Algebraic Multigrid

Adaptivity and Structure-Preservation

Kumulative Habilitationsschrift

Karsten Kahl

Bergische Universität Wuppertal
Fakultät 4 – Mathematik und Naturwissenschaften

2018

The habilitation can be quoted as follows:

urn:nbn:de:hbz:468-20210517-121540-7
[http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20210517-121540-7]

DOI: 10.25926/5q92-sg47
[https://doi.org/10.25926/5q92-sg47]

Preface

Eight years have passed since the day I handed in my doctoral thesis, eight years in which I had time to do a lot of interesting research, teach and advise a handful of gifted students. I feel privileged to be able to spend my time doing mathematical research and teaching classes at the university, while also finding time to spend with my family. What started as a chance encounter evolved into a deep love of algebraic multigrid and some fruitful, as I hope, advances in both theory and application.

My work in the past couple of years can be characterized by a large number of cooperations with various people of different backgrounds. There are two main reasons for my preference of collaborative research. On one hand, I very much enjoy discussing new, radical ideas with peers, which often leads to publications. On the other hand, I consider discussions inevitable in my work conduct. They are of utmost importance to me to reflect and refine my ideas as well as to find the right reference frame for them.

I am grateful to all my collaborators for their positive feedback, and in particular to James Brannick. I first met James during the late stages of my Diplomarbeit. Over the years he not only aquainted me with some of the leading scientists in the United States in the field, but also became the main reflector for my ideas on algebraic multigrid. In frequent discussions we freely share our ideas and develop the main line of our research, the bootstrap algebraic multigrid framework, together.

A second source of collaborations comes from the field of theoretical physics. Being part of the Sonderforschungsbereich "Hadron Physics from Lattice QCD" as an applied mathematician has honed my skills of transdisciplinary research. While we were able to drastically advance the code base of our collaborators, the transdisciplinarity has been very fruitful for my own research as well. The confrontation with the numerical challenges of lattice gauge theory led to new insights in adaptive algebraic multigrid methods and a general broadening of my research interests.

The third and final source of collaborative work originates from my advisory work of students at all stages of their studies. Seeing myself as equal part researcher and teacher I put large effort in the supervision of Bachelor and Master theses as well as a substantial investment of time to accompany doctoral students in their work. Confronting an unprepossessed mind with one of my current research problems has led to some remarkable theses, which also resulted in joint publications.

I would like to thank all my co-workers at Bergische Universität Wuppertal who make the work in the Arbeitsgruppe Angewandte Informatik such an enjoyable everyday experience. Especially I would like to thank my mentor Andreas Frommer for his counsel, his trust and his undying faith in my work. Last but not least I would like to thank my family and friends for their patience and support.

July 2018 Karsten Kahl

Collection of publications

[K1] M. Bolten, A. Brandt, J. Brannick, A. Frommer, K. Kahl, and I. Livshits. "A bootstrap algebraic multilevel method for Markov chains". *SIAM J. Sci. Comput.* 33.6 (2011), pp. 3425–3446. DOI: 10.1137/100791816.

[K2] A. Brandt, J. Brannick, K. Kahl, and I. Livshits. "Bootstrap AMG". *SIAM J. Sci. Comput.* 33.2 (2011), pp. 612–632. DOI: 10.1137/090752973.

[K3] J. Brannick and K. Kahl. "Bootstrap algebraic multigrid for the 2d Wilson Dirac system". *SIAM J. Sci. Comput.* 36.3 (2014), B321–B347. DOI: 10.1137/130934660.

[K4] A. Frommer, K. Kahl, S. Krieg, B. Leder, and M. Rottmann. "Adaptive aggregation-based domain decomposition multigrid for the lattice Wilson-Dirac operator". *SIAM J. Sci. Comput.* 36.4 (2014), A1581–A1608. DOI: 10.1137/130919507.

[K5] A. Brandt, J. Brannick, K. Kahl, and I. Livshits. "Algebraic distance for anisotropic diffusion problems: multilevel results". *Electron. Trans. Numer. Anal.* 44 (2015), pp. 472–496. URL: https://www.elibm.org/article/10006370.

[K6] M. Bolten, K. Kahl, and S. Sokolović. "Multigrid methods for tensor structured Markov chains with low rank approximation". *SIAM J. Sci. Comput.* 38.2 (2016), A649–A667. DOI: 10.1137/140994447.

[K7] J. Brannick, A. Frommer, K. Kahl, B. Leder, M. Rottmann, and A. Strebel. "Multigrid preconditioning for the overlap operator in lattice QCD". *Numer. Math.* 132.3 (2016), pp. 463–490. DOI: 10.1007/s00211-015-0725-6.

[K8] M. Bolten, K. Kahl, D. Kressner, F. Macedo, and S. Sokolović. "Multigrid methods combined with low-rank approximation for tensor structured Markov chains". *Electron. Trans. Numer. Anal. (accepted)* (2018). DOI: 10.1553/etna_vol48s348.

[K9] J. Brannick, F. Cao, K. Kahl, R. D. Falgout, and X. Hu. "Optimal interpolation and compatible relaxation in classical algebraic multigrid". *SIAM J. Sci. Comput.* 40.3 (2018), A1473–A1493. DOI: 10.1137/17M1123456.

[K10] K. Kahl and N. Kintscher. "Geometric multigrid for the tight-binding Hamiltonian of graphene." *SIAM J. Numer. Anal.* 56.1 (2018), pp. 499–519. DOI: 10.1137/16M1102033.

[K11] K. Kahl and M. Rottmann. "Least angle regression coarsening in bootstrap algebraic multigrid". *SIAM J. Sci. Comput.* 40.6 (2018), A3928–A3954. DOI: 10.1137/18M1168182.

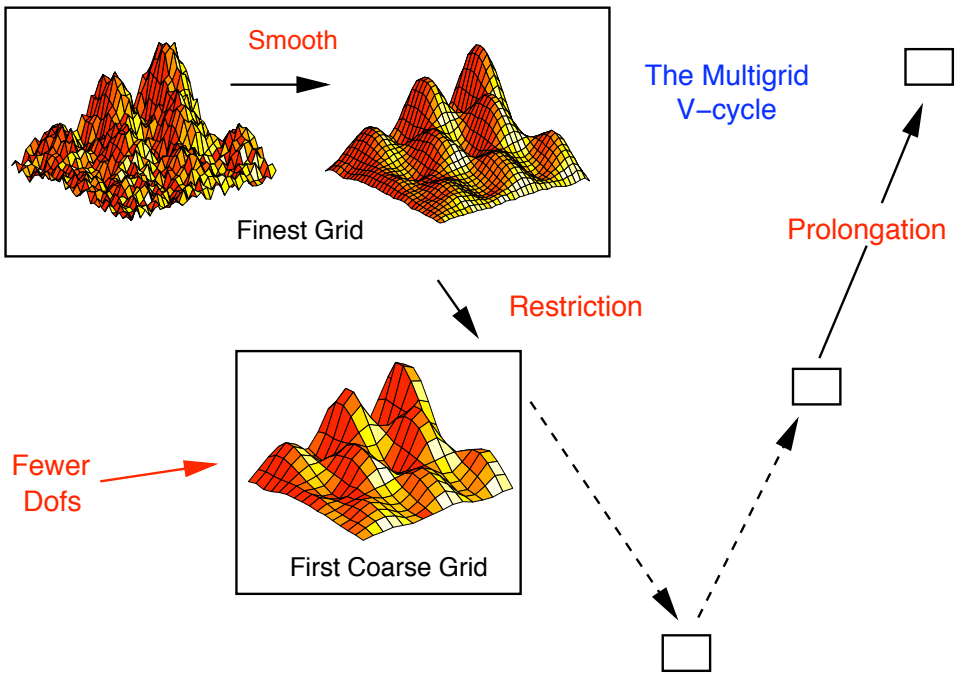


Figure 1: Multigrid in a nutshell.

1 Introduction

Multigrid methods can be viewed as the ultimate answer to the question how one should solve linear systems of equations

$$Ax = b, \;\; A \in \mathbb{C}^{n \times n},$$

which originate from the discretization of particular partial differential equations. This perspective is justified by their potential for optimal asymptotic complexity $\mathcal{O}(n)$ as well as their numerical robustness. This unparalleled efficiency of multgrid methods is obtained by a multi-scale resolution of the iterative process, such that only local computations need to be done on each scale.

Generically, multigrid methods consist of two components. First, a local relaxation process that can be formulated independent of scale, which "smoothes" an arbitrary error. Second, a way to transfer information between scales that is accurate on "smoothed" data and, in addition, a representation of the problem on different scales. In fig. 1 the multigrid principle is illustrated and the aforementioned effectiveness of the multigrid cycle to resolve an arbitrary error is demonstrated. It is worth noting that neither of the two components of a multigrid method would result in an efficient method on its own. What makes multigrid methods so effective is the complementarity of its two components.

The essential observation of a smoothing property of certain relaxation schemes for elliptic problems dates back at least to Southwell [61], but it took decades to complement this observation with suitable descriptions of the problem at different scales to formulate a multigrid method. This is surprising, as the formulation of the problem at different scales is readily available in the discretization of partial differential equations. It lies in the very nature of a discretization that a partial

differential equation can be approximated on different scales. Nevertheless, the realization that the essential difference between discretizations on two different scales is comprised of local information and is thus amenable to smoothing occured only in works by Fedorenko [30] and Bakhvalov [4]. Shortly after these first works, the topic of multgrid methods gained traction by the seminal works of Brandt [9, 10] and Hackbusch [34] and the first conferences dedicated to multigrid methods [35]. Around the same time, starting with Nicolaides [49, 50], multigrid methods were also introduced in the framework of finite element discretizations. This again is a quite natural process due to the fact that finite element methods not only provide descriptions of the continuous problem on different scales, but also readily provide projection and interpolation operators between finite dimensional spaces of functions at different scales.

All these early works on multigrid methods had one feature in common. The construction of the multigrid hierarchy is essentially based on the underlying discretization scheme and thus intimately coupled to it. Thus, to incorporate multigrid solvers in simulation packages requires changes in the core routines and deepest layers of the implementation, a task that often is deemed too daunting or expensive to be considered. This problem has been the main motivation for the introduction of algebraic multigrid methods [7, 13, 60, 68].

The essential difference of geometric and algebraic multigrid methods is outlined in fig. 2. On one hand, *geometric multigrid methods* are intimitely tied to the partial differential equation and an appropriate discretization scheme, so that the multigrid hierarchy can be automatically generated, and transfer operators between scales can be easily derived. In this context, one has to find a suitable relaxation scheme, which is complementary to the given multigrid hierarchy. On the other hand, *algebraic multigrid methods* do not (want to) make any assumption on the origin of the linear system of equations, but only assume that a (convergent) relaxation process can be constructed. It is the task of the method to construct a suitable multigrid hierarchy and scale-bridging transfer operators based on this information alone. Clearly, one would expect that, applied to the same problem, a geometric multigrid construction is superior to an algebraic one as it uses a lot more information on the underlying problem than the algebraic construction can possibly do. While this is indeed often true, the algebraic construction offers more flexibility in terms of independence from the actual discretization process and is thus much easier to incorporate into existing codes. It can also be applied to problems where geometric multigrid constructions are (not yet) available (e.g. lattice gauge theories) or might be used as a proof-of-concept to motivate the development of geometric approaches or a reimplementation of the code-base, where a geometric multigrid method is built in from the start. Furthermore, algebraic multigrid methods can readily be applied to discretizations of systems of partial differential equations.

