# Modified Tangential Frequency Filtering Decomposition and its Fourier Analysis

Qiang Niu, Laura Grigori, Pawan Kumar, Frédéric Nataf

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Modified Tangential Frequency Filtering Decomposition and its Fourier Analysis*

Qiang Niu — Laura Grigori — Pawan Kumar — Frédéric Nataf

**N° 6662**

September 2008

Thème NUM


ISSN 0249-6399 ISRN INRIA/RR--6662--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY - ÎLE-DE-FRANCE

# Modified Tangential Frequency Filtering Decomposition and its Fourier Analysis

Qiang Niu *, Laura Grigori†, Pawan Kumar ‡ , Frédéric Nataf §

Thème NUM — Systèmes numériques
Équipes-Projets Grand-Large

Rapport de recherche n° 6662 — September 2008 — 24 pages

**Abstract:** In this paper, a modified tangential frequency filtering decomposition (MTFFD) preconditioner is proposed. The optimal order of the modification and the optimal relaxation parameter are determined by Fourier analysis. With this choice of the optimal order of modification, the Fourier results show that the condition number of the preconditioned matrix is $\mathcal{O}(h^{-\frac{2}{3}})$, and the spectrum distribution of the preconditioned matrix can be predicted by the Fourier results. The performance of MTFFD is compared with tangential frequency filtering (TFFD) preconditioner on a variety of large sparse matrices arising from the discretization of PDEs with discontinuous coefficients. The numerical results show that the MTFFD preconditioner is much more efficient than the TFFD preconditioner.

**Key-words:** preconditioner; linear system; tangential frequency filtering decomposition; GMRES

* School of Mathematical Sciences, Xiamen University, Xiamen, 361005, P.R. China; The work of this author was performed during his visit to INRIA, funded by China Scholarship Council; (Email:kangniu@gmail.com)

† INRIA Saclay - Ile de France, Laboratoire de Recherche en Informatique Universite Paris-Sud 11, France (Email:laura.grigori@inria.fr).

‡ INRIA Saclay - Ile de France, Laboratoire de Recherche en Informatique Universite Paris-Sud 11, France (Email:kumar@lri.fr).

§ Laboratoire J. L. Lions, CNRS UMR7598, Universite Paris 6, France; (Email: nataf@ann.jussieu.fr)

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

# Préconditionnement à base de filtrage tangentiel modifié et son analyse de Fourier

**Résumé :** Dans ce papier nous proposons une modification du préconditionnement à base de filtrage tangentiel (MTFFD). Les valeurs optimales des paramètres de la modification sont déterminées par une analyse de Fourier. Avec ce choix des paramètres, l'analyse de Fourier montre que le conditionnement de la matrice préconditionnée est de l'ordre de $\mathcal{O}(h^{-\frac{2}{3}})$. Les résultats numériques présentés montrent que MTFFD est plus efficace que le préconditionnement à base de filtrage tangentiel TFFD.

**Mots-clés :** préconditionnement, systèmes linéaires, GMRES

# 1 Introduction

In this paper, we investigate preconditioning techniques for solving the linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

with

$$\mathbf{A} = \begin{bmatrix} D_1 & U_1 & & \\ L_1 & D_2 & \ddots & \\ & \ddots & \ddots & U_{n_x-1} \\ & & L_{n_x-1} & D_{n_x} \end{bmatrix} \in \mathcal{R}^{N\times N}, \quad \mathbf{b} \in \mathcal{R}^N,$$

which often arises from the discretization of many PDEs by finite difference or finite volume schemes. When preconditioned iterative methods are used for solving (1), the convergence rate of an iterative method heavily depends on the property of the preconditioner [36]. Therefore, developing efficient preconditioners has been one of the major research interests in many applications. Algebraic multigrid (AMG) methods work well for many problems in practice [21, 35]. However, conventional AMG methods may suffer from relatively expensive setup time and large memory requirements, particularly for three dimensional problems [18]. Another more general preconditioner is the multilevel incomplete block factorization [9, 37]. A theoretical comparison of algebraic multigrid methods and algebraic multilevel methods are carried out by Y. Notay [33] for symmetric positive definite (SPD) matrices, and generalized by C. Mense and R. Nabben [29, 30] for nonsymmetric matrices.

