M. Griebel, P. Oswald

Greedy and Randomized Versions of the Multiplicative Schwarz Method

Greedy and Randomized Versions of the Multiplicative Schwarz Method

Michael Griebel[a], Peter Oswald[b]

[a]*Institut für Numerische Simulation, Universität Bonn, Wegelerstr.6, D-53115 Bonn, and Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, D-53754 Sankt Augustin*
[b]*Jacobs University Bremen, Research I, 600 Campus Ring 1, D-28759 Bremen*

Abstract

We consider sequential, i.e., Gauss-Seidel type, subspace correction methods for the iterative solution of symmetric positive definite variational problems, where the order of subspace correction steps is not deterministically fixed as in standard multiplicative Schwarz methods. Here, we greedily choose the subspace with the largest (or at least a relatively large) residual norm for the next update step, which is also known as the Gauss-Southwell method. We prove exponential convergence in the energy norm, with a reduction factor per iteration step directly related to the spectral properties, e.g., the condition number, of the underlying space splitting. To avoid the additional computational cost associated with the greedy pick, we alternatively consider choosing the next subspace randomly, and show similar estimates for the expected error reduction. We give some numerical examples, in particular applications to a Toeplitz system and to multilevel discretizations of an elliptic boundary value problem, which illustrate the theoretical estimates.

Keywords: subspace correction, Gauss-Southwell, multiplicative Schwarz, greedy, randomized
2000 MSC: 65F10, 65F08, 65N22, 65H10

1. Introduction

With this note, we continue our collaboration [7, 17, 9] on iterative solvers for variational problems based on Hilbert space splittings, so-called additive and multiplicative Schwarz methods. While in standard multiplicative Schwarz methods the subproblems are traversed in a fixed cyclic order, the new element now is to choose the ordering in a greedy fashion according to residual information, or randomly. For linear systems and Gauss-Seidel methods (a special instance of multiplicative Schwarz methods) the greedy ordering goes back to Gauss and Seidel, and has been popularized by Southwell [24]. The method has been theoretically analyzed in the framework of coordinate descent methods for convex optimization methods [30, 14]. It is also a specific instance of a projection onto convex sets (POCS) algorithm. Lately, it has been revived in the context of sparse approximation and compressed sensing [6, 3, 4]. Randomized versions are also used but investigated little (see [25, 15, 5] for the theory of a randomized Kaczmarz iteration). Moreover, Hilbert space splittings underlying the theory of Schwarz methods have reappeared as fusion frames (see [19] for a more detailed account on this connection).

Given these more recent developments, we decided to formulate and prove convergence results for a greedy version of the multiplicative Schwarz method (called Southwell-Schwarz method) for the case of splittings into finitely many subproblems. The main result, an upper bound for the error reduction in energy norm per iteration step, reveals the role of the spectral bounds characterizing the space splitting and shows that greedy strategies can help to improve and stabilize the performance of multiplicative Schwarz methods. We admit that the result itself can be recovered from the optimization literature but hope that stating it in the context of abstract subspace correction methods will make it more accessible to readers working on large-scale PDE discretizations. Similar bounds for the expected convergence

rate are proved if the subproblem ordering is appropriately randomized. This result includes results on the randomized Kaczmarz iteration from [25] as a partial case.

The remainder of this paper is organized as follows. In the next section we briefly summarize the necessary facts on stable space splittings and abstract iterative Schwarz methods, before introducing two new versions called Southwell-Schwarz and random-order Schwarz iteration, respectively. In Section 3, the main convergence results for these methods are formulated and proved. We compare them with the convergence estimates for standard multiplicative Schwarz iterations, and discuss the relation to the recent results for the Kaczmarz method from [25]. In Section 4 some numerical tests are presented for a Toeplitz system and for multilevel discretizations of an elliptic boundary value problem, which illustrate our theoretical findings.

Note added in revision. After submitting this paper, we learned about the recent papers [13, 16, 20] in which the authors considered randomized block-coordinate decent methods for certain classes of convex optimization problems. When restricted to the quadratic case considered here, their results are closely related to ours for non-overlapping space splittings.

2. Space Splittings and Multiplicative Schwarz Algorithms

Consider a (possibly infinite-dimensional but separable) Hilbert space V, let $a(\cdot, \cdot)$ be a continuous symmetric positive definite bilinear form on V, and let F be a bounded linear functional on V. We use the notation V_a if we consider V as Hilbert space with the scalar product given by this bilinear form $a(\cdot, \cdot)$. To solve the variational problem

(A) Find $u \in V$ such that
$$a(u, v) = F(v) \qquad \forall v \in V,$$

we use an iterative subspace correction scheme based on stable space splittings. Let now the separable Hilbert space V_a with associated scalar product $a(\cdot, \cdot)$ be represented by a finite number of Hilbert spaces V_{a_i} with associated scalar products $a_i(\cdot, \cdot)$ and corresponding linear bounded operators $R_i : V_{a_i} \to V_a$, $i = 1, \ldots, N$, as follows:

$$V_a = \sum_{i=1}^{N} R_i V_{a_i} := \{v = \sum_{i=1}^{N} R_i v_i : v_i \in V_{a_i}, \ i = 1, \ldots, N\}. \tag{1}$$

We allow for redundancy, i.e., we do not assume that V_a is the direct sum of its subspaces $R_i V_{a_i}$. We call (1) a stable space splitting, if

$$0 < \lambda_{\min} := \inf_{u \in V_a} \frac{a(u, u)}{\|\|u\|\|^2} \ \leq \ \lambda_{\max} := \sup_{u \in V_a} \frac{a(u, u)}{\|\|u\|\|^2} < \infty, \tag{2}$$

where

$$\|\|u\|\|^2 := \inf_{v_i \in V_{a_i} : \ u = \sum_{i=1}^{N} R_i v_i} \sum_{i=1}^{N} a_i(v_i, v_i).$$

The constants $\lambda_{\min}$ and $\lambda_{\max}$ are called lower and upper stability constants, and $\kappa := \lambda_{\max}/\lambda_{\min}$ is called the condition number of the space splitting (1), respectively. It is easy to see that frames and fusion frames [19] are special cases of this definition, where $a(\cdot, \cdot) = (\cdot, \cdot)_V$, the V_{a_i} are closed subspaces of $V = V_a$, the scalar products $a_i(\cdot, \cdot) = w_i^2(\cdot, \cdot)_V$ are modified by weights $w_i > 0$, and R_i denote the natural embeddings. In the frame case, the V_{a_i} are one-dimensional and spanned by individual frame elements.