Given the prominent role in my work, let us now take a closer look at the components of algebraic multigrid methods and the challenges in their construction. As sketched in fig. 2, algebraic multigrid methods start with the algebraic description of the sparse linear system of equations

$$Ax = b,\ A \in \mathbb{C}^{n\times n}. \tag{1}$$

Generically, the error propagator of an algebraic two-grid method is given by

$$E_{2g} = (I - M_{\text{post}}A)(I - PA_c^{-1}RA)(I - M_{\text{pre}}A)\,, \tag{2}$$

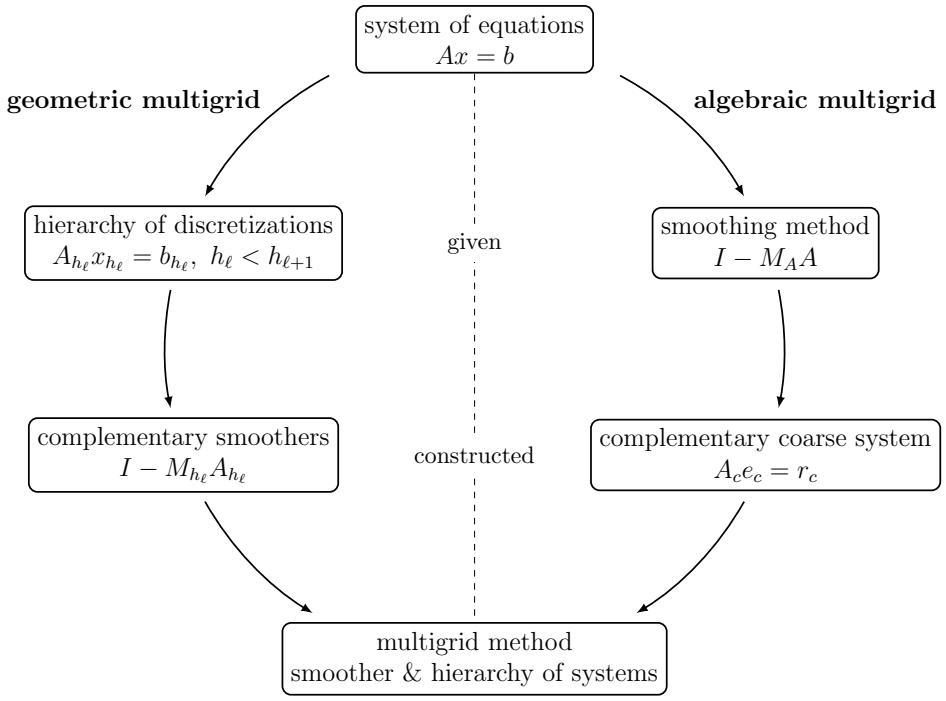


Figure 2: Illustration of the difference of geometric and algebraic multigrid methods (cf. [69, p. 416]).

where $M_\cdot$ defines a simple, i.e., cheap, relaxation process and $I - PA_c^{-1}RA$ denotes the coarse-grid correction. In here, A_c can be thought of as the coarse representation of A, and P denotes the interpolation, R the restriction operator. A "true", i.e., multi-level, algebraic multigrid method is then obtained by simply recursing on the two-grid construction.

Most of the time the coarse representation A_c is chosen variationally, i.e., $A_c = RAP$ even though deviations from this rule have been considered in the literature, e.g., in [2, 28, 71], mainly motivated to reduce fill-in in A_c. Assuming that the coarse representation A_c of A is chosen variationally, the problem of defining a multigrid hierarchy in the algebraic approach reduces to constructing P and R. In case A is Hermitian, R is typically chosen as P^H in order to preserve Hermiticity in A_c. This further reduces the problem to only requiring a suitable definition of interpolation P. The general appeal of the variational construction can be seen when considering positive definiteness of A. In this case the coarse grid correction

$$I - P\left(RAP\right)^{-1} RA,$$

with $R = P^H$ and $A_c = P^H AP$ becomes an A-orthogonal projection.

Now that we boiled down the construction of an algebraic multigrid method to the construction of P, let us consider the general design goals in its definition from a high-level point of view. First, in order to obtain an efficient method, the range of interpolation has to be chosen complementary to the space where the smoother is efficient. That is, it needs to represent the space where the smoother is inefficient. Note that in the case that A_c is chosen variationally, the range of P uniquely defines the action of the coarse-grid correction as any change of basis $P \to P \cdot G$, G non-singular, cancels out in eq. (2). Choosing $G^{-1} := P_{\mathcal{C}}$, i.e., the inverse of the square

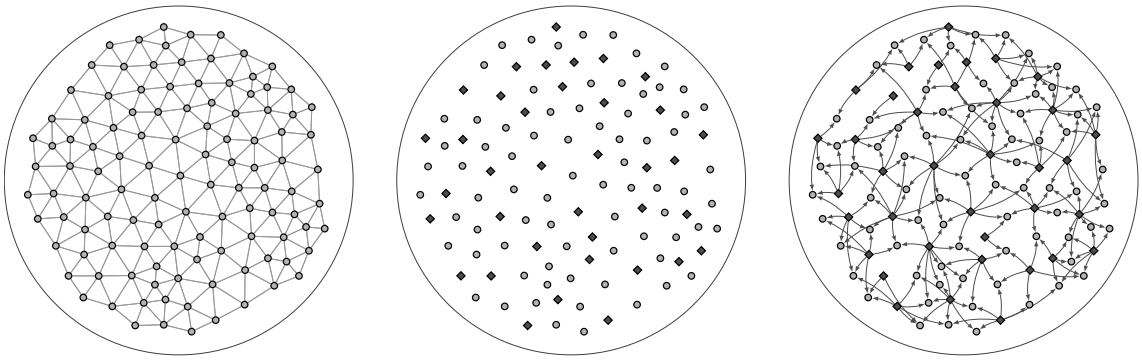


Figure 3: Algebraic coarsening in a nutshell: (left) adjacency graph of A; (middle) choice of $\mathcal{C}$, i.e., coarse variables ◆; (right) interpolation relations, ingoing arrows for ○ correspond to $\mathcal{C}_i$.

submatrix of P corresponding to a subset $\mathcal{C}$ of variables, we can always transform P such that $P_{\mathcal{C}} = \mathrm{id}_{\mathcal{C}}$. That is, interpolation is the identity on $\mathcal{C}$, the so-called coarse degrees of freedom, and the remainder of P can be interpreted as interpolation from variables in $\mathcal{C}$ to variables not in $\mathcal{C}$. While this transformation can be unpractical due to fill-in in P, it allows us to treat coarsening on the basis of a variable splitting, which is convenient for this presentation. Specifying a suitable set $\mathcal{C}$ that allows for the accurate representation of the aforementioned space is thus part of the construction of interpolation P as well as the computation of suitable entries of P. In here, sparsity of P and with it sparsity of $A_c = P^H AP$ is beneficial in order to be able to recursively apply the construction to obtain a multigrid hierarchy of operators, to reduce communication in parallel implementations and also to obtain a significant reduction in complexity along the multigrid hierarchy.

Achieving these goals is typically split into two parts. Finding $\mathcal{C}$ and the sparsity pattern of P is often referred to as the *coarsening* problem, while determining the entries of P, and thus defining its range, is known as the *interpolation* problem. These terms will become clear when we go into more detail in section 2. Often, an artificial geometric interpretation of the algebraic equations (1) in terms of the corresponding adjacency graph

$$\mathcal{G}_A = \{V, E\},\ V = \{1, \ldots, n\},\ E = \{(i,j) \in V \times V \mid a_{ij} \neq 0\}$$

is used in the setup process of the algebraic multigrid method. Guided by the graph of the matrix $\mathcal{G}_A$ a set of coarse variables $\mathcal{C}$ is chosen and sets $\mathcal{C}_i \subset \mathcal{C}$ are determined to define interpolating variables for every variable $i \notin \mathcal{C}$. In fig. 3 the process from $\mathcal{G}_A$ to $\mathcal{C}$ and finally $\mathcal{C}_i$, $i \notin \mathcal{C}$ is depicted.

The earliest constructions that fit into this framework of an algebraic solution to either the coarsening or interpolation problem are aggregation approaches [6, 23], reduction-based approaches [58, 59], which use approximations of Schur complements to construct a multigrid hierarchy for the problem, and the classical "Ruge-Stüben" algebraic multigrid method [13, 7, 60, 68]. While the early aggregation approaches mainly focus on the construction of suitable coarsenings, as the entries of P are known from context, i.e., based on knowledge about the underlying partial differential equations, the classical algebraic multigrid method can be seen as the first approach to tackle the algebraic multigrid construction in its entirety.

In the classical approach both the coarsening and the interpolation problem are solved using the entries of A, and its convergence analysis shows that this is appropriate as long as A is an M-matrix.[1] While this classical approach to algebraic multigrid can be extended to some degree [21, 24, 27, 38], its applicability is limited due to implicit assumptions on the underlying problem, in particular, the importance of the constant vector, which is exactly interpolated. Since then, many attempts have been made to further extend the applicability of algebraic multigrid methods beyond their initial limitations.

An important step in this direction has been the introduction of adaptivity to construct multigrid hierarchies, which is discussed in detail in section 2. Developing adaptive ideas for the construction of efficient algebraic multigrid methods has been a central topic of my work. As such, I use section 2 to reflect on my contributions as well. These contributions to algebraic multigrid methods have been fueled by the collaborations with theoretical physicists working on lattice gauge theories. Due to the inherent randomness in the considered models, no geometric multigrid construction is known up to date, but my collaborators and I were able to show that efficient methods can be constructed using adaptive algebraic multigrid methods. One crucial aspect that came up in this exploration has been the importance of structure preservation. Systems of partial differential equations typically possess peculiar symmetry structures that need to be considered in the construction of multigrid methods to ensure robustness, the preservation of spectral properties and thus recursive applicability. I review the topic of structure preserving multigrid methods in general and specifically the development of methods for the operators of lattice gauge theory in detail in section 3. I conclude with a statement about my contributions to the publications which comprise this cumulative Habilitation in section 4.

[1] M-matrices arise naturally in the finite difference discretization of elliptic partial differential equations.

2 Adaptive algebraic multigrid methods

In the context of algebraic multigrid methods, adaptivity is understood as an ability of the method to adapt itself to the problem at hand. Based on the fundamental assumption of algebraic multigrid, i.e., that a smoother is given and a complementary coarse grid has to be constructed, first adaptive approaches have been introduced shortly after formulating the concept of algebraic multigrid itself.

Early works that use adaptivity [13] to improve the performance of an algebraic multigrid method modify classical algebraic multigrid interpolation by introducing a representative of algebraically smooth error, i.e., error components that cannot be effectively treated by relaxation. Motivated by the observation that the error components that are not effectively reduced by the relaxation process oftentimes strongly correlate to the lower part of the spectrum of the system operator A, early works also advocate the use of spectral information to adapt algebraic multigrid methods [47, 60]. The development gained momentum in the first decade of the 21st century with the introduction of adaptive smoothed aggregation (αSA) in [19, 20] and adaptive algebraic multigrid (αAMG) in [18], which is an adaptive extension of the classical algebraic multigrid method. Many of these approaches add adaptivity specifically to solve the interpolation problem, i.e., to adaptively adjust the space spanned by interpolation, without touching the coarsening problem. Another approach to solve the interpolation problem that relies on energy minimizing coarse grid basis functions and is thus related to the (smoothed) aggregation approch has been developed in [15, 45, 54, 70, 72]. To some degree these methods can be viewed as solving the interpolation and coarsening problem at the same time by obtaining suitable basis functions of the coarse grid without having to rely on particular coarsening strategies.

