Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

M. Griebel, A. Hullmann, and P. Oswald

Optimal scaling parameters for sparse grid discretizations

INS Preprint No. 1314

August 2013

OPTIMAL SCALING PARAMETERS FOR SPARSE GRID DISCRETIZATIONS

M. GRIEBEL*, A. HULLMANN*, AND P. OSWALD‡

Abstract. We apply iterative subspace correction methods to elliptic PDE problems discretized by generalized sparse grid systems. The involved subspace solvers are based on the combination of *all* anisotropic full grid spaces that are contained in the sparse grid space. Their relative scaling is at our disposal and has significant influence on the performance of the iterative solver. In this paper, we follow three approaches to obtain close-to-optimal or even optimal scaling parameters of the subspace solvers and thus of the overall subspace correction method. We employ a Linear Program that we derive from the theory of additive subspace splittings, an algebraic transformation that produces partially negative scaling parameters which result in improved asymptotic convergence properties, and finally we use the OptiCom method as a variable non-linear preconditioner.

Key words. generalized sparse grids, additive Schwarz preconditioner, subspace splittings, diagonal scaling, OptiCom, variable preconditioning

1. Introduction. In many large-scale numerical problems, e.g., the function approximation in higher dimensions or the numerical solution of partial differential equations, discretization techniques exploit specific a priori assumptions on the solution to increase efficiency. Especially for high dimensional problems, a conventional discretization approach, which is based on isotropic uniform grids, is doomed due to the curse of dimensionality [Bel61]. Then, sparse grid discretizations [BG04] are an important way to circumvent this problem, at least to some extent, provided that additional regularity conditions hold. This way, improved error decay rates with respect to the required degrees of freedom are obtained which depend only logarithmically or, in the best case, not at all on the dimension. Next, to guarantee overall efficiency, the corresponding systems of discrete equations must be solved in a fast way. Normally, some iterative method is employed which results in the need of preconditioning.

In this paper, our focus is on additive (sometimes coined "parallel" or "asynchronous") Schwarz preconditioners for generalized sparse grid discretizations of symmetric H-elliptic variational problems

$$a(u,v) = F(v) \quad \forall\, v \in H\ , \tag{1.1}$$

where H is a Hilbert-space and F is a bounded linear functional on H. Equivalently, the linear variational problem (1.1) can be cast as a quadratic minimization problem

$$\phi(u) \to \min_{u\in H} \quad \text{with} \quad \phi(u) := \frac{1}{2}a(u,u) - F(u)\ .$$

Regular sparse grid spaces [Gri91, Zen91, Bun92a] have been around for a long time for the efficient solution of PDEs. They can be described by a non-direct sum of anisotropic full grid spaces. Other, more general sparse grids with different sets of full grid spaces have been studied in [BG99, GK09, GH13] for deriving theoretically optimal error bounds in various smoothness norms and classes. Moreover, in [GG03, Feu10, BG12] dimension-adaptive discretizations have been employed.

*Institute for Numerical Simulation, University of Bonn, Wegelerstr. 6, D-53115 Bonn, Germany ({griebel, hullmann}@ins.uni-bonn.de)

‡Jacobs University Bremen, 600 Campus Ring 1, D-28759 Bremen, Germany (p.oswald@jacobs-university.de)

These approaches can be summarized by the term 'generalized sparse grid spaces'. Finally, there are locally adapted sparse grid spaces [BG04, Feu10, Bun92b, Gri98]. They all have in common that they offer great flexibility in enriching the respective discretization.

We apply a subspace correction method to (1.1), where the subspace solvers are based on all the possible anisotropic full grid spaces that are contained in the sparse grid space under consideration. Their relative scaling amounts to a diagonal scaling of the operator matrix of the discretized system and is in principle at our disposal.

Our aim is now to find optimal or close-to-optimal scaling parameters. To this end, we follow three different approaches: One is based on a Linear Program (LP) that minimizes the relative splitting condition number with respect to a subspace splitting based on orthogonal complement spaces. Here, we prove that, for H^t-elliptic problems, the best possible set of *positive* scaling parameters results in condition numbers that grow by $\Theta(J^{d-2})$, where J is the discretization level of the sparse grid and d is the space dimension. The second method is based on the observation that *negative* scaling parameters can still result in a positive definite operator on the sparse grid space, even though this case is not covered by the classical theory of subspace splittings. With an algebraic transformation that produces partially negative scaling parameters, we obtain an optimal iterative scheme with error contraction rates that are bounded independently of the level J and even *independently* of the dimension d in case of the Laplacian. The third method is a variable non-linear preconditioner which, in the general setting of subspace correction and domain decomposition methods, dates back to [JN99]. In the context of data mining using sparse grids, it was independently introduced in [Heg03, Gar06, HGC07]. This approach successively computes the best possible scaling in every iteration step. We show that the cost of such an OptiCom iteration step is log-linear with respect to the degrees of freedom if a fast matrix-vector multiplication with the operator matrix is available. This is a significant improvement over previous applications of OptiCom, which typically involved quadratic costs. We also present conjugate gradient (CG) versions of all considered iteration schemes, which shows that there is further cost reduction potential.

The remainder of the paper is organized as follows. In Section 2, we give a short overview of space splitting theory and state the problem of finding sets of scaling parameters in the associated subspace correction methods. In Section 3, we introduce sparse grid discretizations and describe methods to find optimal scaling parameters by an LP, an algebraic transformation and the OptiCom. In Section 4, we discuss the efficient implementation of the described methods. In Section 5, we deal with a standard H_0^1-elliptic Laplacian test problem in dimensions $d \leq 10$ and present convergence plots that support the theoretical findings from Section 3. Results of the CG versions are also presented. We give some concluding remarks in Section 6.

2. Scaling parameters in subspace correction methods. In this section, we first recall some facts about subspace correction methods [Xu92, Osw94] for solving (1.1) in a Hilbert space H. Then, we slightly depart from this standard theory and introduce scaling parameters for the subspace solvers. Finally, we discuss the optimal choice of such scaling parameters in the general case.

2.1. Subspace splitting theory. Let $H_i, i \in I$, be auxiliary Hilbert spaces with an at most countable index set I (if we discuss algorithmical issues, we silently assume that I is finite). Each H_i carries a symmetric H_i-elliptic bilinear form $b_i(u_i, v_i)$, writing $\{H_i; b_i\}$ indicates that we use this bilinear form as scalar product on H_i. In general, the H_i, $i \in I$, are not assumed to be subspaces of H, and in order to relate

them with H, we define bounded linear embedding operators $R_i : H_i \to H$. We call the formal decomposition

$$\{H; a\} = \sum_{i \in I} R_i \{H_i, b_i\} \tag{2.1}$$

a stable space splitting if for any $u \in H$ there is at least one H-converging representation of the form

$$u = \sum_{i \in I} R_i v_i, \qquad v_i \in H_i, \quad i \in I \,, \tag{2.2}$$

and

$$0 < \lambda_{\min} := \inf_{u \in H} \frac{a(u,u)}{|||u|||^2} \;\le\; \lambda_{\max} := \sup_{u \in H} \frac{a(u,u)}{|||u|||^2} < \infty \,, \tag{2.3}$$

where

$$|||u|||^2 := \inf_{v_i \in H_i:\ u = \sum_{i \in I} R_i v_i} \sum_{i \in I} b_i(v_i, v_i) \,. \tag{2.4}$$

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 (2.1), respectively. It is easy to see that frames and fusion frames [CK04, Osw09] are special cases of this definition, where $a(\cdot,\cdot) = (\cdot,\cdot)_H$, the H_i are closed subspaces of H, the scalar products $b_i(\cdot,\cdot) = w_i \cdot (\cdot,\cdot)_V$ are modified by weights $w_i > 0$, and the R_i denote the natural embeddings $H_i \subset H$ for $i \in I$. In the frame case, the H_i are one-dimensional spaces and are spanned by individual frame elements. Specific examples of stable space splittings related to sparse grid discretizations will be given in Section 3.

To formally define the core subspace correction methods associated with a stable space splitting (2.1), we define the adjoint operators $R_i^* : H \to H_i$ by

$$b_i(R_i^* u, v_i) = a(u, R_i v_i) \qquad \forall\, v_i \in H_i \,, \tag{2.5}$$

and eventually set $T_i := R_i R_i^* : H \to H$, $i \in I$, and $P = \sum_{i \in I} T_i$.

The *additive Schwarz method* for (2.1) (also called *parallel* or *asynchronous subspace correction method*) is then given by the iteration

$$u^{(m+1)} = u^{(m)} + \omega \sum_{i \in I} T_i e^{(m)} = u^{(m)} + \omega P e^{(m)}, \quad m = 0, 1, \dots \,, \tag{2.6}$$

where the single relaxation parameter $\omega > 0$ is to be chosen appropriately and $e^{(m)} := u - u^{(m)}$ denotes the current error. The essential work to be done is to compute all $u_i^{(m)} := R_i^* e^{(m)} \in H_i$ by solving the subproblems

$$\begin{aligned} b_i(u_i^{(m)}, v_i) &= a(u - u^{(m)}, R_i v_i) \\ &= F(R_i v_i) - a(u^{(m)}, R_i v_i) \quad \forall\, v_i \in H_i, \quad i \in I \,. \end{aligned}$$

The theoretically best value of ω is given by $\omega^* = 2/(\lambda_{\min} + \lambda_{\max})$, since the operator $P = \sum_{i \in I} T_i$ satisfies the identity

$$a(Pu, u) = \sum_{i \in I} b_i(R_i^* u, R_i^* u) = |||Pu|||^2 \,.$$

Together with (2.3), this implies that $\lambda_{\min}(P) = \lambda_{\min}$, $\lambda_{\max}(P) = \lambda_{\max}$, and $\kappa(P) = \kappa$. Thus, the best possible error reduction factor in the energy norm for the linear iteration (2.6) is given by

$$\rho := \inf_{\omega>0} \|\mathrm{Id} - \omega P\|_a = \|\mathrm{Id} - \omega^* P\|_a = 1 - \frac{2}{1+\kappa} \, . \tag{2.7}$$

This simple result has appeared in many papers, see [Xu92, Osw94, GO95a].

Since in practice the value ω^* is hardly accessible, one often determines in each iteration the value

$$\omega^{(m)} = \frac{a(Pe^{(m)}, e^{(m)})}{a(Pe^{(m)}, Pe^{(m)})} \tag{2.8}$$

which corresponds to finding $u^{(m+1)}$ by solving the minimization problem

$$\phi(u^{(m)} + \omega^{(m)} Pe^{(m)}) \to \min_{\omega^{(m)}} \, ,$$

or, equivalently, by minimizing the energy error

$$\|u - u^{(m)} - \omega^{(m)} Pe^{(m)}\|_a^2 \to \min_{\omega^{(m)}} \tag{2.9}$$

with respect to the parameter $\omega^{(m)} > 0$. The iterative method with the parameter choice $\omega^{(m)}$ from (2.8) can also be interpreted as steepest descent method for the quadratic minimization problem associated with the linear variational problem

$$a(Pu, v) = f(Pv) \qquad \forall\, v \in H \, , \tag{2.10}$$

which is a preconditioned version of (1.1). Consequently, since

$$\|e^{(m+1)}\|_a = \|u - (u^{(m)} + \omega^{(m)} Pe^{(m)})\|_a \le \inf_{\omega>0} \|\mathrm{Id} - \omega P\|_a \|e^{(m)}\|_a = \rho \|e^{(m)}\|_a \, ,$$

this method is as good as any linear method (2.6).

An alternative to the above additive Schwarz method is the *multiplicative Schwarz method* (or *synchronous subspace correction method*), where in the n-th step only one index $i = i^{(n)} \in I$ is picked, the corresponding subproblem is solved and used to immediately update the iterate according to

$$u^{(n+1)} = u^{(n)} + \omega T_{i^{(n)}} e^{(n)}, \qquad n = 0, 1, \dots \, . \tag{2.11}$$

Here, various rules for choosing the next subproblem index $i^{(n)}$ (cyclic deterministic rules, random choices, greedy pick, and their combinations), and block updates (intermediate between the additive and multiplicative versions) have been proposed, see for example [GO12].

The convergence theory of multiplicative Schwarz methods is a bit more intricate. Generally speaking they often are slightly faster than additive Schwarz methods (to achieve a fair comparison, usually $\#I$ steps of the multiplicative method are combined into one step). A potential drawback is that the multiplicative method is less straightforward for parallelization. Since the focus of this paper is the choice of scaling parameters in additive Schwarz methods, we do not want to go into detail but refer to the literature, see [Xu92, Osw94, Osw09, GO95a, GO12, Gri94a, XZ02, Gri94b].

