

# A New Discrete Optimization of the Damping Parameter in Space-Time Multigrid with Block Jacobi Relaxation for the Heat Equation

Martin J. Gander<sup>[0000-0001-8450-9223]</sup>,  
Aušra Pogoželskytė<sup>[0009-0004-5446-4721]</sup>

## 1 Introduction

Parallel-in-Time methods have been developed to reduce time-to-solution when space parallelisation has reached its limits, which is particularly important for computations on million-core supercomputers [1]. Parallel-in-Time methods are particularly attractive as some [8, 4] can be applied without changing the existing code base. However, this convenience often limits scalability. While more intrusive, the Space-Time Multigrid (STMG) algorithm has great scalability properties [5]. This makes it an excellent candidate for large-scale applications where time-to-solution is critical.

To achieve the full potential of the STMG algorithm, its parameters need to be optimized, in particular its damping parameter  $\alpha$  used in the smoothing procedure. For the heat equation, the paper introducing STMG with block-Jacobi smoothing [5] suggests to use  $\alpha = 1/2$  in order to guarantee that time coarsening is always possible, and thus overcoming a drawback of earlier multigrid methods for solving parabolic problems [6, 7]. It had then been shown in [3] that for the discretization considered in this paper this choice can be improved in the case where  $\nu\Delta t/\Delta x^2 < 1/\sqrt{2}$ . Here, we show that it is possible to compute an even better choice of the damping parameter  $\alpha$  which leads in our numerical experiments to substantially faster convergence when STMG is used in practice, especially on the coarser meshes in the multigrid hierarchy.

As many multigrid algorithms, STMG is analyzed using Local Fourier Analysis (LFA) [9, Chapter 3]. This method is not rigorous and assumes that the problem is infinitely large, so the boundary conditions can be neglected and periodic boundary conditions can be considered instead. Neglecting the boundary conditions makes the

---

Aušra Pogoželskytė  
University of Geneva, rue du Conseil-Général 7-9, Geneva, Switzerland, e-mail:  
ausra.pogozelskyte@unige.ch

Martin J. Gander  
University of Geneva, rue du Conseil-Général 7-9, Geneva, Switzerland, e-mail:  
martin.gander@unige.ch

analysis much simpler, and gives rather good predictions for the actual convergence of STMG, as it was shown in [5, Figure 3 and 4]. We show here that performing a discrete LFA allows us to take into account that the constant error modes are actually not present in the STMG solver, and the resulting new optimized damping parameter further speeds up STMG in practice, while we still have a closed form formula for it. The insight that constant error modes are not present also holds in higher spatial dimensions and for more complex space-time problems, and the approach we present here can be generalized to STMG for these cases by including higher dimensional discrete LFA ranges without difficulty.

## 2 The Heat Equation and its Discretization

Consider the one-dimensional heat equation with Dirichlet boundary conditions and initial condition  $u_0$ ,

$$\partial_t u(x, t) = \nu \partial_{xx} u(x, t), \quad (x, t) \in [0, L] \times [0, T].$$

We discretize the heat equation using centered finite differences in space and Backward Euler in time as in [3]. Denoting by  $L_{n_x}$  the discrete Laplacian, we get

$$(I - \Delta t L_{n_x}) \mathbf{u}_{n+1} = \mathbf{u}_n + \mathbf{f}_{n+1}, \quad n = 0, \dots, n_t - 1. \quad (1)$$

Equation (1) can be written as an all-at-once system  $A\mathbf{x} = \mathbf{b}$ , namely

$$\underbrace{\begin{pmatrix} (I - \Delta t L_{n_x}) & & & & \\ -I & (I - \Delta t L_{n_x}) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & -I & (I - \Delta t L_{n_x}) \end{pmatrix}}_{=:A} \underbrace{\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{n_t} \end{pmatrix}}_{=:x} = \underbrace{\begin{pmatrix} \mathbf{f}_1 + \mathbf{u}_0 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_{n_t} \end{pmatrix}}_{=:b}, \quad (2)$$

where we chose  $\mathbf{x}$  and  $\mathbf{b}$  to avoid a notation conflict below in multigrid.