For the setup of iterative Schwarz methods we will additionally assume that we have upper bounds for the norms of the operators $R_i : V_{a_i} \to V_a$, i.e., we know positive constants γ_i such that

$$a(R_i v_i, R_i v_i) \leq \gamma_i a_i(v_i, v_i), \qquad \forall \, v_i \in V_{a_i}, \quad i = 1, \ldots, N. \tag{3}$$

2

It is easy to see that then the upper stability bound $\lambda_{\max}$ of a stable splitting satisfies the inequalities

$$\max_i \|R_i\|^2_{V_{a_i} \to V_a} \leq \lambda_{\max} \leq \sum_i \|R_i\|^2_{V_{a_i} \to V_a} \leq \sum_i \gamma_i. \tag{4}$$

We define operators $T_i : V_a \to V_{a_i}$ via the variational problems

$$a_i(T_i v, v_i) = a(v, R_i v_i) \qquad \forall\, v_i \in V_{a_i}, \tag{5}$$

to be solved for given $v \in V_a$ on the spaces V_{a_i}, $i = 1, \ldots, N$. Using these T_i, analogs of the classical Jacobi-Richardson and Gauss-Seidel iterations, called additive and multiplicative Schwarz methods associated with the stable space splitting (1) can be defined and investigated, pretty much along the lines of the standard methods, see [7, 8, 9, 17, 19, 31, 32]. The multiplicative Schwarz methods considered in these papers traverse the subspaces one by one in a fixed, problem-independent order.

In the present note, we consider two versions of the multiplicative Schwarz method for solving (A), where the order of subspace correction steps depends either on residual information or is randomized.

The first scheme is called *Southwell-Schwarz iteration* associated with (1), and represents a version of a greedy algorithm in a Hilbert space, see [28, 27]. To this end, let $0 < \omega < 2$ (relaxation parameter) and $\beta \in (0, 1]$ (weakness parameter) be fixed, start with some initial guess $u^{(0)} \in V_a$, and repeat the following steps for $m = 0, 1, \ldots$, until a stopping criterion is satisfied:

1. **Residual computation**: Determine $r_i^{(m)} := T_i e^{(m)}$, $i = 1, \ldots, N$, where $e^{(m)} := u - u^{(m)}$. This can be done since the right-hand side in the corresponding subproblems (5) reads

$$a(e^{(m)}, R_i v_i) = F(R_i v_i) - a(u^{(m)}, R_i v_i),$$

   and does not depend on knowledge about u.

2. **Greedy pick**: Choose an index i^* such that

$$\frac{1}{\gamma_{i^*}} a_{i^*}(r_{i^*}^{(m)}, r_{i^*}^{(m)}) \geq \beta^2 \max_{i=1,\ldots,N} \frac{1}{\gamma_i} a_i(r_i^{(m)}, r_i^{(m)}). \tag{6}$$

3. **Linear update**: Set

$$u^{(m+1)} = u^{(m)} + \frac{\omega}{\gamma_{i^*}} R_{i^*} r_{i^*}^{(m)}. \tag{7}$$

In contrast to fixed-order multiplicative Schwarz methods, where in each step only one residual needs to be computed, in the Southwell-Schwarz method all residuals $r_i^{(m)}$ need to be computed in the first substep, and compared with each other in the second substep. To clarify this difference, consider the classical case of solving a linear system $Ax = b$ with a symmetric positive definite $N \times N$ matrix A, and a subspace splitting associated with the basis of unit coordinate vectors e_i. Set $V = \mathbb{R}^N$, $a(x, y) = y^T A x$, $F(x) = b^T x$, define $V_{a_i} = \mathbb{R}$ with scalar product $a_i(x_i, y_i) = x_i y_i$, and let $R_i(x_i) = x_i e_i$ be the injections from $V_{a_i} = \mathbb{R}$ into V_a, $i = 1, \ldots, N$. Obviously, the stability constants $\lambda_{\min}$, $\lambda_{\max}$, and the condition number κ of this space splitting coincide with the extreme eigenvalues and the spectral condition number of the matrix A, respectively, and in (3) we can set

$$\gamma_i = \|R_i\|^2_{V_{a_i} \to V_a} = a_{ii}.$$

It is easy to see that $T_i y = (Ay)_i = \sum_{j=1}^{N} a_{ij} y_j$, and that a single update step in the above methods coincides with a standard Gauss-Seidel/SOR update:

$$x_i^{(m+1)} = x_i^{(m)} + \frac{\omega}{a_{ii}} (b - Ax^{(m)})_i, \qquad x_j^{(m+1)} = x_j^m, \quad j \neq i.$$

In this particular situation, the above Southwell-Schwarz iteration reduces to the classical Gauss-Seidel/SOR-Southwell (for short GS-Southwell) method. The main difference to the standard Gauss-Seidel/SOR method (for short GS method) is the operation count per single update step. Indeed, for

dense A a GS-Southwell update takes $O(N^2)$ elementary operations due to the need for computing the whole residual vector $r^{(m)} = b - Ax^{(m)}$, while in the GS method only $O(N)$ operations are needed since just a single residual component is needed. For sparse A the comparison is more subtle, see the remarks in the last section.

In the second scheme, which we call *random-order Schwarz iteration*, the first two of the above substeps are replaced by a single residual computation for a randomly picked subspace:

1'. **Random pick and residual computation**: Choose an index $i \in \{1, 2, \ldots, N\}$ randomly with probability $p_i = \gamma_i / (\gamma_1 + \ldots + \gamma_N)$ (we assume independence for different m), denote the result by i^*, and compute the corresponding residual $r_{i^*}^{(m)}$.

2'. **Linear update**: Set

$$u^{(m+1)} = u^{(m)} + \frac{\omega}{\gamma_{i^*}} R_{i^*} r_{i^*}^{(m)}. \tag{8}$$

Note that the random-order Schwarz iteration does not suffer from a cost penalty. In the above example of solving a linear system with dense matrix A based on the standard coordinate splitting, each update step of the resulting random-order GS method can be done in $O(N)$ operations, the cost for picking i^* does not exceed $O(\log N)$, and is negligible. A combination of greedy and random pick ideas will be introduced in the next section, as indicated by the numerical tests reported in Section 4, it can considerably improve the convergence of the random-order Schwarz iteration at little extra cost.

3. Convergence Results and Discussion

Our main theoretical result is stated in the following theorem.

Theorem 1 *Assume that (1) is a stable space splitting of the Hilbert space V_a, with lower stability constant $\lambda_{\min}$ given by (2), and assume that the constants $\gamma_i > 0$, $i = 1, \ldots, N$, satisfy (3).*
a) The relative error reduction of the Southwell-Schwarz iteration for solving the variational problem (A) with initial guess $u^{(0)}$ and parameters $0 < \omega < 2$, $0 < \beta \leq 1$ is bounded by

$$\|u - u^{(m)}\|_a^2 \leq \left(1 - \frac{\beta^2 \omega (2 - \omega) \lambda_{\min}}{\sum_i \gamma_i}\right)^m \|u - u^{(0)}\|_a^2, \qquad m \geq 1. \tag{9}$$

b) The random-order Schwarz iteration converges with the expected error decay rate

$$E(\|u - u^{(m)}\|_a^2) \leq \left(1 - \frac{\omega (2 - \omega) \lambda_{\min}}{\sum_i \gamma_i}\right)^m \|u - u^{(0)}\|_a^2, \qquad m \geq 1. \tag{10}$$