2.2. Introducing multiple scaling parameters. From now on we slightly depart from the above theory and ask the question if the introduction of individual scaling parameters ω_i offers additional improvements. Note that a similar question has been discussed for frames in [KOPT13]. We keep the basic setup of given auxiliary problems in $\{H_i, b_i\}$, and consider the family of iterations

$$u^{(m+1)} = u^{(m)} + \sum_{i\in I} \omega_i T_i e^{(m)} = u^{(m)} + P_{\boldsymbol{\omega}} e^{(m)}, \quad m = 0, 1, \dots , \tag{2.12}$$

where $P_{\boldsymbol{\omega}} = \sum_{i\in I} \omega_i T_i$ and $\boldsymbol{\omega}$ stands from now on for a set of scaling parameters $(\omega_i)_{i\in I}$. Convergence of this linear iterative method is guaranteed if and only if $P_{\boldsymbol{\omega}}$ is strictly positive definite ($\lambda_{\min}(P_{\boldsymbol{\omega}}) > 0$) and $\lambda_{\max}(P_{\boldsymbol{\omega}}) < 2$. We note that enforcing the upper bound $\lambda_{\max}(P_{\boldsymbol{\omega}}) < 2$ is not a major concern. For this, we can, at little extra work, determine an additional relaxation parameter using the steepest descent approach, and guarantee an error reduction of at least $1 - 2/(1 + \kappa(P_{\boldsymbol{\omega}}))$. Thus, the question about guaranteeing best convergence rates for (2.12) is essentially equivalent to determining positive definite $P_{\boldsymbol{\omega}}$ with (close to) minimal condition numbers. To this end, let Ω denote the family of all parameter sets $\boldsymbol{\omega}$ such that $P_{\boldsymbol{\omega}}$ is bounded and strictly positive definite. Then, the optimal error reduction rate ρ^* can be expressed as

$$\rho^* = \inf_{\boldsymbol{\omega}\in\Omega} \|\mathrm{Id} - P_{\boldsymbol{\omega}}\|_a = 1 - \frac{2}{1+\kappa^*}, \qquad \kappa^* = \inf_{\boldsymbol{\omega}\in\Omega} \kappa(P_{\boldsymbol{\omega}}) \, . \tag{2.13}$$

Note that Ω may contain parameter sets with some negative or zero ω_i. To cover such situations, the theory of subspace correction methods based on stable space splittings (2.1) is not of immediate help, as it can only deal with the case $\omega_i > 0$. Indeed, the operator P associated with the space splitting (2.1) becomes $P_{\boldsymbol{\omega}}$ if the auxiliary bilinear forms $b_i(u_i, v_i)$ are replaced by their weighted versions $\omega_i^{-1} \cdot b_i(u_i, v_i)$, $i \in I$. Thus, whenever the space splitting is stable, and $0 < \omega_{\min} \le \omega_i \le \omega_{\max} < \infty$, the space splitting with these modified auxiliary scalar products is also stable according to (2.3), and satisfies

$$\frac{\omega_{\min}}{\omega_{\max}} \kappa(P) \le \kappa(P_{\boldsymbol{\omega}}) \le \frac{\omega_{\max}}{\omega_{\min}} \kappa(P) \, .$$

For $\omega_i > 0$ and finite I, this rough estimate guarantees $\boldsymbol{\omega} \in \Omega$, but does not help with minimizing $\kappa(P_{\boldsymbol{\omega}})$, nor with dealing with sets $\boldsymbol{\omega} \in \Omega$ which contain some negative ω_i.

2.2.1. A priori choice of scaling parameters. We want to find a set $\boldsymbol{\omega}^* \in \Omega$ that realizes or at least comes close to the optimal error reduction rate (2.13), i.e. $\rho^* = \rho(P_{\boldsymbol{\omega}^*})$ and $\kappa^* = \kappa(P_{\boldsymbol{\omega}^*})$, respectively.

Even though we do not believe that this leads to a practically useful approach in this generality, we mention that the problem of finding $\boldsymbol{\omega}^*$ can be formulated as semi-definite program (if H is finite dimensional, and I is finite). To this end, we set

$$\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & -\mathrm{Id} \end{pmatrix}, \quad \mathbf{A}_i = \begin{pmatrix} -T_i & 0 \\ 0 & T_i \end{pmatrix}, \quad \mathbf{A}' = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix}$$

and minimize λ with respect to the vector of variables $(\boldsymbol{\omega}, \lambda)$ subject to the constraint

$$\mathbf{A} + \sum_{i\in I} \omega_i \mathbf{A}_i + \lambda \mathbf{A}' \ge 0 \, , \tag{2.14}$$

where the inequality ≥ 0 means that the 2×2-matrix of operators on the left-hand side of (2.14) needs to be positive semi-definite. For more information on semi-definite programs, we refer the reader to [VB96]. The resulting set of weights $\boldsymbol{\omega}^*$ realizes the optimal error reduction rate (2.13). Indeed, if $(\tilde{\boldsymbol{\omega}}, \tilde{\lambda})$ is a minimizer of λ satisfying (2.14), we have

$$\kappa^* = \tilde{\lambda}, \qquad \omega_i^* = \frac{2\tilde{\omega}_i}{1+\tilde{\lambda}}, \quad i \in I \, .$$

In Section 3, we will tackle this problem for sparse grid discretizations in a less general but more efficient way by incorporating knowledge on the tensor product structure of sparse grid spaces and a norm equivalence (3.4).

2.2.2. Variable and non-linear choice of scaling parameters: OptiCom. There is a nonlinear iterative method which generalizes the line search (2.9) of the steepest descent method for (2.10), and provides a safe lower bound for the best possible error reduction factor ρ^* from (2.13). It was introduced in [JN99] and used in the context of subspace correction methods for L_2-data approximation with sparse grids in [Heg03]. It was later called OptiCom [Gar06, HGC07]. In the following, we essentially recall some general results and observations from [JN99]. We will apply this approach to our sparse grid setting in Section 3.

The update formula of OptiCom is the same as in (2.12), i.e.,

$$u^{(m+1)} = u^{(m)} + \sum_{i \in I} \omega_i^{(m)} T_i e^{(m)} = u^{(m)} + P_{\boldsymbol{\omega}^{(m)}} e^{(m)}, \quad m = 0, 1, \dots \, , \tag{2.15}$$

however, the parameter set $\boldsymbol{\omega}^{(m)} = (\omega_i^{(m)})_{i \in I}$ now depends on $u^{(m)}$: We obtain $\boldsymbol{\omega}^{(m)}$ by solving the quadratic minimization problem

$$\| u - u^{(m)} - \sum_{i \in I} \omega_i^{(m)} T_i e^{(m)} \|_a^2 \to \min_{\boldsymbol{\omega}^{(m)}} \tag{2.16}$$

in each iteration step m. This is typically done by solving a system

$$\tilde{\mathbf{A}}^{(m)} \boldsymbol{\omega}^{(m)} = \tilde{\mathbf{b}}^{(m)}$$

of linear equations, where the system matrix $\tilde{\mathbf{A}}^{(m)} \in \mathbb{R}^{\#I \times \#I}$ is positive semi-definite with

$$(\tilde{\mathbf{A}}^{(m)})_{ij} = a(T_i e^{(m)}, T_j e^{(m)}) \quad \text{for} \quad i, j \in I$$

and the right-hand side $\tilde{\mathbf{b}}^{(m)} \in \mathbb{R}^{\#I}$ is given by

$$(\tilde{\mathbf{b}}^{(m)})_i = a(e^{(m)}, T_i e^{(m)}) = F(T_i e^{(m)}) - a(u^{(m)}, T_i e^{(m)}) \quad \text{for} \quad i \in I \, .$$

The OptiCom iteration converges at least as fast as any stationary additive Schwarz iteration, and thus provides a lower bound for the convergence rate of the latter. The following theorem can also be found in [JN99].

THEOREM 2.1. *The error $e^{(m)} = u - u^{(m)}$ of the OptiCom iteration* (2.15) *with $\boldsymbol{\omega}^{(m)}$ from* (2.16) *for the space splitting (2.1) decays in energy norm according to*

$$\| e^{(m+1)} \|_a \leq \rho^* \| e^{(m)} \|_a, \qquad m \geq 0 \, ,$$

where ρ^* is the optimal error reduction factor (2.13) for additive Schwarz methods based on the same space splitting.

Proof. As above, denote by $\tilde{\boldsymbol{\omega}}^{(m)}$ the solution of the minimization problem (2.16). Then $e^{(m+1)} = e^{(m)} - \sum_{i\in I} \tilde{\omega}_i^{(m)} T_i e^{(m)}$, and, for any fixed parameter set $\boldsymbol{\omega}$, we have

$$\begin{aligned}\|e^{(m+1)}\|_a &= \|u - u^{(m)} - \sum_{i\in I} \tilde{\omega}_i^{(m)} T_i e^{(m)}\|_a \le \|e^{(m)} - \sum_{i\in I} \omega_i T_i e^{(m)}\|_a \\ &= \|(I - P_{\boldsymbol{\omega}})e^{(m)}\|_a \le \|I - P_{\boldsymbol{\omega}}\|_a \|e^{(m)}\|_a \ .\end{aligned}$$

It remains to take the infimum over all $\boldsymbol{\omega} \in \Omega$ to get the claimed bound for the error reduction factor of the OptiCom iteration. ∎

To conclude this short subsection, we point to the CG version of OptiCom. It could further reduce the dependence of the convergence estimates on the condition of the additive Schwarz splitting from an average reduction factor per step of $(1 - \mathcal{O}((\kappa^*)^{-1}))$ to $(1 - \mathcal{O}((\kappa^*)^{-1/2}))$. Suppose that we have, starting from $u^{(0)} = u^{(-1)} = 0$, already computed $u^{(1)}, \ldots, u^{(m)}$, and that $\boldsymbol{\omega}^{(m)} = (\omega_i^{(m)})_{i\in I}$ and $\eta^{(m)}$ are to be determined as solutions of the slightly modified minimization problem

$$\|u - u^{(m)} - \sum_{i\in I} \omega_i^{(m)} T_i e^{(m)} - \eta^{(m)}(u^{(m)} - u^{(m-1)})\|_a^2 \to \min_{\boldsymbol{\omega}^{(m)}, \eta^{(m)}} \ . \tag{2.17}$$

Then,

$$u^{(m+1)} = u^{(m)} + \sum_{i\in I} \omega_i^{(m)} T_i e^{(m)} + \eta^{(m)}(u^{(m)} - u^{(m-1)}) \tag{2.18}$$

realizes a CG-OptiCom iteration step. We note that, for fixed $\boldsymbol{\omega}$, solving the two-parameter ($\tau^{(m)}$ and $\eta^{(m)}$) minimization problem

$$\|u - u^{(m)} - \tau^{(m)} \sum_{i\in I} T_i e^{(m)} - \eta^{(m)}(u^{(m)} - u^{(m-1)})\|_a^2 \to \min_{\tau^{(m)}, \eta^{(m)}} \tag{2.19}$$

is equivalent to the usual PCG-iteration for solving (1.1) with a preconditioner derived from the additive Schwarz operator $P_{\boldsymbol{\omega}}$, thus the name CG-OptiCom. By including the parameter set $\boldsymbol{\omega}^{(m)}$ into the minimization (2.17) we incorporate the scaling of the subproblems. This makes the convergence analysis more difficult, since the preconditioner is no longer fixed but changes from iteration to iteration. This difficulty was discussed in a slightly different setting in [KL07]. Nevertheless, it is easy to see from the proof of Theorem 2.1 (just set $\eta^{(m)} = 0$) that the CG-OptiCom has at least the same error reduction factor per step as the OptiCom. In [JN99] this version and other CG variants of a variable preconditioner were presented, but no stronger convergence estimates could be proven. However, the numerical experiments in Section 5 suggest a still significant speed-up by using the CG-OptiCom (2.18) over the plain OptiCom, i.e. the update (2.15) with the parameter set from (2.16).

3. Optimal scaling parameters for sparse grid discretizations. In Section 2, we recalled the general theory of additive subspace correction methods and introduced the scaling parameters. Now, we concentrate on sparse grid discretizations of elliptic PDEs and discuss three approaches that produce optimal or close-to-optimal scaling parameters and convergence rates.

Sparse grid discretizations assume a tensor product structure. Even though our main examples are discretizations of standard elliptic boundary value problems on

product domains in $\mathbb{R}^d$, we want to develop our approach in a slightly more abstract setting. More precisely, we assume that we have a Hilbert space $V := V^{(1)} \otimes \ldots \otimes V^{(d)}$ given as the tensor product of Hilbert spaces $V^{(i)}$ each of which is the closure of the union of an increasing ladder of its nontrivial finite-dimensional subspaces

$$V_0^{(i)} \subset V_1^{(i)} \subset \ldots , \qquad i = 1, \ldots, d \, . \tag{3.1}$$

Introducing the orthogonal complement spaces $W_l^{(i)} = V_l^{(i)} \ominus_{V^{(i)}} V_{l-1}^{(i)}$ for $l \geq 1$, and setting $W_0^{(i)} := V_0^{(i)}$, we can obviously write V as the $(\cdot,\cdot)_V$-orthogonal (from now on V-orthogonal) sum $V = \bigoplus_{\mathbf{l} \in \mathbb{N}^d} W_{\mathbf{l}}$ of the subspaces

$$W_{\mathbf{l}} = W_{l_1}^{(1)} \otimes \ldots \otimes W_{l_d}^{(d)}, \qquad \mathbf{l} = (l_1, \ldots, l_d) \in \mathbb{N}^d \, . \tag{3.2}$$

Similarly, any finite-dimensional space $V_{\mathbf{l}} = V_{l_1}^{(1)} \otimes \ldots \otimes V_{l_d}^{(d)} \subset V$, $\mathbf{l} \in \mathbb{N}^d$, is the V-orthogonal sum $V_{\mathbf{l}} = \bigoplus_{\mathbf{k} \leq \mathbf{l}} W_{\mathbf{k}}$, where the inequality $\mathbf{k} \leq \mathbf{l}$ is meant componentwise.

In what follows we assume that the positive definite bilinear form $a(u, v)$ is well-defined on all $V_{\mathbf{l}}$, and that we can define the Hilbert space H_a as the $a(\cdot,\cdot)$-closure of $\operatorname{span}\{V_{\mathbf{l}} : \ \mathbf{l} \in \mathbb{N}^d\}$. Of course, for a computational treatment of the problem (1.1), we have to restrict ourselves to finite-dimensional spaces. To this end, we define the so-called generalized sparse grid spaces

$$V_{\mathcal{I}} := \sum_{\mathbf{l} \in \mathcal{I}} V_{\mathbf{l}} \subset H_a \, , \tag{3.3}$$

where the index set $\mathcal{I} \subset \mathbb{N}^d$ describes the anisotropic full grid spaces $V_{\mathbf{l}}$ that are to be included in the discretization. Note that the free choice of the index set $\mathcal{I}$ allows for dimension-adaptivity, whereas space-adaptivity is not possible since the subspaces $V_{\mathbf{l}}$ are either fully included or excluded by $\mathbf{l} \in \mathcal{I}$ or $\mathbf{l} \notin \mathcal{I}$, respectively. Since $V_{\mathbf{k}} \subset V_{\mathbf{l}}$ whenever $\mathbf{k} \leq \mathbf{l}$, we always silently assume that $\mathcal{I}$ satisfies the monotonicity condition that $\mathbf{l} \in \mathcal{I}$ and $\mathbf{k} \leq \mathbf{l}$ implies $\mathbf{k} \in \mathcal{I}$. Then it is easy to verify that

$$V_{\mathcal{I}} = \bigoplus_{\mathbf{k} \in \mathcal{I}} W_{\mathbf{k}}$$

can be written as V-orthogonal sum of the subspaces $W_{\mathbf{k}}$ with $\mathbf{k} \in \mathcal{I}$.