## 3 The Space-Time Multigrid (STMG) Algorithm

We then solve the system (2) in a multigrid fashion using a block Jacobi smoother,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha D^{-1}(\mathbf{b} - A \mathbf{x}^k), \quad k = 0, \dots, \nu - 1. \quad (3)$$

First, consider a hierarchy of grids  $\Omega_L \subset \dots \subset \Omega_2 \subset \Omega_1$ . On each of these grids, we define a problem  $A_i \mathbf{x}_i = \mathbf{b}_i$  where  $A_1 \mathbf{x}_1 = \mathbf{b}_1$  is given by (2) and subsequent problems, with  $i = 2, \dots, L$ , are defined by rediscretizing (2).

**Algorithm 1:** Algorithm: Space-Time Multigrid

---

```

1 Function STMG(l, x, b):
2   if l = L then
3      $x = A_L^{-1}b_L$  // Solve coarsest problem
4   x = Smoother(x, b,  $\nu_1$ ) // Damped block Jacobi pre-smoother
5   x = x + P · STMG(l + 1,  $\mathbf{0}$ ,  $R(\mathbf{b} - A_l \mathbf{x})$ ) // Coarse grid correction
6   x = Smoother(x, b,  $\nu_2$ ) // Damped block Jacobi post-smoother
7   return x

```

---

The coarsening is done either in time only or both in space and in time depending on the ratio  $\sigma := \nu \Delta t / \Delta x^2$  [3]. For analysis purposes, we also consider coarsening in space only. Define the spatial restriction and prolongation operators by  $R_x$  and  $P_x$ , as well as their time counterparts by  $R_t$  and  $P_t$ . For simplicity of notation, we omit to index those operators by the level  $i$ . Denoting by  $I_n$  the  $n \times n$  identity matrix, the space-time restriction and prolongation operators are then defined by a Kronecker product depending on the type of coarsening:

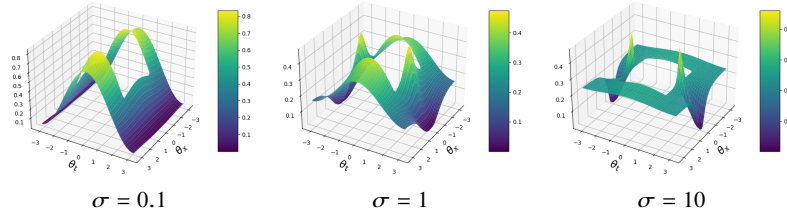
1. Semi-coarsening in time:  $R = R_t \otimes I_{n_x}$  and  $P = P_t \otimes I_{n_x}$ ,
2. Semi-coarsening in space:  $R = I_{n_t} \otimes R_x$  and  $P = I_{n_t} \otimes P_x$ ,
3. Full-coarsening:  $R = R_t \otimes R_x$  and  $P = P_t \otimes P_x$ .

## 4 Analysis of the Block Jacobi Smoother

Local Fourier Analysis (LFA) can be used to study the smoothing steps (lines 4 and 6) in the STMG Algorithm 1 separately from the coarse-grid correction (line 5). We focus here on the analysis of the block Jacobi smoother. The idea is to look at the approximate solution in Fourier space, that is, in function of its modes. To define them, we introduce  $\Theta := \{(2\pi\kappa/n_x, 2\pi\tau/n_t) \mid \kappa, \tau \in \mathbb{N}, 0 \leq \kappa < n_x, 0 \leq \tau < n_t\}$  the discrete set containing the normalized angular frequencies for our discretization. Note that we can recover the continuous set of normalized angular frequencies usually considered in LFA by considering the limit  $n_x, n_t \rightarrow \infty$ .