Proof. Both parts a) and b) use the same recursive error bound which can be deduced from the update formula (7). Consider this formula for any $i = 1, \ldots, N$ (i.e., replace i^* by a generic i), use the definition of T_i in (5) together with the residual formula $T_i e^{(m)} = r_i^{(m)}$ and (3) as follows:

$$
\begin{aligned}
\|e^{(m+1)}\|_a^2 &= \|e^{(m)} - (u^{(m+1)} - u^{(m)})\|_a^2 \\
&= \|e^{(m)}\|_a^2 - 2\frac{\omega}{\gamma_i} a(e^{(m)}, R_i r_i^{(m)}) + \frac{\omega^2}{\gamma_i^2} \|R_i r_i^{(m)}\|_a^2 \\
&\leq \|e^{(m)}\|_a^2 - \frac{\omega(2-\omega)}{\gamma_i} a_i(r_i^{(m)}, r_i^{(m)}) \\
&= \|e^{(m)}\|_a^2 \left(1 - \frac{\omega(2-\omega)}{\gamma_i} \frac{a_i(r_i^{(m)}, r_i^{(m)})}{\|e^{(m)}\|_a^2}\right).
\end{aligned}
$$

4

For part a), we observe that the maximum of a sequence of non-negative numbers always exceeds any of their convex combinations, thus by (6)

$$\frac{1}{\gamma_{i*}} a_{i*}(r_{i*}^{(m)}, r_{i*}^{(m)}) \geq \beta^2 \max_i \frac{1}{\gamma_i} a_i(r_i^{(m)}, r_i^{(m)}) \geq \beta^2 \frac{\sum_i a_i(r_i^{(m)}, r_i^{(m)})}{\sum_i \gamma_i},$$

if we use $\gamma_i / \sum_j \gamma_j$ as weights. Consequently,

$$\|e^{(m+1)}\|_a^2 \leq \|e^{(m)}\|_a^2 \left(1 - \frac{\beta^2 \omega(2-\omega)}{\sum_i \gamma_i} \frac{\sum_i a_i(r_i^{(m)}, r_i^{(m)})}{\|e^{(m)}\|_a^2}\right).$$

It now remains to observe (see [17] or [19, Theorem 1]) that

$$\sum_i a_i(T_i v, T_i v) \geq \lambda_{\min} a(v, v) = \lambda_{\min} \|v\|_a^2, \qquad v \in V_a, \tag{11}$$

where the lower stability bound $\lambda_{\min}$ of the underlying space splitting (1) comes into play. Indeed, consider an arbitrary representation $v = \sum_i R_i v_i$, then

$$a(v, v) = \sum_i a(v, R_i v_i) = \sum_i a_i(T_i v, v_i) \leq (\sum_i a_i(T_i v, T_i v))^{1/2} (\sum_i a_i(v_i, v_i))^{1/2}.$$

Squaring this estimate and taking the infimum with respect to all representations, we arrive at

$$
\begin{aligned}
a(v, v)^2 &\leq (\sum_i a_i(T_i v, T_i v)) \inf_{v = \sum_i R_i v_i} \sum_i a(v_i, v_i) \\
&= (\sum_i a_i(T_i v, T_i v)) \|v\|^2 \leq \lambda_{\min}^{-1} (\sum_i a_i(T_i v, T_i v)) a(v, v).
\end{aligned}
$$

This establishes (11), and applying the latter with $v = e^{(m)}$ and $r_i^{(m)} = T_i e^{(m)}$, we get

$$\|e^{(m+1)}\|_a^2 \leq \left(1 - \frac{\lambda_{\min} \beta^2 \omega(2-\omega)}{\sum_i \gamma_i}\right) \|e^{(m)}\|_a^2.$$

This finishes the proof of (9).

For part b), we similarly estimate the expectation of $\|e^{(m+1)}\|_a^2$ conditioned to given $u^{(m)}$ (and thus $\|e^{(m)}\|_a^2$):

$$
\begin{aligned}
E(\|e^{(m+1)}\|_a^2 | u^{(m)}) &\leq \sum_i \frac{\gamma_i}{\sum_j \gamma_j} \|e^{(m)}\|_a^2 \left(1 - \frac{\omega(2-\omega)}{\gamma_i} \frac{a_i(r_i^{(m)}, r_i^{(m)})}{\|e^{(m)}\|_a^2}\right) \\
&= \|e^{(m)}\|_a^2 \left(1 - \frac{\omega(2-\omega)}{\sum_i \gamma_i} \frac{\sum_i a_i(r_i^{(m)}, r_i^{(m)})}{\|e^{(m)}\|_a^2}\right) \\
&\leq \left(1 - \frac{\omega(2-\omega)\lambda_{\min}}{\sum_i \gamma_i}\right) \|e^{(m)}\|_a^2.
\end{aligned}
$$

Finally, by the assumed independence of picking i^* for different m, we get the desired recursion for the unconditional expectations

$$E(\|e^{(m+1)}\|_a^2) \leq \left(1 - \frac{\omega(2-\omega)\lambda_{\min}}{\sum_i \gamma_i}\right) E(\|e^{(m)}\|_a^2).$$

This proves (10).

We add some comments on the role of the various parameters entering the convergence estimates of Theorem 1, and on the relation to known results for standard Schwarz iterations and other subspace correction methods.

5

- First of all, the estimates suggest that $\omega = 1$ is the best choice for the relaxation parameter, even though it is well-known that for certain applications, over- ($\omega > 1$) or under-relaxation ($\omega < 1$) pays off. The question of choosing ω is intertwined with our choice of γ_i in (3). Since only in special cases, e.g., in the frame case, where the V_{a_i} are one-dimensional, the norms of the operators R_i are exactly known, we have to live with rough guesses for the γ_i. Equivalently, this can be recast as the problem of optimal scaling of the subproblems in V_{a_i} which does not have a trivial solution.

- A comparison of the Southwell-Schwarz iteration with the standard iterative Schwarz methods is in order. If we choose equal values $\gamma_i = \lambda_{\max}$ (this is admissible according to (4)) then (9) implies for $\beta = \omega = 1$ the estimate

$$\|u - u^{(m)}\|_a^2 \le (1 - \frac{1}{N\kappa})^m \|u - u^{(0)}\|_a^2, \qquad m \ge 1.$$

