SIAM J. Numer. Anal. **63** (2025), 262–287.

[3] F. Cornalba, J. Fischer, J. Ingmanns, C. Raithel, *Density fluctuations in weakly interacting particle systems via the Dean-Kawasaki equation*, to appear in Ann. Probab. (2025), arXiv:2303.00429.

[4] F. Cornalba, T. Shardlow, J. Zimmer, *A regularized Dean-Kawasaki model: derivation and analysis*, SIAM J. Math. Anal. **51** (2019), 1137–1187.

[5] D. Dean, *Langevin equation for the density of a system of interacting Langevin processes*, J. Phys. A **29** (1996), L613–L617.

[6] N. Dirr, B. Fehrman, B. Gess, *Conservative stochastic PDE and fluctuations of the symmetric simple exclusion process*, arXiv:2012.02126 (2020).

[7] A. Djurdjevac, H. Kremp, N. Perkowski, *Weak error analysis for a nonlinear spde approximation of the Dean-Kawasaki equation*, Stochastics and Partial Differential Equations: Anal. Comp. **12** (2022), 2330–2355.

[8] B. Fehrman and B. Gess, *Non-equilibrium large deviations and parabolic-hyperbolic PDE with irregular drift*, Invent. Math. **234** (2023), 573–636.

[9] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer. **24** (2015) 259–328.

[10] K. Kawasaki, *Microscopic analyses of the dynamical density functional equation of dense fluids*, J. Statist. Phys. **93** (1998), 527–546.

[11] V. Konarovskyi, T. Lehmann, M.-K. von Renesse, *Dean-Kawasaki dynamics: ill-posedness vs. triviality*, Electron. Commun. Probab. **24** (2019), 8.

[12] L. D. Landau and E. M. Lifshitz, *Course of theoretical physics. Vol. 6. Fluid mechanics*, Pergamon Press, Oxford, second ed., 1987.

[13] H. Spohn, *Large scale dynamics of interacting particles*, Springer Science & Business Media, 2012.

Identification of subwavelength microstructural information from macroscopic boundary measurements in elastodynamics

Malte A. Peter

(joint work with Tanja Lochner)

For durability assessment of structures in use, which have a microstructure, e.g. building structures such as concrete bridges, it is of vital importance to obtain information about the condition of the microstructure, e.g. thinning of reinforcements due to corrosion or the existence of (microscopic) cracks. On the other

hand, measurements of quantities allowing conclusions about the state of the microstructure can typically only be obtained macroscopically and at the surface of the structure. We derive a method for the identification of the geometry of the microstructure, from probing with macroscopic (long) waves, having in mind that the wavelength regime is much larger than the characteristic length of the microstructure. This has the potential of monitoring the structure while in use. The idea is to exploit the macroscopic anisotropy induced by the microstructure to identify microscopic geometrical information, in order to draw conclusions about the state of the structure.

In mathematical terms, we present an inverse homogenisation method for linearly elastic wave propagation allowing the identification of parametrised microstructural geometrical parameters from macroscopic surface displacement data. The derived inverse homogenisation method relies on constraining the parameter-identification technique using mappings from fine-scale parameters to effective parameters obtained by (forward) homogenisation.

The results on the inverse problem are derived under the assumption that the periodicity cell $Y \subset \mathbb{R}^3$ consists of two solids and we investigate the minimisation problem

$$\operatorname*{argmin}_{\tau \in K} \mathcal{J}(\tau) := \frac{1}{2} \|u[\tau] - u_{\mathrm{m}}\|^2_{[L^2(S\times\partial\Omega)]^3} \tag{1}$$

where the parameter vector $\tau \in K$, $K \subset \mathbb{R}^N$ compact, parameterises the geometry of the microstructure, $u[\tau]\colon S \times \Omega \to \mathbb{R}^3$ is the spatio-temporal displacement field for a given τ and u_{m} is the time evolution of the displacement measured on $\partial\Omega = \Gamma_{\mathrm{D}} \dot\cup \Gamma_{\mathrm{N}}$.

The displacement field satisfies the upscaled linear elastodynamics problem, which is given by

$$\begin{cases} \partial_t(\mathcal{M}_Y(\rho)\partial_t u) - \operatorname{div}(A^{\mathrm{hom}} e(u)) = f & \text{in } S \times \Omega, \\ u = 0 & \text{on } S \times \Gamma_{\mathrm{D}}, \\ A^{\mathrm{hom}} e(u) n = g & \text{on } S \times \Gamma_{\mathrm{N}}, \end{cases}$$

and appropriate initial conditions, where the linearised strain tensor $e(u)$ is the symmetric gradient of the displacement field, i.e.

$$e(u) := \frac{1}{2}(\nabla u + (\nabla u)^T)$$

and $A^{\mathrm{hom}} = (A^{\mathrm{hom}}_{ijkl})_{1\le i,j,k,l\le 3}$ given by

$$A^{\mathrm{hom}}_{ijkl}(x) = \frac{1}{|Y|} \int_Y B(x,y) e_{ij}(e_{kl} - e_y(w^{kl})(x,y))\,\mathrm{d}y$$

for a.e. $x \in \Omega$ describes the elastic properties of the homogenised material, where B is the unfolded elasticity tensor (before homogenisation) and the e_{kl} constitute the canonical basis of symmetric matrices. The functions w^{kl}, $k, l \in \{1, 2, 3\}$, are

the unique solutions in $[L^\infty(\Omega, H^1_{\mathrm{per},0}(Y))]^3$ of the cell problems

$$\int_Y B(x,y)\left(e_y(w^{kl})(x,y) - e_{kl}\right) e_y(v)(y)\,\mathrm{d}y = 0$$

for all $v \in [H^1_{\mathrm{per},0}(Y)]^3$, where the subscript "per, 0" denotes periodicity and vanishing mean. Moreover,

$$\mathcal{M}_Y(\rho) := \frac{1}{|Y|}\int_Y \rho(x,y)\,\mathrm{d}x$$

is the local average of the (unfolded) density. The derivation of this upscaled system of equations from the classic linear elastodyamics problem follows by standard periodic homogenisation techniques, see e.g. [1]. In particular, for given data, the solution $u[\tau]$ is uniquely determined.

In the talk, based on [2, 3], we present a method for finding the structure of the unit cell for given f, g and some measured data u_{m}. For this purpose, we define the linear and continuous input–output operator

$$\mathcal{L}_\tau : [L^2(S\times\Omega)]^3 \times H^1(S; [L^2(\Gamma_{\mathrm{N}})]^3) \to [L^2(S\times\partial\Omega)]^3$$

with $(f,g) \mapsto u[\tau]|_{\partial\Omega}$, where $u[\tau]$ is the solution of the homogenised problem above for given τ. We then consider the inverse problem of finding $\tau \in K$ such that for given measured data $u_{\mathrm{m}} \in [L^2(\partial\Omega)]^3$, when forces (f,g) are applied, τ is the solution of the minimisation problem (1).

We prove the existence of at least one solution of the inverse homogenisation problem. Furthermore, to facilitate the use of gradient-based algorithms, we derive the Gâteaux derivative of the functional of the inverse problem making use of shape calculus, where the Gâteaux derivative of the homogenised tensor is computed based on the Lagrangian method of Céa, see [4]. Moreover, we present some numerical experiments for ellipsoidal microstructures showcasing the functioning of the method and showing its robustness with respect to noise.

References

[1] D. Cioranescu, A. Damlamian, G. Griso, *The periodic unfolding method*, Springer, 2018.
[2] T. Lochner, M. A. Peter, *Identification of microstructural information from macroscopic boundary measurements in steady-state linear elasticity*, Math. Meth. Appl. Sci. **46** (2023), 1295–1316.
[3] T. Lochner, M. A. Peter, *Identification of subwavelength microstructural information from macroscopic boundary measurements in elastodynamics*, Proc. Roy. Soc. A **480** (2024), 20240090.
[4] G. Allaire, F. Jouve, N. Van Goethem, *Damage and fracture evolution in brittle materials by shape optimization methods*, J. Comput. Phys. **23** (2011), 5010–5044.

Finite element approximation of Fokker–Planck–Kolmogorov equations with application to numerical homogenization

Timo Sprekeler
(joint work with Endre Süli, Zhiwen Zhang)

In the introductory part of the talk, we consider periodic homogenization of the prototypical elliptic nondivergence-form problem

$$-A\left(\frac{\cdot}{\varepsilon}\right) : D^2 u_\varepsilon = f \quad \text{in } \Omega, \qquad u_\varepsilon = g \quad \text{on } \partial\Omega, \tag{1}$$

posed on a bounded convex domain $\Omega \subset \mathbb{R}^n$, and with a discontinuous diffusion coefficient $A \in L^\infty_{\mathrm{per}}(Y; \mathbb{R}^{n\times n}_{\mathrm{sym}})$ satisfying the Cordes condition. Here, $Y := (0,1)^n$. We present a convergence result for the sequence of strong solutions $(u_\varepsilon)_{\varepsilon>0}$ of (1) to the solution u of an effective constant-coefficient problem, and we discuss corrector bounds as well as optimal convergence rates. See [1, 2] for more details. In particular, the effective diffusion matrix $\overline{A} := \int_Y Ar$ can be obtained from an invariant measure $r \in L^2_{\mathrm{per}}(Y)$, that is, the solution to a stationary Fokker–Planck–Kolmogorov (FPK) equation on Y subject to periodic boundary conditions and a unit integral constraint.

In the main part of the talk, we study the well-posedness and the finite element approximation of "very weak" solutions $r \in L^2_{\mathrm{per}}(Y)$ to the stationary FPK problem

$$-D^2 : (Ar) + \nabla \cdot (br) = 0 \quad \text{in } Y, \qquad r \text{ is } Y\text{-periodic}, \qquad \int_Y r = 1, \tag{2}$$

where the pair of coefficients $(A, b) \in L^\infty_{\mathrm{per}}(Y; \mathbb{R}^{n\times n}_{\mathrm{sym}}) \times L^\infty_{\mathrm{per}}(Y; \mathbb{R}^n)$ is assumed to satisfy the Cordes-type condition

$$\exists\, \delta \in (\delta_0, 1] : \qquad \frac{|A|^2 + |b|^2}{(\mathrm{tr}(A))^2} \le \frac{1}{n - 1 + \delta} \quad \text{a.e.}$$

with $\delta_0 := 0$ if $\|b\|_{L^\infty(Y)} = 0$, and $\delta_0 := \frac{1}{1+4\pi^2}$ otherwise. Here, $|A|^2 := A : A$. The coefficient b in (2) arises in periodic homogenization when the nondivergence-form operator in (1) is modified to include a large drift term $\frac{1}{\varepsilon} b(\frac{\cdot}{\varepsilon}) \cdot \nabla$. We show existence and uniqueness of a very weak solution $r \in L^2_{\mathrm{per}}(Y)$ to the FPK problem (2), and that the solution is given by

$$r = C \frac{\mathrm{tr}(A)}{|A|^2 + |b|^2}(1 - \Delta\psi)$$

for some constant $C > 0$ and some function ψ that solves a Lax–Milgram-type problem in $H^2_{\mathrm{per},0}(Y)$ (subscript 0 denotes zero mean). Using a mixed approach with a $(\mathrm{curl}, \mathrm{curl})_{L^2}$-type stabilization, we will see that $\Delta\psi = \nabla \cdot \rho$ with ρ being the solution of a Lax–Milgram-type problem in $H^1_{\mathrm{per},0}(Y; \mathbb{R}^n)$. This allows for approximating r via an H^1-conforming finite element approximation of ρ. See [3] for more details.

Finally, we conclude the talk with some remarks regarding ongoing work on time-dependent FPK problems.

References

[1] X. Guo, T. Sprekeler, H. V. Tran, *Characterizations of diffusion matrices in homogenization of elliptic equations in nondivergence-form*, Calc. Var. Partial Differential Equations **64** (2025), Paper No. 1.

[2] T. Sprekeler, *Homogenization of Nondivergence-Form Elliptic Equations with Discontinuous Coefficients and Finite Element Approximation of the Homogenized Problem*, SIAM J. Numer. Anal. **62** (2024), 646–666.

[3] T. Sprekeler, E. Süli, Z. Zhang, *Finite Element Approximation of Stationary Fokker–Planck–Kolmogorov Equations with Application to Periodic Numerical Homogenization*, SIAM J. Numer. Anal., in press.