The basis for all further considerations in this paper is that we assume a set of fixed positive weights $\beta_{\mathbf{k}}, \mathbf{k} \in \mathbb{N}^d$, and an equivalence of norms

$$\|u\|_{H_a}^2 \eqsim \sum_{\mathbf{k} \in \mathbb{N}^d} \beta_{\mathbf{k}} \|w_{\mathbf{k}}\|_V^2 \, , \tag{3.4}$$

where $w_{\mathbf{k}} \in W_{\mathbf{k}}, \mathbf{k} \in \mathbb{N}^d$, denote the components of the unique V-orthogonal decomposition of $u \in H_a$, i.e., $u = \sum_{\mathbf{k} \in \mathbb{N}^d} w_{\mathbf{k}}$. In other words, we assume that

$$\{H_a, a(\cdot,\cdot)\} = \sum_{\mathbf{k} \in \mathbb{N}^d} \{W_{\mathbf{k}}, \beta_{\mathbf{k}}(\cdot,\cdot)_V\} \tag{3.5}$$

is a stable subspace splitting with a finite condition number denoted by κ^W. Note that in (3.5) we have a decomposition of H_a into a direct sum of V-orthogonal *sub*spaces, which allows us to omit the trivial embedding operators $R_{\mathbf{k}}^W : W_{\mathbf{k}} \to H_a$ that correspond to the R_i in (2.1) and (2.2).

Of course, for any generalized sparse grid space $V_{\mathcal{I}} \subset H_a$, the norm equivalence (3.4) implies an associated stable subspace splitting

$$\{V_{\mathcal{I}}, a(\cdot,\cdot)\} = \sum_{\mathbf{k}\in\mathcal{I}} \{W_{\mathbf{k}}, \beta_{\mathbf{k}}(\cdot,\cdot)_V\} \tag{3.6}$$

with condition numbers $\kappa_{\mathcal{I}}^W$ uniformly bounded by κ^W. Thus, if $\#\mathcal{I} < \infty$, we arrive at subspace correction methods for solving (1.1) on generalized sparse grid spaces $H = V_{\mathcal{I}}$ with convergence rates that are uniform with respect to $\mathcal{I}$. The computational cost per step of these optimally converging methods essentially depend on the involved subproblem solvers $T_{\mathbf{k}}^W : V_{\mathcal{I}} \to W_{\mathbf{k}}$. Since $W_{\mathbf{k}} \subset H$, we can write (2.5) for $b_i(\cdot,\cdot) = (\cdot,\cdot)_V$ directly in the form

$$(T_{\mathbf{k}}^W u, w_{\mathbf{k}})_V = a(u, w_{\mathbf{k}}) \qquad \forall\, w_{\mathbf{k}} \in W_{\mathbf{k}}\ .$$

Of course, we have to account for the $\beta_{\mathbf{k}}$-weights in (3.6), which is done in the update step of the resulting subspace correction method

$$u^{(m+1)} = u^{(m)} + \tau \sum_{\mathbf{k}\in\mathcal{I}} \omega_{\mathbf{k}}^W T_{\mathbf{k}}^W e^{(m)}, \qquad m = 0, 1, \ldots \tag{3.7}$$

by setting $\omega_{\mathbf{k}}^W = \beta_{\mathbf{k}}^{-1}, \mathbf{k} \in \mathcal{I}$. Here, an appropriately selected relaxation parameter $\tau > 0$ guarantees convergence rates uniformly in $\mathcal{I}$. Note that methods like steepest descent or conjugate gradients do not need τ to be fixed a priori but instead determine a parameter $\tau = \tau^{(m)}$ automatically in every iteration step m.

However, computing with the spaces $W_{\mathbf{k}}$ is often not as convenient as computing with the original anisotropic full grid spaces $V_{\mathbf{l}}$. This is why we now turn to the splitting

$$\{V_{\mathcal{I}}, a(\cdot,\cdot)\} = \sum_{\mathbf{l}\in\mathcal{I}} \{V_{\mathbf{l}}, \gamma_{\mathbf{l}}(\cdot,\cdot)_V\} \tag{3.8}$$

with positive weights $\gamma_{\mathbf{l}}, \mathbf{l} \in \mathcal{I}$, which only needs subspace solvers $T_{\mathbf{l}}^V : V_{\mathcal{I}} \to V_{\mathbf{l}} \subset V_{\mathcal{I}}$,

$$(T_{\mathbf{l}}^V u, v_{\mathbf{l}})_V = a(u, v_{\mathbf{l}}) \qquad \forall\, v_{\mathbf{l}} \in V_{\mathbf{l}}\ ,$$

that operate on full grid subspaces. The subspace correction method, which is based on the V-splitting corresponding to (3.8), is then of the form

$$u^{(m+1)} = u^{(m)} + \tau \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V T_{\mathbf{l}}^V e^{(m)}, \qquad m = 0, 1, \ldots\ , \tag{3.9}$$

with $\omega_{\mathbf{l}}^V = \gamma_{\mathbf{l}}^{-1}$. We then pose the question for $\gamma_{\mathbf{l}}$-weights for a V-splitting based method (3.9) such that the resulting convergence rates are competitive with the rates of the W-splitting based method (3.7) for given $\beta_{\mathbf{k}}, \mathbf{k} \in \mathcal{I}$.

The described setup is motivated by the discretization of H^t-elliptic problems by regular sparse grid spline spaces $V_{\mathcal{J}}$ over the d-dimensional unit cube, where the defining index set

$$\mathcal{J} = \big\{\mathbf{l} \in \mathbb{N}^d : |\mathbf{l}|_1 \le J\big\} \tag{3.10}$$

is of size $\Theta(J^d)$. The restriction on t is $|t| < r + 3/2$ if we use C^r spline spaces of fixed degree $m \geq r + 1$ over dyadic partitions of step-size 2^{-l} as the building blocks $V_l^{(i)}$. Then, the equivalence of norms (3.4) based on the W-splitting is given by

$$\|u\|_{H^t}^2 \simeq \sum_{\mathbf{k}\in\mathbb{N}^d} 2^{2t|\mathbf{k}|_\infty} \|w_{\mathbf{k}}\|_{L_2}^2, \qquad |\mathbf{k}|_\infty = \max_{i=1,\ldots,d} k_i \,, \tag{3.11}$$

where the $w_{\mathbf{k}} \in W_{\mathbf{k}}, \mathbf{k} \in \mathbb{N}^d$, denote the L_2-orthogonal components of the function $u \in H^t$, see [Osw94] for such kind of results. This, of course, implies the W-splitting (3.5) with $\beta_{\mathbf{k}} = 2^{2t|\mathbf{k}|_\infty}, \mathbf{k} \in \mathbb{N}^d$. The resulting condition number κ^W is a uniform upper bound for the splitting condition number $\kappa_{\mathcal{I}}^W$ of (3.6) with $\mathcal{I} = \mathcal{J}$. We note that, for some elliptic PDE problems with sums of tensor product operators as discussed for instance in [GK09, GO95b], the $\beta_{\mathbf{k}}$-weights can be modified such that the splitting condition number κ^W is independent of the dimension d (see [GH14] for the case $t = 1$). We revisit this model problem in the next subsections as a possible application and provide numerical experiments for the H_0^1-elliptic Laplace problem (the inclusion of Dirichlet boundary conditions is not essential) with linear splines ($r = 0$, $m = 1$) in Section 5.

In the following subsection, we present a method that explores the theory of stable subspace splittings and shows how to determine (positive) $\gamma_{\mathbf{l}}$-weights by solving a linear optimization problem such that the condition number of the splitting is in some sense minimal. We show that, in the H^t-elliptic case, the resulting condition numbers are *not* bounded independently of the sparse grid level J or the dimension d, but grow as $\Theta(J^{d-2})$.

Then, in Subsection 3.2, we see that there is an algebraic transformation that allows us to obtain optimal convergence rates, i.e., the same rates we would expect from a W-splitting based method given the norm equivalence (3.4). To this end, we have to circumvent the subspace splitting theory since partially negative scaling parameters occur. We explicitly state for which sets of positive *and* negative scaling parameters the resulting subspace splitting operator remains positive definite.

Finally, in Subsection 3.3, we apply the OptiCom from Subsection 2.2.2 to the sparse grid discretization. The OptiCom delivers the best possible scaling (including negative values) in each step of the iteration and results in a convergence that is at least as good as any fixed choice of scaling parameters, unfortunately at the extra cost of setting up and solving an auxiliary system of linear equations in every iteration step. It is therefore not competitive if an explicit norm equivalence (3.4) and the algebraic transformation is known. Nevertheless, the OptiCom does not take the detour via the W-splitting (3.6) and poses a lower bound on the convergence rate that we can achieve by optimizing fixed a-priori weights. So the OptiCom can be used to check whether the fixed scaling parameters obtained by other methods are close-to-optimal.

3.1. Stable space splittings with positive scaling parameters. Now, we want to determine weights $\gamma_{\mathbf{l}} > 0, \mathbf{l} \in \mathcal{I}$, such that the splitting number $\kappa_{\mathcal{I}}^V$ of (3.8) is small. To this end, we take a detour: Instead of estimating $\kappa_{\mathcal{I}}^V$ by comparing $a(u,u) = \|u\|_a^2$ and the squared splitting norm

$$|||u|||_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}^2 = \inf_{u=\sum_{\mathbf{l}\in\mathcal{I}} u_{\mathbf{l}}} \sum_{\mathbf{l}\in\mathcal{I}} \gamma_{\mathbf{l}} \|u_{\mathbf{l}}\|_V^2 \tag{3.12}$$

associated with (3.8) directly, we concentrate on comparing $|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}$ and the squared splitting norm

$$|||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W} = \sum_{\mathbf{k}\in\mathcal{I}} \beta_{\mathbf{k}} \|w_{\mathbf{k}}\|_V^2 \ , \qquad u = \sum_{\mathbf{k}\in\mathcal{I}} w_{\mathbf{k}}, \quad w_{\mathbf{k}} \in W_{\mathbf{k}} \tag{3.13}$$

associated with (3.6). We define

$$0 < \lambda^{WV}_{\min} := \inf_{u\in V_{\mathcal{I}}} \frac{|||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W}}{|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}} \ \le \ \lambda^{WV}_{\max} := \sup_{u\in V_{\mathcal{I}}} \frac{|||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W}}{|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}} < \infty \ ,$$

and try to minimize $\kappa^{WV}_{\mathcal{I}} := \lambda^{WV}_{\max}/\lambda^{WV}_{\min}$ which is the relative condition number between the two splittings

$$\sum_{\mathbf{k}\in\mathcal{I}} \{W_{\mathbf{k}}, \beta_{\mathbf{k}}(\cdot,\cdot)_V\} \quad \text{and} \quad \sum_{\mathbf{l}\in\mathcal{I}} \{V_{\mathbf{l}}, \gamma_{\mathbf{l}}(\cdot,\cdot)_V\} \tag{3.14}$$

based on the same underlying generalized sparse grid space $V_{\mathcal{I}}$. Since

$$\frac{a(u,u)}{|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}} = \frac{a(u,u)}{|||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W}} \cdot \frac{|||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W}}{|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}} \ ,$$

we get an upper and lower estimate of the condition number $\kappa^V_{\mathcal{I}}$ of (3.8) as

$$\max(\kappa^W_{\mathcal{I}}/\kappa^{WV}_{\mathcal{I}}, 1) \le \kappa^V_{\mathcal{I}} \le \kappa^W_{\mathcal{I}} \kappa^{WV}_{\mathcal{I}} \ . \tag{3.15}$$

Thus, under the assumption that an orthogonal splitting (3.6) (with a preferably low $\kappa^W_{\mathcal{I}}$) is available, the minimization of $\kappa^{WV}_{\mathcal{I}}$ results in tight upper and lower bounds for $\kappa^V_{\mathcal{I}}$.

In the remainder of this subsection, we express $\kappa^{WV}_{\mathcal{I}}$ in explicit form as a function of the parameters $\omega^V_{\mathbf{l}} = \gamma_{\mathbf{l}}^{-1} \ge 0$, $\mathbf{l} \in \mathcal{I}$, which can then be minimized using an LP. Note that we do not exclude the case $\omega^V_{\mathbf{l}} = 0$ (formally, this corresponds to $\gamma_{\mathbf{l}} = \infty$) for which the corresponding subspaces $V_{\mathbf{l}}$ in (3.12) and thus the corresponding operators $T^V_{\mathbf{l}}$ in (3.9) are "switched off". As a first step, we express the norm $|||u|||_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}$ in terms of V-orthogonal decompositions into the subspaces $W_{\mathbf{l}}$.