Tangential Frequency Filtering Decomposition (TFFD) proposed in [2] is a special kind of incomplete block factorization preconditioner. Similar to some popular preconditioning techniques discussed in [1, 3, 8, 12, 13, 38, 39, 40, 42], the TFFD preconditioner can be used as a preconditioner or a smoother for multigrid methods. The preconditioner has the feature of filtering property, i.e. $(\mathbf{M}-\mathbf{A})\mathbf{f} = 0$ for a vector $\mathbf{f}$, where $\mathbf{M}$ is the TFFD preconditioner. For $\mathbf{A} \succeq 0$ (symmetric positive semidefinite), the preconditioner satisfies $\mathbf{M} - \mathbf{A} \succeq 0$, i.e. $\mathbf{M}$ is a compensative matrix of $\mathbf{A}$ [4, 5, 24]. This is an important property that preconditioners of SPD coefficient matrix should possess. By combining TFFD and ILU(0) in a multiplicative way [2], the combinative preconditioner is shown to be very efficient on several challenging problems. Therefore, it is important to give a deep understanding of the TFFD preconditioner. In [7, 8], the authors have presented some nice ways to analyze the properties of general block factorization preconditioners. For frequency filtering decomposition type preconditioners, some analysis have been done in [1, 13, 14, 38, 39]. The bounds on the spectral radii or condition numbers are derived by a couple of complicated inequalities. These results are useful in understanding the behavior of the preconditioners. However, the results have difficulty in outlining the range and clustering of the spectrum of the preconditioned matrix, especially when the matrix dimension is large. Fourier analysis is a powerful tool for analyzing the properties of a preconditioner [15]. It has been applied successfully to local model analysis in multigrid methods [10, 11, 41], and popularized by T. F. Chan and H. C. Elman [15] for analyzing algebraic preconditioners and classical iterative methods. Fourier analysis has been recognized as a standard tool for estimating the convergence rate of preconditioned iterative methods, see e.g. T. F. Chan et.al [16, 17], R J. Le Veque and L. N. Trefethen [23], K. Otto [34]. For

point-wise incomplete factorization type preconditioners like ILU(0), MILU(0) and RILU preconditioners, Fourier analysis has been done in [15, 16, 17]. The Fourier analysis of block ILU and MILU factorization preconditioners is considered in [34] for a time-dependent hyperbolic PDE problem.

The original aim of the present work is to analyze the TFFD preconditioner by means of Fourier analysis. Whereas, later we find that Fourier analysis for TFFD preconditioner is not feasible for our model problem. This is because of an exact cancelation in the denominator of a parameter, which is determined by symbolic computation (this will be shown in Section 3). But this does not mean that in practice TFFD is not a good preconditioner. This issue leads us to the derivation of the Modified Tangential Frequency Filtering Decomposition (MTFFD) preconditioner, in which the recursion formula of TFFD is modified by adding a term $c\Lambda_i h^q$. This idea of modification comes from the MILU preconditioner [20], where an additional term of order $\mathcal{O}(h^{-2})$ ($c \neq 0$) is added to the diagonal along with dropped fill-in. For problems arising from the discretizations of second-order elliptic partial differential equations, it is known [6, 20] that the modification is able to reduce the condition number of the preconditioned matrix by MILU from $\mathcal{O}(h^{-2})$ ($c = 0$) to $\mathcal{O}(h^{-1})$ ($c \neq 0$). Using a two dimensional Poisson equation as a model problem, we perform the Fourier analysis of the MTFFD preconditioner. The optimal choice of $q$ and $c$ are determined by this analysis, which shows that $q = \frac{4}{3}$ and $c = (4\pi^2)^{\frac{2}{3}}$ are the optimal choices as $h$ tends to 0. The optimality of these parameters is illustrated by the numerical tests. When the optimal choice of modification order $q$ is used, the Fourier analysis reveals that the condition number of the preconditioned matrix is $\mathcal{O}(h^{-\frac{2}{3}})$. This bound is better compared with other incomplete factorization type preconditioners (c.f. [15]). To compare the preconditioning effect of MTFFD with TFFD, we present tests on large sparse matrices arising from the discretization of PDEs with discontinuous coefficients. The results show that the MTFFD preconditioner is much more efficient, and MTFFD preconditioned GMRES needs less than half of the iteration numbers of TFFD preconditioned GMRES.

We use $ctrid_m(\alpha, \beta, \gamma)$ and $circ_m(\gamma_1, \ldots, \gamma_m)$ to denote the tridiagonal circulant matrix and circulant matrix of order $m$, i.e.

$$ctrid_m(\alpha,\beta,\gamma) = \begin{bmatrix} \beta & \gamma & & \alpha \\ \alpha & \ddots & \ddots & \\ & \ddots & \ddots & \gamma \\ \gamma & & \alpha & \beta \end{bmatrix}, circ_m(\gamma_1,\ldots,\gamma_m) \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{m-1} & \gamma_m \\ \gamma_m & \gamma_1 & \gamma_2 & \cdots & \gamma_{m-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma_3 & \cdots & \gamma_m & \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_3 & \cdots & \gamma_m & \gamma_1 \end{bmatrix}$$

We also use $trid_m(\alpha, \beta, \gamma)$ and $Btrid_m(L, T, U)$ to denote the $m \times m$ tridiagonal and $mk \times mk$ block tridiagonal matrix with each diagonal block of size $k \times k$ respectively, i.e.

$$trid_m(\alpha,\beta,\gamma) = \begin{bmatrix} \beta & \gamma & & \\ \alpha & \ddots & \ddots & \\ & \ddots & \ddots & \gamma \\ & & \alpha & \beta \end{bmatrix} \quad and \quad Btrid_m(L,T,U) = \begin{bmatrix} T & U & & \\ L & \ddots & \ddots & \\ & \ddots & \ddots & U \\ & & L & T \end{bmatrix}.$$

The paper is organized as follows, In Section 2, a model problem is described, which will be used for Fourier analysis. In Section 3, we present the modified TFFD preconditioner and carry out the Fourier analysis for the MTFFD preconditioner. In Section 4, the performance of the MTFFD preconditioner is compared with the TFFD preconditioner by several examples. Finally, we conclude the paper in Section 5.