In recent years, adaptivity has also been used in new ways to solve the coarsening problem. This has been done either by replacing the definition of strength of connection in a classical algebraic multigrid sense by an adaptive measure of strength as in [55, 57], by using compatible relaxation [8, 14] or by considering binary relations of variables to construct suitable matchings in $\mathcal{G}_A$ as in [17, 42, 48, 52]. Almost all of these approaches make use of the concept that the given smoother reveals the coarse subspace due to the need of complementarity. That is, what is not reduced by the smoother needs to be in the coarse space.

Together with my collaborators, I was substantially involved in developing the bootstrap algebraic multigrid framework which tries to solve both the interpolation and coarsening problem by leveraging a similarity of the construction of a suitable coarse space to a statistical learning process. While I will focus in the following on publications that I have been involved in, other groups have picked up and developed components of the bootstrap algebraic multigrid framework as well (cf. [22, 46, 62]).

The bootstrap algebraic multigrid framework

The bootstrap algebraic multigrid framework is comprised of two main ideas. The first of them is to use statistical information of algebraically smooth error in least squares interpolation. In order to get a good approximation of algebraically smooth error, least squares interpolation uses local weighted least squares fits to define the entries of the interpolation operator based on stochastic samples of algebraically

smooth error. Assuming that no a priori information about algebraically smooth error is available, the least squares interpolation approach can be interpreted as building a local model of the nature of algebraically smooth error by using samples that are initiated with random values and smoothed afterwards. Combined with a weighting that favors global algebraic smoothness, we showed in [K1, K2] that least squares interpolation yields a robust method for the adaptive computation of interpolation operators in algebraic multigrid methods. In addition we demonstrated in [K5, K11] how the concept of least squares interpolation can be used to find solutions to the coarsening problem.

The second concept, introduced in the bootstrap algebraic multigrid framework is the bootstrap setup cycle. Its main idea is the computation of global samples of algebraically smooth error by exploiting the evolving multigrid hierarchy. Based on the observation that the formulation of global algebraic smoothness can be transferred to coarse grids, we demonstrated that samples of globally algebraic smooth error can efficiently be obtained and fed back into the subsequent setup iterations. True to its name, our proposed bootstrap setup is able to start from scratch and use the currently best available method and information to improve itself.

My contribution to the development of the bootstrap algebraic multigrid framework has been published in the following four papers, each of which is summarized shortly.

[K2] Bootstrap AMG. Achi Brandt described a rough outline of a bootstrap setup cycle and least squares interpolation for algebraic multigrid methods in his seminal work [11]. In [K2] we elaborate this outline and discuss the two main ideas of the bootstrap algebraic multigrid framework, least squares interpolation and the bootstrap setup cycle, in detail.

Given a set of test vectors $v^{(1)}, \ldots, v^{(K)}$, which are samples of algebraically smooth error, i.e., either stochastic samples (initially random smoothed vectors), a priori known information on algebraically smooth error or samples computed in the booststrap setup cycle, the entries of p_{ij} of the interpolation operator P are defined by the local weighted least squares fits

$$\min_{p_{i,\cdot}} \sum_{k=1}^{K} \omega_k \left(v_i^{(k)} - \sum_{j \in \mathcal{C}_i} p_{ij} v_j^{(k)} \right)^2 \quad \text{for all } i \notin \mathcal{C}.$$

In here ω_k is chosen such that vectors which are algebraically smooth globally, i.e., $||Av|| \ll ||v||$, get a larger weight in the least squares fit.

The bootstrap setup cycle is motivated by the observation that for a given multigrid hierarchy, defined by interpolation operators $P_{i+1}^i,\ i = 0, 1, \ldots, L-1$ and coarse grid operators $A_{i+1} = (P_{i+1}^i)^H A_i P_{i+1}^i,\ A_0 = A$, we find with $P_i := P_1^0 \cdot P_2^1 \cdot \ldots \cdot P_i^{i-1}$ that

$$\frac{\langle A_i x, x \rangle_2}{\langle x, x \rangle_2} = \frac{\langle A P_i x, P_i x \rangle_2}{\langle P_i x, P_i x \rangle_2}.$$

That is, eigenvectors to small eigenvalues of the generalized eigenvalue problem

$$P_L^H \left(\lambda P_L x - A P_L x \right) = 0 \quad \Longleftrightarrow \quad A_L x = \lambda P_L^H P_L x$$

are good candidates of algebraically smooth errors that can be computed or approximated on coarse grids. Leveraging this spectral connection of the multigrid hierarchy allows us to formulate the bootstrap setup cycle depicted in fig. 4.

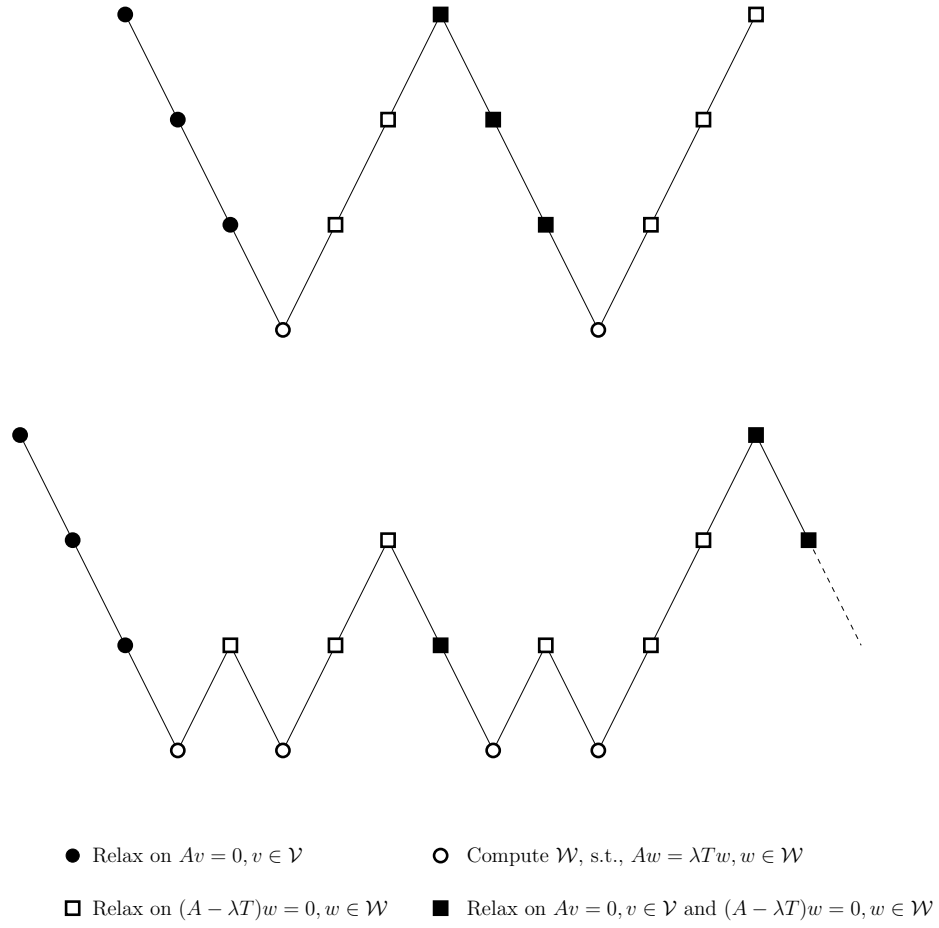


Figure 4: Bootstrap AMG V^2-cycle and W-cycle setup schemes (cf. [K2]).

Finally, we demonstrate the effectiveness and robustness of the bootstrap setup cycle and least squares interpolation in numerical experiments.

[K1] A bootstrap algebraic multilevel method for Markov chains. Motivated by the success of the bootstrap algebraic multigrid method and the numerous publications that dealt with the construction of algebraic multigrid methods for discrete time Markov chain problems at that time (cf. [64, 65, 66, 67]), we apply and adapt the bootstrap algebraic multigrid framework to the non-symmetric singular operators arising in discrete time Markov chain modelling in [K1].

Discrete time Markov chains describe the behaviour of a random process by transition probabilities between distinct states of the system. Due to the probabilistic nature of the model, the resulting matrix $A \in \mathbb{R}^{n \times n}$, where n denotes the number of possible states, is column stochastic, i.e., $\mathbb{1}^T A = \mathbb{1}^T$. In order to calculate the steady-state distribution (the probability distribution of long-term behaviour) we have to solve the eigenvalue problem

$$Ax = x \quad \Longleftrightarrow \quad (I - A)\, x = 0$$

for its (up to a scale factor) unique solution x. Due to the non-symmetric nature of the problem, the algebraic multigrid method has to consider the construction of restriction R independently of interpolation P. While we are able to apply least-squares interpolation with only miniscule modifications (mainly in the choice of

weights) to the construction of P, the construction of R has to be reconsidered. We show in this work that it is necessary to preserve the column-stochastic nature of the system matrix on coarser grids in order to guarantee consistency of the singular coarse grid systems. This can be enforced by constructing R such that $\mathbb{1}^T R = \mathbb{1}^T$, e.g., by using classical algebraic multigrid recipes.

By combining this construction of R with least squares interpolation P, the bootstrap algebraic multigrid cycle and compatible relaxation coarsening, we are able to show scalability of the approach for Markov chains on random planar graphs based on Delaunay triangulation and other, uniform, Markov chain models. Due to the fact that the presented approach makes use not only of multiplicative updates from the bootstrap cycle, but also from additive updates by preconditioned GMRES it outperforms state-of-the-art algebraic multigrid methods for this application.

[K5] Algebraic distance for anisotropic diffusion problems: Multilevel results. While the interpolation problem and overall setup are well developed within the bootstrap algebraic multigrid framework, the proper solution of the coarsening problem within the framework has been an open problem up to this point. Thus, motivated by yet another remark in [11] and the overall good performance of least squares interpolation we explore the idea of algebraic distances in this paper. The concept of algebraic distance solves the coarsening problem by defining an algebraic strength measure

$$\mu_{ij} := \left(\min_{p_{ij}} \sum_{k=1}^{K} \omega_k \left(v_i^{(k)} + \frac{1}{a_{ii}} r_i^{(k)} - p_{ij} v_j^{(k)} \right) \right)^{-2},$$

which is simply least squares interpolation with an additional local residual correction (cf. [22]) restricted to pairs of variables. Measuring algebraic distance in a local neighborhood of each variable yields a strength graph that is used in a compatible relaxation based coarsening algorithm [14]. By limiting the problem to pairs of variables the cost of this algebraic strength measure scales linearly with the number of variables and the size of the local neighborhoods. Due to the fact that the test vectors of the bootstrap setup cycle can be used to define the algebraic strength measure it can be seamlessly integrated into the bootstrap framework.

In numerical experiments with anisotropic diffusion problems we are able to show that the bootstrap algebraic multigrid method, using algebraic distances to solve the coarsening problem, yields scalable convergence for a large range of anisotropy strength and different angles of anisotropy with respect to the orientation of the underlying uniform grid. This is largely due to the fact that the algebraic distance measure is able to detect the direction of anisotropy even at long range, thus enabling coarsening in that direction as depicted in fig. 5.

[K11] Least angle regression coarsening in bootstrap algebraic multigrid. Considering the current explosion of techniques for machine learning, it is interesting to understand that the bootstrap algebraic multigrid framework can be interpreted as a statistical learning method. The observation that least squares interpolation learns a local model of algebraically smooth error by examining samples of it is at the heart of the development of another solution ansatz for the coarsening problem within the bootstrap algebraic multigrid framework that we develop in [K11].