Therefore, after N steps of the Southwell-Schwarz iteration the square energy error reduction is bounded by a constant factor

$$(1 - 1/(N\kappa))^N \approx e^{-1/\kappa} \approx 1 - 1/\kappa \tag{12}$$

which is controlled by the condition number κ of the underlying stable space splitting. Qualitatively, this is at least as good as any of the existing estimates for standard Schwarz iterations with a fixed order of subproblem traversal can guarantee. Each iteration step of these methods is formally equivalent to overall N Southwell-Schwarz steps if we neglect the cost for computing the next Southwell-Schwarz index i^*. E.g., the optimal convergence rate for the additive Schwarz (AS) method

$$u^{(n+1)} = u^{(n)} + \omega^* \sum_{i=1}^{N} R_i r_i^{(n)}, \qquad \omega^* := \frac{2}{\lambda_{\min} + \lambda_{\max}},$$

is given by the sharp estimate

$$\frac{\|u^{(n+1)} - u\|_a}{\|u^{(n)} - u\|_a} \le 1 - \frac{2}{1 + \kappa}, \qquad n \ge 0, \tag{13}$$

where κ is the condition number of the underlying stable space splitting. Obviously, the AS method generalizes the Jacobi-Richardson iteration. This shows that the estimate (9) is superior to (13) if $\sum_i \gamma_i << N\lambda_{\max}$. Recall that for the classical example of solving a linear system with symmetric positive definite matrix and the canonical space splitting discussed before, we automatically have

$$\sum_i \gamma_i = \sum_i a_{ii} = \operatorname{tr}(A) \le N\lambda_{\max}.$$

Similarly, the best available general convergence estimate [9] for multiplicative Schwarz (MS) methods is

$$\frac{\|u^{(n+1)} - u\|_a}{\|u^{(n)} - u\|_a} \le 1 - \frac{C}{\log_2(2N)\kappa}, \qquad n \ge 0, \tag{14}$$

where the constant C depends on the appropriate choice of the relaxation parameter ω. As for the AS method, one step of the MS method corresponds to N single update steps of the form (7), whereas each subspace problem is solved once and in a fixed order (such as $i = 1, 2, \ldots, N$). Since the additional log-factor in (14) cannot be removed in general [18], this shows that there are situations where the convergence estimate (9) for the Southwell-Schwarz method is superior. We refer to [32] for a detailed study of convergence of fixed-order MS methods. Note that these remarks also apply to the comparison of the random-order Schwarz iteration with standard iterative Schwarz methods.

- If we set $\beta = \omega = 1$, the estimate (9) for the deterministic Southwell-Schwarz iteration is identical to the estimate (10) for the expected error of the random-order Schwarz iteration in part b). Experiments reported in the next section show that, with respect to the actual error decay, the more costly Southwell-Schwarz iteration often performs better than the random-order Schwarz iteration. This motivates the introduction of the following algorithm. Fix an integer parameter $1 \le k \le N$.

  1". **Block random pick and greedy residual computation**: Choose randomly and independently k indices $i_1, \ldots, i_k \in \{1, \ldots, N\}$ according to the fixed probabilities $p_i = \gamma_i / \sum_j \gamma_j$, and compute the associated residuals $r_{i_l}^{(m)}$, $l = 1, \ldots, k$. Among these indices, pick i^* greedily according to

$$\frac{1}{\lambda_{i^*}} a_{i^*}(r_{i^*}^{(m)}, r_{i^*}^{(m)}) = \max_{l=1,\ldots,k} \frac{1}{\lambda_{i_l}} a_{i_l}(r_{i_l}^{(m)}, r_{i_l}^{(m)}). \tag{15}$$

  2'. **Linear update**: Set

$$u^{(m+1)} = u^{(m)} + \frac{\omega}{\gamma_{i^*}} R_{i^*} r_{i^*}^{(m)}.$$

  The case $k = 1$ corresponds to the random-order Schwarz iteration, while $k = N$ is basically equivalent to the Southwell-Schwarz iteration. As will be demonstrated in the next section, already for relatively small k the new combined method converges as fast as the Southwell-Schwarz iteration.

  Note that, at no additional cost, the convergence of this combined method can be further boosted by replacing the single update step (15) by a Jacobi update using all k computed residuals at once:

$$u^{(m+1)} = u^{(m)} + \sum_{l=1}^{k} \frac{\omega}{\gamma_{i_l}} R_{i_l} r_{i_l}^{(m)}. \tag{16}$$

  Consequently, Step 1" reduces to the computation of k randomly chosen residuals. Following the proof of Theorem 1 using $\gamma_i = \lambda_{\max}$ and the uniform probability distribution, an improved expected error decay estimate

$$E(\|u - u^{(m)}\|_a^2) \le \left(1 - \frac{\omega(2 - \omega)k}{N\kappa}\right)^m \|u - u^{(0)}\|_a^2, \qquad m \ge 1, \tag{17}$$

  can be obtained for the new modification, showing that the increased cost is compensated for. Obviously, for $k = N$ we recover the deterministic additive Schwarz iteration, and the estimate (17) is up to a constant the same as (13). Since in this paper we concentrate on variants of the multiplicative Schwarz method, we will not go further into this extension.

- The theory of iterative Schwarz methods was originally developed as a framework for investigating numerical algorithms such as domain decomposition and multilevel solvers for large-scale discretizations for elliptic boundary value problems and integral equations. The following prototypical example and the multilevel algorithm proposals in [21, 22] were the initial motivation for our study. Consider the Dirichlet problem

$$-\Delta u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{18}$$

  with the associated bilinear form and linear functional defined on $H_0^1(\Omega)$ by

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx, \qquad F(v) = \int_\Omega f v \, dx$$

  respectively. Let V_j denote the subspace in $H_0^1(\Omega)$ consisting of C^0 Lagrange finite element functions of degree $r \ge 1$ with zero boundary conditions over a quasi-uniform (triangular/simplicial or

7

rectangular) partition $\mathcal{T}_j$ of Ω with element-size $h_j \approx 2^{-j}$, $j \geq 0$. For simplicity, let $\Omega \subset \mathbb{R}^d$ be a bounded polyhedral domain, which is exactly resolved by these partitions. We assume nestedness $V_j \subset V_{j+1}$ which is satisfied if the partitions are obtained by regular dyadic refinement. Each of these spaces V_j is spanned by a set of nodal basis functions $\Phi_j := \{\phi_{j,i}\}$ in the usual way. For a suitably given index set Λ, we want to use collections

$$\Phi_\Lambda := \{\phi_{j,i} : (j,i) \in \Lambda\}$$

of such nodal basis functions to generate discretization spaces

$$V_{\Lambda,a} = \operatorname{span} \Phi_\Lambda.$$

To be precise, $V_{\Lambda,a}$ is the closure of the algebraic span of Φ_Λ in $H_0^1(\Omega)$, which is equipped with the scalar product given by $a(\cdot,\cdot)$. The associated space splittings will be generated directly from the generating set Φ_Λ, by canonically considering all one-dimensional subspaces of $V_{\Lambda,a}$ generated by the nodal basis functions $\phi_{j,i} \in \Phi_\Lambda$. The important fact is that there is a large collection of Λ (e.g., those Λ relevant for adaptive finite element methods) such that these subspace splittings have moderate condition number κ_Λ independently of the size N_Λ of Λ (i.e., the number of subspaces in the splitting), see [17, Section 4] for some overview and references. A notable fact is that, if we take the whole multilevel system $\Phi := \cup_{j \geq 0} \Phi_j$ as Φ_Λ, then our $V_{\Lambda,a}$ is $H_0^1(\Omega)$, and Φ represents a frame in $H_0^1(\Omega)$ with moderate condition number. This assumes that we normalize the nodal basis functions accordingly. A common choice is $a(\phi_{j,i}, \phi_{j,i}) = 1$, which implies $\gamma_{j,i} = 1$ for the constants in (3).