## 2 Description of the model problem

We consider the 2-D Poisson equation as the model problem, i.e.

$$-\Delta u = f \tag{2}$$

posed on the unit square $\Omega = 0 \leq x, y \leq 1$ with Dirichlet boundary conditions

$$u(x, y) = 0.$$

This model problem is also used in [8, 15, 17]. Discretizing this problem by the standard second-order finite difference (FE) scheme on a uniform grid with step size $h_d = \frac{1}{m+1}$ in each direction, we can obtain a linear system of order $m^2$

$$\tilde{A}u = \tilde{b}, \tag{3}$$

where

$$\tilde{A} = I_m \otimes \tilde{D} + \kappa_2 \tilde{S} \otimes I_m,$$
$$\tilde{D} = trid_m(-\kappa_1, d, -\kappa_1),$$
$$\tilde{S} = trid_m(-1, 0, -1).$$

and $d = 2(\kappa_1 + \kappa_2)$, $\kappa_1 = \kappa_2 = 1$.

Fourier analysis can only be performed on constant coefficient problems with periodic boundary conditions [15]; hence we also introduce the discretization of equation (2) with periodic boundary conditions

$$u(x, 0) = u(x, 1) \qquad u(0, y) = u(1, y).$$

According to the argument in [15], we assume the discretization step size $h_p$ for the periodic case to be half that of Dirichlet case, i.e. $h_p = \frac{1}{n+1} = \frac{1}{2(m+1)}$. Then we have a linear system of order $n^2$

$$Au = b \tag{4}$$

where

$$A = I_n \otimes D + \kappa_2 S \otimes I_n,$$
$$D = ctrid_n(-\kappa_1, d, -\kappa_1),$$
$$S = ctrid_n(-1, 0, -1),$$

and the values of $d$, $\kappa_1$ and $\kappa_2$ are the same as in (3).

The main idea of Fourier analysis to use the theoretical results obtained for a periodic problem to predict the convergence results of the corresponding

Dirichlet problem. Therefore, we present in this paper the Fourier analysis of the linear system (4). In the following discussion, the equation (3) and (4) will be referred to as Dirichlet problem and periodic problem respectively, and we always use subscript $d$ and $p$ to distinguish the parameters for the Dirichlet case and the periodic case.

Firstly, the Fourier eigenvalue of the coefficient matrix $A$ is given by the following equation [15]

$$Au^{(j,k)} = \lambda_A u^{(j,k)},$$

where $u^{(j,k)}$ is defined by

$$u_{s,t}^{(j,k)} = e^{\mathtt{i}s\theta_j} e^{\mathtt{i}t\phi_k} \tag{5}$$

with

$$\theta_j = \frac{2\pi j}{n+1}, \qquad \phi_k = \frac{2\pi k}{n+1}, \qquad 1 \le j,k \le n. \tag{6}$$

and $\mathtt{i}$ is the imaginary unit.

Substituting the expression of $u_{s,t}^{(j,k)}$ into the grid-equation related to $A$ (refer to [15]), we have

$$\begin{aligned} Au^{(j,k)} &= du_{s,t} - \kappa_1 u_{s+1,t} - \kappa_1 u_{s-1,t} - \kappa_2 u_{s,t-1} - \kappa_2 u_{s,t+1} \\ &= (d - \kappa_1 e^{\mathtt{i}\theta_j} - \kappa_1 e^{-\mathtt{i}\theta_j} - \kappa_2 e^{\mathtt{i}\phi_k} - \kappa_2 e^{-\mathtt{i}\phi_k}) e^{\mathtt{i}s\theta_j} e^{\mathtt{i}t\phi_k} \\ &= (4 - 2\cos(\theta_j) - 2\cos(\phi_k)) e^{is\theta_j} e^{\mathtt{i}t\phi_k} \\ &= 4(\kappa_1 \sin^2(\tfrac{\theta_j}{2}) + \kappa_2 \sin^2(\tfrac{\phi_k}{2}))u^{(j,k)}. \end{aligned} \tag{7}$$

Thus, the Fourier eigenvalue of $A$ corresponding to the Fourier eigenvector $u^{(j,k)}$ are

$$\lambda_A = \lambda_{j,k}(A) = 4(\kappa_1 \sin^2(\frac{\theta_j}{2}) + \kappa_2 \sin^2(\frac{\phi_k}{2})). \tag{8}$$

The expression of the Fourier eigenvalue of $A$ will be used later in the Fourier analysis.