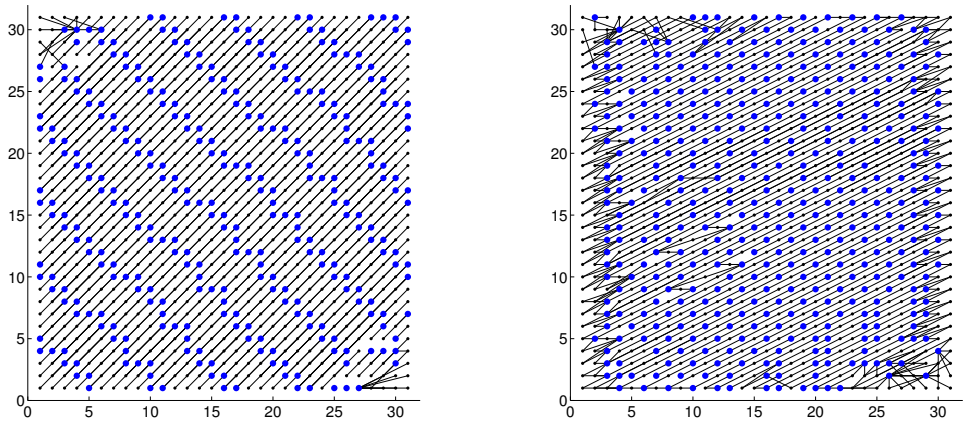


Figure 5: Coarse grids and caliber $c = 2$ interpolation patterns resulting from algebraic distance coarsening for anistotropic diffusion problems with anisotropy angle $\alpha = \frac{\pi}{4}$ (left) and $\alpha = \frac{\pi}{8}$ (right) (cf. [K5]).

Due to the fact that the model of algebraically smooth error built by least squares interpolation (in terms of the calculated coefficients) reflects correlations of variables, i.e., how well one or more variables are able to describe others, it can be considered natural to use this information to define interpolation relations in the bootstrap algebraic multigrid framework. However, the brute-force application of this approach, e.g., to check all possible sets of three variables in the (local) neighborhood of a variable in order to find the best set of three variables to interpolate from, is clearly too costly. In this paper we find an elegant solution to this problem by first generalizing the notion of locality in least squares interpolation and then introducing a natural sparsification modification to least squares interpolation to resolve the combinatorial complexity problem.

The first contribution of this paper is the generalization of the notion of locality in least squares interpolation by the introduction of a kernel function K_η that penalizes distance (e.g. graph distance) in a flexible manner. The resulting least squares problems that are used to determine the correlations between variable i and its neighboring variables read

$$\min_{p_{i,\cdot}} \sum_{k=1}^{K} \omega_k \left(v_i^{(k)} - \sum_{j \neq i} p_{ij} K_\eta(i,j) v_j^{(k)} \right)^2 .$$

Typically, $|\{j \mid K_\eta(i,j) \neq 0\}|$ by far exceeds the number of test vectors K used in the bootstrap framework which poses a serious threat of over-fitting if one is unwilling to increase K. That is, without further limiting the number of calculated coefficients or increasing the number of test vectors K, the coefficients should not be trusted in the decision making process. Due to the fact that the number of test vectors largely influences the cost of the boostrap framework, we are rather unwilling to increase their number, especially when a small number of test vectors is enough to achieve accurate interpolation once the coarsening problem is solved.

In order to deal with this problem, i.e., the reduction of the number of coefficients that need to be calculated, we introduce yet another tool from statistical learning,

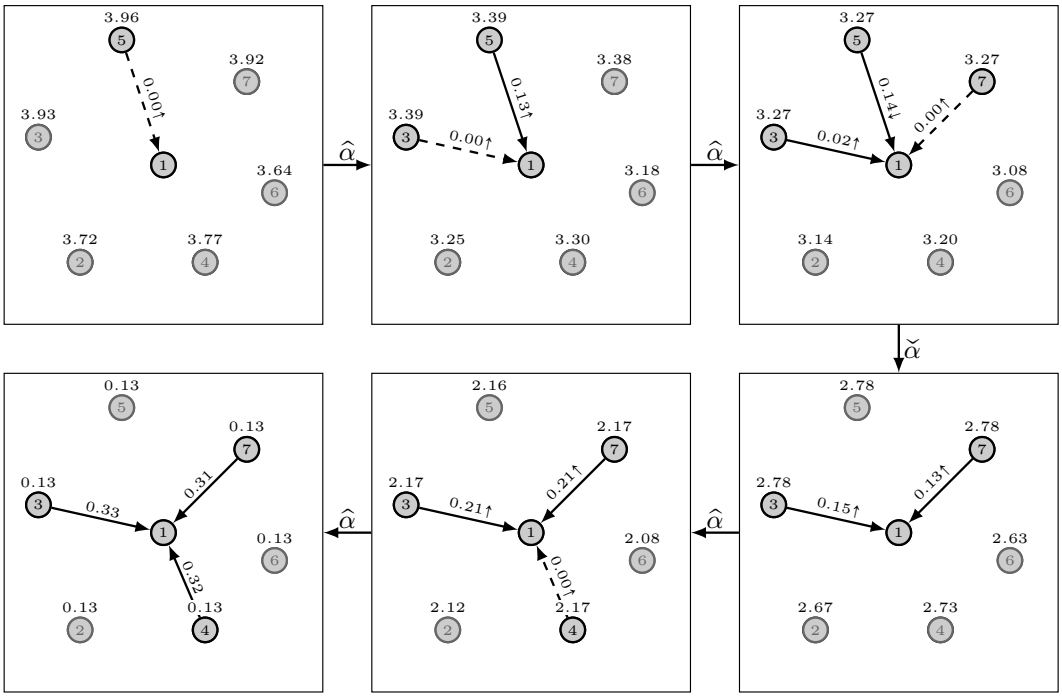


Figure 6: Illustration of the local behaviour of least angle regression applied to Poisson's equation, i.e., in a situation where no directionality of correlation is to be expected. Least angle regression is able to determine the "expected" triangular set of variables as highly correlated (as a set) (cf. [K11]).

an ℓ_1 constraint, to automatically sparsify the least squares coefficients,

$$\min_{p_{i,\cdot}} \sum_{k=1}^{K} \omega_k \left(v_i^{(k)} - \sum_{j \neq i} p_{ij} K_\eta(i,j) v_j^{(k)} \right)^2 + \lambda \| p_{i,\cdot} \|_1 \ .$$

Fortunately, this modification, which turns the least squares problems into a non-linear optimization problem, does not pose a complexity threat as it can be solved by least angle regression for all $\lambda \in [0, \infty)$ that correspond to switching points of coefficients $p_{i,\cdot}$, i.e., where either a coefficient that has been zero becomes non-zero or vice versa.

The most remarkable property of this approach is the ability to determine appropriate strength relations of sets of points in situations where no preference is given on the basis of pairs of points as depicted in fig. 6. In addition the approach is able to automatically determine the necessary number of interpolatory variables and thus reduces operator complexity of the resulting multigrid hierarchy.

In combination with a weighted independent set ansatz and some additional post-processing in the coarsening routine we are able to obtain scalable multigrid results for discretizations of anisotropic diffusion problems on unstructured meshes for both point-smoothers such as Gauss-Seidel, but also for domain decomposition smoothers, where the nature of the smoother is reflected in the coarsening as can be seen in fig. 7.

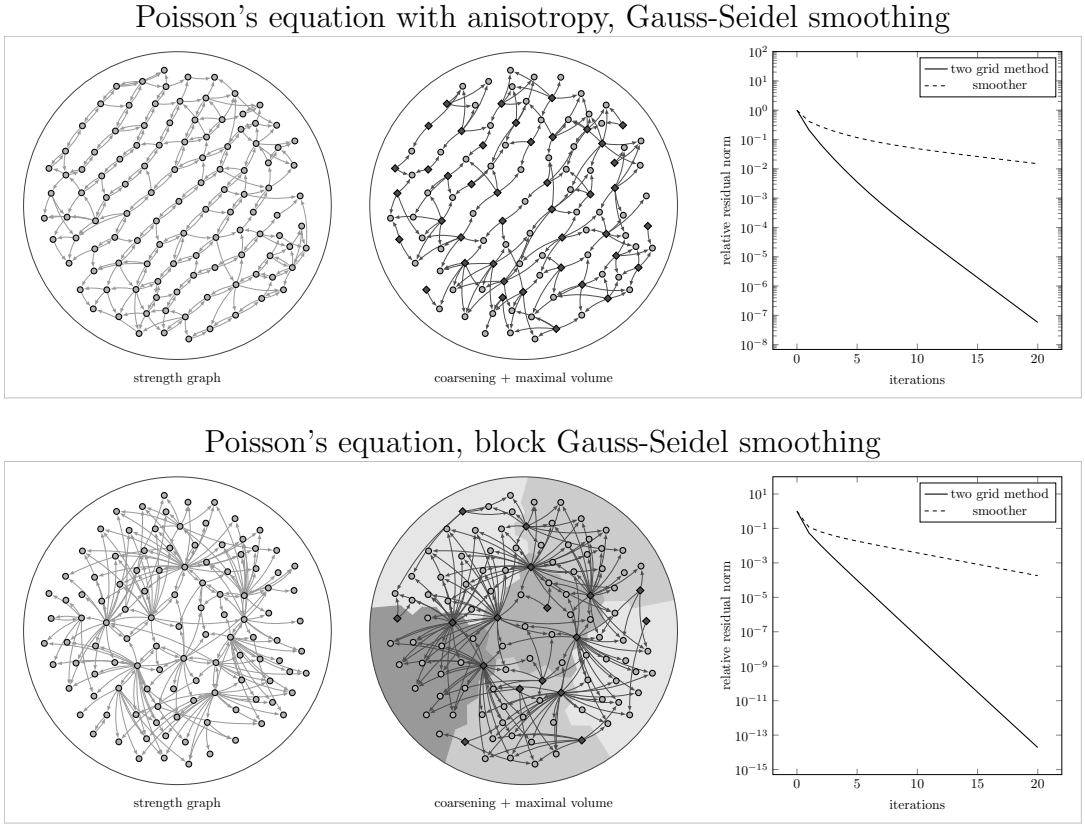


Figure 7: Illustration of the interpolation coupling structure computed by least angle regression coarsening for two different cases. (top) Poisson with anisotropy ($\alpha = \frac{\pi}{4}$, $\varepsilon = 0.01$); (bottom) Poisson w/o anisotropy, block Gauss-Seidel (6 blocks, 4 colors) (cf. [K11]).

Other advances in multigrid methods

It has been the long standing mantra, which I have cited many times in this manuscript already, that

> adaptive algebraic multigrid methods have to construct coarse grid corrections that are able to reduce error that cannot be efficiently reduced by the smoother

without having a thorough theoretical backing for it. While the mantra is certainly well motivated and ultimately also theoretically founded, the theory could not reflect it for a long time due to the fact that a fundamental assumption has been to neglect the influence of the smoother in the construction of coarse grids and interpolation. This assumption has been postulated and used in almost all algebraic multigrid constructions. That is, it is assumed that the spectral information of A can be used to construct efficient algebraic multigrid methods. While this assumption is valid for simple smoothers applied to simple problems (e.g., Jacobi iteration applied to an elliptic problem with constant coefficients) it is no longer appropriate when considering more complex problems and/or smoothers, e.g. block smoothers. Adaptivity has been able to hide this flaw of the original algebraic multigrid construction to some degree due to the "correct" source of information (algebraically smooth error, i.e., not efficiently reduced by the smoother) used in the construction of coarse

spaces, but the constructions and setup heuristics were still largely founded on the simplified assumption. In [K9] we were able to derive a formula for the optimal interpolation in algebraic multigrid for any given smoother. This formula encodes the relationship between the role of interpolation and smoothing and is a precise statement on the required complementarity of smoother and coarse grid correction. While not of immediate practical use, it is an important result for the understanding and advancement of (adaptive) algebraic multigrid methods and backs the general mantra of adaptive algebraic multigrid methods.