For such $V_{\Lambda,a} \subset H_0^1(\Omega)$ and their associated splittings, the standard AS and MS iterations converge provably well independently of N_Λ. In particular, the deterioration due to the logarithmic factor in (14) can be removed. In other words, comparing with the upper bounds derived in Theorem 1, we would not necessarily expect substantial improvements when using Southwell-Schwarz and random-order Schwarz iterations instead. However, in practice one observes a reduction of iteration steps by about a factor of three when comparing the standard MS method and the Southwell-Schwarz iteration for the same target accuracy. Moreover, for the combination (15) of random and greedy ordering one already gains for moderate values of $k \geq 3$ improved convergence in comparison to the standard MS method. Also, our convergence result (17) sheds new light on asynchronous domain decomposition and multigrid methods for elliptic problems [10, 11, 12, 26]. For further comments in this direction, we refer to the next section.

- The results of [25] concerning the expected convergence rate of a randomized Kaczmarz method for solving consistent over-determined linear systems $Ax = b$ with full-rank $M \times N$ matrix A ($M \geq N$) appear to be a very special case of part b) of our main theorem (the inconsistent or noisy case is dealt with in [15]). The reason is that the Kaczmarz iteration step is equivalent to a single Gauss-Seidel step for the symmetric semi-definite problem $AA^*\hat{x} = b$ (and thus to a single subspace update of a multiplicative Schwarz method based on the splitting of $\mathbb{R}^M$ into its coordinate spaces). This can be seen as follows. Set $V = \mathbb{R}^N$ and take $a(x,y) = x^T y$. In other words, the variational problem (A) corresponds to solving a trivial linear system. For the space splitting, we set $V_{a_i} = \mathbb{R}$, $a_i(\hat{x}_i, \hat{y}_i) = \hat{x}_i \hat{y}_i$, and the operators R_i are given by the columns $\hat{a}_i^T \in \mathbb{R}^N$ of A^T (or, equivalently, the rows $\hat{a}_i$ of A):

$$R_i \hat{x}_i = \hat{x}_i \hat{a}_i^T, \qquad i = 1, \ldots, M.$$

We can choose $\gamma_i = \|R_i\|_{V_{a_i} \to V_a}^2 = \|\hat{a}_i\|_2^2$. Consequently,

$$\sum_{i=1}^M \gamma_i = \operatorname{tr}(AA^*) = \|A\|_F^2 = \sum_{j=1}^N \sigma_j^2,$$

where $\sigma_1 \geq \ldots \geq \sigma_N > 0$ denote the non-zero singular values of A. Finally, since $x = \sum_{i=1}^{M} R_i \hat{x}_i = A^T \hat{x}$ for $\hat{x} := A(A^T A)^{-1} x$, we have

$$\|\|x\|\|^2 = \inf_{x = A^T \hat{x}} \|\hat{x}\|_2^2 \leq \|A(A^T A)^{-1} x\|_2^2 = x^T (A^T A)^{-1} x \leq \sigma_N^{-2} a(x, x).$$

This implies the estimate $\lambda_{\min} \geq \sigma_N^2$ for the lower stability constant of the introduced space splitting. Putting things together, we see that the expected convergence rate of the random-order Schwarz iteration with this splitting is given by

$$E(\|x - x^{(m)}\|_2^2) \leq \left(1 - \frac{\sigma_N^2}{\sum_{j=1}^{N} \sigma_j^2}\right)^m \|x - x^{(0)}\|_2^2.$$

This translates one-to-one into the result of [25]. Note that the consistency assumption $Ax = b$ is used when computing

$$T_i e^{(m)} = \hat{a}_i (x - x^{(m)}) = (Ax - Ax^{(m)})_i = b_i - \hat{a}_i x^{(m)}, \qquad i = 1, \ldots, M,$$

and that the probability distribution used for picking i^* in the random-order Schwarz iteration is identical with the one advocated in [25].

Part a) of our Theorem 1 gives a similar estimate for the deterministic Kaczmarz-Southwell iteration which has been mentioned but not investigated in [25], we refer to [5] for an interesting proposal of an approximate Kaczmarz-Southwell iteration. We want to acknowledge that the original version of our note concentrated on the investigation of the deterministic Southwell-Schwarz iteration, and that only after reading the paper [25] we also looked at the random-order Schwarz method in a systematic way.

- As an alternative to [25, 15], one might think of solving the least-squares problem for the over-determined system $Ax = b$ with full-rank A (and not just the consistent case thereof) by the GS-Southwell method or the random-order GS method for the $N \times N$ system of normal equations $A^T A x = A^T b$. The theoretical convergence rates following from our theorem are exactly the same as for the corresponding iterations on $A A^T \hat{x} = b$, i.e., for the above discussed Kaczmarz methods. However, the energy norm error is now $\|Ae^{(m)}\|_2$ instead of $\|e^{(m)}\|_2$. Furthermore, compared to the work to be done for a single subspace update in the Kaczmarz iteration, i.e.

$$x^{(m+1)} = x^{(m)} + \frac{\omega}{\|\hat{a}_i\|_2^2} r_i^{(m)} \hat{a}_i^T, \qquad i \in \{1, \ldots, m\},$$

the computational costs for a single Gauss-Seidel update, i.e.

$$x^{(m+1)} = x^{(m)} + \frac{\omega}{\|\hat{a}^j\|_2^2} ((\hat{a}^j)^T r^{(m)}) e_j, \qquad j \in \{1, \ldots, n\},$$

for solving $A^T A x = A^T b$ are higher since it involves computing the whole residual vector $r = b - Ax$ in each iteration step. Here, $\hat{a}^j$ denotes the j-th column of A.

- So far, we only considered the case of finite space splittings. Investigations on splittings into infinitely many auxiliary spaces are still at their beginning, they can benefit from the theory of greedy algorithms in infinite-dimensional Hilbert and Banach spaces developed by Temlyakov [28, 27] and others. We hope that a better understanding of this topic will also shed new light on adaptive multilevel methods such as the early work by Rüde [21, 22].

4. Numerical Examples

We begin with some illustrating numerical experiments for a linear system from [18]. It was originally constructed to show the necessity of the logarithmic factor in (14), and provides an example where the
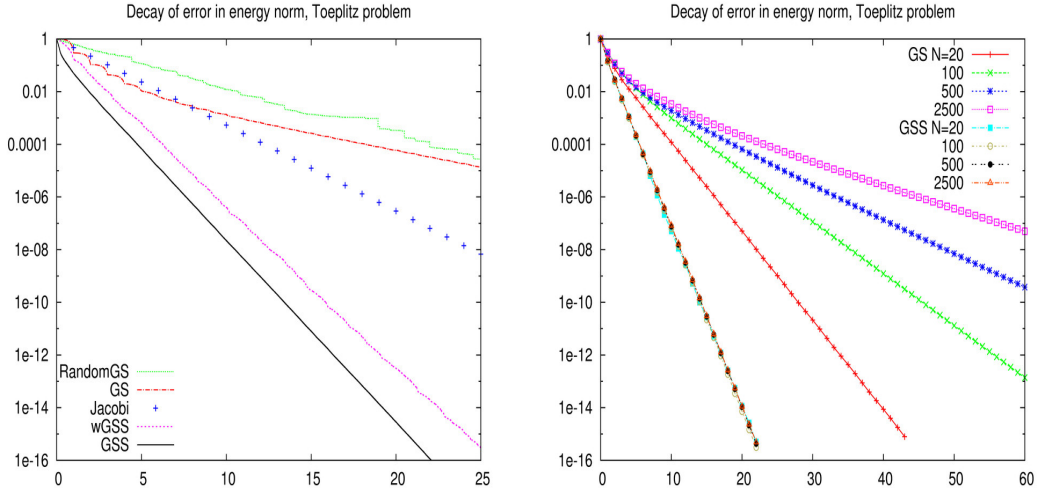
Figure 1: Left: Comparison of the convergence of various Schwarz iterations, Toeplitz problem, $N = 500$. Right: Comparison of the convergence of the GS-Southwell and GS methods for varying N.