For  $(\theta_x, \theta_t) \in \Theta$ , the corresponding modes are given by  $\phi_{\theta_x, \theta_t}(j, n) := e^{ij\theta_x} e^{in\theta_t}$ . In particular, we distinguish the slowly oscillating components (called low modes) from the faster oscillating ones (called high modes), as the high modes will be aliased onto the low modes when the problem is coarsened. We denote by  $\Theta^{\text{high}}$  the set of high modes for the given discretization. Of course, that set is defined differently based on the type of coarsening (full or semi coarsening) that is performed. For semi-discretization in time the set of high modes is given by,

$$\Theta^{\text{high}} = \Theta \cap ([-\pi, \pi) \times ([-\pi, -\pi/2) \cup [\pi/2, \pi))) .$$



**Fig. 1** Symbol  $\hat{S}$  on the set  $\Theta^{\text{high}}$  for various mesh ratios  $\sigma = \nu\Delta t/\Delta x^2$

The goal of smoothing is to damp as much as possible the modes in  $\Theta^{\text{high}}$ . The amount of damping of the high modes is measured using the *smoothing factor*

$$\mu_S(\alpha) := \max\{|\hat{S}(\alpha; \theta_x, \theta_t)|; (\theta_x, \theta_t) \in \Theta^{\text{high}}\}, \quad (4)$$

where  $\hat{S}$  denotes the *symbol* of the block Jacobi smoother  $S$ . In LFA, the symbol is computed as the eigenvalue of the operator corresponding to  $S$  but with periodic conditions in space and time. For solving the heat equation as discretized in (2), the symbol of the block Jacobi smoother is given by

$$\hat{S}(\alpha; \theta_x, \theta_t) = 1 - \alpha + \frac{\alpha e^{-i\theta_t}}{1 + 2\sigma[1 - \cos \theta_x]}. \quad (5)$$

From Figure 1, the maximum of the symbol is attained near  $(\theta_x, \theta_t) = (0, \pi/2)$  or  $(\theta_x, \theta_t) = (\pi/2, 0)$ . An important remark is that the constant error modes where either  $\kappa = 0$  or  $\tau = 0$  do not need to be treated by STMG, since they are not present in the error due to the Dirichlet boundary conditions<sup>1</sup> and the initial condition. In the earlier analyses using a continuous set of frequencies, these modes could not be excluded when optimizing the damping parameter, which resulted in a sub-optimal parameter that was effective also for modes not present in the error. We show now that this change can make an important difference, especially on coarser meshes in the multigrid hierarchy, and leads to substantially faster convergence of STMG. For simplicity of notation, we define  $c_x := \cos(\pi/n_x)$  and  $c_t := \cos(\pi/n_t)$ .

**Lemma 1** *Let  $n_x > 2$  and  $n_t > 2$ . When considering the heat equation discretized by (1) and solved by STMG, the high frequency modes  $(\theta_x^*, \theta_t^*)$  that the block Jacobi smoother damps the least are when considering coarsening in*

1. *time only:*  $(\theta_x^*, \theta_t^*) = (\pi\Delta x/L, \pi/2)$ ;
2. *space only:*  $(\theta_x^*, \theta_t^*) = (\pi/2, \pi\Delta t/T)$ ;
3. *both space and time:*

$$(\theta_x^*, \theta_t^*) = \begin{cases} (\pi/2, \pi\Delta t/T) & \text{if } \alpha < \frac{1}{\frac{2\sigma c_x(1+\sigma(2-c_x))}{c_t(1+2\sigma)(1+2\sigma(1-c_x))^2} + 1}}, \\ (\pi\Delta x/L, \pi/2) & \text{otherwise.} \end{cases}$$

<sup>1</sup> The case of Neumann boundary conditions would be different and will be studied elsewhere.

*Proof.* Items 1. and 2. are obtained by computing  $|\hat{S}|^2$  which has the same maximum as  $|\hat{S}|$ , and maximizing it on the set of high frequencies in space and time to obtain the smoothing factor in (4). The resulting smoothing factors are

$$\mu_{\text{semi-time}}^2(\alpha) = (1 - \alpha)^2 + \frac{\alpha^2}{(1 + 2\sigma(1 - c_x))^2}$$

when coarsening in time only, and

$$\mu_{\text{semi-space}}^2(\alpha) = (1 - \alpha)^2 + \frac{2\alpha(1 - \alpha)c_t}{1 + 2\sigma} + \frac{\alpha^2}{(1 + 2\sigma)^2}$$

when coarsening in space only.