Robust multiscale methods for the Helmholtz equations in highly heterogeneous media

Eric Chung

(joint work with Xingguang Jin, Changqing Ye)

We consider the following Helmholtz equation for heterogeneous media in the bounded space domain $\Omega \subset \mathbb{R}^d$ where $d = 2$ or 3:

$$
\begin{cases} -\nabla \cdot (A\nabla u) - k^2 u = f & \text{in} \quad \Omega, \\ u = 0 & \text{on} \quad \Gamma_D, \\ A\nabla u \cdot \mathbf{n} - \mathrm{i}ku = 0 & \text{on} \quad \Gamma_R, \end{cases} \tag{1}
$$

where $\Gamma = \Gamma_D \cup \Gamma_N \cup \Gamma_R$ is a Lipschitz continuous boundary where $\Gamma_D, \Gamma_N, \Gamma_R$ represents the Dirichlet, Neumann and Robin boundary conditions respectively, $\mathbf{n}$ is the unit outward normal vector to the boundary, $k \in \mathbb{R}$ is a positive wavenumber, $f \in L^2(\Omega)$ represents a harmonic source, i denotes the imaginary unit and the scalar diffusion coefficient A is a piecewise constant with respect to a quadrilateral background mesh $\mathcal{T}_\varepsilon$ with mesh size $O(\varepsilon)$ and $0 < \varepsilon < 1$. On each quadrilateral, A takes either the value ε^2 or 1.

The aim of this paper is to explore an efficient method to solve Heterogeneous Helmholtz problems by using the novel multiscale model reduction skills coming from CEM-GMsFEM, which beyond the need of periodic coefficients or other requirements of the coefficients structures. In the analysis section, we establish a resolution condition and build the inf-sup condition of both global problem and multiscale problem to secure their well-posedness. Subsequently, the exponential decay properties of basis functions are demonstrated, and ultimately, we obtain the error estimate of our multiscale method with the desired the convergence rate. For the first time, we present the evidence supporting of the convergence of CEM-GMsFEM for the Helmholtz equations in heterogeneous media. The numerical simulation section displays the three experiments correspond to three kinds of media, which supports the effectiveness of CEM-GMsFEM and the pollution effect is resolved by using the coarser mesh size to achieve the quasi-optimal convergence. We evaluate the relative error of the CEM-GMsFEM method with respect to different coarse mesh sizes and different oversampling layers. The oversampling layers refer to additional layers of elements or degrees of freedom surrounding the

coarse mesh, which capture the fine-scale details and improve the accuracy of the approximation. The results of the experiments indicate that the relative errors are influenced by the choice of oversampling layers, which distinguishes the CEM-GMsFEM method from traditional FEM. This suggests that the oversampling layers play a significant role in capturing the fine-scale information and reducing the approximation error. We also compare the relative errors obtained with different oversampling layer configurations to demonstrate the impact of these layers on the accuracy of the method.

By highlighting the influence of oversampling layers on the relative errors, we emphasize the advantage of the CEM-GMsFEM method over traditional FEM in capturing fine-scale information and improving the accuracy of the solution. This finding further supports the effectiveness and efficiency of the proposed method in solving Helmholtz equations in heterogeneous domains.

In the following, we use V_{ms} and V_{ms}^* as the new test space and trial space of Petrov-Galerkin method to find the approximated solution, namely, we find $u_{ms} \in V_{ms}$ such that

$$\mathcal{B}(u_{ms}, v) = (f, v), \quad \forall v \in V_{ms}^*. \tag{2}$$

We proved the following theorem.

Theorem 1. *Let u be the solution of the problem* (1) *and u_{ms} be the solution of the multiscale problem* (2)*. Then we have*

$$\|u - u_{ms}\|_a \leq \frac{1}{\varepsilon_0 \sqrt{\Lambda}} \|f\|_{s^{-1}} + C(\Lambda)\sqrt{C_{ol}}\beta^{m/2}(m+1)^{d/2}(C_{inv}\|u^{glo}\|_a + \|\pi u^{glo}\|_s).$$

References

[1] X. Jin, C. Ye, E. Chung, *Robust Multiscale Methods for Helmholtz equations in high contrast heterogeneous media*, arXiv:2407.04364 (2024).

Multiscale Spectral Finite Element Methods: Optimal Spectral Approximation in the Overlaps

Robert Scheichl

(joint work with Christian Alber, Peter Bastian, Moritz Hauck)

Multiscale Spectral Generalized Finite Element Methods (MS-GFEM) [2, 7, 6, 5] are powerful tools for approximating solutions to general variational problems that satisfy a Gårding-type inequality, including strongly non-Hermitian or indefinite problems. The construction of optimal approximation spaces is localised and requires no a priori regularity assumptions on the solution or on the coefficients. The global approximation error is controlled by the local errors, which are rigorously shown to decay nearly exponentially. The optimality hinges on a singular value decomposition of the local restriction operator in a suitable, coefficient-dependent inner product on an oversampled patch. Compactness of this operator in the space of a-harmonic functions guarantees spectral accuracy akin to the Weyl asymptotics for the Laplace eigenvalues. As such, MS-GFEM can be seen as an

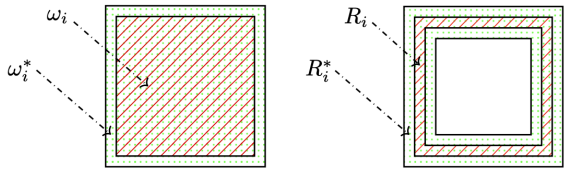


FIGURE 1. Subdomain ω_i with oversampling region ω_i^* (left); overlap R_i with neighbours and oversampling region R_i^* (right).

'hp-version' of Localized Orthogonal Decomposition [8]. Within the generalized FE framework, the optimal local approximation spaces can be used for multiscale discretisation, with links to localized model order reduction method [3]. They can also be used as the coarse space in an overlapping Schwarz preconditioner [9].

However, the construction of the local approximation spaces requires the solution of local eigenproblems, which typically dominate the overall cost of the method, even though they are completely independent and can be solved in parallel without communication. But unless the coarse space is reused multiple times, e.g. in a nonlinear iteration, for multiple right hand sides or for uncertainty quantification, the high cost can make the method less attractive than alternatives.

In [1], we significantly improve the efficiency of the approach for elliptic problems, by restricting the eigenproblems to the vicinity of the overlap while still being able to rigorously prove the nearly exponential decay of the local approximation error. Using ideas in [4], we define a restriction operator to a ring around the subdomain overlap, as shown in Fig. 1, and then solve a modified eigenproblem which provides the best n-dimensional subspace of all a-harmonic functions on this ring. By a-harmonically extending this subspace to the interior of the subdomain we can compute a novel MS-GFEM space. For small overlap and small oversampling the construction is up to $8\times$ cheaper than the optimal MS-GFEM space and still converges with the same nearly exponential rate.

Theoretical results. To give some details, let us consider the model problem

$$a(u,v) = (f,v)_{L^2(\Omega)} \quad \forall\, v \in H_0^1(\Omega), \tag{1}$$

$$\text{with} \qquad a(u,v) = \int_\Omega A(x)\nabla u \cdot \nabla v \,\mathrm{d}x.$$

where $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is bounded and Lipschitz. The only regularity assumptions we make is that $f \in L^2(\Omega)$ and $A \in L^\infty(\Omega)$ is a coercive tensor with eigenvalues bounded between $\alpha_{\min} > 0$ and $\alpha_{\max} < \infty$, uniformly over $x \in \Omega$.

To define the MS-GFEM we consider an overlapping partition $\{\omega_i^*\}_{i=1}^M$ of Ω such that each $x \in \Omega$ belongs to at most $\kappa \in \mathbb{N}$ subdomains, as well as a partition of unity (PoU) $\{\chi_i\}_{i=1}^M \subset W^{1,\infty}(\Omega)$ with $\operatorname{supp}(\chi_i) =: \omega_i \subset \omega_i^*$ with overlap width δ_i and $\|\nabla\chi_i\| = \mathcal{O}(\delta_i^{-1})$. Given local approximation spaces $S_{n_i}(\omega_i) \subset H^1(\omega_i)$ of dimension $n_i \in \mathbb{N}$, the global MS-GFEM approximation $u_n \in S_n(\Omega) \subset H_0^1(\Omega)$

to (1) is sought in the $n = \sum_i n_i$ dimensional generalized FE subspace

$$S_n(\Omega) := \Big\{ \sum\nolimits_i \chi_i \phi_i \,:\, \phi_i \in S_{n_i}(\omega_i) \Big\}.$$

such that $a(u_n, v_n) = (f, v_n)_{L^2(\Omega)}$ for all $v_n \in S_n(\Omega)$. The following is shown in [7].

Theorem 1 (Global MS-GFEM error)**.** *Let $\|v\|_{a,\omega_i^*} := a_{\omega_i^*}(v,v)$ and $a_{\omega_i^*}(\cdot,\cdot)$ be the bilinear form restricted to ω_i^*. Suppose there are functions $\phi_i \in S_{n_i}(\omega_i)$ with*

$$\|\chi_i(u - \phi_i)\|_{a,\omega_i^*} \le \varepsilon_i \|u\|_{a,\omega_i^*}, \quad \textit{for} \;\; i = 1, \cdots, M.$$

Then

$$\|u - u_n\|_a \le \kappa\big(\max_i \varepsilon_i\big)\|u\|_a.$$

To find the optimal n_i-dimensional subspace $S_{n_i}(\omega_i)$, which minimises ε_i, let $\psi_i \in H_0^1(\omega_i^*)$ be the local solution with artificial homogeneous Dirichlet conditions on $\partial\omega_i^*$, i.e. $a_{\omega_i^*}(\psi_i, v) = (f,v)_{L^2(\omega_i^*)}$ for all $v \in H_0^1(\omega_i^*)$. Then $u|_{\omega_i^*} - \psi_i$ is a-harmonic on ω_i^*. The key observation in [7] is that the restriction of a function on ω_i^* by multiplication with the PoU function χ_i to a function in $H_0^1(\omega_i)$ is compact as an operator from the space of a-harmonic functions $H_a(\omega_i^*)$ to $H_0^1(\omega_i)$. Using classical approximation theory for compact linear operators and a suitably defined Kolmogorov n-width we arrive in [7] at the following local approximation result.

Theorem 2 (Optimal local approximation)**.** *Let ψ_i be as defined above and let*

$$S_{n_i}(\omega_i) := \operatorname{span}\big\{\psi_i|_{\omega_i}, v_1|_{\omega_i}, \cdots, v_{n_i-1}|_{\omega_i}\big\}, \tag{2}$$

where v_k denotes the eigenfunction belonging to the kth smallest eigenvalue λ_k of

$$a_{\omega_i^*}(\chi_i v, \chi_i \varphi) = \lambda\, a_{\omega_i^*}(v,\varphi), \quad \forall \varphi \in H_a(\omega_i^*). \tag{3}$$

Then, there exists a $\phi_i \in S_{n_i}(\omega_i)$ such that $\|\chi_i(u-\phi_i)\|_{a,\omega_i^*} \le \lambda_{n_i}^{1/2}\, \|u\|_{a,\omega_i^*}$.

In [1], we show that we can replace ω_i and ω_i^* with R_i and R_i^* (see Fig. 1). We define a new cut-of-function $\chi_i^R := \chi_i - \chi_i^\circ$ by partitioning ω_i into R_i and a suitable chosen interior subdomain ω_i°, such that $\{\chi_i^R, \chi_i^\circ\}_{i=1}^M$ is still a PoU satisfying similar conditions as the original PoU. Exploiting the compactness of the new restriction operator leads again to a local approximation result [1, Thm. 5.1].

Theorem 3 (Optimal local approximation on the overlaps)**.** *Let v_k^R be the eigenfunction belonging to the kth smallest eigenvalue λ_k^R of the new eigenproblem*

$$a_{R_i^*}(\chi_i^R v, \chi_i^R \varphi) = \lambda^R\, a_{R_i^*}(v,\varphi), \quad \forall \varphi \in H_a(R_i^*). \tag{4}$$

Furthermore, we define $v_k \in H_a(\omega_i^*)$ by setting $v_k = v_K^R$ on $\omega_i^* \backslash \omega_i^\circ$ and by extending $v_k^R|_{\partial\omega_i^\circ}$ a-harmonically to the interior ω_i° of ω_i^*. Then, with $S_{n_i}(\omega_i)$ as defined in (2), there exists a $\phi_i \in S_{n_i}(\omega_i)$ such that $\|\chi_i(u-\phi_i)\|_{a,\omega_i^*} \le \big(\lambda_{n_i}^R\big)^{1/2}\, \|u\|_{a,\omega_i^*}$.

Using the same arguments as in [5] based on Weyl asymptotics, we show an analogous nearly exponential convergence result for the local error in [1, Thm. 5.3].

Theorem 4 (Subexponential decay of eigenvalues)**.** *Assume that R_i and R_i^* are concentric rings with dist$(R_i, \partial R_i^*) \geq \delta_i^* > 0$. There exist constants $C_{d,\delta_i^*}, c_{d,\delta_i^*} > 0$ and $n_0 \in \mathbb{N}$ (explicit) such that*

$$\lambda_{n_i}^R \leq C_{d,\delta_i^*} e^{-c_{d,\delta_i^*} n_i^{1/d}} \quad \forall n_i \geq n_0.$$

Similar results to Theorems 3–4 can also be proved for more general boundary conditions, other elliptic PDEs or fully discrete MS-GFEM restricted to the overlaps, following [6, 5]. The results also carry over to the analysis of MS-GFEM as a coarse spaces within a restricted additive Schwarz (RAS) preconditioner [9].