Jacobi method performs better than SOR and SSOR methods, even after optimally adjusting the relaxation parameter ω. Let A be a symmetric $N \times N$ Toeplitz matrix with entries $a_{ij} = b_{i-j}$ given by the formulas $b_0 = 1$, $b_{2k+1} = c(-1)^k(2k+1)^{-1}$, $b_{2k+2} = 0$, $k = 0, 1, \ldots$, which for $|c| < 2/\pi$ turns out to be positive definite with $\kappa(A)$ bounded independently of N. In the following tests, we set $c = 0.3$.

Figure 1 (left) shows the error decay in the energy norm of the GS-Southwell method with $\omega = \beta = 1$ (legend: GSS) for solving the equation $Au = 0$ with $N = 500$. The iteration process was started with a randomly chosen unit energy norm vector $x^{(0)}$, and carried out for $25N$ single update steps. For comparison, we show similar error plots for the Jacobi (AS) method (indicated by crosses, note that one Jacobi step corresponds to N single update steps of the Schwarz methods), the standard Gauss-Seidel (MS) method (legend: GS), the GS-Southwell iteration with weakness parameter $\beta = 0.1$ and $\omega = 1$ (legend: wGSS), and the random-order GS method. While the latter does not seem to be competitive with respect to energy norm error decay per update step, the proposed iterative methods based on greedy orderings are superior to standard Gauss-Seidel and Jacobi methods in this example.

In Figure 1 (right) we give the energy error decay for the GS-Southwell method with $\omega = \beta = 1$ and for the standard Gauss-Seidel method for varying system size N. While the convergence of the latter indeed deteriorates with rising N due to the logarithmic factor in (14) as the theory predicts, it is independent of N for the GS-Southwell iteration, compare (12). Moreover it stays perfectly the same, i.e. the error plots overlay exactly.

Next, we illustrate the performance improvements if we use the algorithm combining random and greedy strategies for index selection, as described in the previous section. Figure 2 shows error plots for different values of k for the same linear system as before. Recall that k is the number of randomly chosen indices from which the next update index i^* is picked by the greedy strategy. Because $\gamma_i = 1$ for our matrix A, indices i are chosen with equal probability $p_i = 1/N$. For each k, we have repeated the iteration ten times (the error plots are represented by gray lines), and computed from them an approximation to the expected error (shown by a thicker solid line). The dashed line corresponds to the upper bound for the (squared) energy error decay predicted by Theorem 1. The experimental results show that, already for small values of k, the combined method becomes competitive with the GS-Southwell iteration with respect to error decay, but is only slightly more expensive than any fixed-order Gauss-Seidel (MS) method.

As a second example, we consider the multilevel discretization of the Poisson problem (18), see the corresponding remark in the previous section for the used notation. For simplicity, let $d = 2$, $\Omega = [0, 1]^2$, define the sequence $\{\mathcal{T}_j\}$ of uniform partitions into squares of side-length $h_j = 2^{-j}$, $j \geq 1$, created from
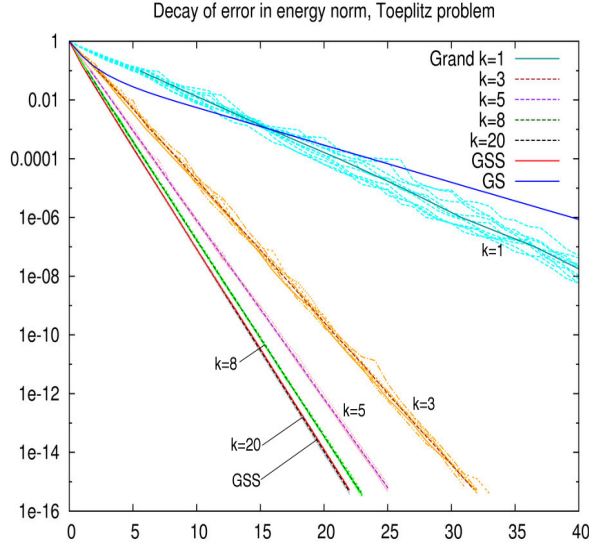
Figure 2: Schwarz iterations combining random and greedy ordering, Toeplitz problem.

an initial partition $\mathcal{T}_0 = \Omega$ by uniform dyadic refinement, and use bilinear finite elements with zero boundary conditions. For a final level $J \geq 1$, the standard multilevel generating system $\Phi_J := \Phi_{\Lambda_J}$ for the resulting subspace $V_J \subset H_0^1(\Omega)$ of bilinear finite element functions on the partition $\mathcal{T}_J$ is given by the index set

$$\Lambda_J = \{(j, i_1, i_2): \ j = 1, \ldots J, i_1 = 1, \ldots, 2^j - 1, i_2 = 1, \ldots, 2^j - 1\},$$

where ϕ_{j,i_1,i_2} denotes the standard bilinear hat function associated with the nodal point $(i_1 2^{-j}, i_2 2^{-j})$ in $\mathcal{T}_j$.

The use of the multilevel generating system Φ_J (instead of just the nodal basis on the finest level) for V_J results in a non-unique representations of $u_J \in V_J$

$$u_J := \sum_{j=1}^J v_j \qquad \text{with} \quad v_j(x) = \sum_{i_1, i_2} x_{j,i_1,i_2} \phi_{j,i_1,i_2}(x_1, x_2), \quad x_{j,i_1,i_2} \in \mathbb{R}, \tag{19}$$

via coefficient vectors $x_J = \{x_{j,i_1,i_2}\}$, i.e. there exist infinitely many coefficient sets x_J which give the same function u_J. Consequently, the matrix of the resulting linear system

$$A_J x_J = f_J \tag{20}$$

is no longer positive definite, but merely positive semi-definite. Its dimension is by a factor of about 4/3 larger than its rank, i.e., the dimension of V_J. Nevertheless, the system is solvable, since the right hand side is build in a consistent manner. It possesses infinitely many solutions, one of which can be gained as fix point in an iterative method. Moreover, it can be shown that a classical Gauss-Seidel iteration with a level-wise traversal ordering from coarse to fine over the set of unknowns resembles a multigrid method, more precisely, a (0,1)-V-cycle with one post-smoothing step by a Gauss-Seidel smoother. Furthermore, one classical Jacobi iteration for (20) resembles just the BPX-preconditioner, for further details see [7, 8, 9, 17]. Altogether, by directly using Φ_J in the discretization process, the application of classical iterative methods to (20) results in modern multilevel methods.