In the case of full space-time coarsening, the worst case from the items 1. and 2. dominates. For this, we compare the smoothing factors for semi-coarsening in time and in space. Isolating  $\alpha$  in  $\mu_{\text{semi-space}}^2(\alpha) > \mu_{\text{semi-time}}^2(\alpha)$  gives the criterion on  $\alpha$ .  $\square$

We can then minimize the smoothing factor (4) for each ratio  $\sigma = v\Delta t/\Delta x^2$ .

**Theorem 1** Let  $n_t > 3$  and  $n_x > 3$ . The optimal damping parameter  $\alpha^*$  when solving (2) with STMG is given by

1.  $\alpha^* = \frac{(1+2\sigma(1-c_x))^2}{1+(1+2\sigma(1-c_x))^2}$  when coarsening in time only;
2.  $\alpha^* = \min \left\{ \frac{1}{2} + \frac{\sigma+\sigma^2}{(1-c_t)+2\sigma(1-c_t)+2\sigma^2}, 1 \right\}$ , when coarsening in space only;
3. when coarsening both in space and time,

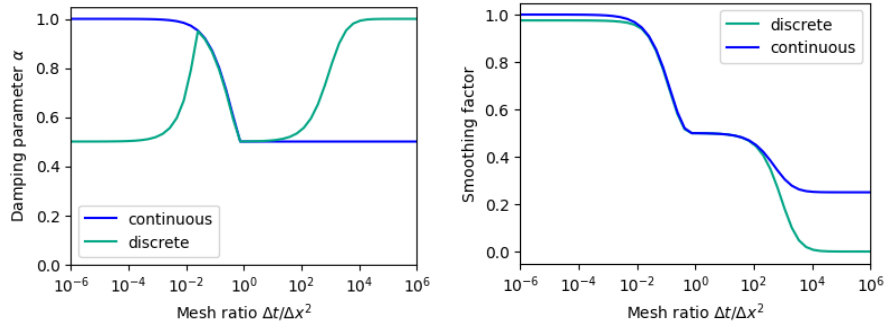
$$\alpha^* = \begin{cases} \min \left\{ \frac{1}{2} + \frac{\sigma+\sigma^2}{(1-c_t)+2\sigma(1-c_t)+2\sigma^2}, 1 \right\} & \text{for } \sigma < \sigma_1, \\ \left( 1 + \frac{2\sigma c_x(1+\sigma(2-c_x))}{c_t(1+2\sigma)(1+2\sigma(1-c_x))^2} \right)^{-1} & \text{for } \sigma_1 \leq \sigma \leq \sigma_2, \\ 1 - \left( (1 + 2\sigma(1 - c_x))^2 \right)^{-1} & \text{for } \sigma > \sigma_2, \end{cases}$$

where,  $\sigma_1$  is the positive real root of  $a_0 + a_1\sigma + a_2\sigma^2 + a_3\sigma^3 = 0$  with

$$\begin{aligned} a_0 &= 2c_x^2 + c_x(c_t - 2) + (1 - 2c_t), \\ a_1 &= (c_x - 2c_t)(6 - 4c_x) + 4(1 - 2c_x), \\ a_2 &= 2c_x(2c_x - 3) + 4(1 - c_x)(3 - c_x)(c_x - 2c_t), \\ a_3 &= 8(1 - c_x)^2(c_x - 2c_t), \end{aligned}$$

$$\text{and } \sigma_2 = \frac{c_t - c_x + \sqrt{c_x^2 + c_t^2 + 2c_x c_t(1 - c_x)}}{2c_x(2 - c_x)}.$$

*Proof.* Items 1. and 2. are obtained by optimizing the smoothing factor obtained in each case, that is, solving  $d\mu^2(\alpha)/d\alpha = 0$ . For item 3., we compute the *domain of time dominance* which is the domain where the smoothing factor for semi-coarsening



**Fig. 2** Comparison between the optimal discrete (with  $n_x = 100$  and  $n_t = 10$ ). Left: continuous damping parameters. Right: the corresponding smoothing factors.