Numerical results. Within an iterative method the number of iterations with the new MS-GFEM coarse space are almost identical to those with the original MS-GFEM [1, Table 1 & 2]. But due to the simpler connectivity pattern of the resulting matrices and (to a lesser degree) also their size, the eigenproblems in (4) can be solved much faster than (3), up to 8× for an example on a 3D cubic domain [1, Fig. 7.3]. However, even though the subexponential convergence rate in general is the same, roughly 2× as many eigenfunctions are required to achieve the same approximation error [1, Fig. 7.2].

References

[1] C. Alber, P. Bastian, M. Hauck, R. Scheichl, *Optimal spectral approximation in the overlaps for generalized finite element methods*, Preprint, 2025.

[2] I. Babuska, R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul. **9** (2011), 373–406.

[3] A. Buhr, K. Smetana. *Randomized local model order reduction*, SIAM J. Sci. Comput. **40** (2018), A2120–A2151.

[4] A. Heinlein, A. Klawonn, J. Knepper, O. Rheinbach, *Adaptive GDSW coarse spaces for overlapping Schwarz methods in 3D*, SIAM J. Sci. Comput. **41** (2019), A3045–A3072.

[5] C. Ma, *A unified framework for multiscale spectral generalized FEMs and low-rank approximations to multiscale PDEs*, Found. Comput. Math. (2025).

[6] C. Ma, R. Scheichl, *Error estimates for discrete generalized FEMs with locally optimal spectral approximations*, Math. Comput. **91** (2022), 2539–2569.

[7] C. Ma, R. Scheichl, T. Dodwell, *Novel design and analysis of generalized FE methods based on locally optimal spectral approximations*, SIAM J. Numer. Anal. **60** (2022), 244–273.

[8] A. Målqvist, D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp. **83** (2014), 2583–2603.

[9] A. Strehlow, C. Ma, R. Scheichl, *Fast-convergent two-level restricted additive Schwarz methods based on optimal local approximation spaces*, arXiv:2408.16282 (2024).

On Edge Multiscale Space based Hybrid Schwarz Preconditioner for Helmholtz Problems with Large Wavenumbers

Guanglian Li

(joint work with Shubin Fu, Shihua Gong, Yueqi Wang)

We develop a novel hybrid Schwarz method, termed as edge multiscale space based hybrid Schwarz (EMs-HS), for solving the Helmholtz problem with large wavenumbers. The problem is discretized using H^1-conforming nodal finite element methods on meshes of size h decreasing faster than k^{-1} such that the discretization error remains bounded as the wavenumber k increases. EMs-HS consists of a one-level Schwarz preconditioner (RAS-imp) and a coarse solver in a multiplicative way. The RAS-imp preconditioner solves local problems on overlapping subdomains with impedance boundary conditions in parallel, and combines the local solutions using partition of unity. The coarse space is an edge multiscale space proposed in [2]. The key idea is to first establish a local splitting of the solution over each subdomain by a local bubble part and local Helmholtz harmonic extension part, and then to derive a global splitting by means of the partition of unity. This facilitates representing the solution as the sum of a global bubble part and a global Helmholtz harmonic extension part.

We prove that the EMs-HS preconditioner leads to a convergent fixed-point iteration uniformly for large wavenumbers, by rigorously analyzing the approximation properties of the coarse space to the global Helmholtz harmonic extension part and to the solution of the adjoint problem. Distinctly, the theoretical convergence analysis are valid in two extreme cases: using minimal overlapping size among subdomains (of order h), or using coarse spaces of optimal dimension (of magnitude k^d, where d is the spatial dimension). We provide extensive numerical results on the sharpness of the theoretical findings and also demonstrate the method on challenging heterogeneous models. The development of a coarse problem with dimension grows only linearly with respect to the wavenumber k is beyond this current work.

References

[1] S. Fu, S. Gong, G. Li, Y. Wang, *On Edge Multiscale Space based Hybrid Schwarz Preconditioner for Helmholtz Problems with Large Wavenumbers*, arXiv:2408.08198 (2024).

[2] S. Fu, G. Li, R. Craster, S. Guenneau, *Wavelet-based edge multiscale finite element method for Helmholtz problems in perforated domains*, Multiscale Model. Simul. **19** (2021), 1684–1709.

A higher-order localized orthogonal decomposition strategy

Roland Maier

In this contribution, we discuss the extension of the localized orthogonal decomposition (LOD) method (see, e.g., [6, 1]) to a higher-order multiscale method, making use of key properties of the LOD approach and adopting ideas from the technique known as gamblets, cf. [7].

Elliptic multiscale problem. First, we consider an elliptic setting in weak form. In a Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, we seek the solution $u \in H_0^1(\Omega)$ to

$$a(u, v) := \int_\Omega A\nabla u \cdot \nabla v \,\mathrm{d}x = \int_\Omega fv \,\mathrm{d}x =: (f, v)_{L^2(\Omega)} \tag{1}$$

for all $v \in H_0^1(\Omega)$. Here, $A \in L^\infty(\Omega)$ denotes a rough and potentially strongly varying diffusion coefficient that is bounded from above and below by positive constants and f is a given right-hand side. It is well-known that classical finite element methods perform poorly for such problems if the scale on which A varies is not sufficiently resolved. To overcome this issue, many multiscale strategies have been developed and we aim at constructing a multiscale approach that achieves higher-order convergence rates under minimal regularity assumptions.

The main ingredient to derive a higher-order multiscale method is the definition of appropriate coarse-scale quantities that a discrete solution should preserve. More precisely, we aim at constructing a discrete space $\tilde{V}_H \subset H_0^1(\Omega)$ such that the Galerkin approximation $\tilde{u}_H \in \tilde{V}_H$ to (1) fulfills $\Pi\tilde{u}_H = \Pi u$, where u is the unique solution to (1) and $\Pi\colon L^2(\Omega) \to V_H$ the L^2-projection into a discontinuous piecewise polynomial space V_H. For simplicial meshes $\mathcal{T}_H$ with coarse mesh size H, the space V_H is given by piecewise polynomials of total degree p, whereas piecewise polynomials of partial degree p are considered if the mesh elements are d-rectangles. Moreover, we want to have that $\dim(\tilde{V}_H) = \dim(V_H)$. If the above conditions are satisfied, we directly calculate using the Galerkin orthogonality

$$\begin{aligned} \|\nabla(u - \tilde{u}_H)\|_{L^2(\Omega)}^2 &\lesssim a(u - \tilde{u}_H, u - \tilde{u}_H) = a(u, u - \tilde{u}_H) \\ &= (f, u - \tilde{u}_H)_{L^2(\Omega)} \\ &= (f - \Pi f, (u - \tilde{u}_H) - \Pi(u - \tilde{u}_H))_{L^2(\Omega)}. \end{aligned} \tag{2}$$

With classical approximation results of the L^2-projection, we can estimate

$$\|(u - \tilde{u}_H) - \Pi(u - \tilde{u}_H)\|_{L^2(\Omega)} \lesssim H\|\nabla(u - \tilde{u}_H)\|_{L^2(\Omega)}$$

and

$$\|f - \Pi f\|_{L^2(\Omega)} \lesssim H^{p+1}|f|_{H^{p+1}(\mathcal{T}_H)}$$

provided that f is (piecewise) regular enough. Therefore, we obtain an error bound of the form

$$\|\nabla(u - \tilde{u}_H)\|_{L^2(\Omega)} \lesssim H^{p+2}|f|_{H^{p+1}(\mathcal{T}_H)}. \tag{3}$$

That is, we achieve higher orders of convergence from (piecewise) regularity of the right-hand side only. In particular, we do not need the coefficient A or the solution u to fulfill higher-order regularity assumptions. Nonetheless, the error estimate (3) is only useful if we are able to construct the space $\tilde{V}_H$ with the desired properties explicitly. It turns out that $\tilde{V}_H$ can be built with the basis functions

$$\tilde{\Lambda}_j = \arg\min_{v \in H_0^1(\Omega)\,:\,\Pi v = \Lambda_j} a(v, v),$$

where $\{\Lambda_j\}_j$ is a basis of V_H. Practically, these basis functions are approximated in a fine finite element space and their computation is restricted to local patches due to favorable decay properties, see [5] for details. Moreover, the method can

be formulated in the framework of the LOD, which is particularly useful to obtain improved localization results, cf. [2].

Parabolic multiscale problem. As a second step, we apply the higher-order approach to time-dependent problems. For simplicity, we restrict ourselves to an idealized setting, excluding aspects regarding localization and fine-scale discretization. Exemplary, we consider a parabolic problem that seeks $u \in L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$ such that

$$(\dot{u},v)_{L^2(\Omega)} + a(u,v) = (f,v)_{L^2(\Omega)} \tag{4}$$

for almost all $t \in (0,T]$ and all $v \in H_0^1(\Omega)$ with an appropriate right-hand side f and a rough and potentially strongly varying coefficient A as before. Unfortunately, the space $\tilde{V}_H$ constructed above turns out to not be a suitable approximation space for the spatial discretization of (4). This is related to the fact that the estimate derived in (2) is tailored to the elliptic case. That is, even for the a-orthogonal projection of u at any time t into the space $\tilde{V}_H$ we obtain a suboptimal rate of order two only. This has been observed in [4] for the wave equation, but the same observations can be made for the parabolic problem (4) as well. To illustrate the issue, let $\mathcal{P}\colon H_0^1(\Omega) \to \tilde{V}_H$ be the a-orthogonal projection onto $\tilde{V}_H$. Further, we note that $v - \mathcal{P}v \in W := \ker(\Pi|_{H_0^1(\Omega)})$ for any $v \in H_0^1(\Omega)$ and $a(\tilde{V}_H, W) = \{0\}$ by construction. Then for any time t (omitting the explicit time dependence in the following estimates), we have analogously to (2)

$$\begin{aligned}
\|\nabla(u-\mathcal{P}u)\|_{L^2(\Omega)}^2 &\lesssim a(u-\mathcal{P}u, u-\mathcal{P}u) = a(u, u-\mathcal{P}u) \\
&= (f, u-\mathcal{P}u)_{L^2(\Omega)} - (\dot{u}, u-\mathcal{P}u)_{L^2(\Omega)} \\
&= (f-\Pi f, (u-\mathcal{P}u) - \Pi(u-\mathcal{P}u))_{L^2(\Omega)} \\
&\qquad - (\dot{u}-\Pi\dot{u}, (u-\mathcal{P}u) - \Pi(u-\mathcal{P}u))_{L^2(\Omega)}.
\end{aligned} \tag{5}$$

The last line in (5) now leads to suboptimal rates, as spatial regularity of $\dot{u}$ beyond $H^1(\Omega)$ cannot be expected in the general case where $A \in L^\infty(\Omega)$. In particular, the argument to derive higher orders in (2) is not fully applicable. To still retain optimal rates, we require an appropriate enrichment of the space $\tilde{V}_H$. To this end, we define the operator $\mathcal{D}\colon L^2(\Omega) \to W$ by

$$a(\mathcal{D}v, w) = (v,w)_{L^2(\Omega)}$$

for all $w \in W$. One can show that replacing $\mathcal{P}u$ by $\mathcal{P}u - \mathcal{D}\dot{u}$ in (5) cancels the problematic term on the right-hand side. That is, optimal rates in the sense of (3) can be obtained. These observations motivate the use of $\tilde{V}_H \cup \mathcal{D}\tilde{V}_H$ as an enriched multiscale space for the spatial discretization of (4). This essentially corresponds to an approximation of $\dot{u}$ in the discrete space $\tilde{V}_H$, which appears to be reasonable. However, choosing $\tilde{V}_H \cup \mathcal{D}\tilde{V}_H$ as the spatial discretization space combined with a suitable time-stepping scheme does not remove the reduced rates completely. Instead, one obtains that the reduced rates are now of order four and the relevant threshold where reduced rates are observed is lowered as well. Practically, this is sufficient as the threshold is small enough to not be problematic.

From a theoretical point of view, one can extend the above argument and further enrich the space with $\mathcal{D}^2\tilde{V}_H$, $\mathcal{D}^3\tilde{V}_H$, etc. until the convergence rate of order $p+2$ is retained.

A thorough theoretical analysis of this idea (including a suitable localization strategy) and a set of numerical experiments will be presented and discussed in [3].

References

[1] R. Altmann, P. Henning, D. Peterseim, *Numerical homogenization beyond scale separation*, Acta Numer. **30** (2021), 1–86.

[2] Z. Dong, M. Hauck, R. Maier, *An improved high-order method for elliptic multiscale problems*, SIAM J. Numer. Anal. **61** (2023), 1918–1937.

[3] B. Kalyanaraman, F. Krumbiegel, R. Maier, S. Wang, *Optimal higher-order convergence rates for parabolic multiscale problems*, In preparation (2025).