An important tool in the quantitative analysis of multigrid methods on structured grids is the so-called local Fourier Analysis introduced in [10, 12]. In case an operator, defined on a lattice, can be described locally due to shift-invariance, it is block diagonalized by the Fourier transform. Utilizing this block diagonalization for all components of a multi-level method allows for precise bounds on the convergence speed of these methods as has been demonstrated in countless works, e.g. [31, 32, 33, 36, 37, 44, 73]. In [K10] we contributed to the literature on local Fourier analysis by adding a geometric multigrid method for the tight-binding Hamiltonian of graphene, which is a maximally indefinite[2] operator. Most interestingly the definition of interpolation is solely guided by findings made using local Fourier analysis.

In the following I want to review these two contributions to the bigger landscape of multigrid methods in more detail.

[K9] Optimal interpolation and compatible relaxation in classical algebraic multigrid. Complementarity of smoother and coarse grid correction have been established and analyzed for a long time using the simplifying assumption that algebraically smooth error is represented by eigenvectors corresponding to small eigenvalues of A, i.e., neglecting the potential influence of the smoother when analyzing complementarity. In many situations this assumption turns out to be "good enough", but especially when considering complex problems or complex smoothers it can be outright misleading in the sense that while the number of small eigenvalues of A is large, the number of dominant vectors in algebraically smooth error is small.

In this work we present a precise statement about the relation between smoothing and coarse grid correction. We formulate the representation of optimal interpolation for a given smoother M, defined by its error propagator

$$I - M^{-1}A$$

and its symmetrized version $\widetilde{M} := M\left(M + M^H - A\right)M^H$. In [29] the convergence of a two-grid method with pre-smoothing by M and post-smoothing by M^H has been shown to fulfill

$$\|E_{2g}(P)\|_A^2 = 1 - \frac{1}{\sup_v \kappa(P,v)}, \quad \kappa(P,v) := \frac{\| \left(I - \pi_{\widetilde{M}}(P)\right) v\|_{\widetilde{M}}^2}{\|v\|_A^2},$$

where $\pi_{\widetilde{M}}(P)$ denotes the $\widetilde{M}$-orthogonal projection onto range(P). Based on this result we show that the optimal interpolation of an algebraic multigrid method, which minimizes the A-norm of the corresponding two-grid error propagator, is given by

$$P_{\text{opt}} = \begin{bmatrix} v_1 & v_2 & \dots & v_{n_c} \end{bmatrix},$$

[2]Having the same number of negative and positive eigenvalues

where $v_1, v_2, \ldots, v_{n_c}$ are the eigenvectors corresponding to the smallest eigenvalues of the generalized eigenvalue problem

$$Av_i = \lambda_i \widetilde{M} v_i \ .$$

Note that, assuming non-singularity of $\widetilde{M}$, this generalized eigenvalue problem can be rewritten as

$$\left(I - \widetilde{M}^{-1} A\right) v_i = (1 - \lambda_i) v_i \ .$$

That is, algebraically smooth vectors v_i correspond precisely to the eigenvectors of the error propagator of the (symmetrized) smoother belonging to eigenvalues close to one, i.e., they parametrize the space where the symmetrized smoother is slowest to converge.

This observation not only justifies the mantra of adaptive algebraic multigrid, but also yields a lower bound on the convergence speed of any (adaptive) algebraic multigrid method using the smoother M, as we obtain

$$\|E_{2g}(P_{\text{opt}})\|_A^2 = 1 - \lambda_{n_c+1} \ .$$

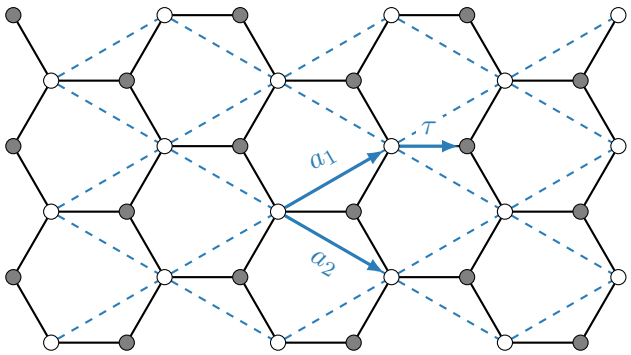
Furthermore, we show by introducing the concept of a maximal volume basis for the coarse space, that the optimal interpolation can give us information about the existence of a sparse interpolation operator that achieves near optimal convergence rates. To illustrate the potential of this result for adaptive algebraic multigrid methods we describe how to build it into the bootstrap framework, especially its setup cycle and present some fairly preliminary results of this method. A more extensive study on how to incorporate the result into the bootstrap framework is in progress.

[K10] Geometric multigrid for the tight-binding Hamiltonian of graphene. The carbon-allotrope graphene consists of carbon atoms arranged in a hexagonal lattice, which can be described as a triangular lattice $\mathbb{L}_T$ generated by the lattice vectors

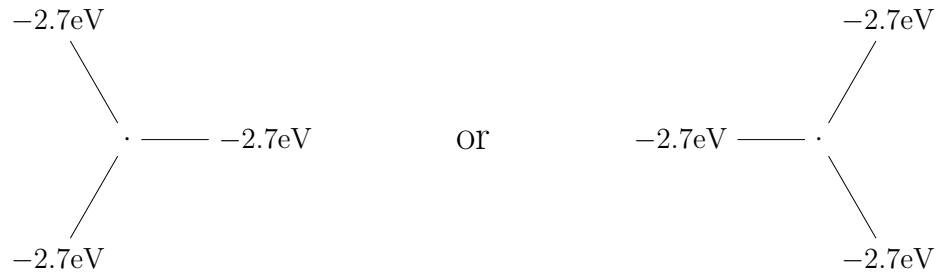
$$a_1 = \begin{bmatrix} \frac{3a}{2} & \frac{\sqrt{3}a}{2} \end{bmatrix} \quad \text{and} \quad a_1 = \begin{bmatrix} \frac{3a}{2} & -\frac{\sqrt{3}a}{2} \end{bmatrix}$$

with lattice spacing a and a unit cell attached to each lattice site defined by

$$\mathbb{L}_G := \{x + \delta\tau \mid x \in \mathbb{L}_T, \delta \in \{0, 1\}\}, \ \tau = \begin{bmatrix} a & 0 \end{bmatrix} \ .$$



A commonly used tool to simulate electron movement in solid state physics is the so-called tight-binding Hamiltonian. In short, it simply formulates the energy that it takes for an electron at one site of the material to hop to a neighboring site. In case of graphene, due to the highly symmetric lattice structure, the simplest tight-binding Hamiltonian takes the form:



It is known that this operator is maximally indefinite with two double-zeroes at the so-called Dirac points, which are located in phase space at

$$K_1 = \frac{2\pi}{9a}\begin{bmatrix}1 & \frac{\sqrt{3}}{3}\end{bmatrix} \quad \text{and} \quad K_2 = \frac{2\pi}{9a}\begin{bmatrix}1 & -\frac{\sqrt{3}}{3}\end{bmatrix}.$$

Close to these points the tight-binding Hamiltonian has a linear dispersion relation, which is largely responsible for the interesting properties of graphene.

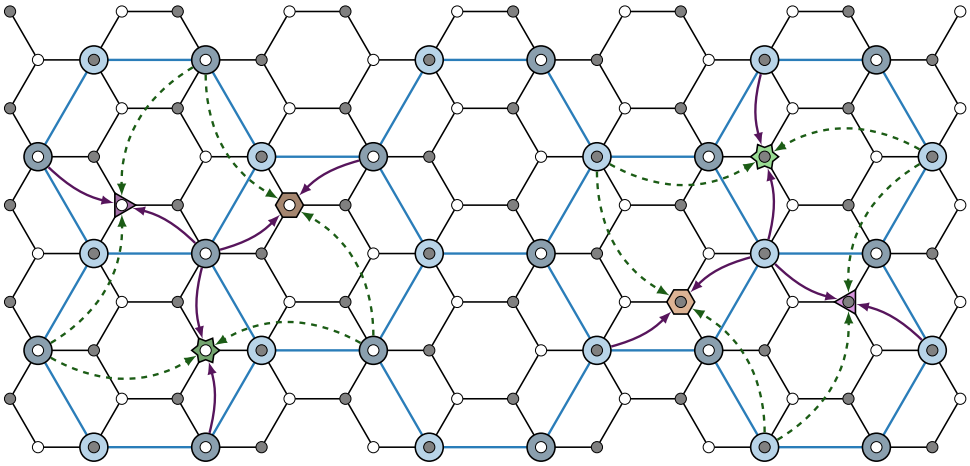


Figure 8: Illustration of interpolation using weights w_s (⟶) and w_ℓ (⇢) [K10].

In order to formulate a geometric multigrid method for this operator we introduce a coarsening that results in a coarse hexagonal lattice at a coarsening rate of $\frac{1}{4}$. Further we specify interpolation that preserves the spectral structure of the operator at the Dirac points as depicted in fig. 8. Denoting the short range interpolation weights by ω_s and long range interpolation weights ω_ℓ, we show that preserving the Dirac points requires

$$\omega_s = 2\omega_\ell - 1$$

and a stable coarse grid operator is obtained for $\omega_\ell \in (\frac{1}{6}, \frac{1}{3})$.

Using a generalized version of local Fourier analysis we are able to prove convergence and scalability of this approach when using Kaczmarz relaxation. As a follow-up of this work a further generalization of local Fourier analysis is in preparation. This work constructs local Fourier analysis solely on position space information, i.e., in terms of shift-invariance of operators. Using this approach, we are able to use local Fourier analysis for operators that live on arbitrary crystal structures with a minimal amount of required user input. This allows us to analyze complex operators, e.g., (overlapping) block smoothers or systems of partial differential equations, in a unified framework.

3 Structure preserving algebraic multigrid

The preservation of structure, be it in terms of symmetries, geometry or quite literally structure, is a recurring topic in numerical analysis. Modelling physical, biological and mechanical processes in mathematical terms usually involves special treatment of important structures of the problem. Sometimes these can be used to simplify the model, sometimes they have to be carefully integrated into the model to ensure that the obtained numerical solution bears resemblance with the actual real world and sometimes structure is automatically encoded in the model. In the latter two cases it is advisable that the inherent structure of the model is not lost in the successive numerical treatment, i.e., in linearization and discretization. Not abiding to this advice might result in anything from catastrophic failure due to numerical instabilities to a poorer quality of the obtained solution or no detrimental effect at all.

While ignoring structure on the subsequent and last level of simulation, i.e., at the level of numerical linear algebra that follows linearization and discretization seldomly has catastrophic effects, trying to preserve structure on this last and most basic level of a simulation seems obvious for the aforementioned reasons. Using as much of the information and structure of a model might be beneficial and can potentially reduce time-to-solution of the resulting linear algebra problems, which make up a large chunk of computing time in many simulation packages.