# 3 Modified Tangential Frequency Filtering Decompositions and its Fourier analysis

For a general block tridiagonal linear system (3), we introduce the *Modified Tangential Frequency Filtering Decomposition* (MTFFD) preconditioner $\tilde{M}$ as follows

$$\tilde{M} = \begin{bmatrix} \tilde{T}_1 & & & \\ L_1 & \tilde{T}_2 & & \\ & \ddots & \ddots & \\ & & L_{m-1} & \tilde{T}_m \end{bmatrix} \begin{bmatrix} \tilde{T}_1^{-1} & & & \\ & \tilde{T}_2^{-1} & & \\ & & \ddots & \\ & & & \tilde{T}_m^{-1} \end{bmatrix} \begin{bmatrix} \tilde{T}_1 & U_1 & & \\ & \tilde{T}_2 & \ddots & \\ & & \ddots & U_{m-1} \\ & & & \tilde{T}_m \end{bmatrix}. \tag{9}$$

The diagonal blocks $\tilde{T}_i$ in MTFFD are computed by the following recursion formula

$$\tilde{T}_i = \begin{cases} D_1 + c\Lambda_1 h^q, & i = 1, \\ D_i - L_{i-1}(2\beta_i - \beta_i \tilde{T}_{i-1} \beta_i) U_{i-1} + c\Lambda_i h^q, & 1 < i \le m. \end{cases} \tag{10}$$

where $\Lambda_i$, $1 \leq i \leq m$, are diagonal matrices, parameter $q$ is the order of modification, and $c$ is a relaxation parameter. The optimal choice of $q$ and $c$ will be discussed later. The matrix $\beta_i$ is an approximation to the inverse of $\tilde{T}_{i-1}$, and it can be determined by enabling $\tilde{M}$ to have a filtering condition. The analysis in [2] shows that it reduce to solving

$$\beta_i(U_{i-1}f) = \tilde{T}_{i-1}^{-1}U_{i-1}f, \tag{11}$$

where $f$ is a filtering vector.

We can see that the MTFFD preoconditioner differs from the TFFD preconditioner in that an additional term $c\Lambda_i h^q$ is added in the recursion formula (10). If $\Lambda_i = I_m$, then the modification is similar to those done in the modified ILU factorization [6, 20], where the modification is $ch^2 I_m$. The modification in (10) is quite similar to the shifted iteration methods discussed in [43], where the ILU factorization of a shifted coefficient matrix is constructed and used as a preconditioner for the original problem. For analysis purpose, we will fix $\Lambda_i = I_m$, $1 \leq i \leq m$, and the filtering vector is chosen as $\mathbf{1} = [1, \ldots, 1]^T$.

As mentioned before, Fourier analysis can be performed only on the constant coefficient problems with periodic boundary conditions. According to the theory developed in [15], there are several assumptions on which our analysis will be based,

• The grid size $h_p = \frac{1}{2}h_d$ should be used in order to relate the Fourier analysis results to that of the Dirichlet problems. We have made this assumption be satisfied when the discretization of (2) is done.

• For the linear system (4) generated by the discretization of (2) with periodic boundary conditions, the MTFFD preconditioner $\hat{M}$ is forced to have constant diagonals, i.e. the MTFFD preconditioner $\hat{M}$ for periodic system (4) should take the form of

$$\hat{M} = (L + \hat{T})\hat{T}^{-1}(\hat{T} + U), \tag{12}$$

where $\hat{T}$ has the same diagonal blocks, i.e.

$$\hat{T} = I_n \otimes \hat{T}_0$$

and each diagonal block $\hat{T}_0$ is circulant, i.e.

$$\hat{T}_0 = circ_n(\hat{d}, -\hat{\kappa}_1, 0, \ldots, 0, -\hat{\kappa}_1)$$

with parameters $\hat{d}$ and $\hat{\kappa}_1$ to be determined by the recursion formula (10).

Using the assumptions above and the recursion formula (10), we now construct MTFFD preconditioner for which we will perform Fourier analysis.

Firstly, the parameters $\hat{d}$ and $\hat{\kappa}_1$ can be computed by solving

$$\begin{aligned}
\hat{T}_i &= D_i - L_{i-1}(2\beta_{i-1} - \beta_{i-1}\hat{T}_{i-1}\beta_{i-1})U_{i-1} + ch^q I_n \\
&= \begin{bmatrix} d & -\kappa_1 & & -\kappa_1 \\ -\kappa_1 & d & \ddots & \\ & \ddots & \ddots & -\kappa_1 \\ -\kappa_1 & & -\kappa_1 & d \end{bmatrix} + \frac{\kappa_2^2}{(\hat{d}-2\hat{\kappa}_1)^2} \begin{bmatrix} \hat{d} & -\hat{\kappa}_1 & & -\hat{\kappa}_1 \\ -\hat{\kappa}_1 & \hat{d} & \ddots & \\ & \ddots & \ddots & -\hat{\kappa}_1 \\ -\hat{\kappa}_1 & & -\hat{\kappa}_1 & \hat{d} \end{bmatrix} - \frac{2\kappa_2^2}{(\hat{d}-2\hat{\kappa}_1)} I_n + ch^q I_n \\
&= \begin{bmatrix} d - \frac{\kappa_2^2\hat{d} - 2\kappa_2^2(\hat{d}-2\hat{\kappa}_1)^2}{(\hat{d}-2\hat{\kappa}_1)} & -\kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} & & -\kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} \\ -\kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} & d - \frac{\kappa_2^2\hat{d} - 2\kappa_2^2(\hat{d}-2\hat{\kappa}_1)^2}{(\hat{d}-2\hat{\kappa}_1)} & \ddots & \\ & \ddots & \ddots & -\kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} \\ -\kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} & & \kappa_1 - \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2} & d - \frac{\kappa_2^2\hat{d} - 2\kappa_2^2(\hat{d}-2\hat{\kappa}_1)^2}{(\hat{d}-2\hat{\kappa}_1)} \end{bmatrix} + ch^q I_n. \end{aligned} \tag{13}$$