[4] F. Krumbiegel, R. Maier, *A higher-order multiscale method for the wave equation*, IMA J. Numer. Anal. (2024), published online.

[5] R. Maier, *A high-order approach to elliptic multiscale problems with general unstructured coefficients*, SIAM J. Numer. Anal. **59** (2021), 1067–1089.

[6] A. Målqvist, D. Peterseim, *Numerical homogenization by localized orthogonal decomposition*, volume 5 of SIAM Spotlights, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2020.

[7] H. Owhadi, C. Scovel, *Operator-adapted wavelets, fast solvers, and numerical homogenization*, volume 35 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2019.

Super-localized numerical homogenization

Moritz Hauck

(joint work with Philip Freese, Tim Keil, Daniel Peterseim)

On a polygonal Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d \in \{1,2,3\}$, we consider the prototypical elliptic multiscale problem $-\operatorname{div}(A\nabla u) = f$ subject to homogeneous Dirichlet boundary conditions, with a coefficient $A \in L^\infty(\Omega)$ such that $\alpha \leq A(x) \leq \beta$ for all $x \in \Omega$ and a source term $f \in L^2(\Omega)$. No additional structural assumptions are imposed on the coefficient; in particular, smoothness, periodicity, and scale separation are not required. This report, which is based on the works [1, 2], presents a numerical homogenization method for the considered elliptic model problem that achieves uniformly optimal approximation rates, regardless of the roughness of the coefficient. For this, the method utilizes problem-adapted basis functions whose supports grow mildly with the desired accuracy. Specifically, to obtain an approximation error of $\mathcal{O}(H)$, the diameter of the supports of the basis functions practically need to be increased as $\mathcal{O}(H|\log H|^{(d-1)/d})$. This represents a significant improvement over the best-known localization results for the Localized Orthogonal Decomposition (LOD) method [3], which require supports of diameter $\mathcal{O}(H|\log H|)$. Other notable methods that also achieve optimal approximation rates independent of coefficient regularity include, e.g., the AL basis [4] and generalized finite element methods [5, 6], which yield qualitatively similar results.

1. Optimal operator-dependent approximation

The weak formulation of the elliptic model problem is based on the bilinear form $a(u,v) := \int_\Omega (A\nabla u)\cdot\nabla v\,\mathrm{d}x$, which defines an inner product on the Sobolev space $H^1_0(\Omega)$ with induced norm $\|\cdot\|_a^2 := a(\cdot,\cdot)$. For any $f \in L^2(\Omega)$, the unique weak solution $u \in H^1_0(\Omega)$ satisfies $a(u,v) = (f,v)_{L^2}$ for all $v \in H^1_0(\Omega)$, and we denote by $\mathcal{A}^{-1}\colon L^2(\Omega) \to H^1_0(\Omega)$ the corresponding solution operator. To construct optimal approximation spaces, we introduce a coarse mesh $\mathcal{T}_H$ over Ω, which typically does not resolve the fine-scale variations of A. The space is given by $V_H := \mathcal{A}^{-1}\mathbb{P}^0(\mathcal{T}_H)$, where $\mathbb{P}^0(\mathcal{T}_H)$ denotes the space of $\mathcal{T}_H$-piecewise constant functions. The discrete solution operator $\mathcal{A}_H^{-1}\colon L^2(\Omega) \to V_H$, which we define as the Galerkin projection onto V_H, satisfies the error bound

$$\sup_{f\in L^2(\Omega)} \frac{\|\mathcal{A}^{-1}f - \mathcal{A}_H^{-1}f\|_a}{\|f\|_{L^2}} \le \alpha^{-1/2}\pi^{-1}H; \tag{1}$$

see, e.g., [1, Lem. 3.2]. On the contrary, classical finite elements would require H^2-regularity of the solution, typically unavailable for rough coefficients.

2. Novel localization strategy

The canonical basis functions $\{\mathcal{A}^{-1}\mathbf{1}_K : K \in \mathcal{T}_H\}$ of the operator-adapted space V_H are inherently non-local, with only algebraic decay. As a result, their direct localization introduces large errors, rendering this naive method computationally impractical. To address this, we introduce a novel localization strategy that selects $\mathcal{T}_H$-piecewise constant source terms yielding rapidly decaying (or even local) responses under the solution operator $\mathcal{A}^{-1}$. Given an element $K \in \mathcal{T}_H$ and oversampling parameter $\ell \in \mathbb{N}$, we denote by $\omega_{K,\ell}$ the patch of ℓ coarse layers around K. The associated ideal basis function $\varphi_{K,\ell} \in V_H$ is defined as $\varphi_{K,\ell} := \mathcal{A}^{-1}g_{K,\ell}$ with $g_{K,\ell} \in \mathbb{P}^0(\mathcal{T}_H \cap \omega_{K,\ell})$ to be determined. A localized approximation of it is obtained by applying the localized solution operator $\mathcal{A}_{K,\ell}^{-1}\colon L^2(\omega_{K,\ell}) \to H^1_0(\omega_{K,\ell})$, giving $\varphi_{K,\ell}^{\mathrm{loc}} := \mathcal{A}_{K,\ell}^{-1}g$. Although $\varphi_{K,\ell}^{\mathrm{loc}}$ in general is a poor approximation to $\varphi_{K,\ell}$, a suitable choice of $g_{K,\ell}$ yields a highly accurate approximation in the energy norm.

To characterize such a $g_{K,\ell}$, we introduce the trace operator $\gamma_{K,\ell}\colon H^1(\omega_{K,\ell}) \to H^{1/2}(\partial\omega_{K,\ell})$ and its right inverse $\gamma_{K,\ell}^{-1}\colon H^{1/2}(\partial\omega_{K,\ell}) \to H^1(\omega_{K,\ell})$, given by the A-harmonic extension. Using the definition of $\varphi_{K,\ell}$, the A-harmonic extension property, and the locality of $g_{K,\ell}$, we obtain, for all $v \in H^1_0(\Omega)$, the key identity

$$a(\varphi_{K,\ell} - \varphi_{K,\ell}^{\mathrm{loc}}, v) = (g_{K,\ell}, v)_{L^2(\omega_{K,\ell})} - a_{\omega_{K,\ell}}(\varphi_{K,\ell}^{\mathrm{loc}}, v) = (g_{K,\ell}, \gamma_{K,\ell}^{-1}\gamma_{K,\ell}v)_{L^2(\omega_{K,\ell})}.$$

It shows that the localization error is characterized by the L^2-orthogonality of $g_{K,\ell}$ to the space $Y := \operatorname{im}\gamma_{K,\ell}^{-1}$ of A-harmonic functions on $\omega_{K,\ell}$, i.e.,

$$g_{K,\ell} \in \operatorname*{arg\,min}_{q\in\mathbb{P}^0(\mathcal{T}_H\cap\omega_{K,\ell})\colon \|q\|_{L^2(\omega_{K,\ell})}=1} \ \sup_{v\in Y\colon \|v\|_{H^1(\omega_{K,\ell})}=1} (q,v)_{L^2(\omega_{K,\ell})} \tag{2}$$

represents an optimal choice of the local source term. The function $g_{K,\ell}$ can be computed via a (random) singular value decomposition of the L^2-projection operator onto piecewise constants, restricted to Y. The corresponding localization error, which is the minimal value associated with $g_{K,\ell}$ in (2), is denoted by $\sigma_{K,\ell}$.

3. Super-localized orthogonal decomposition

The localized approximation space of the proposed method, referred to as the super-localized orthogonal decomposition (SLOD), is defined by

$$V_{H,\ell} := \operatorname{span}\{\varphi_{K,\ell}^{\mathrm{loc}} : K \in \mathcal{T}_H\}$$

with associated Galerkin projection $\mathcal{A}_{H,\ell}^{-1}\colon L^2(\Omega) \to V_{H,\ell}$. To estimate the approximation error, we define the maximal localization error as $\sigma := \max_{K\in\mathcal{T}_H} \sigma_{K,\ell}$. A key assumption to ensure the stability of the method is that $\{g_{K,\ell} : K \in \mathcal{T}_H\}$ forms a Riesz basis of $\mathbb{P}^0(\mathcal{T}_H)$. That is, there exists a constant $C_{\mathrm{rb}} > 0$, depending algebraically on H and ℓ, such that, for all $(c_K)_{K\in\mathcal{T}_H}$, it holds that

$$C_{\mathrm{st}}^{-1} \sum_{K\in\mathcal{T}_H} c_K^2 \le \Big\| \sum_{K\in\mathcal{T}_H} c_K g_{K,\ell} \Big\|_{L^2}^2 \le C_{\mathrm{st}} \sum_{K\in\mathcal{T}_H} c_K^2, \tag{3}$$

where we recall that the local source terms $g_{K,\ell}$ are L^2-normalized by construction.

In [1, Thm. 6.1] we proved the following error estimate for the localized method

$$\sup_{f\in L^2(\Omega)} \frac{\|\mathcal{A}^{-1}f - \mathcal{A}_{H,\ell}^{-1}f\|_a}{\|f\|_{L^2}} \le C\big(H + C_{\mathrm{st}}^{1/2}(H,\ell)\ell^{d/2}\sigma(H,\ell)\big),$$

where the first term on the right-hand side is the error of the ideal method given in (1). The second term on the right-hand side, denoted by ε, decays rapidly with increasing oversampling parameter ℓ. Numerical experiments in Figure 1 for $d \in \{2,3\}$ suggest a super-exponential decay of the form $\exp(-c\ell^{d/(d-1)})$, which appears linear under the chosen axis scaling. In contrast, classical LOD methods exhibit only exponential decay and significantly larger errors. While a rigorous proof of this super-exponential behavior remains open, classical LOD arguments guarantee at least exponential decay of σ; see [1, Lem. 6.4]. Regarding the stability constant C_{st}, we have proposed an algorithm that ensures a robust selection of the local source terms in the sense of (3) which works for a wide range of multiscale problems. However, numerical experiments in [2, Sec. 7] reveal its failure in cases with high-contrast channeled coefficients. To address this, we have introduced a method combining the present approach with partition of unity techniques. It retains the super-localization properties and achieves LOD-like error estimates without requiring Riesz stability of the basis; see [2, Thm. 5.3 & Thm. 6.1].

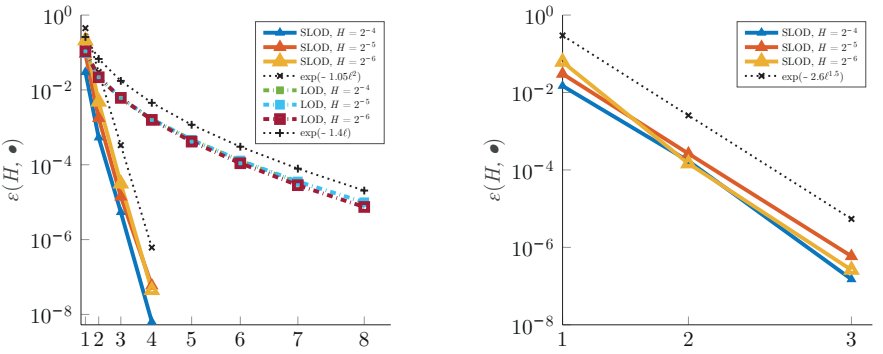


FIGURE 1. Decay of ε in two (left) and three (right) spatial dimensions. The dashed lines indicate the respective LOD errors

REFERENCES

[1] M. Hauck, D. Peterseim, *Super-localization of elliptic multiscale problems*, Math. Comp. **92** (2023), 341–981.

[2] P. Freese, M. Hauck, T. Keil, D. Peterseim, *A super-localized generalized finite element method*, Numer. Math. **156** (2024), 205–235

[3] A. Målqvist, D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp. **83** (2014), 2583–2603.

[4] L. Grasedyck, T. Greff, S. Sauter, *The AL basis for the solution of elliptic problems in heterogeneous media*, Multiscale Model. Simul. **10** (2012), 245–258.

[5] I. Babuška, R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul. **9** (2011), 373–406.

[6] C. Ma, R. Scheichl, T. Dodwell, *Novel Design and Analysis of Generalized Finite Element Methods Based on Locally Optimal Spectral Approximations*, SIAM J. Numer. Anal. **60** (2022), 244-273.

Multicontinuum homogenization and applications

YALCHIN EFENDIEV

(joint work with Wing T. Leung)

Many problems in nature have complex multiscale nature and high contrast. Their numerical solutions pose significant challenges due to resolving scales and contrast via subgrid models. Many approaches are developed to account for subgrid effects. These approaches, e.g., [3, 6, 4, 5, 9, 1, 8] include the construction of multiscale basis functions that are supported in domain larger than the target coarse block. Among these approaches, the CEM-GMsFEM [5] is related to the multicontinuum approach presented here.