In my research I have largely explored structure preservation in algebraic multigrid methods in terms of lattice gauge theory, e.g., lattice Quantumchromodynamics, but many of the arguments developed for this particular application can possibly be extended to other applications as well, such as systems of partial differential equations with special symmetries or problems with structure that originate in the geometry of the problem.

I also explored another, completely different, venue of structure preservation in nature of tensor approximations, where the preserved tensor structure is of artificial nature. That is, rather than being a natural choice, tensor approximations are rather a tool to make simulation of the presented interacting systems viable in the first place. This is due to what is known as the curse of dimensionality. In models which describe the interaction of many submodels, the size of the overall system grows exponentially with the number of submodels. This in turn renders a full solution of these models impossible due limited storage and computing resources. However due to their particular structure approximations in a tensor-format are widely accepted to be the method of choice. Structure preservation, i.e., its tensor-format, in algebraic multigrid methods for such models is thus not an option, but a neccessity.

Symmetry preservation in lattice gauge theory

The Dirac equation of Quantumchromodynamics that is simulated in lattice gauge theory is a system of first order partial differential equations, which depends on a mass m and gauge field $A_\mu \in \mathfrak{su}(3)$, i.e., the Lie algebra of the special unitary group SU(3), by

$$\mathcal{D} = m + \sum_{\mu=0}^{3} \gamma_\mu \otimes (\partial_\mu + A_\mu) \ .$$

In here, the matrices $\gamma_\mu \in \mathbb{C}^{4\times 4}$, $\mu = 0, \ldots, 3$ fulfill the anticommutation relations $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu} I$, i.e., they generate a matrix representation of a Clifford algebra. In abstract terms one can think of the Dirac equation as the formal square root of the Klein-Gordon equation. Based on the structure of the γ-matrices and defining $\gamma_5 := i\gamma_0\gamma_1\gamma_2\gamma_3$ such that $\gamma_\mu\gamma_5 + \gamma_5\gamma_\mu = 0$, there are two immediate and important symmetries of the Dirac equation. The γ_5-symmetry, which reads

$$\gamma_5 \mathcal{D} \gamma_5 = \mathcal{D}^*$$

and in addition *chiral* symmetry in the case $m = 0$, defined by

$$e^{i\alpha\gamma_5} \mathcal{D} e^{i\alpha\gamma_5} = \mathcal{D}\,.$$

Due to its origin in the structure of the system of partial differential equations, the γ_5 symmetry can be easily preserved in the discretization using finite covariant differences, i.e., approximating $\partial_\mu + A_\mu$ by

$$(\partial_\mu + A_\mu)\,\psi(x) = \frac{1}{2a}\left(U_\mu^{x-a}\right)^H \psi(x - ae_\mu) - U_\mu^x \psi(x + ae_\mu) + \mathcal{O}(a^2)\,,$$

where $U_\mu^x \approx \int_x^{x+ae_\mu} e^{A_\mu(x)} dx$. This so-called naive discretization is numerically unstable as it has kernel vectors that have no physical meaning. In order to get rid of the numerical instability it has been proposed to add a stabilizing term of second order to obtain what is known as the Wilson formulation

$$\sum_{\mu=0}^{3} \gamma_\mu \otimes (\partial_\mu + A_\mu) \;\longrightarrow\; \sum_{\mu=0}^{3} \gamma_\mu \otimes (\partial_\mu + A_\mu) + a\partial_{\mu\mu}\,.$$

While this discretization is numerically stable, it explicitely breaks chiral symmetry even for $m = 0$. As according to the infamous no-go theorem [51] there exists no local discretization of the Dirac operator that is at the same time numerically stable and preserves chiral symmetry, this is the discretization that is widely used in lattice gauge simulations.

Over the years there have been numerous attempts at finding a multigrid method for the solution of the systems arising in simulations of lattice gauge theories, but none had lasting effect [5, 39, 40, 41]. That is, until the development of adaptive algebraic multigrid methods [3, 16, 43] made a huge impact and structure preservation played a major role in this.

My contributions to the development of structure preserving adaptive algebraic multigrid methods for operators of lattice gauge theory have been published in the following three papers, each of which is summarized shortly.

[K3] Bootstrap algebraic multigrid for the 2D Wilson Dirac operator. In order to reduce the computational overhead of the 4-dimensional lattice gauge theory of Quantumchromodynamics we first considered a 2-dimensional variant associated with Quantumelectrodynamics in [K3].

In this publication we first generalize the bootstrap algebraic multigrid framework for non Hermitian matrices, i.e., we consider and develop the Petrov-Galerkin case where R is not equal to P^H anymore. This generalization is based on the singular value decomposition $A = U\Sigma V^H$ and the assumption that interpolation

P should capture the nature of right-singular vectors to small singular values and restriction R the corresponding left singular vectors. The bootstrap method is then constructed using the equivalence of the singular value decomposition with a Hermitian eigenvalue problem of twice the dimension

$$\underbrace{\begin{bmatrix} 0 & A \\ A^H & 0 \end{bmatrix}}_{\widehat{A}} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} = \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} .$$

Applying the Galerkin bootstrap construction to the Hermitian matrix $\widehat{A}$ immediately implies that smoothing for interpolation test vectors should be based on A for interpolation test vectors and A^H for restriction test vectors and the weighting of these test vectors in least squares interpolation has to be adjusted accordingly. Furthermore, the coarsest grid generalized eigenvalue problem that is used in the bootstrap setup cycle becomes

$$\underbrace{\begin{bmatrix} 0 & A_L \\ A_L^H & 0 \end{bmatrix}}_{\widehat{A}} \begin{bmatrix} U_L & U_L \\ V_L & -V_L \end{bmatrix} = \begin{bmatrix} R_L & 0 \\ 0 & P_L \end{bmatrix} \begin{bmatrix} U_L & U_L \\ V_L & -V_L \end{bmatrix} \begin{bmatrix} \Sigma_L & 0 \\ 0 & -\Sigma_L \end{bmatrix} ,$$

where $R_L = R_1^L (R_1^L)^H$, $P_L = (P_L^1)^H P_L^1$ with $R_L^1 = R_{L-1}^L \cdot R_{L-2}^{L-1} \cdot \ldots \cdot R_1^2$ and $P_L^1 = P_2^1 \cdot P_3^2 \cdot \ldots \cdot P_L^{L-1}$.

If on the other hand interpolation is constructed such that it preserves Γ_5-symmetry, i.e.,

$$\Gamma_5 P = P \widetilde{\Gamma}_5, \quad \Gamma_5 = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = \widetilde{\Gamma}_5,$$

we are able to show that the Petrov-Galerkin generalization of the bootstrap setup cycle is equivalent to a Galerkin bootstrap setup cycle for the Hermitian (indefinite) operator $\Gamma_5 A$. Thus preserving the Γ_5-structure reduces the bootstrap algebraic multigrid setup again to just finding a suitable interpolation operator and eliminates the requirement of using two sets of test vectors.

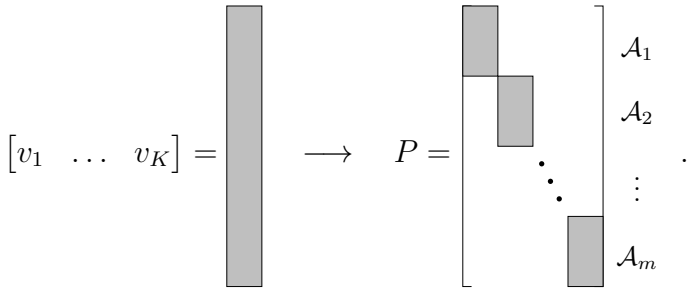
Finally, we show that both the generalized non Hermitian and the structure preserving Hermitian bootstrap algebraic multigrid methods yield robust and scalable solvers for the operators of 2-dimensional lattice gauge theory.

[K4] Adaptive aggregation-based domain decomposition multigrid for the Wilson Dirac operator. Partly based on the findings of [K3] and the previous work in [3, 16] we develop an aggregation based multigrid method with domain decomposition smoother for the 4-dimensional Wilson Dirac operator in [K4].

Due to the 4-dimensional lattice structure and the fact that 12 variables[3] exist at each lattice site the size of the resulting operator grows rapidly and particular care has to be taken in terms of parallelizability of the proposed method. This focus on parallelism leads us to choose a rather simple structure for interpolation that, given vectors of algebraically smooth error, constructs the interpolation operator by

[3] All combinations of 4 spin and 3 color components.

splitting these vectors according to aggregates



In here structure preservation, i.e., preserving the Γ_5-symmetry, is guaranteed by grouping spin components accordingly. The main benefit of the use of an aggregation based coarsening approach is the exact replication of the operator structure on the coarse grid with minimal fill-in.

A similar argument based on the potential of parallelism leads us to choose a simple domain decomposition smoother. While this smoother is not guaranteed to converge on its own, we are able to numerically demonstrate its smoothing property. The setup is done using a multigrid inverse iteration, where the evolving algebraic multigrid method is used to approximate the inverse, such that it resembles the bootstrap algebraic setup cycle. Using numerous optimizations in the implementation (e.g., vectorization, single-precision) and a K-cycle strategy [53] for the resulting solver we are able to show that the developed method is able to outperform standard Krylov subspace methods by a large margin and edge out state-of-the-art two- and multilevel methods. Our algebraic multigrid approach is especially well performing with respect to the measure of core hours per solution, an important energy efficiency measure, due to its ability to employ many coarse levels thanks to symmetry preservation.

In contrast to the multilevel method developed in [3, 16] we opt to choose a more powerful smoother, that in addition can be parallelized more efficiently, by using a domain decomposition method instead of a polynomial smoother. This, combined with a different setup routine, leads to improvements in performance. On the other hand, based on our discussion of the fundamental differences between our method and the deflation ansatz developed in [43], this method has been changed to a true two-grid method, but it does not use structure preserving interpolation, which does not impair its efficiency due to its limitation to a purely two-grid structure.

Nowadays, almost all big research collaborations make use of one of the adaptive algebraic two- or multilevel methods [3, K4, 43] due to their demonstrated potential to speed up simulations in the field of lattice gauge theories. A high performance implementation of our approach is publically available on GitHub[4]. Other open source high performance implementations that include an algebraic multigrid solver have been developed by Nvidia (QUDA[5]) and the OpenQCD[6] collaboration. In addition, the algebraic multigrid approach has been extended to other fermion formulations such as twisted-mass fermions [1], domain wall fermions [25]. Further extensions to staggered fermions and beyond the standard model physics like super symmetry are under investigation.

[4] `github.com/DDalphaAMG`
[5] `github.com/lattice/quda`
[6] `luscher.web.cern.ch/luscher/openQCD`

[K7] Multigrid preconditioning for the overlap operator in lattice QCD. A special operator in lattice gauge theory is the overlap operator as it preserves a discrete version of the chiral symmetry of the Dirac equation. More precisely, it is a solution to the Ginsparg Wilson relation

$$\Gamma_5 D + D\Gamma_5 = aD\Gamma_5 D$$

and can be expressed based on a kernel operator (e.g., the Wilson Dirac operator D_W) by

$$D = \rho I + \Gamma_5 \operatorname{sign}\left(\Gamma_5 D_W(m_{\text{ker}})\right) \ .$$

Thus, in comparison to simulations with the kernel operator itself, simulations with the overlap operator are much more expensive due to the requirement to evaluate the matrix sign function.