in time is larger than for semi-coarsening in space. Its boundary is obtained by isolating  $\alpha$  in  $\mu_{\text{semi-time}}^2(\alpha) = \mu_{\text{semi-space}}^2(\alpha)$  and is given by

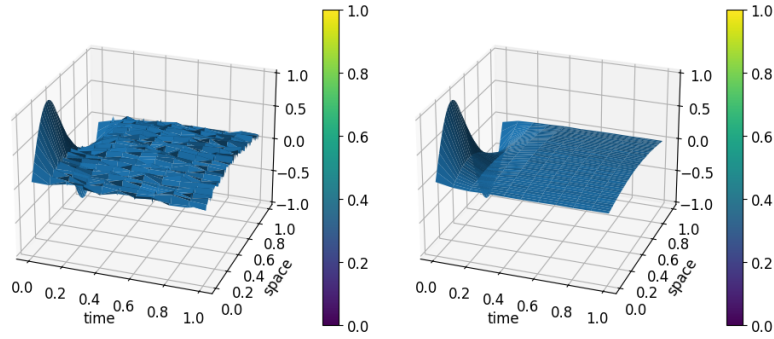
$$\alpha = \left( 1 + \frac{4\sigma c_x(1 + \sigma(2 - c_x))}{2c_t(1 + 2\sigma)(1 + 2\sigma(1 - c_x))^2} \right)^{-1}. \quad (6)$$

Next, we evaluate the smoothing factor with the optimal damping parameters for semi-coarsening in space and time (from item 1. and 2.). We observe that these curves cross the domain of time dominance only once. This means that for  $\sigma < \sigma_1$ , we can use the optimal damping parameter for semi-coarsening in space and for  $\sigma_2 < \sigma$  the one for time. For  $\sigma_1 \leq \sigma \leq \sigma_2$ , the optimal damping parameter is given by the boundary of the domain of time dominance as it minimizes the smoothing factor. As  $a_0 > 0$  and  $a_1, a_2, a_3 < 0$ , the existence and uniqueness of  $\sigma_1$  is given by the Descartes rule of signs: as the successive coefficients  $a_i$  change sign only once, our polynomial has a single positive real root.  $\square$

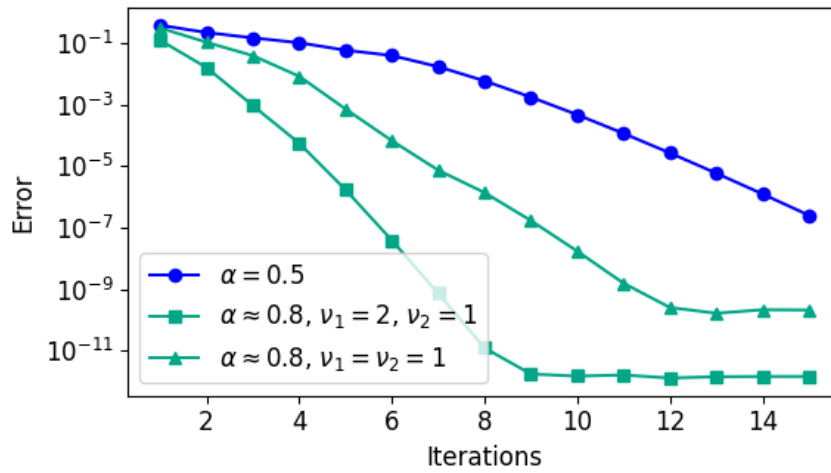
*Remark 1* More details for the proofs of Lemma 1 and Theorem 1 can be found in [2, Lemma 6.2.7 and Theorem 6.2.8].

*Remark 2* Note that in the case  $n_x, n_t \rightarrow \infty$ , the root  $\sigma_1$  degenerates to 0 and  $\sigma_2$  tends to the value  $1/\sqrt{2}$ , which is consistent with results from the continuous analysis in [5].