Multicontinuum homogenization is formally introduced in [7, 2]. In this method, several macroscopic variables at each macroscale point are defined and the resulting multicontinuum equations are formulated. The main idea of multicontinuum

approach is to develop macroscopic models using smoothly varying macroscopic variables. The derivation of multicontinuum approach consists of several main parts. In the first part, we propose a general expansion, where the solution is expressed via the product of multiple macro variables and associated cell problems. The second part consists of formulating the cell problems. The cell problems are formulated as saddle point problems with constraints for each continua. Defining the continua via test functions, we set the constraints as an integral representation. Finally, substituting the expansion to the original system, we obtain multicontinuum systems.

The motivation for the derivation comes from previously developed approaches such as GMsFEM, NLMC, CEM-GMsFEM. If we denote U to consist of all macroscopic variables and Π^{CEM} is a projection of macroscopic variables, then

$$\Pi^{CEM}(u) \approx \Pi^{CEM}(\psi_i U_i).$$

If we consider U_i are smooth and Π^{CEM} to be local, then we further obtain

$$\begin{aligned}\Pi^{CEM}(u) &\approx \Pi^{CEM}(\psi_i \mathbf{1} U_i(x_0^*) + \psi_i (x - x_0^*) \cdot \nabla U_i(x^0)) \\ &\approx \Pi^{CEM}(\psi_i \mathbf{1}) U_i(x_0^*) + \Pi^{CEM}(\psi_i (x - x_0^*)) \cdot \nabla U_i(x_0^*) \\ &= \phi_i U_i(x_0^*) + \phi_i^m \partial_m U_i(x_0^*),\end{aligned}$$

in the region near x_0^* where $\phi_i := \Pi^{CEM}(\psi_i \mathbf{1})$ and $\phi_i^m := (\Pi^{CEM}(\psi_i (x - x_0^*)))_m$.

Multicontinuum homogenization is applied for different problems, where new macroscopic models are derived. In the talk, we discussed an example of poroelastic problem consisting of coupled flow and mechanics problems. We derive the corresponding multicontinuum displacement and pressure equations using multicontinuum expansions. As a result, we obtain a general multicontinuum poroelasticity model for an arbitrary number of continua, which generalizes existing multinetwork poroelasticity models.

One of our goals with the proposed approach is to derive multicontinuum models for complex nonlinear problems, such as two-phase flow and so on, where continua evolves according to some physical laws. The main idea is to use dynamic multicontinua concept, where one of the unknowns define the multicontinua. A simple example consists of flow and transport, $-\nabla \cdot (\lambda(c)\nabla p) = q$, $c_t + v\nabla c = 0$, $v = -\lambda(c)\nabla p$. The problem is set up such that the concentration defines continua (which is smooth), e.g., ith continua is defined when the value of concentration belongs to a certain interval. With this definition of the continua, we can derive new macroscopic models, where microscale heterogeneities are hidden in macroscopic variables. Using dynamic multicontinua concept, we plan to derive various models that are not been known before.

References

[1] R. Altmann, P. Henning, D. Peterseim, *Numerical homogenization beyond scale separation*, Acta Numer. **30** (2021), 1–86.

[2] E. Chung, Y. Efendiev, J. Galvis, W. T. Leung, *Multicontinuum homogenization. general theory and applications*, J. Comput. Phys. **510** (2024), 112980.

[3] E. Chung, Y. Efendiev, T. Y. Hou, *Adaptive multiscale model reduction with generalized multiscale finite element methods*, J. Comput. Phys. **320** (2016), 69–95.
[4] E. Chung, Y. Efendiev, T. Y. Hou, *Multiscale Model Reduction*, Springer, 2023.
[5] E. T. Chung, Y. Efendiev, W. T. Leung, *Constraint energy minimizing generalized multiscale finite element method*, Comput. Methods Appl. Mech. Engrg. **339** (2018), 298–319.
[6] Y. Efendiev, J. Galvis, T. Y. Hou, *Generalized multiscale finite element methods (gmsfem)*, J. Comput. Phys. **251** (2013), 116–135.
[7] Y. Efendiev, W. T. Leung, *Multicontinuum homogenization and its relation to nonlocal multicontinuum theories*, J. Comput. Phys. **474** (2023), 111761.
[8] A. Målqvist, D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp. **83** (2014), 2583–2603.
[9] H. Owhadi, L. Zhang, *Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic odes/pdes with rough coefficients*, J. Comput. Phys. **347** (2017), 99–128.

Randomized Multiscale Methods for Heterogeneous Nonlinear PDEs

Kathrin Smetana

(joint work with Charles Beall, Tommaso Taddei, Marissa Whitby, Zhiyu Yin)

For linear elliptic problems it was shown e.g., in [1, 12, 18] that the local ansatz basis functions for multiscale methods can be chosen as the leading left singular vectors of a linear compact so called transfer operator that acts on the space of all local solutions of the partial differential equation (PDE). This operator maps arbitrary functions on the boundary of an oversampling domain that is strictly larger than the target subdomain for which we wish to construct our multiscale ansatz space, to the corresponding (local) solution of the PDE restricted to the target subdomain. It can then be shown [1, 12, 14, 18] that these local ansatz spaces are optimal in the sense of Kolmogorov [11] and minimize the local approximation error among all spaces of the same dimension. Optimal local approximation spaces for parabolic problems have been derived in [15, 16].

For nonlinear PDEs we also consider a transfer operator that maps arbitrary admissible boundary data on the boundary of the oversampling domain to the respective (local) solution on the target subdomain [19]. As the concept of optimality used for linear PDEs is only defined for linear operators, we try to approximate the set of all local solutions on the target subdomain. Interpreting the boundary data as some input parameter, we can view this set of local solutions as a set of solutions depending on a parameter. This motivates using methods from model order reduction such as the proper orthogonal decomposition (POD) [9, 17] or the Greedy algorithm [22] to approximate this set. As several results (see e.g., [4, 7, 9, 10]) show that these algorithms can provide a quasi-optimal approximation to this set, we can hope that this approach might lead to multiscale methods that converge at a quasi-optimal rate.

However, both the POD and the Greedy algorithm rely on a training set of finite cardinality that is chosen such that every point in the admissible parameter set is close to a point in the training set. Therefore, both algorithms suffer from the curse of dimensionality [2, 3] requiring a potentially computationally prohibitive large

training set if the dimension of the parameter set is high. This is the case in our setting as the dimension of the parameter set equals the dimension of the finite element space on the boundary of the oversampling domain. We thus employ randomization and consider the parameter (here: boundary data) as a random variable with values in a Hilbert space. By choosing a suitable distribution we can then exploit the concentration of measure phenomenon (see e.g., [13, 5]) that is also sometimes called the "blessing of dimensionality" to break the curse. One example of the concentration of measure phenomenon that we are using is that a Lipschitz function of a standard Gaussian random vector has a sub-Gaussian distribution (see e.g., [5]), meaning that the tails of the probability density function decay at least as fast as that of the Gaussian, resulting in many favorable properties.

In [19] we suggested to use a randomized POD to construct the local multiscale ansatz spaces for nonlinear elliptic problems. Here, we use (as usual) a Monte-Carlo approximation of the correlation operator. Numerical results for a parametric nonlinear diffusion problem in Fig. 1 show an exponential convergence of the approximation error for the randomized POD; for details on the test case see [19]. Moreover, we derived an a priori error bound for the randomized POD that shows that the expectation of the projection error for the POD using the Monte-Carlo approximation converges (nearly) with the same rate as the eigenvalues of the randomized POD which uses the exact correlation operator (integral in the expectation) and thus (quasi-)optimally; see [20] and references therein. One drawback of the randomized POD is that it does not allow us to assess whether the distribution of the random variable on the boundary of the oversampling domain was chosen well enough. Therefore, we have also been investigating a randomized Greedy algorithm.

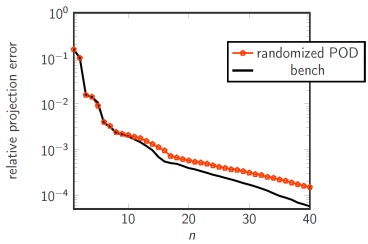


FIGURE 1. Benchmark local ansatz space is computed from global solutions of problem.

We were able to derive a randomized Greedy algorithm that provides a certification for the whole parameter set rather than only for the parameters in the training set, which, to the best of our knowledge, sets our result apart from existing approaches. The basic idea of the Greedy algorithm [22] is that in each iteration we search for the parameter in the training set whose corresponding solution is worst approximated by the current reduced space. Having some way of measuring the error or estimating it via an a posteriori error estimator is thus crucial. In our setting, we face the problem that we cannot compute the projection error in the standard norm as this would require us to compute the supremum of the error over the whole parameter set, which is, except for very rare special cases, not possible. As for linear problems [6, 8] we are thus trying to estimate the supremum norm by the maximum over the evaluations of the error for certain parameter values exploiting so-called Marcinkiewicz inequalities (see e.g., [21]).

References

[1] I. Babuška, R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul. **9** (2011), 373–406.

[2] R. Bellman, *Dynamic programming*, Princeton University Press, Princeton, NJ, 1957.

[3] R. Bellman, *Adaptive control processes: A guided tour*, Princeton University Press, Princeton, N.J., 1961.

[4] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*, SIAM J. Math. Anal. **43** (2011), 1457–1472.

[5] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities. A nonasymptotic theory of independence*, Oxford University Press, Oxford, 2013.

[6] A. Buhr, K. Smetana, *Randomized Local Model Order Reduction*, SIAM J. Sci. Comput. **40** (2018), A2120–A2151.

[7] R. DeVore, G. Petrova, P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constr. Approx. **37** (2013) 455–466.

[8] N. Halko, P. G. Martinsson, J. A. Tropp, *Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev. **53** (2011), 217–288.

[9] P. Holmes, J. L. Lumley, G. Berkooz, C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, 2012.

[10] M. Kahlbacher, S. Volkwein, *Galerkin proper orthogonal decomposition methods for parameter dependent elliptic systems*, Discuss. Math. Differ. Incl. Control Optim. **27** (2007), 95–117.

[11] A. Kolmogoroff, *Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse.* Ann. of Math. **37** (1936), 107–110.

[12] C. Ma, R. Scheichl, T. Dodwell, *Novel design and analysis of generalized finite element methods based on locally optimal spectral approximations*, SIAM J. Numer. Anal. **60** (2022), 244–273.

[13] M. Ledoux, *The concentration of measure phenomenon*, volume 89 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2001.

[14] A. Pinkus, *N-Widths in Approximation Theory*, Ergeb. Math. Grenzgeb. **7**, Springer, Berlin, 1985.

[15] J. Schleuß, K. Smetana, *Optimal local approximation spaces for parabolic problems*, Multiscale Model. Simul. **20** (2022), 551–582.

[16] J. Schleuß, K. Smetana, L. ter Maat, *Randomized quasi-optimal local approximation spaces in time*, SIAM J. Sci. Comput. **45** (2023), A1066–A1096.

[17] L. Sirovich, *Turbulence and the dynamics of coherent structures. I. Coherent structures*, Quart. Appl. Math. **45** (1987), 561–571.

[18] K. Smetana, A. T. Patera, *Optimal local approximation spaces for component-based static condensation procedures*, SIAM J. Sci. Comput. **38** (2016), A3318–A3356.

[19] K. Smetana, T. Taddei, *Localized Model Reduction for Nonlinear Elliptic Partial Differential Equations: Localized Training, Partition of Unity, and Adaptive Enrichment*, SIAM J. Sci. Comput. **45** (2023), A1300–A1331.

[20] K. Smetana, T. Taddei, M. Whitby, Z. Yin, *Probabilistic Error Analysis of Proper Orthogonal Decomposition and its Application to Multiscale Problems*, in preparation, 2025.

[21] V. N. Temlyakov, *The Marcinkiewicz-type discretization theorems*, *Constr. Approx.* **48** (2018), 337–369.

[22] K. Veroy, C. Prud'homme, D. V. Rovas, A. T. Patera, *A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, In Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, volume 3847, 2003.