In this paper we show that the overlap operator can be effectively preconditioned by a shifted version of its kernel operator. We develop the theoretical foundation for this approach in three steps. First, we show that the overlap operator is indeed a discretization of the Dirac equation in the numerical sense, which, by the argument of auxiliary space preconditioners, makes it plausible that it can be preconditioned by another discretization (in this case the Wilson discretization). Second, under the assumption that the kernel operator is normal, we are able to show a precise bound on the spectrum of the preconditioned operator. Assuming $D_W(0) = X\Lambda X^H$ we find

$$D = X\left(\rho I + \operatorname{csign}\left(\Lambda + mI\right)\right)X \ ,$$

i.e., D_W and D share the same eigenvectors and thus

$$\operatorname{spec}(DD_W(m_{\text{prec}})^{-1}) = \left\{ \frac{\rho + \operatorname{csign}(\lambda + m_{\text{ker}})}{\lambda + m_{\text{prec}}},\ \lambda \in \operatorname{spec}(D_W(0)) \right\} \ .$$

Although D_W is not normal, we are able to show in the third part of the theoretical justification of the preconditioner that its deviation from normality is proportional to the pure gauge action, i.e.,

$$\|D_W D_W^H - D_W^H D_W\|_F^2 = 16 \sum_x \sum_{\mu<\nu} \operatorname{Re}\left(\operatorname{tr}\left(I - Q_x^{\mu,\nu}\right)\right) \ ,$$

where $Q_x^{\mu,\nu} = (U_\nu^x)^H (U_\mu^{x+e_\nu})^H U_\nu^{x+e_\mu} U_\mu^x$ denotes the so-called plaquette. [7] This implies that frequently used gauge smearing techniques, which are known to reduce the pure gauge action, also improve normality of the Wilson Dirac operator. This in turn implies an improved accuracy of our theory and in turn improved performance of the preconditioner when using smeared configurations.

In extensive numerical studies we are able to verify all theoretical arguments that show the connection between the overlap and its kernel operator, which motivated the idea of preconditioning. Using our adaptive algebraic multigrid method as the preconditioner that implements D_W^{-1}, we are able to demonstrate great improvements over the state-of-the-art methods for overlap simulations.

[7] A product of gauge links around a 2 dimensional face of the 4 dimensional lattice.

Structure preservation for tensor structures

Interacting Markov chain models, although they might appear not overly complex at first sight, e.g., a system of a handful of consecutive queues, are very difficult to simulate due to the staggeringly large number of different states that have to be considered. Assuming that each queue has K slots and there are d queues in total, the number of states of the whole system of queues is K^d, which rapidly exceeds the domain of possible complete simulation. This explosion of the number of states is one occurence of what is known as the curse of dimensionality.

Leveraging that the description of such systems can be cast into a sum of tensor-product operators, the approximation of the solution in a tensor approximation format like the tensor train format [56] is the only way to simulate such systems. Of particular interest is the long term behaviour of such Markov-chain models which is encoded in its steady-state vector. Typically methods based on alternating least-squares, e.g., the alternating minimal energy method [26], are employed to compute tensor approximations to the steady-state vector. In [63] an algebraic multigrid method has been considered to accelerate the alternating least squares approach.

In two papers, which I will review in the following, we developed a stand-alone algebraic multigrid method for the approximation of the steady state vector, which is then supplemented by a coarsest grid solver based on the alternating minimal energy method.

[K6] Multigrid methods for tensor structured Markov chains with low rank approximation. Tensor structured Markov chains arise when describing models where Markov chain submodels are synchronised by transitions between the submodels. The generator matrix A can then be written as

$$A = \sum_{t=1}^{T} \bigotimes_{j=1}^{J} E_j^t \,,$$

where $E_j^j \in \mathbb{R}^{n_j \times n_j}$ describe the local transitions within the j-th submodel and $E_j^t,\ j \neq t$ the synchronised transitions between two submodels. Instead of calculating the full steady-state vector $x \in \mathbb{R}^{\prod_{j=1}^{J} n_j}$, which has a prohibitive large memory footprint, one approximates it by a tensor approximation, e.g., in canonical tensor format given by

$$x = \sum_{r=1}^{R} \bigotimes_{j=1}^{J} x_j^r \,.$$

Due to the fact that preserving the tensor structure is of utmost importance in order to be able to carry out any computation at all for these problems, the construction of the algebraic multigrid method has to preserve it by all means. This is achieved by considering interpolation and restriction operators of the form

$$P = \bigotimes_{j=1}^{J} P_j \quad \text{and} \quad R = \bigotimes_{j=1}^{J} R_j \,,$$

respectively. That is, coarsening and interpolation are essentially built for and motivated by the structure and properties of the J submodels. The corresponding

coarse grid operator is then given by

$$A_c = RAP = \sum_{t=1}^{T} \bigotimes_{j=1}^{J} R_j E_j^t P_j \,.$$

Depending on the structure and properties of the submodels and transitions we construct P_j and R_j using the whole arsenal of algebraic multigrid constructions available, ranging from aggregation based and classical "Ruge-Stüben" algebraic multigrid approaches to adaptive algebraic multigrid constructions. Aside from their structure preserving nature, the particular structure of interpolation and restriction has the added benefit of not increasing the tensor rank of a vector represented in tensor format, when applied to it.

In numerical experiments we are able to show that, especially in the situation where the size n_j of the submodels is large and the number J of submodels is only of moderate size, the algebraic multigrid method outperforms the commonly used ALS method. In addition we are able to show scalability of the approach as a preconditioner to a tensorized GMRES iteration.

[K8] Multigrid methods combined with low-rank approximation for tensor structured Markov chains. A remaining drawback of the algebraic multigrid method developed in [K6] is its lower bound on the coarsest level dimension. Due to the fact that fast convergence requires coarsest levels in each submodel dimension of at least 2, the curse of dimensionality is only offset, but not completely removed. For large numbers of dimensions (i.e., submodels) the method slows down as the solution of the coarsest grid becomes too expensive.

The alternating minimal energy method [26], which is a gradient-enriched version of alternating least squares for the tensor-train format, is limited by large submodel dimensions. It thus seems attractive to employ our algebraic multigrid construction and replace the solver on the coarsest grid, which has small submodel dimension but still contains all submodels, by the alternating minimal energy method.

In the numerical tests we consider a wide range of problems ranging from chemistry to network modelling, that the complementary nature of the algebraic multigrid construction and the alternating minimal energy coarsest grid solver yields solver performances that eclipse both methods individually in the case of both large number of submodels J and large submodel size n_j. Thus drastically extending the range of models that can be efficiently simulated.

4 Contributions to the publications by the author

Due to the fact that all publications of this cumulative habilitation are collaborative pieces of work, where all authors contributed to the success of the paper, I want to briefly reflect on my perceived personal contribution to these publications.

The bootstrap algebraic multigrid framework

[K2] Bootstrap AMG

- Development of the algorithm and setup cycle based on remark in [11]
- Idea to formulate coarse grid generalized eigenvalue problem to be used in bootstrap setup
- Implementation and numerical tests

[K1] A bootstrap algebraic multilevel Method For Markov Chains

- Idea to apply bootstrap AMG to Markov chain problems
- Design of coarsening and restrictions for particular problems
- Implementation and numerical tests

[K5] Algebraic distance for anisotropic diffusion problems: Multilevel results

- Idea to use least squares interpolation to determine strength of connection motivated by remark in [8]
- Implementation and numerical tests

[K11] Least angle regression coarsening in bootstrap algebraic multigrid

- Idea to use ℓ_1-norm penalty term to sparsify least squares interpolation and to solve with least angle regression
- Design of the coarsening algorithm
- Joint implementation and numerical tests

Advances in multigrid methods

[K9] Optimal interpolation and compatible relaxation in classical algebraic multigrid

- Joint development of the proof of optimality
- Joint proof of equivalence with ideal interpolation in case of F-smoothing
- Idea to use max-vol algorithm for sparsification/localization of optimal interpolation
- Implementation and numerical tests

[K10] Geometric multigrid for the tight-binding Hamiltonian of graphene

- Idea to construct geometric multigrid method for graphene operator
- Joint development of theory and algorithm
- Idea to generalize theory to arbitrary grid/crystal operators

Symmetry preservation in lattice gauge theory

[K3] Bootstrap algebraic multigrid for the 2D Wilson Dirac operator

- Idea to derive non-Hermitian algebraic multigrid based on twice-as-large Hermitian problem
- Joint development of the theoretical foundation
- Implementation and numerical tests

[K4] Adaptive aggregation-based domain decomposition multigrid for the Wilson Dirac operator

- Outline of the algebraic multigrid method
- Design of the algorithm

[K7] Multigrid preconditioning for the overlap operator in lattice QCD

- Idea to use Wilson Dirac operator to precondition overlap operator
- Proof of spectral relation between Wilson Dirac and overlap operator
- Proof that deviation from normality of Wilson Dirac operator is proportional to pure gauge action
- Joint development of algorithm and implementation

Structure preserving algebraic multigrid for tensor structures

[K6] Structure preservation for tensor structures

- Conceptual development and design (in terms of interpolation and restriction) of structure preserving algebraic multigrid methods for tensor-structured problems
- Joint development of implementation

[K8] Multigrid methods combined with low-rank approximation for tensor structured Markov chains

- Idea to combine algebraic multigrid with AMEn as a coarsest grid solver
- Joint design of algebraic multigrid methods for various problems
- Joint development of implementation

References

[1] C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann. "Adaptive aggregation-based domain decomposition multigrid for twisted mass fermions". *Phys. Rev. D* 94 (2016), p. 114509.

[2] S. F. Ashby and R. D. Falgout. "A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations". *Nucl. Sci. Eng.* 124.1 (1996), pp. 145–159.

[3] R. Babich et al. "Adaptive multigrid algorithm for the lattice Wilson-Dirac operator". *Phys. Rev. Lett.* 105 (2010), p. 201602.

[4] N. Bakhvalov. "On the convergence of a relaxation method with natural constraints on the elliptic operator". *USSR Comput. Math. Math. Phys.* 6.5 (1966), pp. 101–135.

[5] R. Ben-Av et al. "Parallel-transported multi-grid for inverting the Dirac-operator: variants of the method and their efficiency". *Nuclear Physics B* 405.2-3 (1993), pp. 623–666.

[6] D. Braess. "Towards algebraic multigrid for elliptic problems of second order." *Computing* 55.4 (1995), pp. 379–393.

[7] A. Brandt. "Algebraic multigrid theory: the symmetric case". *Appl. Math. Comput.* 19.1–4 (1986), pp. 23–56.

[8] A. Brandt. "General highly accurate algebraic coarsening". *Electron. Trans. Numer. Anal.* 10.1 (2000), p. 21.

[9] A. Brandt. "Multi-level adaptive solutions to boundary value problems". *Math. Comp.* 31 (1977), pp. 333–390.

[10] A. Brandt. "Multi-level adaptive technique (mlat) for fast numerical solution to boundary value problems". In: *Proceedings of the Third International Conference on Numerical Methods in Fluid Mechanics*. Springer, 1973.

[11] A. Brandt. "Multiscale scientific computation: review 2001". *Multiscale and Multiresolution Methods*. Ed. by T. J. Barth, T. Chan, and R. Haimes. Berlin, Heidelberg: Springer, 2002, pp. 3–95.

[12] A. Brandt. "Rigorous quantitative analysis of multigrid, I. constant coefficients two-level cycle with l_2-norm". *SIAM J. Numer. Anal.* 31.6 (1994), pp. 1695–1730.