LEMMA 3.1. *For weights $\gamma_{\mathbf{l}}, \mathbf{l} \in \mathcal{I}$, and $\alpha_{\mathbf{k}}, \mathbf{k} \in \mathcal{I}$, with*

$$\alpha_{\mathbf{k}} := \Big(\sum_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1}\Big)^{-1} \quad \textit{for} \quad \mathbf{k} \in \mathcal{I} \ , \tag{3.16}$$

the norm (3.12) *is given by*

$$|||u|||_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V} = |||u|||_{\{\alpha_{\mathbf{k}}\}_{\mathcal{I}},W} \ . \tag{3.17}$$

Proof. We obtain the result by the following rearrangements

$$\begin{aligned}
|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V} &= \inf_{\substack{u_{\mathbf{l}}\in V_{\mathbf{l}}, \mathbf{l}\in\mathcal{I} \\ u=\sum_{\mathbf{l}\in\mathcal{I}} u_{\mathbf{l}}}} \sum\nolimits_{\mathbf{l}\in\mathcal{I}} \gamma_{\mathbf{l}} \|u_{\mathbf{l}}\|_V^2 \\
&= \inf_{\substack{w_{\mathbf{l},\mathbf{k}}\in W_{\mathbf{k}}, \mathbf{l}\in\mathcal{I}, \mathbf{k}\le\mathbf{l} \\ u=\sum_{\mathbf{l}\in\mathcal{I},\mathbf{k}\le\mathbf{l}} w_{\mathbf{l},\mathbf{k}}}} \sum\nolimits_{\mathbf{l}\in\mathcal{I}} \gamma_{\mathbf{l}} \sum_{\mathbf{k}\le\mathbf{l}} \|w_{\mathbf{l},\mathbf{k}}\|_V^2 \\
&= \inf_{\substack{w_{\mathbf{l},\mathbf{k}}\in W_{\mathbf{k}}, \mathbf{k}\in\mathcal{I}, \mathbf{l}\ge\mathbf{k} \\ u=\sum_{\mathbf{k}\in\mathcal{I},\mathbf{l}\ge\mathbf{k}} w_{\mathbf{l},\mathbf{k}}}} \sum_{\mathbf{k}\in\mathcal{I}} \sum_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}} \|w_{\mathbf{l},\mathbf{k}}\|_V^2 \ ,
\end{aligned}$$

where we first replaced each $u_{\mathbf{l}}$ by its unique V-orthogonal decomposition $u_{\mathbf{l}} = \sum_{\mathbf{k}\le\mathbf{l}} w_{\mathbf{l},\mathbf{k}}$ with $w_{\mathbf{l},\mathbf{k}} \in W_{\mathbf{k}}$, and then changed the order of summation. Note that $w_{\mathbf{k}} = \sum_{\mathbf{l}\ge\mathbf{k}} w_{\mathbf{l},\mathbf{k}}$ must hold for all $\mathbf{k} \in \mathcal{I}$. Thus the infimum and the summation over $\mathbf{k}$ commute, and we get

$$\begin{aligned} |||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V} &= \sum_{\mathbf{k}\in\mathcal{I}} \inf_{\substack{w_{\mathbf{l},\mathbf{k}}\in W_{\mathbf{l}},\mathbf{l}\ge\mathbf{k}\\ w_{\mathbf{k}}=\sum_{\mathbf{l}\ge\mathbf{k}} w_{\mathbf{l},\mathbf{k}}}} \sum\nolimits_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}} \|w_{\mathbf{l},\mathbf{k}}\|_V^2 \\ &= \sum_{\mathbf{k}\in\mathcal{I}} \Big(\sum\nolimits_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1}\Big)^{-1} \|w_{\mathbf{k}}\|_V^2. \end{aligned} \tag{3.18}$$

The equality (3.18) follows from solving a quadratic minimization problem, see also [GO95b]. Note that we can conclude $\alpha_{\mathbf{l}} > 0$ for all $\mathbf{l} \in \mathcal{I}$ if $\gamma_{\mathbf{l}}$ is finite for at least all maximal subspaces $V_{\mathbf{l}}$ in $V_{\mathcal{I}}$ (i.e., those $V_{\mathbf{l}}$ for which $V_{\mathbf{l}} \subset V_{\mathbf{k}} \subset V_{\mathcal{I}}$ implies $\mathbf{k} = \mathbf{l}$). But this must be true for any splitting (3.8), which concludes the proof. ∎

Lemma 3.1 says that the norm $|||\cdot|||_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}$ is also the norm of a weighted space splitting of $V_{\mathcal{I}}$ into V-orthogonal subspaces $W_{\mathbf{k}}$ with the $\alpha_{\mathbf{k}}$-weights from (3.16). Since W-splittings are direct sum splittings this means we can easily compute the splitting condition number $\kappa_{\mathcal{I}}^{WV}$ between the V-splitting and the W-splitting in (3.14) by directly comparing $\beta_{\mathbf{k}}$ and $\alpha_{\mathbf{k}}$ for $\mathbf{k} \in \mathcal{I}$. The following theorem is therefore an immediate consequence of Lemma 3.1.

THEOREM 3.2. *Let $\mathcal{I}$ be an index set, and consider the splittings in* (3.14) *with weights $(\beta_{\mathbf{k}} > 0)_{\mathbf{k}\in\mathcal{I}}$ and $(\gamma_{\mathbf{l}} > 0)_{\mathbf{l}\in\mathcal{I}}$, where $\gamma_{\mathbf{l}} < \infty$ or $\omega_{\mathbf{l}}^V = \gamma_{\mathbf{l}}^{-1} > 0$ for at least all maximal subspaces $V_{\mathbf{l}}$ in $V_{\mathcal{I}}$. Then, the best constants in the norm equivalence*

$$\lambda_{min}|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V} \le |||u|||^2_{\{\beta_{\mathbf{k}}\}_{\mathcal{I}},W} \le \lambda_{max}|||u|||^2_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}},V}\,,$$

valid for any $u = \sum_{\mathbf{k}\in\mathcal{I}} w_{\mathbf{k}} \in V_{\mathcal{I}}$, where $w_{\mathbf{k}} \in W_{\mathbf{k}}$, $\mathbf{k} \in \mathcal{I}$, are given by

$$\lambda_{max} = \max_{\mathbf{k}\in\mathcal{I}} \beta_{\mathbf{k}} \sum\nolimits_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1} \qquad \textit{and} \qquad \lambda_{min} = \min_{\mathbf{k}\in\mathcal{I}} \beta_{\mathbf{k}} \sum\nolimits_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1}\,.$$

In particular, it holds

$$\kappa_{\mathcal{I}}^{WV} = \frac{\max_{\mathbf{k}\in\mathcal{I}} \beta_{\mathbf{k}} \sum_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1}}{\min_{\mathbf{k}\in\mathcal{I}} \beta_{\mathbf{k}} \sum_{\mathbf{l}\ge\mathbf{k}} \gamma_{\mathbf{l}}^{-1}}\,. \tag{3.19}$$

Now, we want to minimize $\kappa_{\mathcal{I}}^{WV}$ with respect to $(\gamma_{\mathbf{l}})_{\mathbf{l}\in\mathcal{I}}$. To this end, we write $\kappa_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}^{WV}$ instead of $\kappa_{\mathcal{I}}^{WV}$ to indicate the dependence of the condition number on the set of $\gamma_{\mathbf{l}}$-weights and state the following LP.

LEMMA 3.3. *We can formulate the problem of finding the optimal weights and scaling parameters*

$$(\gamma_{\mathbf{l}}^*)_{\mathbf{l}\in\mathcal{I}} := \operatorname*{arg\,min}_{(\gamma_{\mathbf{l}}>0)_{\mathbf{l}\in\mathcal{I}}} \kappa_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}^{WV} \quad \textit{with} \quad \omega_{\mathbf{l}}^V = \gamma_{\mathbf{l}}^{-1}, \mathbf{l} \in \mathcal{I}, \tag{3.20}$$

as the LP

$$\begin{aligned} &\textit{Minimize} && \lambda \\ &\textit{subject to} && \sum_{\mathbf{l}\ge\mathbf{k}} \omega_{\mathbf{l}}^V \ge \beta_{\mathbf{k}}^{-1} \quad \forall\, \mathbf{k} \in \mathcal{I} && (3.21)\\ &\textit{and} && \beta_{\mathbf{k}}^{-1}\lambda - \sum_{\mathbf{l}\ge\mathbf{k}} \omega_{\mathbf{l}}^V \ge 0 \quad \forall\, \mathbf{k} \in \mathcal{I} && (3.22)\\ &\textit{and} && \lambda \ge 0,\ \omega_{\mathbf{k}}^V \ge 0 \quad \forall\, \mathbf{k} \in \mathcal{I}\,. && (3.23) \end{aligned}$$

The vector of unknowns $\left(\lambda, (\omega_{\mathbf{l}}^V)_{\mathbf{l}\in\mathcal{I}}\right)$ *of this LP is of size* $\#\mathcal{I}+1$ *and the parameter* λ *represents an upper bound for* $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}$.

Proof. Let us check that the above LP is indeed minimizing $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}$. The inequality constraints (3.21) are equivalent to

$$\beta_{\mathbf{k}} \sum\nolimits_{\mathbf{l}\geq\mathbf{k}} \gamma_{\mathbf{l}}^{-1} \geq 1, \qquad \mathbf{k}\in\mathcal{I}\,,$$

which ensures that for any feasible vector the denominator of (3.19) is at least 1. Moreover, the inequalities (3.21) imply that for any $\mathbf{k}\in\mathcal{I}$ there is at least one $\mathbf{l}\geq\mathbf{k}$ in $\mathcal{I}$ such that $\gamma_{\mathbf{l}}<\infty$, i.e., feasible $\left(\lambda, (\omega_{\mathbf{l}}^V)_{\mathbf{l}\in\mathcal{I}}\right)$ create admissible weight sets such that $V_{\mathcal{I}} = \sum_{\substack{\mathbf{l}\in\mathcal{I}\\ \gamma_{\mathbf{l}}<\infty}} V_{\mathbf{l}}$. The other set of constraints (3.22) can be rewritten as

$$\beta_{\mathbf{k}} \sum\nolimits_{\mathbf{l}\geq\mathbf{k}} \gamma_{\mathbf{l}}^{-1} \leq \lambda\,, \qquad \mathbf{k}\in\mathcal{I}\,,$$

which implies that for any feasible vector we guarantee $\lambda \geq \kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}$ according to (3.19). Altogether, this specifies an LP with non-empty feasibility set, and any optimal solution provides a set of weights $(\gamma_{\mathbf{l}}^*)_{\mathbf{l}\in\mathcal{I}}$ minimizing $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{I}}}$. ∎

The LP in Lemma 3.3 can be solved by one of the common LP algorithms. No strongly polynomial-time algorithm is known [Sma98], but in our numerical experiments the computational cost of solving the LP was negligible compared to solving the variational problem (1.1).

We now return to our model problem from the beginning of the current Section 3. In [GO94], a set of $\gamma_{\mathbf{l}}$-weights for a regular sparse grid space splitting (i.e., (3.8) with $\mathcal{I}=\mathcal{J}$ from (3.10)) was constructed which resulted in a condition number $\kappa^V_{\mathcal{J}} = \mathcal{O}(J^{d-2})$ for the special case of H^1-elliptic problems discretized by linear splines. Theorem 3.4 gives a similar result for general H^t-elliptic problems. Note that it is asymptotically sharp, i.e., $\kappa^V_{\mathcal{J}}$ possesses a matching lower bound.

THEOREM 3.4. *Let* $d\geq 2$*, and consider the discretization of a* H^t*-elliptic problem with regular sparse grid spaces* $V_{\mathcal{J}}$ *based on* C^r *splines of degree* $m\geq r+1$*, where* $0<t<r+3/2$*. We denote the condition number of the splitting* (3.8) *with* $\mathcal{I}=\mathcal{J}$ *and the set of weights* $(\gamma_{\mathbf{l}})_{\mathbf{l}\in\mathcal{I}}$ *by* $\kappa^V_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}}$*. For optimal* $\gamma_{\mathbf{l}}$*-weights, the condition number* $\kappa^V_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}}$ *grows as* $\Theta(J^{d-2})$ *in* J*, i.e., it holds*

$$\max(cJ^{d-2},1) \leq \inf_{(\gamma_{\mathbf{l}}>0)_{\mathbf{l}\in\mathcal{J}}} \kappa^V_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}} \leq CJ^{d-2}$$

with constants $0<c<C<\infty$ *independently of* $J\geq 1$.

Proof. According to (3.15), it is sufficient to prove this result for the optimal value of $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}}$ of (3.14), where the underlying weights $\beta_{\mathbf{k}} = 2^{2t|\mathbf{k}|_\infty}, \mathbf{k}\in\mathcal{J}$, originate from the well-known norm equivalence (3.4) for H^t. To obtain upper and lower bounds, we use the characterization of the optimal $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}}$ via the LP in Lemma 3.3. The value of λ associated with any feasible solution of the LP gives an upper bound for $\kappa^{WV}_{\{\gamma_{\mathbf{l}}\}_{\mathcal{J}}}$. We choose the set of weights $(\gamma_{\mathbf{l}})_{\mathbf{l}\in\mathcal{J}}$ from [GO94] to define the $\omega_{\mathbf{l}}^V$ as follows:

$$\omega_{\mathbf{l}}^V = \begin{cases} 2^{-2t|\mathbf{l}|_\infty}, & |\mathbf{l}|_1 = J\,, \\ 2^{-2tl}, & \mathbf{l}=(l,\ldots,l),\ \ l=0,\ldots,\lfloor J/d\rfloor\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{3.24}$$

It is easy to check that, with this choice for the $\omega_{\mathbf{l}}^V$, the inequalities (3.21) are automatically satisfied since, for any $\mathbf{k}\in\mathcal{J}$, there is a $\mathbf{l}\in\mathcal{J}$ with $\mathbf{l}\geq\mathbf{k}$ such that

$\omega_{\mathbf{l}}^V = \beta_{\mathbf{k}}^{-1}$. Indeed, if $k := |\mathbf{k}|_\infty \le J/d$, then $\mathbf{l} = (k, \dots, k) \in \mathcal{J}$ will do, if $k > J/d$ then we can find a $\mathbf{l} \ge \mathbf{k}$ with $|\mathbf{l}|_\infty = k$, and $|\mathbf{l}|_1 = J$ since $k \le |\mathbf{k}|_1 \le J$ and $dk > J$.

The inequalities in (3.22) are satisfied by determining a suitable λ. To this end, we have to bound the maximum of the quantities

$$\lambda_{\mathbf{k}} := \beta_{\mathbf{k}} \sum_{\mathbf{l} \in \mathcal{J} \colon \mathbf{l} \ge \mathbf{k}} \omega_{\mathbf{l}}^V, \qquad \mathbf{k} \in \mathcal{J} \ .$$

Since $\beta_{\mathbf{k}} = 2^{2t|\mathbf{k}|_\infty}$ depends on $k = |\mathbf{k}|_\infty$ only, the maximum of the $\lambda_{\mathbf{k}}$ with the same value k is obtained for $\mathbf{k} = (k, 0, \dots, 0)$. Thus, for each $k = 0, \dots, J$, we must ensure that

$$\lambda \ge \lambda_{(k,0,\dots,0)} = \sum_{l=k}^{\lfloor J/d \rfloor} 2^{2t(k-l)} + \sum_{l=k}^{J} 2^{2t(k-l)} \sum_{\substack{\mathbf{l} \ge (k,0,\dots,0) \\ |\mathbf{l}|_1 = J,\, |\mathbf{l}|_\infty = l}} 1 \ .$$

The first sum stems from the $\omega_{\mathbf{l}}^V > 0$ with $\mathbf{l} = (l, \dots, l) \ge (k, 0, \dots, 0)$ (if such indices exists in $\mathcal{J}$), and is uniformly bounded with respect to k and J since $t > 0$. The second sum stems from the remaining non-zero $\omega_{\mathbf{l}}^V$ with $\mathbf{l} \ge (k, 0, \dots, 0)$ and $|\mathbf{l}|_1 = J$, ordered by their value $l := |\mathbf{l}|_\infty \ge k$. For a rough estimate of the counting sum, observe that

$$\{\mathbf{l} \ge (k, 0, \dots, 0) : \ |\mathbf{l}|_1 = J,\, |\mathbf{l}|_\infty = l\} \subset \bigcup_{i=1}^{d} \{\mathbf{l} : \ l_i = l,\, \sum_{j \ne i} l_j = J - l\} \ .$$

Thus,

$$\sum_{\substack{\mathbf{l} \ge (k,0,\dots,0) \\ |\mathbf{l}|_1 = J,\, |\mathbf{l}|_\infty = l}} 1 \le d \cdot \#\{\mathbf{n} \in \mathbb{N}^{d-1} : \ |\mathbf{n}|_1 = J - l\} \le C' d J^{d-2}$$

with C' independent of l, k, and J. This shows that the first and second sum together, and thus the maximum of the $\lambda_{\mathbf{k}}$, are bounded by CJ^{d-2}, with a constant C independent of k and J. As a consequence, we can always choose some $\lambda \le CJ^{(d-2)}$ to arrive at a feasible vector $\big(\lambda, (\omega_{\mathbf{l}}^V)_{\mathbf{l} \in \mathcal{I}}\big)$. This gives the desired upper bound.