Numerical simulation of beam network models

Axel Målqvist
(joint work with Morgan Görtz, Moritz Hauck, Fredrik Hellman, Andreas Rupp)

Many applications in science and engineering involve geometrically complex structures composed of slender, effectively one-dimensional objects. A prominent example is fiber-based materials [5], see Figure 1 for an illustration. For such problems, resolving all microscopic details in a three-dimensional computer simulation is computationally demanding. Therefore, it is appropriate to describe the geometry by a spatial network represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ of nodes and edges which is embedded into a domain $\Omega \subset \mathbb{R}^3$. The resulting spatial network model then

Figure 1. Fiber network model of paper at the millimeter scale.

involves one-dimensional systems of ordinary differential equations on each edge, coupled by algebraic constraints at the graph nodes. Here we considers the elastic deformation of fiber-based materials, such as paper or cardboard, as a model problem. The spatial network underlying this problem is constructed as follows: Nodes are placed at the intersections of fibers, and an edge connects two nodes if a fiber is connecting them. Depending on the intersection area of the two fibers, additional nodes and edges may be added to strengthen the connection between the fibers. To obtain a spatial network model, we equip each edge with a Timoshenko beam model, cf. [1], describing the elastic deformation of the corresponding fiber. On each edge $\mathfrak{e}$ in the network we solve

$$-C_{\vec{n}}(\partial_x \vec{u}_{\mathfrak{e}} + \vec{i}_{\mathfrak{e}} \times \vec{r}_{\mathfrak{e}}) = \vec{n}_{\mathfrak{e}}, \qquad -C_{\vec{m}}\partial_x \vec{r}_{\mathfrak{e}} = \vec{m}_{\mathfrak{e}},$$
$$\partial_x \vec{n}_{\mathfrak{e}} = \vec{f}_{\mathfrak{e}}, \qquad \partial_x \vec{m}_{\mathfrak{e}} + \vec{i}_{\mathfrak{e}} \times \vec{n}_{\mathfrak{e}} = \vec{g}_{\mathfrak{e}}.$$

Here $\vec{u}_{\mathfrak{e}}$ denotes centerline displacement, $\vec{r}_{\mathfrak{e}}$ is crosssection rotation, and $\vec{i}_{\mathfrak{e}}$ is the unit vector in the direction of the edge. The stresses resulting from normal and shear forces are denoted $\vec{n}_{\mathfrak{e}}$ and the moments resulting from torsion and bending is denoted $\vec{m}_{\mathfrak{e}}$. Material parameters are represented by the matrices $C_{\vec{n}}$ and $C_{\vec{m}}$ and, finally, $\vec{f}_{\mathfrak{e}}$ and $\vec{g}_{\mathfrak{e}}$ denotes the distributed forces and moments.

A well-posed problem is obtained by enforcing the continuity of displacements and rotations and the balance of forces and moments at the nodes of the spatial

network. The displacements and rotations are prescribed in Dirichlet nodes and balance conditions ensure an equilibrium of forces and moments at non-Dirichlet nodes. This means that for concentrated forces and moments at the nodes $\vec{f}_{\mathfrak{n}}, \vec{g}_{\mathfrak{n}} \in \mathbb{R}^3$ it should hold that

$$-\left[\!\left[\vec{n}_{\mathfrak{e}}\nu_{\mathfrak{e}}\right]\!\right]_{\mathfrak{n}} = -\vec{f}_{\mathfrak{n}}, \qquad -\left[\!\left[\vec{m}_{\mathfrak{e}}\nu_{\mathfrak{e}}\right]\!\right]_{\mathfrak{n}} = -\vec{g}_{\mathfrak{n}},$$

where $[\![\cdot]\!]_{\mathfrak{n}}$ is a sum over all values that are attained at $\mathfrak{n}$ and $\nu_{\mathfrak{e}} = \pm 1$ represents the normal.

We propose a hybridizable discontinuous Galerkin (HDG) method to discretize the spatial network model. The use of such discretization is motivated by the study in [6], where the authors discretize diffusion-type problems on networks of hypersurfaces. In the special case of graphs, the nodes at which the equations are coupled are zero-dimensional objects, whence the spatial network problem can be equivalently reformulated as a symmetric positive definite system of linear equations posed on the network nodes. For such problems, an HDG discretization can achieve arbitrary convergence orders without increasing the number of globally coupled degrees of freedom. Following the paradigm of HDG methods that local solves are essentially for free since the corresponding problems are small and they can be solved in parallel, cf. [2], the proposed method can achieve arbitrary orders of convergence at (almost) constant computational cost. We prove the following a priori error bound for the HDG method, see [4]. The bar notation refers to the quantities computed using the HDG method.

Theorem 1 (Convergence of HDG method)**.** *Suppose that for any edge $\mathfrak{e} \in \mathcal{E}$ it holds that $\vec{u}_{\mathfrak{e}}, \vec{r}_{\mathfrak{e}}, \vec{n}_{\mathfrak{e}}, \vec{m}_{\mathfrak{e}} \in H^{p+1}(\mathfrak{e})$. Then, given a stabilization parameter which scales like $h_{\mathfrak{e}}^s$ for some $s \in \{-1, 0, 1\}$, the HDG approximation converges to the solution of the Timoshenko beam network model with the error estimates*

$$\left[\sum_{\mathfrak{e}\in\mathcal{E}} \left[\|\vec{u}_{\mathfrak{e}} - \bar{\vec{u}}_{\mathfrak{e}}\|_{\mathfrak{e}}^2 + \|\vec{r}_{\mathfrak{e}} - \bar{\vec{r}}_{\mathfrak{e}}\|_{\mathfrak{e}}^2\right]\right]^{1/2} \leq Ch^{p+1-s^+},$$

$$\left[\sum_{\mathfrak{e}\in\mathcal{E}} \left[\|\vec{n}_{\mathfrak{e}} - \bar{\vec{n}}_{\mathfrak{e}}\|_{\mathfrak{e}}^2 + \|\vec{m}_{\mathfrak{e}} - \bar{\vec{m}}_{\mathfrak{e}}\|_{\mathfrak{e}}^2\right]\right]^{1/2} \leq Ch^{p+1-|s|},$$

where we denote $s^+ := \max(s, 0)$.

Due to the complex geometry of the spatial network and possibly highly varying material coefficients, the linear system of equations obtained by the HDG method is typically very poorly conditioned. Numerical experiments demonstrate that standard black-box preconditioners, such as many algebraic multigrid variants (see, e.g., the review article [7]), can typically not significantly speed up convergence. This is because they do not sufficiently consider the underlying problem's geometry. To overcome these difficulties, this paper employs a preconditioner based on the observation that the network can be treated essentially as a continuous object at sufficiently coarse scales. This allows one to place an artificial coarse mesh over the network and use finite element techniques on this mesh to introduce a two-level overlapping Schwarz preconditioner similar to [3]. We prove

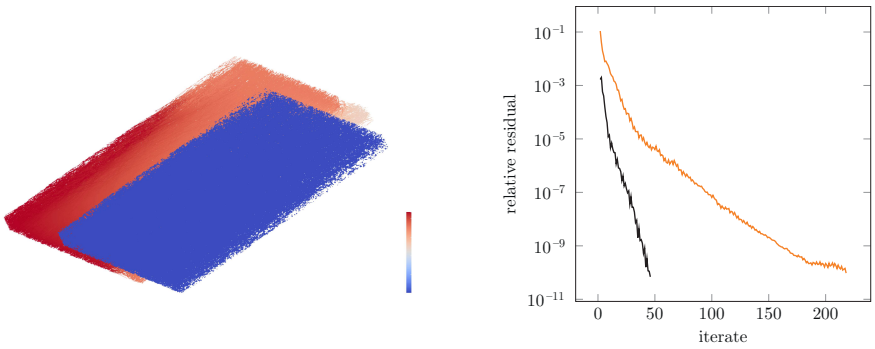


FIGURE 2. The network before (blue) and after (red) deformation (left). Relative residual as a function of the iteration number for constant (black) and realistic (orange) material parameters (right).

that the global system matrix is spectrally equivalent to an edge-length weighted graph Laplacian (in each component). Under certain network connectivity assumptions, this allows us to prove the uniform convergence of the corresponding preconditioned conjugate gradient method.

Finally, we present numerical experiments where our methodology is applied to a realistic example. We consider the elastic deformation of a 4 mm x 4 mm piece of paper. The spatial network consists of about 615K edges and 424K nodes. We consider the stretching of the paper caused by inhomogeneous Dirichlet boundary conditions at nodes at the lateral boundary, see Figure 2 (left). We use a HDG discretization with polynomial degree 5 and penalty 1 for all edges. The resulting linear system of equations is then solved by the preconditioned conjugate gradient method with the domain decomposition preconditioner. The results are presented in Figure 2 (right). We detect rapid convergence both for constant (orange) and realistic (black) constant material parameters.

REFERENCES

[1] E. Carrera, G. Giunta, M. Petrolo, *Beam Structures: Classical and Advanced Theories,* Wiley, 2011.

[2] B. Cockburn, J. Gopalakrishnan, R. Lazarov, *Unified Hybridization of Discontinuous Galerkin, Mixed, and Continuous Galerkin Methods for Second Order Elliptic Problems,* SIAM J. Numer. Anal. **47** (2009) 1319–1365.

[3] M. Görtz and F. Hellman, A. Målqvist, *Iterative solution of spatial network models by subspace decomposition*, Math. Comp. **93** (2024), 233–258.

[4] M. Hauck, A. Målqvist, A. Rupp, *Arbitrary order approximations at constant cost for Timoshenko beam network models,* arXiv:2407.14388 (2024).

[5] G. Kettil, A. Målqvist, A. Mark, M. Fredlund, K. Wester, F. Edelvik, *Numerical upscaling of discrete network models*, BIT Numer. Math. **60** (2020), 67–92.

[6] A. Rupp, M. Gahn, G. Kanschat, *Partial differential equations on hypergraphs and networks of surfaces: Derivation and hybrid discretizations,* ESAIM Math. Model. Numer. Anal. **56** (2022), 505–528.

[7] J. Xu, L. Zikatanov, *Algebraic multigrid methods*, Acta Numer. **26** (2017), 591–721.

Localized implicit time stepping for the wave equation

Dietmar Gallistl
(joint work with Roland Maier)

The Crank–Nicolson discretization of the acoustic wave equation

$$\partial_t^2 u - \nabla \cdot (A \nabla u) = f$$

over a suitable domain Ω of $\mathbb{R}^d$ with right-hand side f and initial and boundary conditions is an implicit scheme that involves solving the linear system

$$(u_h^{k+1}, v_h)_{L^2(\Omega)} + \frac{\tau^2}{4}(A \nabla u_h^{k+1}, v_h) = G(v_h) \tag{1}$$

where τ is the time step size and u_h^{k+1} is the discrete solution at time step $k+1$, while G is some linear functional involving data and the discrete solution from prior time steps.

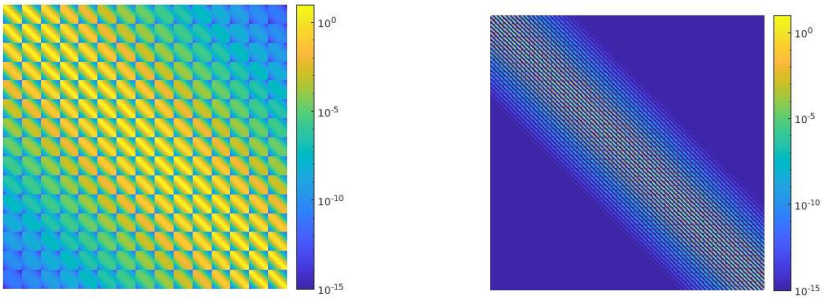


Figure 1. Values of the inverse system matrix of the Crank–Nicolson scheme in two dimensions on a uniform and lexicographically ordered mesh with logarithmic color coding; mesh size (and time step) $h = \tau = 2^{-4}$ (left) and $h = \tau = 2^{-6}$ (right).

The left-hand side of (1) is a discrete elliptic problem, whence the discrete solution will have global support even if G is supported locally. Since the Crank–Nicolson scheme is a convergent method, the finite propagation speed of the wave equation suggests that the artificial amplitudes outside the cone of propagation must be of negligible magnitude. This is illustrated in Figure 1, where the entries of the inverse system matrix are displayed in logarithmic color coding. The observation that there is some exponential decay away from the diagonal can be made rigorous. The result from [1] states that for data supported in a subset $\omega \subseteq \Omega$, the estimate

$$\|\!| u_h^{n+1/2} |\!\|_{\Omega \setminus \mathbb{N}_\ell(\omega)} \leq (\ell+1)^{n/2} \gamma^\ell \max_{k \leq n} \|\!| u_h^{k+1/2} |\!\| \quad \text{for all } n, \ell \in \mathbb{N}$$

holds. The norm $\|\!|\cdot|\!\|$ is the energy norm corresponding to the left-hand side of (1) and $\mathtt{N}_\ell(\omega)$ is the domain arising from adding ℓ layers of neighbouring elements from the finite element triangulation to ω. The constant γ satisfies

$$0 < \gamma = \sqrt{(C_{\tau,h} + \tfrac{1}{2})/(1 + C_{\tau,h})} < 1$$

for some $C_{\tau,h} \eqsim \beta\alpha^{-1}(\tau/h + h/\tau)$, where $\alpha \leq \beta$ are the positive lower and upper bounds of the diffusion coefficient A. This means that the solution decays, exponentially in ℓ, away from the region ω where the sources are supported. The proof of this result is inspired by techniques from [2, 3]. A conclusion of this result is that, although the matrix inversion related to each time step of the implicit scheme transports information globally over Ω, relevant information decays fast and can be captured by solving a system over a smaller sub-domain, which implies less computational cost. Global initial data and sources can be localized through a partition of unity on some coarser spatial scale H and the discrete solution can be defined by superposition of the solutions to local subproblems that can be solved in parallel. After some reset time T, this procedure is then repeated. This procedure involves the additional artificial scales T, H. The approach may be understood as a domain decomposition strategy in space on successive coarse time intervals that does neither require multiple iterations nor a sophisticated definition of boundary conditions between the different sub-regions. It is proven in [1] that such a superposition method produces an approximation that has an overall approximation error comparable to that of the classical Crank–Nicolson, provided ℓ is chosen appropriately, where the involved overhead is logarithmic compared to the natural local domain occupied by the expanding wave until the local time horizon T.