From the above relationship, we have

$$\hat{d} = d - \frac{\kappa_2^2\hat{d} - 2\kappa_2^2(\hat{d}-2\hat{\kappa}_1)^2}{(\hat{d}-2\hat{\kappa}_1)} + ch^q,$$

$$\hat{\kappa_1} = \kappa_1 + \frac{\kappa_2^2\hat{\kappa}_1}{(\hat{d}-2\hat{\kappa}_1)^2},$$

or

$$(\hat{d}-d)(\hat{d}-2\hat{\kappa}_1)^2 = \kappa_2^2(\hat{d}-2\hat{\kappa}_1) - 2\kappa_2^2(\hat{d}-2\hat{\kappa}_1) + 2\hat{\kappa}_1\kappa_2^2, \tag{14}$$

$$(\hat{\kappa_1}-\kappa_1)(\hat{d}-2\hat{\kappa}_1)^2 = \kappa_2^2\hat{\kappa}_1. \tag{15}$$

By using matlab symbolic computation [27], we have

$$\begin{aligned}
\hat{d} - 2\hat{\kappa}_1 &= -1 + \tfrac{1}{2}(d+ch^q) + \tfrac{1}{2}\sqrt{(d+ch)^2 - 4(d+ch^q)} \\
&= 1 + \tfrac{1}{2}ch^q + \tfrac{1}{2}\sqrt{(4+ch^q)ch^q} \\
&= 1 + \eta_h,
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\hat{\kappa}_1 &= \frac{-(d+ch^q) - \sqrt{-4(d+ch^q)+(d+ch^q)^2} + (-1+\frac{1}{2}(d+ch^q) + \frac{1}{2}\sqrt{-4(d+ch^q)+(d+ch^q)^2})(d+ch^q)}{(d+ch^q)(d+ch^q-4)} \\
&= \tfrac{1}{2} + \frac{\sqrt{(d+ch^q)ch^q}(\frac{1}{2}(d+ch^q)-1)}{ch^q(d+ch^q)} \\
&= \tfrac{1}{2} + \frac{\frac{1}{2}ch^q+1}{\sqrt{(4+ch^q)ch^q}} \\
&= \tfrac{1}{2} + \tfrac{1}{2\delta_h},
\end{aligned} \tag{17}$$

where $\eta_h = \frac{1}{2}ch^q + \frac{1}{2}\sqrt{(4+ch^q)ch^q}$, $\delta_h = \frac{\sqrt{(4+ch^q)ch^q}}{ch^q+2}$.

From (16) and (17) we can see that $\hat{\kappa}_1 \to \infty$ as $c \to 0$. Thus, Fourier analysis can not be performed on the original tangential frequency filtering decomposition preconditioner.

By straightforward computation as in (7), the Fourier eigenvalues of $L$, $U$, and $\hat{T}$ are

$$\lambda_L = -\kappa_2 e^{-\mathrm{i}\phi_k},$$

$$\lambda_U = -\kappa_2 e^{\mathrm{i}\phi_k},$$

$$\lambda_{\hat{T}} = \hat{d} - \hat{\kappa}_1 \cos(\theta_j),$$

respectively. Therefore, the Fourier eigenvalues of the MTFFD preconditioner $\hat{M}$ are

$$\begin{aligned} \lambda(\hat{M}) &= (\lambda_L + \lambda_{\hat{T}})\lambda_{\hat{T}}^{-1}(\lambda_U + \lambda_{\hat{T}}) \\ &= \frac{(\hat{d}-2\hat{\kappa}_1\cos(\theta_j)-\kappa_2\cos(\phi_k)+\mathtt{i}\kappa_2\sin(\phi_k))(\hat{d}-2\hat{\kappa}_1\cos(\theta_j)-\kappa_2\cos(\phi_k)-\mathtt{i}\kappa_2\sin(\phi_k))}{\hat{d}-2\hat{\kappa}_1\cos(\theta_j)}. \end{aligned} \tag{18}$$

Letting $\xi = \hat{d} - 2\hat{\kappa}_1\cos(\theta_j)$, we have

$$\begin{aligned} \lambda(\hat{M}) &= \frac{1}{\xi}(\xi - \kappa_2\cos(\phi_k) + \mathtt{i}\kappa\sin(\phi_k)(\xi - \kappa_2\cos(\phi_k) - \mathtt{i}\kappa_2\sin(\phi_k)) \\ &= \frac{1}{\xi}((\xi - \kappa_2\cos(\phi_k))^2 + \kappa_1^2\sin(\phi_k)^2) \\ &= \frac{1}{\xi}((\xi^2 - 2\xi\kappa_2\cos(\phi_k) + \kappa_2^2). \end{aligned} \tag{19}$$

As we know,

$$\lambda_{j,k}(A) = 4(\kappa_1\sin^2(\frac{\theta_j}{2}) + \kappa_2\sin^2(\frac{\phi_k}{2})).$$