A matching lower bound can be found from considering the corresponding dual LP. To formulate it we use two vectors $\mathbf{y} = (y_{\mathbf{k}})_{\mathbf{k} \in \mathcal{J}}$ and $\mathbf{z} = (z_{\mathbf{k}})_{\mathbf{k} \in \mathcal{J}}$ associated to (3.21) and (3.22), respectively. The dual problem then reads

$$\text{Maximize} \quad \sum_{\mathbf{k} \in \mathcal{J}} \beta_{\mathbf{k}}^{-1} y_{\mathbf{k}} \tag{3.25}$$

$$\text{subject to} \quad \sum_{\mathbf{k} \in \mathcal{J}} \beta_{\mathbf{k}}^{-1} z_{\mathbf{k}} \le 1 \tag{3.26}$$

$$\text{and} \quad \sum_{\mathbf{k} \le \mathbf{l}} y_{\mathbf{k}} \le \sum_{\mathbf{k} \le \mathbf{l}} z_{\mathbf{k}} \quad \forall\, \mathbf{l} \in \mathcal{J} \tag{3.27}$$

$$\text{and} \quad \mathbf{y}, \mathbf{z} \ge 0 \ . \tag{3.28}$$

All we have to do is to find a feasible pair of vectors for this dual LP and to evaluate the target function on it. To this end, fix the smallest integer $k \ge J/2$ and set

$$z_{\mathbf{k}} = \begin{cases} 2^{2tk}, & \mathbf{k} = (k, 0, \dots) \ , \\ 0, & \text{otherwise} \ , \end{cases} \qquad y_{\mathbf{k}} = \begin{cases} 2^{2tk}, & \mathbf{k} = (k, k_2, \dots, k_d),\ |\mathbf{k}|_1 = J \ , \\ 0, & \text{otherwise} \ . \end{cases}$$

These vectors trivially fulfil (3.26) and (3.28). For (3.27) observe that $\sum_{\mathbf{k}\le\mathbf{l}} y_{\mathbf{k}}$ contains exactly one non-zero term (namely $y_{\mathbf{k}} = 2^{2tk}$ if $\mathbf{k} = (k, k_2, \ldots, k_d)$ satisfies $|\mathbf{k}|_1 = J$ and equals $\mathbf{l}$). Since $(k, 0, \ldots, 0) \le \mathbf{l}$ for such $\mathbf{l}$, we have

$$2^{2tk} = \sum_{\mathbf{k}\le\mathbf{l}} y_{\mathbf{k}} = z_{(k,0,\ldots,0)} = \sum_{\mathbf{l}\le\mathbf{k}} z_{\mathbf{l}} \,.$$

For all other $\mathbf{l}$, the inequality (3.27) is automatically valid since $\sum_{\mathbf{k}\le\mathbf{l}} y_{\mathbf{k}} = 0$.

Now, the value of the target functional for this feasible pair of vectors is

$$\sum_{\mathbf{k}\in\mathcal{J}} \beta_{\mathbf{k}}^{-1} y_{\mathbf{k}} = \sum_{k_2+\ldots+k_d=J-k} 1 \ge c'(J-k)^{d-2} \ge cJ^{d-2} \,,$$

where we have used the fact that, due to $k \ge J/2$ for all $\mathbf{k}$ with $y_{\mathbf{k}} = 2^{2tk} > 0$, we have $|\mathbf{k}|_\infty = k$, and thus $\beta_{\mathbf{k}}^{-1} y_{\mathbf{k}} = 1$. This proves the lower bound. ∎

3.2. Algebraic transformations. In Subsection 3.1, we have seen that for sparse grid discretizations of standard H^t-elliptic problems ($t > 0$) in dimensions $d \ge 3$, we cannot expect an optimal preconditioner derived from (3.8) using positive weights only. We now present an alternative approach which results in an optimally converging iterative scheme of the form (3.9) by allowing also negative relaxation parameters $\omega_{\mathbf{l}}^V, \mathbf{l} \in \mathcal{I}$. For this, we need the V-orthogonal projectors $Q_{\mathbf{l}}^V : V_{\mathcal{I}} \to V_{\mathbf{l}}$ and $Q_{\mathbf{k}}^W : V_{\mathcal{I}} \to W_{\mathbf{k}}$ with

$$\begin{aligned} (Q_{\mathbf{l}}^V u, v_{\mathbf{l}})_V &= (u, v_{\mathbf{l}})_V \qquad \forall\, v_{\mathbf{l}} \in V_{\mathbf{l}} \,, \quad \text{and} \\ (Q_{\mathbf{k}}^W u, w_{\mathbf{k}})_V &= (u, w_{\mathbf{k}})_V \qquad \forall\, w_{\mathbf{k}} \in W_{\mathbf{k}} \,, \end{aligned}$$

respectively. The following simple Lemma proves helpful for future considerations.

LEMMA 3.5. *For $\mathbf{k} \le \mathbf{l}$, we have $Q_{\mathbf{k}}^V T_{\mathbf{l}}^V = T_{\mathbf{k}}^V$ and $Q_{\mathbf{k}}^W T_{\mathbf{l}}^V = T_{\mathbf{k}}^W$.*

Proof. This immediately follows from

$$(Q_{\mathbf{k}}^V T_{\mathbf{l}}^V u, v_{\mathbf{k}})_V = (T_{\mathbf{l}}^V u, v_{\mathbf{k}})_V = a(u, v_{\mathbf{k}}) = (T_{\mathbf{k}}^V u, v_{\mathbf{k}})_V \qquad \forall\, v_{\mathbf{k}} \in V_{\mathbf{k}} \,,$$

and

$$(Q_{\mathbf{k}}^W T_{\mathbf{l}}^V u, w_{\mathbf{k}})_V = (T_{\mathbf{l}}^V u, w_{\mathbf{k}})_V = a(u, w_{\mathbf{k}}) = (T_{\mathbf{k}}^W u, w_{\mathbf{k}})_V \qquad \forall\, w_{\mathbf{k}} \in W_{\mathbf{k}} \,.$$

∎

The following theorem gives a formula for rewriting the subspace correction scheme based on a W-splitting (3.7) as a subspace correction method based on a V-splitting (3.9).

THEOREM 3.6. *Given the operator $P_{\boldsymbol{\omega}} = \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V T_{\mathbf{l}}^V$, we have*

$$P_{\boldsymbol{\omega}} = \sum_{\mathbf{k}\in\mathcal{I}} \omega_{\mathbf{k}}^W T_{\mathbf{k}}^W$$

if

$$\omega_{\mathbf{l}}^V = \sum_{\substack{\mathbf{e}\in\{0,1\}^d \\ \mathbf{l}+\mathbf{e}\in\mathcal{I}}} (-1)^{|\mathbf{e}|_1} \omega_{\mathbf{l}+\mathbf{e}}^W \,. \tag{3.29}$$

Proof. Given the one-dimensional orthogonal projections $Q_{l,i}^W : V^{(i)} \to W_l^{(i)}$ and $Q_{l,i}^V : V^{(i)} \to V_l^{(i)}$ for $i = 1, \ldots, d$ and $l \in \mathbb{N}$, it is not hard to see that $Q_{l,i}^W = Q_{l,i}^V - Q_{l-1,i}^V$ with $Q_{-1,i}^V := 0$. Then, due to the tensor product structure of V, we have for $\mathbf{k} \in \mathcal{I}$

$$\begin{aligned} Q_{\mathbf{k}}^W &= Q_{k_1,1}^W \otimes Q_{k_2,2}^W \otimes \cdots \otimes Q_{k_d,d}^W \\ &= (Q_{k_1,1}^V - Q_{k_1-1,1}^V) \otimes (Q_{k_2,2}^V - Q_{k_2-1,2}^V) \otimes \cdots \otimes (Q_{k_d,d}^V - Q_{k_d-1,d}^V) \\ &= \sum_{\substack{\mathbf{e} \in \{0,1\}^d \\ \mathbf{k}-\mathbf{e} \in \mathcal{I}}} (-1)^{|\mathbf{e}|_1} Q_{\mathbf{k}-\mathbf{e}}^V \,, \end{aligned}$$

which is the classical combination formula [HGC07, GSZ92, BGRZ94, BGR94] for projections based on tensor product scalar products. It carries over to the subspace solvers by using Lemma 3.5 twice, i.e. we have

$$T_{\mathbf{k}}^W = Q_{\mathbf{k}}^W T_{\mathbf{k}}^V = \sum_{\substack{\mathbf{e} \in \{0,1\}^d \\ \mathbf{k}-\mathbf{e} \in \mathcal{I}}} (-1)^{|\mathbf{e}|_1} Q_{\mathbf{k}-\mathbf{e}}^V T_{\mathbf{k}}^V = \sum_{\substack{\mathbf{e} \in \{0,1\}^d \\ \mathbf{k}-\mathbf{e} \in \mathcal{I}}} (-1)^{|\mathbf{e}|_1} T_{\mathbf{k}-\mathbf{e}}^V \,.$$

Then we can easily write

$$\sum_{\mathbf{k} \in \mathcal{I}} \omega_{\mathbf{k}}^W T_{\mathbf{k}}^W = \sum_{\mathbf{k} \in \mathcal{I}} \omega_{\mathbf{k}}^W \sum_{\substack{\mathbf{e} \in \{0,1\}^d \\ \mathbf{k}-\mathbf{e} \in \mathcal{I}}} (-1)^{|\mathbf{e}|_1} T_{\mathbf{k}-\mathbf{e}}^V = \sum_{\mathbf{l} \in \mathcal{I}} \Big(\sum_{\substack{\mathbf{e} \in \{0,1\}^d \\ \mathbf{l}+\mathbf{e} \in \mathcal{I}}} (-1)^{|\mathbf{e}|_1} \omega_{\mathbf{l}+\mathbf{e}}^W \Big) T_{\mathbf{l}}^V \,.$$

□

Theorem 3.6 shows that our W-splitting based Schwarz operator can be expressed using the operators $T_{\mathbf{l}}^V$ only. A similar idea was already used in [BPX90] for the full grid case, i.e., when the underlying space splitting is based on spaces $V_l := V_{(l,\ldots,l)}$. Often, this case is benign because for scaling parameters $\omega_l^W = 2^{-2tl}$, typical for H^t-elliptic problems with $t > 0$, we have

$$\omega_l^V = \omega_l^W - \omega_{l+1}^W = 2^{-2lt} - 2^{-2t(l+1)} = 2^{-2l}(1 - 2^{-2t}) \simeq \omega_l^W \,,$$

which means that we do not need to form orthogonal complements at all, neither implicitly nor explicitly. In our sparse grid case, the differences involve however 2^d terms, and we generally cannot expect $\omega_{\mathbf{l}}^W \simeq \omega_{\mathbf{l}}^V$ to hold uniformly in $\mathbf{l}$. Moreover, negative $\omega_{\mathbf{l}}^V$ are possible.

If we set $\omega_{\mathbf{k}}^W = 1$ for $\mathbf{k} \in \mathcal{I}$, Theorem 3.6 yields the standard combination technique [GSZ92, BGRZ94, BGR94], a popular method for approximating sparse grid solutions of, e.g. partial differential equations. Our case differs in the respect that our subspace solvers are based on the auxiliary bilinear forms $(\cdot,\cdot)_V$ instead of $a(\cdot,\cdot)$. Furthermore, we have $\omega_{\mathbf{k}}^W = \beta_{\mathbf{k}}^{-1}$ with $\beta_{\mathbf{k}}$ that stem from the norm equivalence (3.4). Finally, we do not stop after one iteration step but converge to the true sparse grid solution.

As already mentioned in Subsection 2.2, it is crucial that the operator $P_{\boldsymbol{\omega}}$ is positive definite. This is guaranteed for positive weights, but Theorem 3.6 suggests that even some weights with partially negative weights may result in a positive definite operator. We characterize these $\boldsymbol{\omega} = (\omega_{\mathbf{l}}^V)_{\mathbf{l} \in \mathcal{I}}$ by the following theorem.