References

[1] D. Gallistl, R. Maier, *Localized implicit time stepping for the wave equation*, SIAM J. Numer. Anal. **62** (2024), 1589–1608.
[2] A. Målqvist, D. Peterseim, Localization of elliptic multiscale problems, Math. Comp. **83** (2014), 2583–2603.
[3] A. Målqvist, D. Peterseim, *Numerical homogenization by localized orthogonal decomposition*, vol. 5 of SIAM Spotlights, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2020.

Offline-Online Approximation of Multiscale Eigenvalue Problems with Random Defects

Dilini Kolombage
(joint work with Barbara Verfürth)

Multiscale materials, such as metamaterials, derive their unique mechanical, acoustic, or electromagnetic properties from finescale structural features. However, real world manufacturing introduces random defects, leading to uncertain material behavior. This gives rise to elliptic eigenvalue problems with multiscale, randomly

perturbed coefficients. Such problems arise in modeling the response of materials and structures where randomness and finescale features are intertwined. The presented material is based on the detailed manuscript [1].

We consider the variation elliptic eigenvalue problem: Find $(\lambda, u) \in \mathbb{R} \times V$ such that

$$a(u, v) = \lambda m(u, v), \quad \forall v \in V,$$

with $a(u, v) = \int_\Omega A\nabla u \cdot \nabla v dx$, and $m(u, v) = \int_\Omega uv dx$. The coefficient A is defined such that $A := A(x, \omega) = A_\varepsilon(x) + b_p(x; \omega)B_\varepsilon(x)$, with A_ε, B_ε deterministic periodic coefficients with periodic length ε, and b_p a random Bernoulli perturbation modeling random defect inclusions.

Standard finite element methods become computationally infeasible in this context because they require extremely fine meshes to resolve the smallest scales. Moreover, when one needs to compute eigenpairs for many random samples, the cost multiplies significantly.

To address these challenges, we develop an efficient offline-online computational multiscale strategy based on the Localized Orthogonal Decomposition (LOD) method. Our approach leverages precomputed local quantities from reference defect configurations, enabling rapid assembly of the system matrices for new random realizations. We also introduce a heuristic modification in the online phase which enhances the performance, particularly in scenarios with high defect probabilities. We provide rigorous a priori error estimates for both eigenfunctions and eigenvalues.

1. The LOD based Offline-online strategy

We use the LOD method to construct a low-dimentional, problem-adapted multiscale space $V_{H,k}^{\mathrm{ms}}$, where finescale corrections are localized on patches $U_k(T) \subseteq \Omega$. Our formulation employees a Petrov-Galerkin variant of the LOD for the stiffness matrix, while the mass matrix is taken from the standard finite element method.

A key challenge is that the multiscale space $V_{H,k}^{\mathrm{ms}}$ depends on the realization A. Our offline-online strategy addresses this by avoiding the need to re-compute the multiscale basis for every realization, thereby significantly reducing the overall computational cost. While the offline phase remains costly, it is only performed once for a single patch and is independent of the realization. In contrast, the online phase – which assembles the global stiffness matrix in real time for any given realization A – is highly efficient.

- **Offline phase:** On a patch $U_k(T)$, construct a basis of reference coefficients A_i using single-defect configurations. Then, for each A_i, precompute and store the corresponding LOD stiffness matrix contributions.
- **Online phase:** Given a random realization A, decompose it locally as $\overline{A} = \sum_i \mu_i A_i$, with $\sum_i \mu_i = 1$ and $\mu_i \in \{0, 1\}$. Next, combine the precomputed stiffness matrix contributions from the offline phase to assemble the global stiffness matrix.

Moreover, for certain coefficients, the constraint $\sum_i \mu_i = 1$ can be generalized to $\sum_i \mu_i = s$ for $s \in \mathbb{R}^+$. A heuristic strategy for selecting s to minimize the consistency error has been proposed and is further detailed in [1].

2. Error Analysis

We derive a priori error estimates using spectral perturbation theory following the Babuŝka-Osborn framework. Denoting by (λ, u) the continuous eigenpairs, and $(\tilde{\lambda}_H, \tilde{u}_H)$ the offline-online approximations, we show in [1] the following holds:

- Eigenfunction error (energy norm):

$$||u - \tilde{u}_H||_{\mathcal{A}} \lesssim H^2 + g(k) + \left(\max_T E_T\right) k^{d/2},$$

- Relative eigenvalue error:

$$\left|\frac{\lambda - \tilde{\lambda}_H}{\lambda}\right| \lesssim H^2 + g(k) + \left(\max_T E_T\right) k^{d/2} + \left(\max_T E_T\right)^2 k^2,$$

where $g(k) \approx H^2$ for sufficiently large patch size k, and E_T is a computable error indicator that quantifies the local consistency error due to offline-online strategy.

3. Numerical Experiments

We validate the method on one- and two-dimensional models with random checkerboard coefficients, where finescale random defects are inserted on a constant background Cartesian grid. As shown in Figure 1, both methods perform almost similarly, yielding less than 2% RMSE for defects probabilities below 10%. However, for moderate to high defect probabilities, the heuristic approach – which adapts s to the defect probability – demonstrates significantly improved accuracy.

References

[1] D. Kolombage, B. Verfürth, *Offline-online approximation of multiscale eigenvalue problems with random defects*, arXiv:2411.19614 (2024).

[2] A. Målqvist, B. Verfürth, *An offline-online strategy for multiscale problems with random defects*, ESAIM Math. Model. Numer. Anal. **56** (2022), 237–260.

[3] A. Målqvist, D. Peterseim, *Numerical homogenization by localized orthogonal decomposition*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2020.

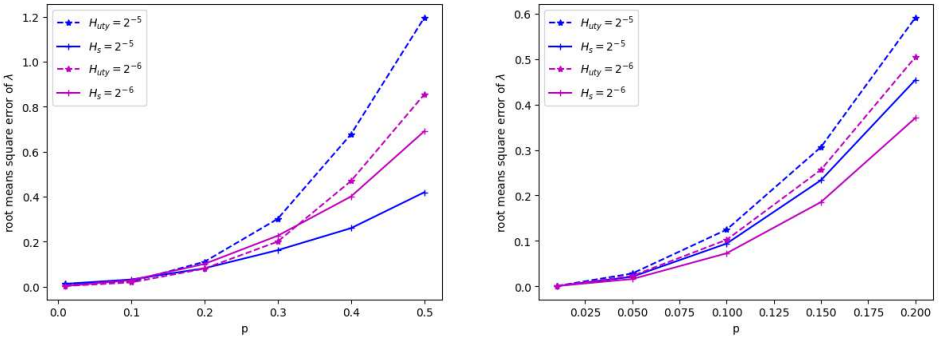


FIGURE 1. The RMSE against probability variation over 200 samples, using a patch size of 3 element layers around a central element, for 1D (*left*) and 2D (*right*) settings. The discretizations H_{uty} and H_{s} correspond to the sum-constraint-one model and the probability-adaptive-s heuristic model respectively.

Resolution of Ginzburg-Landau energy minimizers using multiscale techniques

CHRISTIAN DÖDING

(joint work with Maria Blum, Benjamin Dörich, Patrick Henning)

Superconductivity is the physical phenomenon by which certain materials lose all electrical resistance when cooled to very low temperatures. A superconductor of convex domain Ω in two or three space dimensions is modeled by a minimizing pair $u \in H^1(\Omega, \mathbb{C})$ and $\mathbf{A} \in \mathbf{H}^1(\Omega, \mathbb{R}^d)$ of the Ginzburg-Landau (GL) free energy (cf. [4, 6])

$$E_{\text{GL}}(u, \mathbf{A}) = \frac{1}{2} \int_\Omega \left| \frac{\mathrm{i}}{\kappa} \nabla u + \mathbf{A} u \right|^2 + \frac{1}{2} \left(1 - |u|^2\right)^2 + |\operatorname{curl} \mathbf{A} - \mathbf{H}_{\text{ext}}|^2 + |\operatorname{div} \mathbf{A}|^2 \mathrm{d}x$$

where $\mathbf{H}_{\text{ext}}$ is a suitable externally applied magnetic field. The first component, u, known as the order parameter, characterizes the superconducting state of the material ($|u| = 1$ superconducting, $|u| = 0$ non-superconducting) and the second component, $\mathbf{A}$, denotes the electromagnetic vector potential inside of the superconductor. A key parameter in the model is the Ginzburg–Landau parameter $\kappa > 0$ which is typically large (for type-II superconductors) and gives rise to the exhibition of characteristic vortex structures in the minimizing order parameter – the Abrikosov vortex lattice – see Figure 1. As κ increases, the number of vortices grows while their size decreases and hence κ can be seen as multiscale parameter. From a numerical perspective, this raises two central questions:

- What mesh resolution is required, in terms of κ, to resolve the vortex lattice?
- What are efficient approximation spaces for capturing vortex structures?

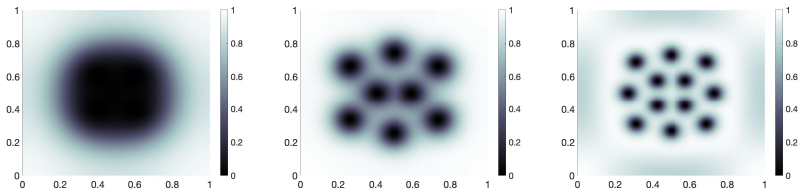


FIGURE 1. Density $|u|^2$ of minimizing order parameters for the GL-parameter $\kappa = 8$ (left), $\kappa = 16$ (middle) and $\kappa = 32$ (right).

To address these questions, we consider a simplified model where a divergence-free vector potential $\mathbf{A} \in \mathbf{L}^\infty(\Omega, \mathbb{R}^d)$ is give. The energy functional reduces to:

$$E(u) = \frac{1}{2}\int_\Omega \left|\frac{\mathrm{i}}{\kappa}\nabla u + \mathbf{A}u\right|^2 + \frac{1}{2}\big(1 - |u|^2\big)^2 \mathrm{d}x.$$

It is known, cf. [3, 4], that a minimizer u exists, is H^2-regular, and satisfies $\|u\|_{H^k} \lesssim \kappa^k$ for $k = 0, 1, 2$. While global uniqueness is unclear, local uniqueness is plausible under phase invariance. More precise, we assume local uniqueness of the minimizer in the subspace $H^1 \cap (\mathrm{i}u)^\perp$, where $(\mathrm{i}u)^\perp$ denotes the L^2-orthogonal complement of the neutral mode arising from the phase symmetry. Under this quasi-uniqueness assumption, the second derivative of the energy is coercive on this subspace, cf. [1, 3]:

$$\langle E''(u)v, v\rangle \geq \rho(\kappa)^{-1}\|v\|^2_{H^1_\kappa} \quad \forall v \in H^1 \cap (\mathrm{i}u)^\perp$$

where $\|v\|^2_{H^1_\kappa} = \kappa^{-2}\|\nabla v\|^2_{L^2} + \|v\|^2_{L^2}$ and $\rho(\kappa)$ is the inverse coercivity constant, which typically grows with κ. Coercivity of the second derivative is crucial for analyzing minimizers; however, the smallness of the coercivity constant poses challenges regarding the approximation properties of minimizers in discrete spaces and their computation with iterative methods.

Resolution condition in classical finite element spaces

The approximation of minimizers using classical finite elements has been studied in [3], emphasizing κ-dependent error estimates. Let V_h be the first-order H^1-conforming Lagrange finite element space over a mesh of size $h > 0$, and define $V_h^\perp = V_h \cap (\mathrm{i}u)^\perp$. Then the best-approximation error satisfies (cf. [3]):

$$\inf_{v \in V_h^\perp} \|u - v\|_{H^1_\kappa} \lesssim \kappa h.$$

This implies a basic resolution condition of $h \lesssim \kappa^{-1}$. However, the actual requirement is more restrictive, as established in the following result:

Theorem 1. [3, Theorem 3.3] *For h sufficiently small let $u_h \in V_h$ be a discrete minimizer of E, i.e., $E(u_h) = \min_{v\in V_h} E(v)$. Then there is a unique exact minimizer $u \in H^1$ of E in a neighborhood of u_h such that $u_h \in (\mathrm{i}u)^\perp$ and*

$$\|u - u_h\|_{H^1_\kappa} \lesssim (1 + \rho(\kappa)\kappa h) \inf_{v\in V_h^\perp} \|u - v\|_{H^1_\kappa} \lesssim \kappa h + \rho(\kappa)\kappa^2 h^2.$$

Thus, a stronger resolution condition $h \lesssim \rho(\kappa)^{-1/2}\kappa^{-1}$ is required, which aligns with numerical observations.