We now apply the Southwell-Schwarz iteration with the splitting generated by Φ_J, i.e. the GS-Southwell method for (20), and compare its convergence in the energy norm with that of the multiplicative Schwarz method, i.e. the Gauss-Seidel iteration with coarse to fine level-wise traversal ordering. We start with an initial random vector of unit norm. Since an iterative method on (20) is only semi-convergent [2], i.e. classically convergent to just one of the many possible solutions, we measure the
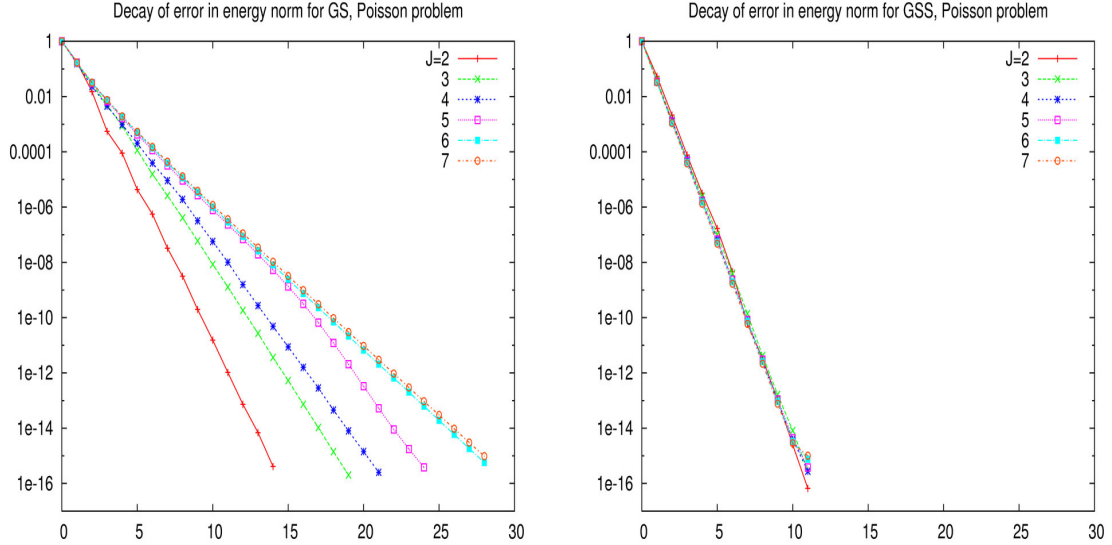
11

Figure 3: Comparison of the convergence of the GS and GS-Southwell method with varying J for the multilevel discretization of the two-dimensional Poisson problem.

error reduction rate by projecting the non-unique representation of an iterate $x_J^{(m)}$ via the projection (19) onto its unique representation with respect to the nodal basis on the finest level J, and computing the iteration error and its energy norm there. The convergence results for different discretization levels J are shown in Figure 3. On the x-axis, we indicate here the number of full cycles of the GS iteration and accordingly the number of $\dim(A_J)$-sized blocks of single steps of the GS-Southwell method for a fair comparison. The error decay for the GS-Southwell iteration is substantially faster than that of the GS iteration with fixed, level-wise traversal order. To be precise, for $J = 6$ we obtain a reduction of the error in the energy norm by a factor of 10^{-16} after about 30 full cycles of the GS method whereas the GS-Southwell method needs only about $10 \cdot \dim(A_J)$ local iterations, i.e. 10 full cycles. Its convergence rate is thus roughly three times better. Moreover, the GS-Southwell iteration perfectly achieves level-independent convergence (for the range of tested values J, the error plots overlay exactly) whereas the convergence rate for the GS method gets practically slightly worse when increasing J (but will asymptotically be bounded independently of J). Note here that we observed the same qualitative behavior also for the anisotropic diffusion problem $-u_{xx} - \varepsilon u_{yy} = f$, with varying values of ε.

For Schwarz iterations (15) combining random and greedy ordering we obtain the results shown in Figure 4. Here, energy error plots are given for different values of k for the matrix A_J with $J = 6$ of the Poisson problem. Again, for each k, we have repeated the iteration ten times (the error plots are represented by gray lines), and computed from them an approximation to the square root of the expected error $\sqrt{E(\|u - u^{(m)}\|_a^2)}$ (shown by a thicker solid line). For comparison, we also give the energy error decay for the GS and GS-Southwell iteration, respectively. Recall that because $a_{ii} = \gamma_i = 1$ for our matrix A_J, the indices i are picked with equal probability $p_i = 1/\dim(A_J)$. From our results, we see a similar behavior as for the Toeplitz problem. Now, in contrast to that, the case $k = 1$ shows a convergence which is slightly worse than that for the GS-iteration. But already for small values of $k \geq 3$, the randomized method again becomes competitive with the GS-Southwell iteration with respect to error decay, but is only slightly more expensive than any fixed-order Gauss-Seidel method. Thus, we have demonstrated that conventional multigrid, i.e. the level-wise Gauss-Seidel approach using the multi-level generating system Φ_J, can indeed be further improved by a combination of random and greedy ordering. Moreover, the asynchronous quality of the random part of our new method may be further exploited in the design of efficient parallel multilevel solvers for elliptic problems on the modern architectures of next-generation supercomputers.
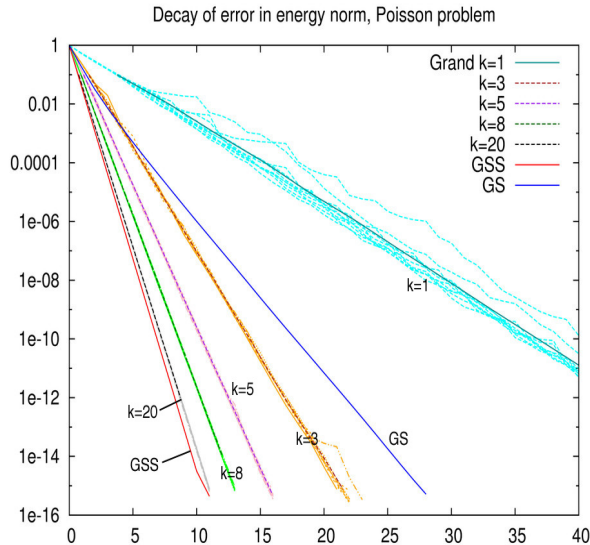
12

Figure 4: Schwarz iterations combining random and greedy ordering for the multilevel discretization of the Poisson problem.