In Figure 2, we show the optimal damping parameter obtained from Theorem 1 compared to the optimal damping parameter obtained in [3]. We observe that the damping parameters are very different when the mesh ratio  $\sigma = \nu\Delta t/\Delta x^2$  is either very large or very small. We can also observe the impact of the choice of these damping parameters on the smoothing factor. In particular, we see that the new choice of damping parameter has most impact for large ratios where choosing the optimal damping parameter leads to extremely good smoothing.



**Fig. 3** Solution after 3 steps of the block Jacobi smoother for  $\sigma = 1000$  with mesh sizes  $n_x = 100$  and  $n_t = 10$  for different damping parameters  $\alpha$ . Left:  $\alpha = 0.5$  from the earlier continuous analysis. Right:  $\alpha \approx 0.8$  from the new discrete analysis.



**Fig. 4** Comparison of the two-grid convergence of STMG for solving the heat equation in the regime of large  $\sigma$  using the optimal damping parameters obtained from the analysis in [3] and the one obtained through our approach.

This is further illustrated by Figure 3 where we compare the smoothing abilities of the optimal damping parameter for the block Jacobi smoother given by the continuous analysis in [3] to the discrete analysis presented here. We observe a much smoother solution using the new damping parameter obtained through our discrete analysis.

Finally, in Figure 4, we illustrate the importance of considering a discrete analysis in particular for the coarser grids in the multigrid hierarchy which are very coarse. For this, we solve the heat equation as discretized in (2) on a grid of size  $n_x = 100$

and  $n_t = 10$  resulting in  $\sigma = 1000$ . We use STMG with  $\nu_1 = 2$  pre-smoothing steps and  $\nu_2 = 1$  post-smoothing steps, because it is the smallest amount of smoothing which results in decent convergence for the damping parameter  $\alpha = 0.5$  from the continuous analysis. Note that with the new optimized parameter from our discrete analysis, it would be largely sufficient to use  $\nu_1 = \nu_2 = 1$  (also shown in Figure 4) to obtain good convergence, resulting in larger savings in computation costs. The solution is transferred between grids using linear interpolation in space and time, full-weighting as restriction in space and injection as restriction in time. We use injection in time instead of full-weighting to limit excessive smoothing in the time dimension. As the damping parameters can be pre-computed given a grid hierarchy, using the damping parameter from either analysis does not increase computing costs. We thus see that using the damping parameter from our new discrete analysis rather than the one obtained from the continuous LFA gives a 2.5 speedup. Indeed, we converge to the same error in 4 iterations instead of 10.

Future work includes the generalization of these results to higher dimensions, including more general boundary conditions, and also to more general parabolic problems.

## References

1. June 2025 | TOP500. URL <https://top500.org/lists/top500/2025/06/>
2. Bronasco, A.: Improving the Efficiency and Theoretical Understanding of Time-Parallel Multigrid Methods. Ph.D. thesis, Université de Genève (2025)
3. Chaudet, B., Pogoželskytė, A., Gander, M.J.: An Optimized Space-Time Multigrid Algorithm for Parabolic PDEs (2023). ArXiv:2302.13881
4. Falgout, R.D., Friedhoff, S., Kolev, T.V., MacLachlan, S.P., Schroder, J.B.: Parallel time integration with multigrid. *SIAM J. Sci. Comp.* **36**(6), C635–C661 (2014)
5. Gander, M.J., Neumüller, M.: Analysis of a new space-time parallel multigrid algorithm for parabolic problems. *SIAM J. Sci. Comp.* **38**(4), A2173–A2208 (2016)
6. Hackbusch, W.: Parabolic multigrid methods. In: Computing methods in applied sciences and engineering, VI (Versailles, 1983), pp. 189–197. North-Holland, Amsterdam (1984)
7. Horton, G., Vandewalle, S.: A space-time multigrid method for parabolic partial differential equations. *SIAM J. Sci. Comp.* **16**(4), 848–864 (1995)
8. Lions, J.L., Maday, Y., Turinici, G.: Résolution d’EDP par un schéma en temps “pararéel”. *C. R. Acad. Sci. Paris Sér. I. Math.* **332**(7), 661–668 (2001)
9. Trottenberg, U., Oosterlee, C.W., Schüller, A.: Multigrid. Academic Press, Inc., San Diego, CA (2001)