Multiscale approximation spaces

To overcome these limitations, we employ multiscale techniques to construct approximation spaces that relax the resolution condition and better capture vortex lattice structures. Consider the bilinear form

$$a_\beta(u, v) = (\tfrac{\mathrm{i}}{\kappa}\nabla u + \mathbf{A}u, \tfrac{\mathrm{i}}{\kappa}\nabla v + \mathbf{A}v)_{L^2} + \beta(u, v)_{L^2}$$

which is induced by the kinetic part of the GL energy and where $\beta \geq 0$ is a stabilization parameter. While $a_\beta(\cdot,\cdot)$ is not coercive on all of H^1 for $\beta = 0$, it is coercive on the kernel of the L^2-projection $P_h : H^1 \to V_h$, allowing us to define a multiscale space in the spirit of the *Localized Orthogonal Decomposition* (cf. [5])

$$V_h^{\mathrm{ms}} = (\mathrm{id} - \mathcal{C}_\beta)V_h, \quad \text{where} \quad \mathcal{C}_\beta : H^1 \to W \text{ solves } a_\beta(\mathcal{C}_\beta v - v, w) = 0 \quad \forall w \in W.$$

Note that $\dim V_h = \dim V_h^{\mathrm{ms}}$ and we proved in [1] the best-approximation estimate

$$\inf_{v\in(V_h^{\mathrm{ms}})^\perp} \|u - v\|_{H^1_\kappa} \lesssim (1 + \beta)\kappa^3 h^3$$

with $(V_h^{\mathrm{ms}})^\perp = V_h^{\mathrm{ms}} \cap (\mathrm{i}u)^\perp$ by using only the H^2-regularity of the minimizer. In addition, we proved in [1] the following estimate for the discrete minimizer in the multiscale space:

Theorem 2. [1, Theorem 3.1] *For h sufficiently small let $u_h^{\mathrm{ms}} \in V_h^{\mathrm{ms}}$ be a discrete minimizer of E, i.e., $E(u_h^{\mathrm{ms}}) = \min_{v\in V_h^{\mathrm{ms}}} E(v)$. Then there is a unique exact minimizer $u \in H^1$ of E in a neighborhood of u_h^{ms} such that $u_h^{\mathrm{ms}} \in (\mathrm{i}u)^\perp$ and*

$$\|u - u_h^{\mathrm{ms}}\|_{H^1_\kappa} \lesssim (1 + \rho(\kappa)\kappa h) \inf_{v\in(V_h^{\mathrm{ms}})^\perp} \|u - v\|_{H^1_\kappa} \lesssim (1 + \beta)(\kappa^3 h^3 + \rho(\kappa)\kappa^4 h^4).$$

This significantly improves the resolution condition to $h \lesssim (1+\beta)^{-1/4}\rho(\kappa)^{-1/4}\kappa^{-1}$ and the choice $\beta = 0$ is preferable. The multiscale framework allows higher-order accuracy and relaxes the resolution condition without requiring additional regularity or adding degrees of freedom. Numerical experiments in [1] confirm that this approach captures the vortex lattice efficiently.

These ideas have been extended to the full GL model in [2], where similar error estimates hold for discrete minimizers $(u, \mathbf{A})$ using a multiscale space for u and standard finite elements for $\mathbf{A}$ (first or second order). The vector potential is observed to exhibit smoother behavior without strong vortex features, so a standard FEM treatment suffices. For solving the full coupled problem, we propose an iterative scheme with an adaptive updating strategy for the multiscale space based on

the current iteration of $\mathbf{A}$, leading to promising computational and approximation results for Ginzburg-Landau energy minimizers in numerical simulations.

References

[1] M. Blum, C. Döding, P. Henning, *Vortex-capturing multiscale spaces for the Ginzburg-Landau equation*, Multiscale Model. Simul. **23** (2025), 339–373.

[2] C. Döding, B. Dörich, P. Henning, *A multiscale approach to the stationary Ginzburg-Landau equations of superconductivity*, arXiv:2409.12023 (2025).

[3] B. Dörich, P. Henning, *Error bounds for discrete minimizers of the Ginzburg-Landau energy in the high κ-regime*, SIAM J. Numer. Anal., **27** (2025), 1313–1343.

[4] Q. Du, M.D. Gunzburger, J.S. Peterson, *Analysis and approximation of the Ginzburg-Landau model of superconductivity*, SIAM Rev., **34** (1992), 54–81.

[5] A. Målqvist, D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp., **83** (2014), 2583–2603.

[6] M. Tinkham, *Introduction to Superconductivity*, McGraw-Hill, New York, 1975.

Reporter: Christian Döding

Participants

Jonah Botvinick-Greenhouse
Cornell University Center for Applied Mathematics
Ithaca, NY 14853
UNITED STATES

Prof. Dr. Susanne C. Brenner
Department of Mathematics
Louisiana State University
303 Lockett Hall
Baton Rouge LA 70803-4918
UNITED STATES

Prof. Dr. Eric T. Chung
Department of Mathematics
The Chinese University of Hong Kong
Room 220, Lady Shaw Building
Shatin, N.T., Hong Kong SAR
CHINA

Matthias Deiml
Institut für Mathematik
Universität Augsburg
Universitätsstraße 12a
86159 Augsburg
GERMANY

Dr. Christian Döding
Institut für Numerische Simulation
Universität Bonn
Friedrich-Hirzebruch-Allee 7
53115 Bonn
GERMANY

Prof. Dr. Yalchin Efendiev
Department of Mathematics
Texas A&M University
College Station, TX 77843-3368
UNITED STATES

Prof. Dr. Björn Engquist
Department of Mathematics
The University of Texas at Austin
2515 Speedway
Austin TX 78712-0257
UNITED STATES

Prof. Dr. Michael Feischl
Institute for Analysis and Scientific Computing
TU Wien
Wiedner Hauptstraße 8-10
1040 Wien
AUSTRIA

Prof. Dr. Julian Fischer
Institute of Science and Technology Austria
(IST Austria)
Am Campus 1
3400 Klosterneuburg
AUSTRIA

Prof. Dr. Christina A. Frederick
Department of Mathematical Sciences
New Jersey Institute of Technology
606 Cullimore Hall
Newark NJ 07102-1982
UNITED STATES

Prof. Dr. Dietmar Gallistl
Institut für Mathematik
Universität Jena
Ernst-Abbe-Platz 2
07743 Jena
GERMANY

Theron Guo
Department of Mechanical Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge MA 02139-4307
UNITED STATES

Moritz Hauck
Karlsruhe Institute of Technology
Institute of applied and numerical mathematics
76131 Karlsruhe
GERMANY

Prof. Dr. Patrick Henning
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum
GERMANY

Martin Hermann
Institut für Mathematik
Universität Augsburg
86135 Augsburg
GERMANY

Cheuk Hin Ho
Department of Mathematics
University of British Columbia
121-1984 Mathematics Road
Vancouver BC V6T 1Z2
CANADA

Prof. Dr. Thomas Y. Hou
Applied Mathematics 9-94
California Institute of Technology
1200 E. California Blvd
Pasadena, CA 91125
UNITED STATES

Junpeng Hu
School of Mathematical Sciences
Shanghai Jiao Tong University
Shanghai Shi 200 240
CHINA

Maher Khrais
Institut für Numerische Simulation
Universität Bonn
Friedrich-Hirzebruch-Allee 7
53115 Bonn
GERMANY

Dilini Kolombage
Institut für Numerische Simulation
Universität Bonn
Friedrich-Hirzebruch-Allee 7
53115 Bonn
GERMANY

Prof. Dr. Ralf Kornhuber
Institut für Mathematik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Prof. Dr. Annika Lang
Department of Mathematical Sciences
Chalmers University of Technology and the University of Gothenburg
412 96 Göteborg
SWEDEN

Prof. Dr. Claude Le Bris
CERMICS - ENPC & INRIA
Cite Descartes, Champs-sur-Marne
6 et 8 Ave. Blaise Pascal
77455 Marne-la-Vallée Cedex 2
FRANCE

Prof. Dr. Frederic Legoll
Laboratoire Navier - ENPC
Cite Descartes, Champs sur Marne
6 et 8 Ave. Blaise Pascal
77455 Marne-la-Vallée Cedex 2
FRANCE

Dr. Guanglian Li
Department of Mathematics
University of Hong Kong
Pokfulam Road
Hong Kong
CHINA

Dr. Xinliang Liu
King Abdullah University of Science and Technology (KAUST) Computer Electrical and Mathematical Science and Engineering Division
Thuwal 23955-6900
SAUDI ARABIA

Prof. Dr. Alexei Lozinski
Laboratoire Mathématiques de Besançon
Université de Franche-Comté
16, route de Gray
25030 Besançon Cedex
FRANCE

Prof. Dr. Mária Lukáčová-Medvidová
Institut für Mathematik
Fachbereich Mathematik/Physik/Informatik
Johannes-Gutenberg-Universität Mainz
Staudingerweg 9
55128 Mainz
GERMANY

JProf. Dr. Roland Maier
Karlsruher Institut für Technologie (KIT)
Institut für Angewandte und Numerische Mathematik
Englerstr. 2
76131 Karlsruhe
GERMANY

Prof. Dr. Axel Malqvist
Department of Mathematics
Chalmers University of Technology and University of Gothenburg
412 96 Göteborg
SWEDEN

Malin Mosquera
Department of Mathematics
Chalmers University of Technology
University of Gothenburg
Göteborg 412 96
SWEDEN

Prof. Dr. Felix Otto
Max-Planck-Institut für Mathematik in den Naturwissenschaften
Inselstrasse 22 - 26
04103 Leipzig
GERMANY

Prof. Dr. Houman Owhadi
Department of Computing and Mathematical Sciences
California Institute of Technology
MC 9-94
1200 E California blvd
Pasadena CA 91125
UNITED STATES

Prof. Dr. Malte Peter
Institut für Mathematik
Universität Augsburg
86135 Augsburg
GERMANY

Prof. Dr. Daniel Peterseim
Institut für Mathematik &
Centre for Advanced Analytics and Predictive Sciences
Universität Augsburg
Universitätsstraße 12a
86159 Augsburg
GERMANY

Prof. Dr. Mira Schedensack
Universität Leipzig
Mathematisches Institut
Postfach 10 09 20
04009 Leipzig
GERMANY

Prof. Dr. Robert Scheichl
Institut für Mathematik
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
GERMANY

Dr. Kathrin Smetana
Dept. of Mathematical Sciences
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ 07030
UNITED STATES

Timo Sprekeler
Department of Mathematics
Texas A & M University
College Station, TX 77843-3368
UNITED STATES

Prof. Dr. Li-yeng Sung
Department of Mathematics
Louisiana State University
208 Lockett Hall
Baton Rouge LA 70803-4918
UNITED STATES

Lucia Swoboda
Department of Mathematics
Chalmers University of Technology
University of Gothenburg
412 96 Göteborg
SWEDEN

Dr. Aretha Teckentrup
School of Mathematics
University of Edinburgh
James Clerk Maxwell Building
Edinburgh EH9 3FD
UNITED KINGDOM

Prof. Dr. Anna-Karin Tornberg
Department of Mathematics
KTH
10044 Stockholm
SWEDEN

Prof. Dr. Ngoc Tien Tran
Institut für Mathematik
Universität Augsburg
Universitätsstraße 2
86159 Augsburg
GERMANY

Prof. Dr. Yen-Hsi Richard Tsai
Department of Mathematics
The University of Texas at Austin
1 University Station C1200
Austin, TX 78712-1082
UNITED STATES

Prof. Dr. Elisabeth Ullmann
Department Mathematik
Technische Universität München
Boltzmannstrasse 3
85748 Garching bei München 85748
GERMANY

Prof. Dr. Barbara Verfürth
Institut für Numerische Simulation
Universität Bonn
Friedrich-Hirzebruch-Allee 7
53115 Bonn
GERMANY

Prof. Dr. Li Wang
School of Mathematics
University of Minnesota
127 Vincent Hall
206 Church Street S. E.
Minneapolis MN 55455-0436
UNITED STATES

Dr. Yangshuai Wang
Department of Mathematics
National University of Singapore
Lower Kent Ridge Road
Singapore 119 260
SINGAPORE

Prof. Dr. Yunan Yang
Department of Mathematics
Cornell University
582 Malott Hall
Ithaca NY 14853-4201
UNITED STATES

Prof. Dr. Lexing Ying
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
UNITED STATES

Prof. Dr. Harry Yserentant
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Lina Zhao
Department of Mathematics
City University of Hong Kong
Tat Chee Avenue, Kowloon Tong
Hong Kong
CHINA