THEOREM 3.7. *Given the operator $P_{\boldsymbol{\omega}} = \sum_{\mathbf{l} \in \mathcal{I}} \omega_{\mathbf{l}}^V T_{\mathbf{l}}^V$, we have*

$$P_{\boldsymbol{\omega}} = \sum_{\mathbf{k} \in \mathcal{I}} \omega_{\mathbf{k}}^W T_{\mathbf{k}}^W$$

if

$$\omega_{\mathbf{k}}^W = \sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \ . \tag{3.30}$$

It follows that $P_{\boldsymbol{\omega}}$ is positive definite if and only if

$$\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V > 0 \qquad \forall\, \mathbf{k} \in \mathcal{I} \ . \tag{3.31}$$

Proof. With Lemma 3.5, we get

$$P_{\boldsymbol{\omega}} = \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V T_{\mathbf{l}}^V = \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V \sum_{\mathbf{k}\leq\mathbf{l}} Q_{\mathbf{k}}^W T_{\mathbf{l}}^V = \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V \sum_{\mathbf{k}\leq\mathbf{l}} T_{\mathbf{k}}^W = \sum_{\mathbf{k}\in\mathcal{I}} \Big(\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \Big) T_{\mathbf{k}}^W$$

and thus

$$a(P_{\boldsymbol{\omega}} u, u) = \sum_{\mathbf{k}\in\mathcal{I}} \Big(\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \Big) a(T_{\mathbf{k}}^W u, u) = \sum_{\mathbf{k}\in\mathcal{I}} \Big(\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \Big) (T_{\mathbf{k}}^W u, T_{\mathbf{k}}^W u)_V \ .$$

Now, if the scaling parameters $\omega_{\mathbf{l}}^V$ satisfy (3.31) and if $u \neq 0$, it follows directly that $a(P_{\boldsymbol{\omega}} u, u) > 0$. Otherwise, if the condition (3.31) is violated for a $\mathbf{k} \in \mathcal{I}$, pick a $z_{\mathbf{k}} \in W_{\mathbf{k}}, z_{\mathbf{k}} \neq 0$ and compute $u \in W_{\mathbf{k}} \subset V_{\mathcal{I}}$ by

$$a(u, w_{\mathbf{k}}) = (z_{\mathbf{k}}, w_{\mathbf{k}})_V \qquad \forall\, w_{\mathbf{k}} \in W_{\mathbf{k}} \ .$$

Then $T_{\mathbf{j}}^W u = \delta_{\mathbf{jk}} z_{\mathbf{k}}$ and $a(P_{\boldsymbol{\omega}} u, u) = \Big(\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \Big) (z_{\mathbf{k}}, z_{\mathbf{k}})_V \leq 0$, which concludes the proof. □

Let us now discuss whether the algebraic scaling parameters $\omega_{\mathbf{l}}^V, \mathbf{l} \in \mathcal{I}$, given by (3.29) satisfy (3.31). Theorem 3.6 and Theorem 3.7 tell us that

$$\sum_{\mathbf{k}\in\mathcal{I}} \omega_{\mathbf{k}}^W T_{\mathbf{k}}^W = \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^V T_{\mathbf{l}}^V = \sum_{\mathbf{k}\in\mathcal{I}} \Big(\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V \Big) T_{\mathbf{k}}^W \ ,$$

so we see that $\sum_{\mathbf{l}\geq\mathbf{k}} \omega_{\mathbf{l}}^V = \omega_{\mathbf{k}}^W = \beta_{\mathbf{k}}^{-1} > 0$ for all $\mathbf{k} \in \mathcal{I}$ and (3.31) is satisfied. The LP in Lemma 3.3 contains the condition (3.31) in (3.21), thus the positivity constraints (3.23) for the $\omega_{\mathbf{l}}^V, l \in \mathcal{I}$, are not necessary to obtain a positive definite operator $P_{\boldsymbol{\omega}}$. In fact, the set of scaling parameters $(\omega_{\mathbf{l}}^V)_{\mathbf{l}\in\mathcal{I}}$ proposed in (3.29) for $\omega_{\mathbf{k}}^W = \beta_{\mathbf{k}}^{-1}, \mathbf{k} \in \mathcal{I}$, would be a solution to the LP with $\lambda = 1$ if we could remove the positivity constraints in (3.23). However, they are an integral part of the derivation using the subspace splitting theory.

3.3. OptiCom. Finally, we turn to another method that chooses the scaling parameters in a non-linear way in every single iteration step. Moreover, they are not restricted to be positive and do not necessarily satisfy (3.31). In Subsection 2.2.2, we gave a general description of how to find an optimal set of scaling parameters in every iteration step by solving an auxiliary minimization problem. This description largely followed [JN99] and was not confined to sparse grids. We now discuss aspects that are specific to the sparse grid case. They are similar to those that arise in the context of subspace correction methods for L_2-data approximation with sparse grids in the so-called OptiCom [Heg03, Gar06, HGC07].

The minimization problem (2.16) reads in the context of sparse grid discretizations as

$$\|u - u^{(m)} - \sum_{\mathbf{l}\in\mathcal{I}} \omega_{\mathbf{l}}^{(m)} T_{\mathbf{l}}^V e^{(m)}\|_a^2 \to \min_{\boldsymbol{\omega}^{(m)}} \tag{3.32}$$

with $\boldsymbol{\omega}^{(m)} = (\omega_{\mathbf{l}}^{(m)})_{\mathbf{l}\in\mathcal{I}}$, and leads in every iteration step m to a new system

$$\tilde{\mathbf{A}}^{(m)} \boldsymbol{\omega}^{(m)} = \tilde{\mathbf{b}}^{(m)} \,, \tag{3.33}$$

of linear equations. Here, the system matrix $\tilde{\mathbf{A}}^{(m)} \in \mathbb{R}^{\#\mathcal{I}\times\#\mathcal{I}}$ is positive semi-definite with

$$(\tilde{\mathbf{A}}^{(m)})_{\mathbf{lk}} = a(T_{\mathbf{k}}^V e^{(m)}, T_{\mathbf{l}}^V e^{(m)})$$

for $\mathbf{l}, \mathbf{k} \in \mathcal{I}$, and the right-hand side $\tilde{\mathbf{b}}^{(m)} \in \mathbb{R}^{\#\mathcal{I}}$ is given by

$$(\tilde{\mathbf{b}}^{(m)})_{\mathbf{l}} = a(e^{(m)}, T_{\mathbf{l}}^V e^{(m)}) = F(T_{\mathbf{l}}^V e^{(m)}) - a(u^{(m)}, T_{\mathbf{l}}^V e^{(m)})$$

for $\mathbf{l} \in \mathcal{I}$. Recall that the $T_{\mathbf{l}}^V e^{(m)} \in V_{\mathbf{l}}, \mathbf{l} \in \mathcal{I}$, are available from the subproblem solves. The OptiCom performs the update step (3.9) with $\tau = 1$ and $\omega_{\mathbf{l}}^V = \omega_{\mathbf{l}}^{(m)}, \mathbf{l} \in \mathcal{I}$, from the system (3.33), which just corresponds to the minimization problem (3.32). In contrast to that, the CG version of OptiCom associated with the minimization problem (2.17) would enlarge the system (3.33) by the unknown $\eta^{(m)}$. This leads to one additional entry on the right-hand side and one additional row and column of the system matrix, which we indicate by the letter η. We have

$$\begin{aligned}
(\tilde{\mathbf{A}}^{(m)})_{\eta,\eta} &= a(u^{(m)} - u^{(m-1)}, u^{(m)} - u^{(m-1)}) \,, \\
(\tilde{\mathbf{b}}^{(m)})_{\eta} &= a(u - u^{(m)}, u^{(m)} - u^{(m-1)}) \,, \\
(\tilde{\mathbf{A}}^{(m)})_{\eta,\mathbf{l}} &= a(T_{\mathbf{l}}^V e^{(m)}, u^{(m)} - u^{(m-1)}) \,, \\
(\tilde{\mathbf{A}}^{(m)})_{\mathbf{l},\eta} &= a(u^{(m)} - u^{(m-1)}, T_{\mathbf{l}}^V e^{(m)})
\end{aligned}$$

for $\mathbf{l} \in \mathcal{I}$. In practice, we use a direct method to solve $\tilde{\mathbf{A}}^{(m)}(\boldsymbol{\omega}^{(m)}, \eta^{(m)}) = \tilde{\mathbf{b}}^{(m)}$. In order to avoid problems with possibly singular system matrices, we use Tikhonov regularization with a very small regularization parameter.

Note that the setup and solution of the auxiliary problem (3.32) or (3.33), respectively, involves additional costs in every iteration step. If the number of scaling parameters in $\boldsymbol{\omega}^{(m)}$ is moderate compared to the total number of degrees of freedom, then the extra work of solving these linear problems can be tolerated. Of course, the extreme case would be the space splitting into one-dimensional spaces (this is the case of frame decompositions), which results in an "auxiliary" system (3.33) which is as large as the original problem. In the sparse grid case, though, the number of subspaces $\#\mathcal{I}$ and the total number of degrees of freedom are well-balanced, as we will see in the next section.

4. Implementation with a generating system. In this section, we describe the implementation of our different sparse grid subspace correction methods on $(V_{\mathbf{l}})_{\mathbf{l}\in\mathcal{I}}$ and discuss the computational complexity of one iteration step under the assumption that the variational form $a(\cdot,\cdot)$ stems from a sum of tensor product operators, e.g., for reaction-diffusion systems with constant coefficients. This allows us to apply the unidirectional principle [Bun92b, BZ96] in due course.

We assume that the $V_l^{(i)}$ in (3.1) can be discretized by dyadically refined linear or higher-order splines with $n_l = \Theta(2^l)$ basis functions on level l. Then, the spaces $V_{\mathbf{l}}$ have each a basis of d-variate local functions $\psi_{\mathbf{l},\mathbf{i}}, \mathbf{i} \in \chi_{\mathbf{l}}$, with

$$\chi_{\mathbf{l}} = \{(i_1, \ldots, i_d) \in \mathbb{N}^d : 1 \le i_j \le n_{l_j}, j = 1, \ldots, d\} .$$

The number of degrees of freedom of $V_{\mathbf{l}}$ is $n_{\mathbf{l}} := \#\chi_{\mathbf{l}} = \prod_{j=1}^d n_{l_j} = \Theta(2^{|\mathbf{l}|_1})$. For representing elements in our generalized sparse grid space $V_{\mathcal{I}}$, we use the generating system

$$\Psi_{\mathcal{I}} = \bigcup_{\mathbf{l}\in\mathcal{I}} \{\psi_{\mathbf{l},\mathbf{i}} : \mathbf{i} \in \chi_{\mathbf{l}}\} .$$

It has several advantages and is preferred over constructing hierarchical or wavelet bases in $V_{\mathcal{I}}$. As $\dim V_{\mathbf{l}} \simeq \dim W_{\mathbf{l}}$ for all $\mathbf{l}$, it follows that the total number of degrees of freedoms associated with representations in the generating system $\Psi_{\mathcal{I}}$ is asymptotically the same as $\dim V_{\mathcal{I}}$ and

$$N_{\mathcal{I}} := \#\Psi_{\mathcal{I}} = \sum_{\mathbf{l}\in\mathcal{I}} n_{\mathbf{l}} = \Theta\Big(\sum_{\mathbf{l}\in\mathcal{I}} 2^{|\mathbf{l}|_1}\Big) .$$

We use $N_{\mathcal{I}}$ and $\#\mathcal{I}$ for evaluating the computational cost later on.

Our discretized system for solving (1.1) on $V_{\mathcal{I}}$ reads

$$\mathbf{A}\mathbf{x} = \mathbf{b} ,$$

where $\mathbf{A} \in \mathbb{R}^{N_{\mathcal{I}}\times N_{\mathcal{I}}}$ is a block-structured matrix with

$$(\mathbf{A})_{(\mathbf{l},\mathbf{i}),(\mathbf{k},\mathbf{j})} = a(\psi_{\mathbf{k},\mathbf{j}}, \psi_{\mathbf{l},\mathbf{i}}) \quad \forall\, \mathbf{i} \in \chi_{\mathbf{l}}, \mathbf{j} \in \chi_{\mathbf{k}}, \mathbf{l}, \mathbf{k} \in \mathcal{I} ,$$

and $\mathbf{b} = (b_{\mathbf{l},\mathbf{i}})_{\mathbf{i}\in\chi_{\mathbf{l}},\mathbf{l}\in\mathcal{I}} \in \mathbb{R}^{N_{\mathcal{I}}}$ is a block-structured vector with

$$b_{\mathbf{l},\mathbf{i}} = F(\psi_{\mathbf{l},\mathbf{i}}) \quad \forall\, \mathbf{i} \in \chi_{\mathbf{l}}, \mathbf{l} \in \mathcal{I} .$$

We assume that $a(\cdot,\cdot)$ stems from a sum of tensor product operators and we can thus use the unidirectional principle to compute the matrix vector product

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad \Longleftrightarrow \quad y_{\mathbf{l},\mathbf{i}} = \sum_{\mathbf{k}\in\mathcal{I}} \sum_{\mathbf{j}\in\chi_{\mathbf{k}}} a(\psi_{\mathbf{k},\mathbf{j}}, \psi_{\mathbf{l},\mathbf{i}}) x_{\mathbf{k},\mathbf{j}} \qquad \forall\, \mathbf{i} \in \chi_{\mathbf{l}}, \mathbf{l} \in \mathcal{I}$$

with a computational complexity that is *linear* in the number of degrees of freedom $N_{\mathcal{I}}$, cf. [Bun92b, BZ96, Zei11].