Hence

$$\begin{aligned} \lambda(\hat{M}^{-1}A) &= \frac{4\xi(\kappa_1\sin^2(\frac{\theta_j}{2})+\kappa_2\sin^2(\frac{\phi_k}{2}))}{\xi^2-2\xi\kappa_2\cos(\phi_k)+\kappa_2^2} \\ &= \frac{4(\hat{d}-2\hat{\kappa}_1\cos(\theta_j))(\kappa_1\sin^2(\frac{\theta_j}{2})+\kappa_2\sin^2(\frac{\phi_k}{2}))}{(\hat{d}-2\hat{\kappa}_1\cos(\theta_j)-\kappa_2)^2+2\kappa_2(\hat{d}-2\hat{\kappa}_1\cos(\theta_j))(1-\cos(\phi_k))} \\ &= \frac{4(\hat{d}-2\hat{\kappa}_1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))(\kappa_1\sin^2(\frac{\theta_j}{2})+\kappa_2\sin^2(\frac{\phi_k}{2}))}{(\hat{d}-2\hat{\kappa}_1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})-\kappa_2)^2+4\kappa_2(\hat{d}-2\hat{\kappa}_1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))\sin^2(\frac{\phi_k}{2})} \\ &= \frac{4(\eta_h+1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))(\sin^2(\frac{\theta_j}{2})+\sin^2(\frac{\phi_k}{2}))}{(\eta_h+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))^2+4\kappa_2(\eta_h+1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))\sin^2(\frac{\phi_k}{2})}. \end{aligned} \tag{20}$$

Thus,

$$\begin{aligned} \lambda^{-1}(\hat{M}^{-1}A) &= \frac{(\eta_h+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))^2+4\kappa_2(\eta_h+1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))\sin^2(\frac{\phi_k}{2})}{4(\eta_h+1+4\hat{\kappa}_1\sin^2(\frac{\theta_j}{2}))(\sin^2(\frac{\theta_j}{2})+\sin^2(\frac{\phi_k}{2}))} \\ &= \frac{16\hat{\kappa}_1^2\sin^4(\frac{\theta_j}{2})+16\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})\sin^2(\frac{\phi_k}{2})+\eta_h^2+8\delta_h\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})+4(1+\eta_h)\sin^2(\frac{\phi_k}{2})}{16\hat{\kappa}_1\sin^4(\frac{\theta_j}{2})+16\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})\sin^2(\frac{\phi_k}{2})+4(1+\eta_h)(\sin^2(\frac{\theta_j}{2})+\sin^2(\frac{\phi_k}{2}))} \\ &= 1 + \frac{16(\hat{\kappa}_1-1)\hat{\kappa}_1\sin^4(\frac{\theta_j}{4})+\eta_h^2+8\hat{\kappa}_1\eta_h\sin^2(\frac{\theta_j}{2})-4(1+\eta_h)\sin^2(\frac{\theta_j}{2})}{16\hat{\kappa}_1\sin^4(\frac{\theta_j}{2})+16\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})\sin^2(\frac{\phi_k}{2})+4(1+\eta_h)(\sin^2(\frac{\theta_j}{2})+\sin^2(\frac{\phi_k}{2}))} \\ &= 1 + \frac{4(\delta_h^{-1}+1)(\delta_h^{-1}-1)\sin^2(\frac{\theta_j}{2})+\frac{\eta_h^2}{\sin^2(\frac{\theta_j}{2})}}{16\hat{\kappa}_1\sin^2(\frac{\theta_j}{2})+16\hat{\kappa}_1\sin^2(\frac{\phi_k}{2})+4(1+\eta_h)+4(1+\eta_h)\frac{\sin^2(\frac{\phi_k}{2})}{\sin^2(\frac{\theta_j}{2})}}. \end{aligned} \tag{21}$$

As $h$ tends to 0, we have $\delta_h^{-1} \geq 1$. Then from (21) it is easy to see that asymptotically

$$\lambda^{-1}(\hat{M}^{-1}A) \geq 1,$$

i.e. $\lambda(\hat{M}^{-1}A) \leq 1$. This is consistent with the theoretical results obtained in [2].

Subsequently, we will derive the upper bound of $\lambda^{-1}(\hat{M}^{-1}A)$ in an analytical way.

Let

$$f(s_1, s_2) = 1 + \frac{\alpha s_1^2 + \frac{\gamma}{s_1^2}}{\beta s_1^2 + \beta s_2^2 + e\frac{s_2^2}{s_1^2} + e}, \tag{22}$$

where we use $s_1 = \sin(\frac{\theta}{2})$, $s_2 = \sin(\frac{\phi}{2})$. It is easy to see that $f(s_1, s_2)$ is a continuous function of $\sin(\frac{\theta}{2})$ and $\sin(\frac{\phi}{2})$, with $(\frac{\theta}{2}, \frac{\phi}{2})$ defined in $(0, \pi) \times (0, \pi)$. In the representation form of $f(s_1, s_2)$, we have set $\alpha = 4(\delta_h^{-1} + 1)(\delta_h^{-1} - 1)$, $\gamma = \eta_h^2$, $\beta = 16\hat{\kappa}_1$, and $e = 4(1 + \eta_h)$.