[13] A. Brandt, S. F. McCormick, and J. W. Ruge. "Algebraic multigrid (AMG) for sparse matrix equations". In: *Sparsity and its Applications*. Ed. by D. J. Evans. Cambridge: Cambridge University Press, 1985, pp. 257–284.

[14] J. Brannick and R. D. Falgout. "Compatible relaxation and coarsening in algebraic multigrid". *SIAM J. Sci. Comput.* 32.3 (2010), pp. 1393–1416.

[15] J. Brannick and L. T. Zikatanov. "Algebraic multigrid methods based on compatible relaxation and energy minimization". *Domain Decomposition Methods in Science and Engineering XVI*. Ed. by O. B. Widlund and D. E. Keyes. Berlin, Heidelberg: Springer, 2007, pp. 15–26.

[16] J. Brannick et al. “Adaptive multigrid algorithm for lattice QCD”. *Phys. Rev. Lett.* 100 (2008), p. 041601.

[17] J. Brannick et al. “An algebraic multigrid method based on matching in graphs”. In: *Domain Decomposition Methods in Science and Engineering XX*. Ed. by R. Bank et al. Berlin, Heidelberg: Springer, 2013, pp. 143–150.

[18] M. Brezina et al. “Adaptive algebraic multigrid.” *SIAM J. Sci. Comput.* 27.4 (2006), pp. 1261–1286.

[19] M. Brezina et al. “Adaptive smoothed aggregation (αSA).” *SIAM J. Sci. Comput.* 25.6 (2004), pp. 1896–1920.

[20] M. Brezina et al. “Adaptive smoothed aggregation (αSA) multigrid.” *SIAM Rev.* 47.2 (2005), pp. 317–346.

[21] M. Brezina et al. “Algebraic multigrid based on element interpolation (AMGe).” *SIAM J. Sci. Comput.* 22.5 (2001), pp. 1570–1592.

[22] M. Brezina et al. “Relaxation-corrected bootstrap algebraic multigrid (rBAMG)”. *Numer. Linear Algebra Appl.* 19.2 (2012), pp. 178–193.

[23] V. E. Bulgakov. “Multi-level iterative technique and aggregation concept with semi-analytical preconditioning for solving boundary-value problems”. *Commun. Numer. Methods Eng.* 9.8 (1993), pp. 649–657.

[24] T. P. Chartier et al. “Spectral AMGe (ρAMGe).” *SIAM J. Sci. Comput.* 25.1 (2003), pp. 1–26.

[25] S. D. Cohen et al. “Multigrid algorithms for domain-wall fermions”. *Proceedings of Science* LATTICE2011 (2011), p. 7.

[26] S. Dolgov and D. V. Savostyanov. “Alternating minimal energy methods for linear systems in higher dimensions.” *SIAM J. Sci. Comput.* 36.5 (2014).

[27] H. V. Emden and P. S. Vassilevski. “Element-free AMGe: general algorithms for computing interpolation weights in AMG”. *SIAM J. Sci. Comput.* 23.2 (2001), pp. 629–650.

[28] R. D. Falgout and J. B. Schroder. “Non-Galerkin coarse grids for algebraic multigrid”. *SIAM J. Sci. Comput.* 36.3 (2014), pp. C309–C334.

[29] R. D. Falgout, P. S. Vassilevski, and L. T. Zikatanov. “On two-grid convergence estimates”. *Numer. Linear Algebra Appl.* 12.5-6 (2005), pp. 471–494.

[30] R. P. Fedorenko. “A relaxation method for solving elliptic difference equations”. *USSR Comput. Math. Math. Phys.* 1.5 (1962), pp. 1092–1096.

[31] S. Friedhoff and S. P. MacLachlan. “A generalized predictive analysis tool for multigrid methods”. *Numer. Linear Algebra Appl.* 22.4 (2015), pp. 618–647.

[32] S. Friedhoff, S. P. MacLachlan, and C. Börgers. “Local fourier analysis of space-time relaxation and multigrid schemes”. *SIAM J. Sci. Comput.* 35.5 (2013), S250–S276.

[33] F. J. Gaspar, J. L. Gracia, and F. J. Lisbona. “Fourier analysis for multigrid methods on triangular grids”. *SIAM J. Sci. Comput.* 31.3 (2009), pp. 2081–2102.

[34] W. Hackbusch. "On the multi-grid method applied to difference equations." *Computing* 20.4 (1978), pp. 291–306.

[35] W. Hackbusch and U. Trottenberg. *Multigrid Methods*. Lecture Notes in Mathematics 960. Springer, 1982, pp. X, 662.

[36] Y. He and S. P. MacLachlan. "Local fourier analysis of block-structured multigrid relaxation schemes for the stokes equations". *Numer. Linear Algebra Appl.* 25.3 (2018). e2147 nla.2147, e2147.

[37] T. K. Huckle and C. Kravvaritis. "Compact Fourier analysis for multigrid methods based on block symbols". *SIAM J. Matrix Anal. Appl.* 33.1 (2012), pp. 73–96.

[38] J. E. Jones and P. S. Vassilevski. "AMGe based on element agglomeration." *SIAM J. Sci. Comput.* 23.1 (2001), pp. 109–133.

[39] T. Kalkreuter. "Idealized multigrid algorithm for staggered fermions". *Phys. Rev. D* 48.5 (1993), R1926.

[40] T. Kalkreuter. "Multigrid methods for propagators in lattice gauge theories". *Journal of Computational and Applied Mathematics* 63.1 (1995). Proceedings of the International Symposium on Mathematical Modelling and Computational Methods Modelling 94, pp. 57–68.

[41] T. Kalkreuter, G. Mack, and M. Speh. "Blockspin and multigrid for staggered fermions in non-abelian gauge fields". *International Journal of Modern Physics C* 3.01 (1992), pp. 121–147.

[42] O. E. Livne and A. Brandt. "Lean algebraic multigrid (LAMG): fast graph Laplacian linear solver". *SIAM J. Sci. Comput.* 34.4 (2012), B499–B522.

[43] M. Lüscher. "Local coherence and deflation of the low quark modes in lattice QCD". *J. High Energy Phys.* 2007.07 (2007), p. 22.

[44] S. P. MacLachlan and C. W. Oosterlee. "Local Fourier analysis for multigrid with overlapping smoothers applied to systems of pdes." *Numer. Linear Algebra Appl.* 18.4 (2011), pp. 751–774.

[45] J. Mandel, M. Brezina, and P. Vaněk. "Energy optimization of algebraic multigrid bases". *Computing* 62.3 (1999), pp. 205–228.

[46] T. A. Manteuffel et al. "Operator-based interpolation for bootstrap algebraic multigrid." *Numer. Linear Algebra Appl.* 17.2-3 (2010), pp. 519–537.

[47] S. F. McCormick and J. W. Ruge. "Algebraic multigrid methods applied to problems in computational structural mechanics" (1989). Ed. by A. K. Noor and J. T. Oden, pp. 237–270.

[48] A. Napov and Y. Notay. "An efficient multigrid method for graph Laplacian systems". *Electron. Trans. Numer. Anal.* 45 (2016), pp. 201–218.

[49] R. A. Nicolaides. "On multiple grid and related techniques for solving discrete elliptic systems". *J. Comput. Phys.* 19.4 (1975), pp. 418–431.

[50] R. A. Nicolaides. "On the l^2 convergence of an algorithm for solving finite element equations". *Math. Comp.* 31.140 (1977), pp. 892–906. URL: `http://www.jstor.org/stable/2006120`.

[51] H. B. Nielsen and M. Ninomiya. "No go theorem for regularizing chiral fermions". *Physics Letters* B105 (1981), p. 219.

[52] Y. Notay. "An aggregation-based algebraic multigrid method". *Electron. Trans. Numer. Anal.* 37 (2010), pp. 123–146.

[53] Y. Notay and P. S. Vassilevski. "Recursive Krylov-based multigrid cycles". *Numer. Linear Algebra Appl.* 15.5 (2007), pp. 473–487.

[54] L. N. Olson, J. B. Schroder, and R. S. Tuminaro. "A general interpolation strategy for algebraic multigrid using energy minimization". *SIAM J. Sci. Comput.* 33.2 (2011), pp. 966–991.

[55] L. N. Olson, J. B. Schroder, and R. S. Tuminaro. "A new perspective on strength measures in algebraic multigrid." *Numer. Linear Algebra Appl.* 17.4 (2010), pp. 713–733.

[56] I. V. Oseledets. "Tensor-train decomposition." *SIAM J. Sci. Comput.* 33.5 (2011), pp. 2295–2317.

[57] B. Philip and T. P. Chartier. "Adaptive algebraic smoothers". *Journal of Computational and Applied Mathematics* 236.9 (2012), pp. 2277–2297.

[58] A. Reusken. "Multigrid with matrix-dependent transfer operators for a singular perturbation problem." *Computing* 50.3 (1993), pp. 199–211.

[59] A. Reusken. "On a robust multigrid solver." *Computing* 56.3 (1996), pp. 303–322.

[60] J. W. Ruge and K. Stüben. "Algebraic multigrid". In: *Multigrid Methods*. Ed. by S. F. McCormick. Frontiers in Applied Mathematics. SIAM, 1987. Chap. 4, pp. 73–130.

[61] R. V. Southwell. *Relaxation Methods in Theoretical Physics*. Oxford: Clarendon Press, 1946, p. 248.

[62] H. D. Sterck. "A self-learning algebraic multigrid method for extremal singular triplets and eigenpairs". *SIAM J. Sci. Comput.* 34.4 (2012), A2092–A2117.

[63] H. D. Sterck and K. Miller. "An adaptive algebraic multigrid algorithm for low-rank canonical tensor decomposition". *SIAM J. Sci. Comput.* 35.1 (2013), B1–B24.

[64] H. D. Sterck et al. "Algebraic multigrid for Markov chains." *SIAM J. Sci. Comput.* 32.2 (2010), pp. 544–562.

[65] H. D. Sterck et al. "Multilevel adaptive aggregation for Markov chains, with application to web ranking." *SIAM J. Sci. Comput.* 30.5 (2008), pp. 2235–2262.

[66] H. D. Sterck et al. "Recursively accelerated multilevel aggregation for Markov chains." *SIAM J. Sci. Comput.* 32.3 (2010), pp. 1652–1671.

[67] H. D. Sterck et al. "Smoothed aggregation multigrid for markov chains." *SIAM J. Sci. Comput.* 32.1 (2010), pp. 40–61.

[68] K. Stüben. "Algebraic multigrid (AMG): experiences and comparisons". *Appl. Math. Comput.* 13.3-4 (1983), pp. 419–451.

[69] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid*. Vol. 33. Texts in Applied Mathematics. Bd. With contributions by A. Brandt, P. Oswald and K. Stüben. San Diego [u.a.]: Academic Press, 2001.

[70] W. Wan, T. Chan, and B. Smith. "An energy-minimizing interpolation for robust multigrid methods". *SIAM J. Sci. Comput.* 21.4 (1999), pp. 1632–1649.

[71] R. Wienands and I. Yavneh. "Collocation coarse approximation in multigrid." *SIAM J. Sci. Comput.* 31.5 (2009), pp. 3643–3660.

[72] J. Xu and L. T. Zikatanov. "On an energy minimizing basis for algebraic multigrid methods". *Computing and Visualization in Science* 7.3 (2004), pp. 121–127.

[73] G. Zhou and S. R. Fulton. "Fourier analysis of multigrid methods on hexagonal grids". *SIAM J. Sci. Comput.* 31.2 (2009), pp. 1518–1538.