In the general theory presented in Section 2, we had some embedding operators $R_i : H_i \to H$. In the sparse grid case, we have not explicitly used them, because the corresponding $R_{\mathbf{l}}^V : V_{\mathbf{l}} \to V_{\mathcal{I}}$ are simply the identity in the subspace case $V_{\mathbf{l}} \subset V_{\mathcal{I}}$ and can be dropped from the notation. In the matrix-vector setting, we need however the associated rectangular matrices $\mathbf{R}_{\mathbf{l}} \in \mathbb{R}^{N_{\mathcal{I}}\times n_{\mathbf{l}}}$ that act as the identity on coefficients that belong to $V_{\mathbf{l}}$, and pad all other entries with zeros. Now, let $\mathbf{x}^{(m)} \in \mathbb{R}^{N_{\mathcal{I}}}$ be the vector representing the current iterate $u^{(m)}$ in the generating system $\Psi_{\mathcal{I}}$. The corresponding residual is then given by

$$\mathbf{r}^{(m)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(m)} = \mathbf{A}\mathbf{e}^{(m)} ,$$

where $\mathbf{e}^{(m)} := \mathbf{x} - \mathbf{x}^{(m)}$ is the current error vector. The auxiliary variational problem (2.7) for the space $V_\mathbf{l}$ translates into

$$\mathbf{M}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)} = \mathbf{R}_\mathbf{l}^T(\mathbf{b} - \mathbf{A}\mathbf{x}^{(m)}) = \mathbf{R}_\mathbf{l}^T\mathbf{r}^{(m)} ,$$

where $\mathbf{M}_\mathbf{l} \in \mathbb{R}^{n_\mathbf{l}\times n_\mathbf{l}}$ is the mass matrix on the subspace $V_\mathbf{l}$ which in the linear spline case is simply a d-fold Kronecker-product of tridiagonal matrices (or band-limited matrices in the higher-order case), and can therefore be inverted with computational costs of $\mathcal{O}(n_\mathbf{l})$ operations. Having $\mathbf{x}_\mathbf{l}^{(m)}$, the vector representation of $u_\mathbf{l}^{(m)} = T_\mathbf{l}^V e^{(m)}$ is simply $\mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}$. Obviously, these tasks can be performed for all $\mathbf{l} \in \mathcal{I}$ with total costs of $\mathcal{O}(N_\mathcal{I})$ operations. All we need to do now is to apply the appropriate scaling parameters. For the cases of the ad hoc chosen scaling parameters (3.24), see also [GO94], LP-optimized scaling parameters (3.20), or algebraic scaling parameters (3.29), the $\omega_\mathbf{l}^V$ are given and fixed, and we can set

$$\mathbf{x}^{(m+1)} = \tau^{(m)} \sum_{\mathbf{l}\in\mathcal{I}} \omega_\mathbf{l}^V \mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)} + \eta^{(m)}(\mathbf{x}^{(m)} - \mathbf{x}^{(m-1)}) \tag{4.1}$$

with optimal $\tau^{(m)}$ and $\eta^{(m)} = 0$ for the steepest descent case, and optimal $\tau^{(m)}$ and $\eta^{(m)}$ for the conjugate gradient case. The computation of the optimal values $\tau^{(m)}$ or $(\tau^{(m)}, \eta^{(m)})$ associated with the minimization problem (2.19) leads to a 1×1 and 2×2 system of linear equations in the steepest descent and conjugate gradient case, respectively. Setting up the system requires only a fixed number of matrix-vector multiplications, which can be computed with $\mathcal{O}(N_\mathcal{I})$ operations. The solution for $\tau^{(m)}$ or $(\tau^{(m)}, \eta^{(m)})$ can be done with costs of $\mathcal{O}(1)$, and the update step (4.1) is linear in the number of degrees of freedom $N_\mathcal{I}$. Altogether, we arrive at a complexity of the order $\mathcal{O}(N_\mathcal{I})$. This holds for any generalized sparse grid space.

The OptiCom-approach is more intricate and additionally requires the setup and solution of (3.33). Then we can perform our update step (4.1) with $\omega_\mathbf{l}^V = \omega_\mathbf{l}^{(m)}$ and $\tau^{(m)} = 1, \eta^{(m)} = 0$ in the steepest descent case, or with $\tau^{(m)} = 1$ and $\eta^{(m)}$ obtained from the solution of $\tilde{\mathbf{A}}^{(m)}(\boldsymbol{\omega}^{(m)}, \eta^{(m)}) = \tilde{\mathbf{b}}^{(m)}$ in the CG-OptiCom case.

We now discuss the additional computational cost associated with the use of OptiCom. First note that

$$(\tilde{\mathbf{A}}^{(m)})_{\mathbf{lk}} = a(u_\mathbf{k}^{(m)}, u_\mathbf{l}^{(m)}) = \left\langle \mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}, \mathbf{A}\mathbf{R}_\mathbf{k}\mathbf{x}_\mathbf{k}^{(m)} \right\rangle \tag{4.2}$$

for all $\mathbf{k}, \mathbf{l} \in \mathcal{I}$ and that

$$\begin{aligned}(\tilde{\mathbf{b}}^{(m)})_\mathbf{l} &= F(u_\mathbf{l}^{(m)}) - a(u^{(m)}, u_\mathbf{l}^{(m)}) = \left\langle \mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}, \mathbf{b} \right\rangle - \left\langle \mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}, \mathbf{A}\mathbf{u}^{(m)} \right\rangle \\ &= \left\langle \mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}, \mathbf{r}^{(m)} \right\rangle \end{aligned} \tag{4.3}$$

for all $\mathbf{l} \in \mathcal{I}$. Thus, for *setting up* the matrix $\tilde{\mathbf{A}}^{(m)}$, we have to compute the matrix-vector products $\mathbf{A}\mathbf{R}_\mathbf{k}\mathbf{x}_\mathbf{k}^{(m)}$ for every $\mathbf{k} \in \mathcal{I}$. Then, the entries $(\tilde{\mathbf{A}}^{(m)})_{\mathbf{lk}}$ are computed by the scalar product with $\mathbf{R}_\mathbf{l}\mathbf{x}_\mathbf{l}^{(m)}$ for all $\mathbf{l} \in \mathcal{I}$. As these scalar products need only to be evaluated for coefficients that belong to the subspace $V_\mathbf{l}$, the costs for all $\mathbf{l} \in \mathcal{I}$ together is $\mathcal{O}(N_\mathcal{I})$ operations. The same argument applies to the entries (4.3) of the right hand side. Thus, we arrive at costs of $\mathcal{O}(N_\mathcal{I})$ operations for every $\mathbf{k} \in \mathcal{I}$, and consequently $\mathcal{O}(\#\mathcal{I} \cdot N_\mathcal{I})$ operations in total for setting up the system (3.33). The

same holds true for the CG version. Solving the system (3.33) by a direct method needs $\mathcal{O}(\#\mathcal{I}^3)$ operations. The subsequent update step (4.1) is again linear in $N_{\mathcal{I}}$. So, the total costs of one OptiCom iteration are $\mathcal{O}(\#\mathcal{I} \cdot N_{\mathcal{I}} + \#\mathcal{I}^3 + N_{\mathcal{I}})$ operations. Since for generalized sparse grid spaces $\#\mathcal{I} \ll N_{\mathcal{I}}$ holds, we can conclude that the total cost complexity of one iteration step is dominated by $\mathcal{O}(\#\mathcal{I} \cdot N_{\mathcal{I}})$. This is by a factor of $\#\mathcal{I}$ more expensive than the costs for the fixed a-priori scaling parameters.

We will now be a little more specific and choose the regular sparse grid setting $\mathcal{I} = \mathcal{J}$ (3.10) for level J and dimension d. It is well-known that the dimension of a regular sparse grid space grows as $N_{\mathcal{J}} = \Theta(J^{d-1}2^J)$ and $\#\mathcal{J} = \Theta(J^d)$. This means that the cost for one OptiCom iteration is $\mathcal{O}(J^d \cdot J^{d-1}2^J) = \mathcal{O}(J^{2d-1}2^J)$, which is log-linear in $N_{\mathcal{J}}$, whereas the methods with fixed a-priori scaling parameters are only linear in $N_{\mathcal{J}}$.

In previous applications of OptiCom to sparse grids in data mining [Gar06], the $a(\cdot,\cdot)$-products of $u_{\mathbf{l}}$ and $u_{\mathbf{k}}$ with $\mathbf{l} \neq \mathbf{k}$ are computed by first embedding $u_{\mathbf{l}}$ and $u_{\mathbf{k}}$ into the much larger subspace $V_{\max(\mathbf{l},\mathbf{k})}$ that contains both functions. This operation is expensive and typically leads to quadratic costs. Nevertheless, it is hardly avoidable since in that particular case the data-based energy norm $a(\cdot,\cdot)$ does not allow the representation as a sum of tensor products, which is a necessary prerequisite of the unidirectional principle. Our contribution is to show that we can achieve log-linear costs rather than quadratic costs if a fast matrix-vector multiplication method like the unidirectional principle is available.

5. Numerical experiments. In this section, we focus on the Poisson problem

$$-\Delta u = f \tag{5.1}$$

on the open unit cube $D = (0,1)^d$ with $f \in L_2(D)$ and zero boundary conditions on ∂D. The weak formulation of (5.1) is a $H_0^1(D)$-elliptic variational problem of the form (1.1), where

$$a(u,v) = \sum_{i=1}^{d} \Big(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i}\Big)_{L_2(D)}, \qquad F(u) = (f,u)_{L_2(D)} \ .$$

For (5.1), a discretization with linear C^0 splines is sufficient. More precisely, for the $V_l^{(i)}$ in (3.1) we use linear spline spaces defined over dyadic partitions of step-size $2^{-(l+1)}$ on $[0,1]$, with homogeneous boundary conditions at the boundary. Note here that, in order to avoid trivial subspaces for $l = 0$, the step-size associated with $V_l^{(i)}$ is chosen as $2^{-(l+1)}$ and not as 2^{-l}. See Fig. 5.1 for an illustration in the one- and two-dimensional cases.

Similar to the H^t-elliptic example discussed in Section 3, the norm equivalence

$$a(u,u) = \|u\|^2_{H_0^1(D)} \simeq \sum_{\mathbf{k}\in\mathbb{N}^d} \Big(\sum_{i=1}^{d} 2^{2k_i}\Big) \|w_{\mathbf{k}}\|^2_{L_2}, \qquad w_{\mathbf{k}} = Q^W_{\mathbf{k}} u, \quad \mathbf{k} \in \mathbb{N}^d \tag{5.2}$$

holds for $H_0^1(D)$ and the above described linear spline spaces. Thus, we employ the weights

$$\beta_{\mathbf{k}} = \sum_{i=1}^{d} 2^{2k_i} \tag{5.3}$$

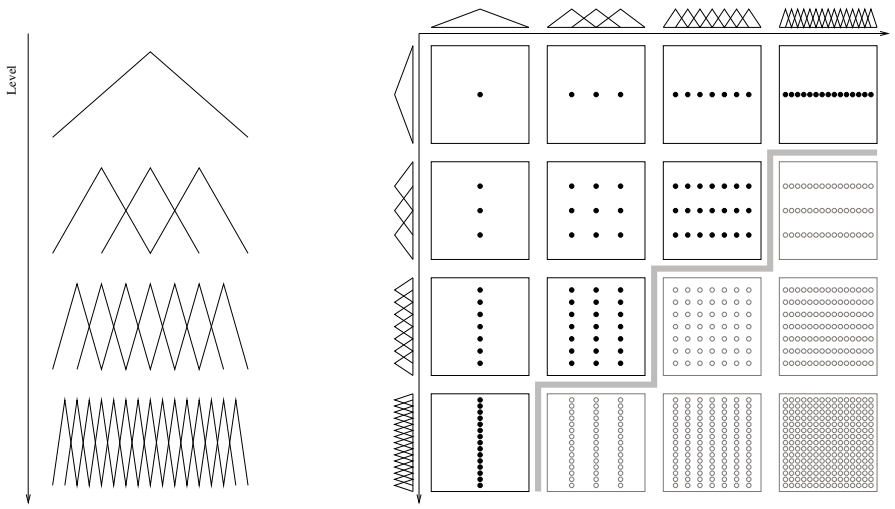

FIG. 5.1. *The first four levels of a one-dimensional multilevel system on linear splines (left). Two-dimensional tensorization and the regular sparse grid subspace (right)*

in the following experiments. Since $\sum_{i=1}^{d} 2^{2k_i} \simeq 2^{2|\mathbf{k}|_\infty}$ independently of $\mathbf{k}$, this is only a slight modification of the weights suggested by (3.11) for $t = 1$. However, this choice leads to splitting condition numbers κ^W for (3.5), which are independent of the space dimension d, cf. [GH14].

5.1. Splitting condition numbers. First, we compare the condition numbers of the operators $P_{\boldsymbol{\omega}}$ associated with the linear iterative methods for the regular sparse grid case (3.10) with different choices of parameter sets $\boldsymbol{\omega}$ discussed above (OptiCom and CG are nonlinear methods, and left out in this comparison). The results are shown in Table 5.1. We see that the LP-optimized scaling parameters (3.20) lead to condition numbers that are improved by a factor of up to 3 compared to the ad hoc scaling parameters (3.24) proposed in [GO94], which shows that weight optimization has a positive impact. However, for both scaling parameters, the condition numbers are of the order $\Theta(J^{d-2})$, and we can clearly observe their growth for increasing J in dimensions $d \geq 3$. Furthermore, we see that the W-splitting (3.6) with the weights (5.3) leads to condition numbers that are bounded independently of J *and* d. Moreover, they even *decrease* with rising dimension for sparse grids (this effect is explained in [GH14]). This condition number is realized by the partially negative scaling parameters $\omega_{\mathbf{l}}^V$ that stem from the algebraic transformation (3.29) with $\omega_{\mathbf{k}}^W = \beta_{\mathbf{k}}^{-1}, \mathbf{k} \in \mathcal{J}$.

5.2. Iteration counts. Now, we solve the test problem (5.1) with a random right-hand side f and a randomly initialized starting vector. We choose $J = 10$ and $d = 4$, and plot the reduction of the initial residual in the Euclidean norm against the iteration count. Figure 5.2 shows the convergence of the V-splitting based methods with ad hoc (3.24), LP-optimized (3.20) and algebraic (3.29) scaling parameters as well as for OptiCom. Both the steepest descent approach and the CG versions are considered. We observe that OptiCom is indeed always at least as good as any linear method, but the graphs also show that the potential gain from further optimizing the algebraic scaling parameters is quite limited. Furthermore, the conjugate gradient approach works for all four methods and roughly halves the

TABLE 5.1
Splitting condition numbers of the V-splittings with ad hoc and LP-optimized scaling parameters and of the W-splitting for different dimensions d and levels J

	V-splitting condition number $\kappa_{\mathcal{J}}^V$ with ad hoc scaling parameters (3.24)									
$d \setminus J$	1	2	3	4	5	6	7	8	9	10
1	2.86	3.87	4.74	5.47	6.06	6.55	6.95	7.29	7.58	7.83
2	3.13	5.08	8.12	8.07	10.81	10.37	11.69	11.14		
3	3.91	7.92	9.30	14.45	16.60	20.43				
4	4.86	12.33	18.84	26.53	38.05					
5	5.85	18.30	34.68	47.97						
6	6.85	26.38	59.33	94.59						
7	7.86	36.92	97.03	175.93						
8	8.87	50.16	153.26							
9	9.88	66.23	234.01							
10	10.88	85.28	345.58							

	V-splitting condition number $\kappa_{\mathcal{J}}^V$ with LP-optimized scaling parameters (3.20)									
$d \setminus J$	1	2	3	4	5	6	7	8	9	10
1	3.40	4.67	5.17	5.84	6.37	6.80	7.16	7.47	7.74	7.96
2	2.99	4.46	4.97	5.75	6.47	7.04	7.77	7.76		
3	2.42	4.05	6.69	8.72	12.31	14.62				
4	2.78	5.68	9.47	16.18	23.17					
5	3.43	7.85	14.58	26.09						
6	4.23	10.92	22.61	41.28						
7	5.08	15.02	34.36	67.90						
8	5.97	20.15	51.11							
9	6.89	26.27	73.92							
10	7.81	33.38	103.76							

	W-splitting condition number $\kappa_{\mathcal{J}}^W$ based on (3.6) with weights (5.3)									
$d \setminus J$	1	2	3	4	5	6	7	8	9	10
1	3.40	4.67	5.17	5.84	6.37	6.80	7.16	7.47	7.74	7.96
2	2.99	4.46	5.06	5.65	6.20	6.65	7.04	7.36		
3	2.71	4.28	5.00	5.49	6.06	6.53				
4	2.51	4.12	4.94	5.35	5.95					
5	2.36	3.97	4.88	5.23						
6	2.24	3.83	4.82	5.17						
7	2.15	3.71	4.77	5.15						
8	2.07	3.60	4.71							
9	2.00	3.50	4.66							
10	1.94	3.41	4.61							

number of necessary iterations (as it should). Furthermore, we see that the LP-optimized scaling parameters are better than the ad hoc scaling parameters, however, both methods converge very slowly. We investigate this effect further in the following experiment, where we vary also J.