We conclude with a few comments on the computational costs. We again use the standard problem of solving $Ax = b$ for this discussion. Of course, the GS-Southwell iterations are more costly due to the evaluation of all residuals and the determination of its maximum in each iteration step. For the case of a full matrix, we already mentioned its $O(N^2)$ cost complexity per update step. For the case of a sparse matrix with a constant number of non-zero entries per row the situation is better: For a naive implementation, the cost to determine all residual values and to pick the largest one is $O(N)$ while the cost for a single update is $O(1)$. However, in a more clever implementation, only the $O(1)$ updates of residual values caused by the previous iteration step need to be handled (and not the whole residual vector). Then, the determination of the index with maximal residual value boils down to simple sorting and inserting into tree-like data structures which results in a further $\log N$ term. Altogether, N steps of the GS-Southwell iteration cost $O(N \log N)$ operations which is not so bad compared to the $O(N)$ operations needed for a full cycle of the GS method.

For the case of the multilevel discretization of an elliptic PDE in V_J via the generating system Φ_J things are principally similar but more involved in the details. First, a naive setting up of the system matrix would involve $O(2^{dJ} J^\mu)$ non-zero entries, where μ depends on the dimension d, and the structure of the problem. This is due to the overlay of the supports of all functions in Φ_J which is analogous to the well-known finger structure of the matrix resulting from a wavelet discretization. However, using the multilevel structure of Φ_J it is possible to implement the GS iteration step with level-wise traversal ordering in only $O(2^{dJ})$ operations, see the zipper algorithm in [7] or compare the cost complexity of a multigrid V-cycle. Here the explicit assembly of the overall system matrix and its subsequent use in the iteration is avoided. In contrast to that, the GS-Southwell approach involves again a further logarithmic term in the cost complexity for an implementation which exploits the above-mentioned tricks for sparse matrices, but now taylored to the more involved system matrices which stem from Φ_J. Altogether, $O(2^{dJ})$ steps of the GS-Southwell iteration, which correspond to one full cycle of the GS method, can be realized with a cost of $O(2^{dJ} J^\nu)$ operations, where ν depends on the dimension, the coefficient functions of the problem, and the respective data structures in the specific implementation. For further details on an efficient implementation of the GS-Southwell method and a discussion of the involved special data structures, see [29].

13

References

[1] A. Barron, A. Cohen, W. Dahmen, R. DeVore, Approximation and learning by greedy algorithms, Annals of Statistics, 3, 1, 2008, 64–94.

[2] A. Berman, R. J. Plemmons, Nonnegative matrices in the mathematical sciences. Academic Press, New York, 1979

[3] T. Blumensath, M. Davies, Gradient pursuits, IEEE Trans. Signal Processing, 56, 2008, 2370–2382.

[4] T. Blumensath, M. Davies, Stagewise weak gradient pursuits, IEEE Trans. Signal Processing, 57, 2009, 4333–4346.

[5] Y. C. Eldar, D. Needell, Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma, Numer. Algorithms, to appear.

[6] R. Gribonval, P. Vandergheynst, On the exponential convergence of matching pursuit in quasi-coherent dictionaries, IEEE Trans. Information Th. 52, 1, 2006, 255–261.

[7] M. Griebel, Multilevelverfahren als Iterationsverfahren über Erzeugendensystemen, Teubner-Skripte zur Numerik, Teubner, Stuttgart, 1994.

[8] M. Griebel, Multilevel algorithms considered as iterative methods on semidefinite systems. SIAM Int. J. Sci. Stat. Comput., 15(3):547-565, 1994.

[9] M. Griebel, P. Oswald, Remarks on the abstract theory of additive and multiplicative Schwarz methods, Numer. Math. 70, 1995, 163–180.

[10] P. K. Jimack, M. A. Walkley. Asynchronous parallel solvers for linear systems arising in computational engineering. In: B. Topping and P. Ivanyi (editors). Computational Technology Reviews, (Saxe-Coburg), Volume 3, 2011, 1–20.

[11] B. Lee, S. McCormick, B. Philip, D. Quinlan, Asynchronous fast adaptive composite-grid methods: numerical results, SIAM J. Sci. Comp. 25 (2003), pp. 682-700.

[12] B. Lee, S. McCormick, B. Philip, D. Quinlan, Asynchronous fast adaptive composite-grid methods for elliptic problems: theoretical foundations, SIAM J. Num. Anal. 42, 2004, 130–152.

[13] D. Leventhal, A. S. Lewis, Randomized methods for linear constraints: Convergence rates and conditioning, Math. Oper. Res. 35, 3, 2010, 641–654.

[14] Z. Q. Luo, P. Tseng, On the convergence of the coordinate descent method for convex differentiable minimization, J. Optim. Th. Appl. 72, 1, 1992, 7–35.

[15] D. Needell, Randomized Kaczmarz solver for noisy linear systems, BIT Numer. Math. 50, 2010, 395–403.

[16] Y. Nesterov, Efficiency of coordinate descent methods on huge-scale optimization problems, CORE Discussion Paper 2/2010, UC Louvain, 2010.

[17] P. Oswald, Multilevel Finite Element Approximation - Theory & Applications. Teubner-Skripte zur Numerik, Teubner, Stuttgart, 1994.

[18] P. Oswald, On the convergence rate of SOR: A worst case estimate, Computing, 52, 1994, 245–255.

[19] P. Oswald, Stable space splittings and fusion frames, in Wavelets XIII (V.K. Goyal, M. Papadakis, D. Van de Ville, eds.), Proc. SPIE Vol. 7446 (SPIE, Bellingham, 2009), 744611.

[20] P. Richtarik, M. Takac, Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function, Manuscript, School of Mathematics, Univ. Edinburgh, April/July 2011.

[21] U. Rüde, Mathematical and Computational Techniques for Multilevel Adaptive Methods, Frontiers in Applied Mathematics vol. 13, SIAM Philadephia, 1993.

[22] U. Rüde, Fully adaptive multigrid methods, SIAM J. Numer. Anal. 30, 1993, 230–248.

[23] J. Sliačan, Residual-based Gauss-Seidel methods, BSc thesis, Jacobs University, 2011.

[24] R. V. Southwell, Relaxation Methods in Engineering Science - A Treatise in Approximate Computation, Oxford Univ. Press, Oxford, 1940.

[25] T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, J. Fourier Anal. Appl. 15, 2009, 262–278.

[26] X. Tai, P. Tseng, Convergence rate analysis of an asynchronous space decomposition method for convex minimization, Math. Comp. 71, 2002, 1105-1135.

[27] V. N. Temlyakov, Nonlinear methods of approximation, Found. Comput. Math. 3, 2003, 33–107.

[28] V. N. Temlyakov, Greedy approximation, Acta Numerica, vol. 17, 2008, 235–409.

[29] R. Thesen. Effiziente adaptive Lösung von Multilevelsystemen. Diplomarbeit, Institut für Numerische Simulation, Universität Bonn, 2010.

[30] P. Tseng, Dual ascent methods for problems with strictly convex cost and linear constraints: A unified approach, SIAM J. Control and Optimization 28, 1,1990, 214–242.

[31] J. Xu, Iterative methods by space decomposition and subspace correction, SIAM Review 34, 1992, 581–613.

[32] J. Xu, L. Zikatanov, The method of alternating projections and the method of subspace corrections in Hilbert space, J. Amer. Math. Soc. 15, 2002, 573–597.