Taking partial derivation of $f(s_1, s_2)$ with $s_2$, we have

$$f'_{s_2}(s_1, s_2) = \frac{-2(\alpha s_1^2 + \frac{\gamma}{s_1^2})(\beta + \frac{e}{s_1^2})s_2 c_2}{(\beta s_1^2 + \beta s_2^2 + e\frac{s_2^2}{s_1^2} + e)^2} \quad \begin{cases} \leq 0, & \phi \in (0, \frac{\pi}{2}), \\ > 0, & \phi \in (\frac{\pi}{2}, \pi), \end{cases} \tag{23}$$

where $c_2 = \cos(\frac{\phi}{2})$.

Thus,

$$\max_{\theta,\phi}(f(s_1, s_2)) = f(s_1, 0) = f(s_1, \pi) = 1 + \frac{\alpha s_1^2 + \frac{\gamma}{s_1^2}}{\beta s_1^2 + e}.$$

Therefore, the maximum value of $\lambda(\hat{M}^{-1}A)$ is attained on the line $(j, k)$ with $k = 1$, or $k = n$.

Also we have

$$\begin{aligned} f'_{s_1}(s_1, 0) &= \frac{4\alpha s_1^3 c_1(\beta s_1^4 + e s_1^2) - (\alpha s_1^4 + \gamma)(4\beta s_1^3 c_1 + 2e s_1 c_1)}{(\beta s_1^4 + e s_1^2)^2} \\ &= \frac{2\alpha e s_1^4 - \gamma(4\beta s_1^2 + 2e)}{(\beta s_1^4 + e s_1^2)^2} s_1 c_1 \\ &= \frac{32(\delta_h^{-2} - 1)(1 + \eta_h)s_1^4 - \frac{32\eta_h^2(1+\delta_h)s_1^2}{\delta_h} - 8\eta_h^2(1+\eta_h)}{(\beta s_1^4 + e s_1^2)^2} s_1 c_1 \\ &\approx \frac{s_1 c_1}{(\beta s_1^4 + e s_1^2)^2}(\frac{32 s_1^4}{\delta_h^2} - \frac{32\eta_h^2 s_1^2}{\delta_h}) \\ &\approx \frac{s_1 c_1}{(\beta s_1^4 + e s_1^2)^2}(\frac{32 s_1^4}{c h^q} - 32\sqrt{c} h^{\frac{q}{2}} s_1^2). \end{aligned} \tag{24}$$

In the above approximation, the high-order terms are ignored as $h$ is assumed to be sufficiently small. Subsequently, we will analyze the sign of $f'_{s_1}(s_1, 0)$ in two cases:

- When $q \geq \frac{4}{3}$, then as $h \to 0$, we have

$$f'_{s_1}(s_1, 0) \quad is \quad \begin{cases} \geq 0, & \theta \in (0, \frac{\pi}{2}], \\ < 0, & \theta \in (\frac{\pi}{2}, \pi), \end{cases} \tag{25}$$

Therefore, the maximum value of $\lambda^{-1}(\tilde{M}^{-1}A)$ is attained whenever $j = \lfloor \frac{n}{2} \rfloor + 1$, and $k = 1$ or $k = n$, where $\lfloor \frac{n}{2} \rfloor$ denotes the largest integer less than $\frac{n}{2}$. At these points $(j, k)$, we have

$$\begin{aligned} \lambda^{-1}_{\lfloor \frac{n}{2} \rfloor + 1, k}(\tilde{M}^{-1}A) &\approx 1 + \frac{\frac{4}{ch^q} + ch^q - 4}{\frac{8}{\sqrt{ch^q} + 4}} \\ &\approx 1 + \frac{1}{2\sqrt{c} h^{\frac{q}{2}}} \\ &\geq \frac{1}{2\sqrt{c} h^{\frac{2}{3}}}. \end{aligned} \tag{26}$$

- When $0 \leq q \leq \frac{4}{3}$, then as $h \to 0$, we have

$$f'_{s_1}(s_1, 0) \quad is \quad \begin{cases} \leq 0, & \theta \in (0, \xi_h], \\ > 0, & \theta \in (\xi_h, \frac{\pi}{2}), \\ \leq 0, & \theta \in [\frac{\pi}{2}, \pi - \xi_h], \\ > 0, & \theta \in (\pi - \xi_h, \pi), \end{cases} \tag{27}$$

where $\xi_h$ the positive angle such $\frac{32s_1^4}{ch^q} = 32\sqrt{c}h^{\frac{q}{2}}s_1^2$, i.e. $s_1^2 = c^{\frac{3}{2}}h^{\frac{3q}{2}}$. From the equality, we have $\xi_h = \arcsin(c^{\frac{3}{4}}h^{\frac{3q}{4}})$. Therefore, in this case, the maximum value of $\lambda^{-1}(\hat{M}^{-1}A)$ is possibly attained at one of the following three points $(j,k) = (1,k)$, $(j,k) = (n,k)$, or $(j,k) = (\lfloor \frac{n}{2} \rfloor + 1, k)$, with $k = 1$ or $k = n$. At the first two points, we have

$$\begin{aligned} \lambda_{1,k}^{-1}(\hat{M}^{-1}A) = \lambda_{n,k}^{-1}(\hat{M}^{-1}A) &\approx 1 + \frac{ch_p^q}{8\pi^2 h^2} \\ &\approx 1 + \frac{c}{8\pi^2 h^{2-q}} \\ &\geq \frac{c}{8\pi^2 h^{\frac{2}{3}}}. \end{aligned} \tag{28}$$