In Fig. 5.3 we observe that the residual reduction takes more steps for higher

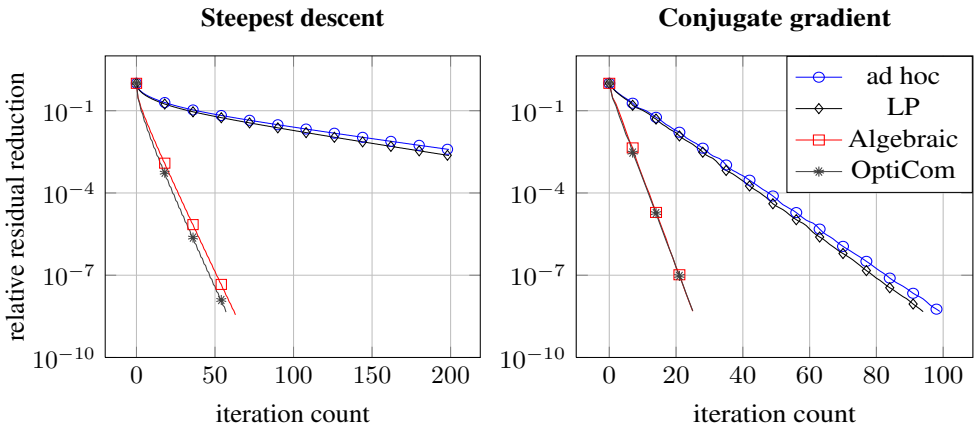

FIG. 5.2. *Residual reduction with ad hoc scaling parameters, LP-optimized scaling parameters, algebraic scaling parameters, and for OptiCom with the steepest descent method (left) and the conjugate gradient extension (right) for $J = 10$ in dimension $d = 4$*

values of J. However, we observe that the residual reduction rate ρ of the algebraic scaling parameters appears to be bounded from above independently of J, whereas the convergence rate of the LP-optimized scaling parameters deteriorates quickly. This is in full agreement with our theory, recall that, according to Theorem 3.4, the growth of the condition number of the underlying $P_{\boldsymbol{\omega}}$ is $\Theta(J^2)$ for LP-optimized scaling parameters in dimension $d = 4$.

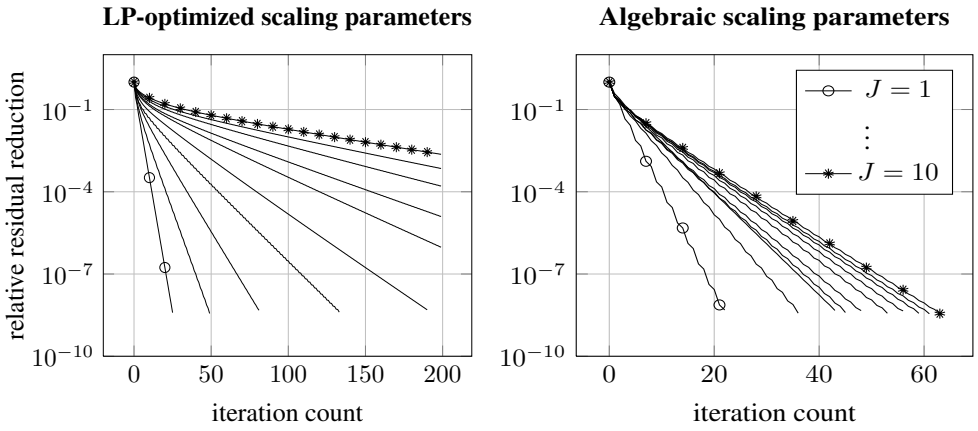

FIG. 5.3. *Residual reduction with LP-optimized scaling parameters (left) and algebraic scaling parameters (right) in dimension $d = 4$ for different values of J*

One final experiment concerns the dimension-dependence of the proposed scaling parameters. Figure 5.4 shows the residual reduction for $J = 10$ in the dimensions $d = 1, \ldots, 4$. The convergence rates of the LP-optimized scaling parameters deteriorate with dimension $d \geq 3$, whereas, as remarked earlier, the condition numbers of the W-splitting and thus the convergence rates of the algebraic scaling parameters are bounded independently of the dimension. This is true for the specific, problem-dependent weights (5.3) with our model problem (5.1), and generalizes to problems

with a sum of tensor product operators similar to the Laplacian. For other weights (such as the standard weights $\beta_{\mathbf{k}} = 2^{2|\mathbf{k}|_\infty}$ for H^1-problems), the d-independence is lost.

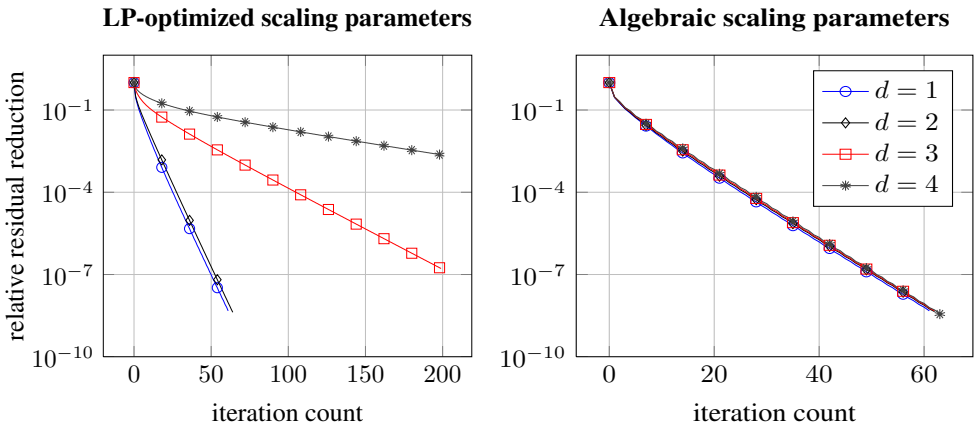

FIG. 5.4. *Residual reduction with LP-optimized scaling parameters (left) and algebraic scaling parameters (right) for $J = 10$ in dimensions $d = 1, \ldots, 4$*

6. Concluding remarks. To find close to optimal scaling parameters for elliptic variational problems discretized by sparse grids, we used approaches based on the theory of subspace splittings, on algebraic transformations and on a non-linear iterative algorithm called OptiCom. It turned out that non-negative scaling parameters are not the optimal choice. But admitting negative scaling parameters by an algebraic transformation leads to low-cost, high-speed linear iterative schemes that scale well with problem size J and dimension d. This approach is significantly cheaper than the variable preconditioning of OptiCom but results in nearly the same convergence rates. Moreover, it allows for a straightforward CG extension. It is noteworthy that the subspace correction method with algebraic scaling parameters has the same convergence properties as one that is based on prewavelet bases, but works with implementations of the operators $T_{\mathbf{l}}^V : V_{\mathcal{I}} \to V_{\mathbf{l}} \subset V_{\mathcal{I}}$ in the standard B-spline bases only. Under the proper circumstances, this can be a non-intrusive way to use existing full grid codes on $V_{\mathbf{l}}$ for $\mathbf{l} \in \mathcal{I}$ in a sparse grid subspace correction method.

Acknowledgements. The authors would like to thank the Deutsche Forschungsgemeinschaft for their partial support through the SFB 611 "Singular Phenomena and Scaling in Mathematical Models" and the SFB 1060 "The Mathematics of Emergent Effects".

REFERENCES

[Bel61] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[BG99] H. Bungartz and M. Griebel. A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. *J. Complexity*, 15:167–199, 1999.

[BG04] H. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.

[BG12] B. Bohn and M. Griebel. An adaptive sparse grid approach for time series predictions. In J. Garcke and M. Griebel, editors, *Sparse grids and applications*, volume 88 of

Lecture Notes in Computational Science and Engineering, pages 1–30. Springer, 2012.

[BGR94] H. Bungartz, M. Griebel, and U. Rüde. Extrapolation, combination, and sparse grid techniques for elliptic boundary value problems. *Comput. Methods Appl. Mech. Eng.*, 116:243–252, 1994.

[BGRZ94] H. Bungartz, M. Griebel, D. Röschke, and C. Zenger. Pointwise convergence of the combination technique for the Laplace equation. *East-West J. Numer. Math.*, 2:21–45, 1994.

[BPX90] J. Bramble, J. Pasciak, and J. Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990.

[Bun92a] H. Bungartz. An adaptive Poisson solver using hierarchical bases and sparse grids. In P. de Groen and R. Beauwens, editors, *Proceedings of the IMACS International Symposium, Brussels, 2.-4. 4. 1991*, pages 293–310, Amsterdam, 1992. Elsevier.

[Bun92b] H. Bungartz. *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*. Dissertation, Fakultät für Informatik, Technische Universität München, 1992.

[BZ96] R. Balder and C. Zenger. The solution of multidimensional real Helmholtz equations on sparse grids. *SIAM J. Sci. Comput.*, 17:631–646, 1996.

[CK04] P. Casazza and G. Kutyniok. Frames of subspaces. In *Wavelets, frames and operator theory*, volume 345 of *Contemp. Math.*, pages 87–113. Amer. Math. Soc., 2004.

[Feu10] C. Feuersänger. *Sparse Grid Methods for Higher Dimensional Approximation*. Dissertation, Institut für Numerische Simulation, Universität Bonn, September 2010.

[Gar06] J. Garcke. Regression with the optimised combination technique. In W. Cohen and A. Moore, editors, *Proceedings of the 23rd ICML '06*, pages 321–328, New York, NY, USA, 2006. ACM Press.

[GG03] T. Gerstner and M. Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1):65–87, 2003.

[GH13] M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Mathematics of Computations*, 82(282):975–994, April 2013.

[GH14] M. Griebel and A. Hullmann. On a multilevel preconditioner and its condition numbers for the discretized Laplacian on full and sparse grids in higher dimensions. In *Singular Phenomena and Scaling in Mathematical Models*. Springer International Publishing Switzerland, 2014.

[GK09] M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. *Mathematics of Computations*, 78(268):2223–2257, 2009.

[GO94] M. Griebel and P. Oswald. On additive Schwarz preconditioners for sparse grid discretizations. *Numer. Math.*, 66:449–464, 1994.

[GO95a] M. Griebel and P. Oswald. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numer. Math.*, 70:163–180, 1995.

[GO95b] M. Griebel and P. Oswald. Tensor product type subspace splitting and multilevel iterative methods for anisotropic problems. *Adv. Comput. Math.*, 4:171–206, 1995.

[GO12] M. Griebel and P. Oswald. Greedy and randomized versions of the multiplicative Schwarz method. *Linear Algebra and its Applications*, 437(7):1596–1610, 2012.

[Gri91] M. Griebel. A parallelizable and vectorizable multi-level algorithm on sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for partial differential equations*, volume 31, pages 94–100, Braunschweig, 1991. Vieweg.

[Gri94a] M. Griebel. Multilevel algorithms considered as iterative methods on semidefinite systems. *SIAM Int. J. Sci. Stat. Comput.*, 15(3):547–565, 1994.

[Gri94b] M. Griebel. *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen*. Teubner Skripten zur Numerik. Teubner, Stuttgart, 1994.

[Gri98] M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.

[GSZ92] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. In P. de Groen and R. Beauwens, editors, *Iterative Methods in Linear Algebra*, pages 263–281. IMACS, Elsevier, North Holland, 1992.

[Heg03] M. Hegland. Additive sparse grid fitting. In *Curve and Surface Fitting (Saint-Malo, 2002)*, Mod. Methods Math., pages 209–218. Nashboro Press, Brentwood, TN, 2003.

[HGC07] M. Hegland, J. Garcke, and V. Challis. The combination technique and some generalisations. *Linear Algebra and its Applications*, 420(2–3):249–275, 2007.

[JN99] M. Jung and S. Nepomnyaschikh. Variable additive preconditioning procedures. *Computing*, 62(2):109–128, 1999.

[KL07] A. Knyazev and I. Lashuk. Steepest descent and conjugate gradient methods with variable preconditioning. *SIAM J. Matrix Analysis Applications*, 29(4):1267–1280, 2007.

[KOPT13] G. Kutyniok, K. Okoudjou, F. Philipp, and E. Tuley. Scalable frames. *Linear Algebra and its Applications*, 438(5):2225–2238, 2013.

[Osw94] P. Oswald. *Multilevel Finite Element Approximation: Theory and Applications*. Teubner Skripten zur Numerik. Teubner, 1994.

[Osw09] P. Oswald. Stable space splittings and fusion frames. *Wavelets XIII (V. Goyal, M. Papadakis, D. Van de Ville, eds.), Proc. SPIE San Diego*, 7446, 2009.

[Sma98] S. Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2):7–15, 1998.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[Xu92] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34:581–613, 1992.

[XZ02] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *Journal of the American Mathematical Society*, 15(3):573–598, 2002.

[Zei11] A. Zeiser. Fast matrix-vector multiplication in the sparse-grid Galerkin method. *J. Sci. Comput.*, 47(3):328–346, 2011.

[Zen91] C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations*, volume 31 of *Notes on Numerical Fluid Mechanics*, pages 241–251. Vieweg, 1991.