At the third point, we have

$$\begin{aligned} \lambda_{\lfloor \frac{n}{2} \rfloor + 1,k}^{-1}(\tilde{M}^{-1}A) &\approx 1 + \frac{\frac{4}{ch^q} + ch^q - 4}{\frac{8}{\sqrt{ch^q}+4}} \\ &\approx 1 + \frac{1}{2\sqrt{c}h^{\frac{q}{2}}} \\ &\leq \frac{1}{2\sqrt{c}h^{\frac{2}{3}}}. \end{aligned} \tag{29}$$

Therefore, in this case the maximum value of $\lambda^{-1}(\hat{M}^{-1}A)$ is attained at $(j,k) = (1,1)$, or $(j,k) = (n,n)$, i.e. the value shown by (28).

As we have shown above, the maximum eigenvalue of the preconditioned matrix is approximately equal to 1. Therefore, the condition number of the preconditioned matrix is approximately given by

$$\begin{aligned} \kappa(\hat{M}^{-1}A) &= \frac{\max_{j,k} \lambda_{j,k}(\hat{M}^{-1}A)}{\min_{j,k} \lambda_{j,k}(\hat{M}^{-1}A)} \\ &\approx \frac{\lambda_{max}^{-1}(\hat{M}^{-1}A)}{1} \end{aligned}$$

where $\lambda_{max}^{-1}(\hat{M}^{-1}A)$ denotes the maximum value of $\lambda^{-1}(\hat{M}^{-1}A)$. For fixed $c$, from the above analysis we can see that the optimal $q$ that minimizes the condition number is attained at $\frac{4}{3}$.

Define

$$\begin{cases} g_1(c) = 1 + \frac{1}{2\sqrt{c}h^{\frac{2}{3}}}, \\ g_2(c) = 1 + \frac{c}{8\pi^2 h^{\frac{2}{3}}}. \end{cases}$$

Then the optimal $c$ can be determined by solving

$$\min_c \max\{g_1(c), g_2(c)\}. \tag{30}$$

This min-max problem can be solved by their plot. In Figure 1, we give the curve of function $g_1(c)$ and $g_2(c)$ when $h^{\frac{2}{3}} = 0.1$. From the figure we can see that (30) is solved whenever

$$g_1(c) = g_2(c).$$

From this equation, we can get the optimal $c = (4\pi^2)^{\frac{2}{3}}$.

Suppose $q$ is chosen as $\frac{4}{3}$, then the condition estimate of the preconditioned linear system is

$$\begin{aligned} \kappa(\hat{M}^{-1}A) &= \frac{\max_{j,k} \lambda_{j,k}(\hat{M}^{-1}A)}{\min_{j,k} \lambda_{j,k}(\hat{M}^{-1}A)} \\ &\approx \frac{\lambda_{max}(\hat{M}^{-1}A)^{-1}}{1} \\ &\leq 1 + \max\{\frac{1}{2\sqrt{c}h^{\frac{2}{3}}}, \frac{c}{8\pi^2 h^{\frac{2}{3}}}\} \\ &= \mathcal{O}(h^{-\frac{2}{3}}). \end{aligned}$$

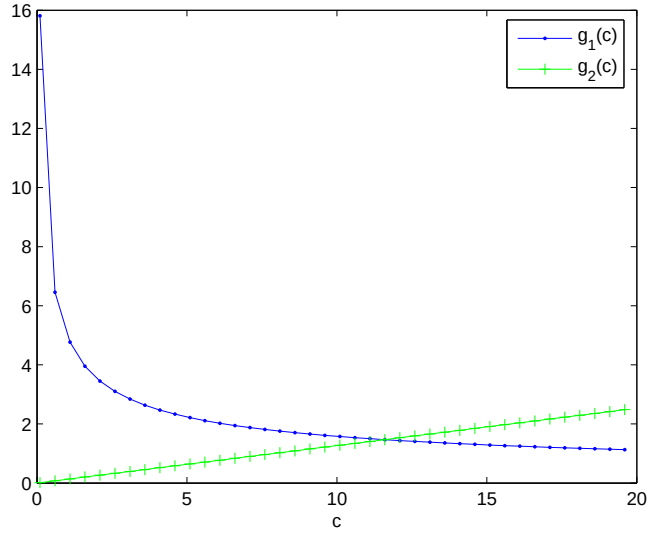


Figure 1: The curves of function $g_1(c)$ and $g_2(c)$.

We remark that the analysis above is for a periodic problem. If subscripts of $c_d$ and $c_p$ are used to distinguish the parameters for Dirichlet and periodic problems, then the optimal $c_p = (4\pi^2)^{\frac{2}{3}}$. From the mesh size relationship of Dirichlet and periodic problems, we have $h_p = \frac{1}{2}h_d$. Hence, For $q = \frac{4}{3}$, the modification should satisfy the relationship [17]

$$c_p h_p^{\frac{4}{3}} = c_d h_d^{\frac{4}{3}},$$

from where we can get the optimal $c_d$, i.e. $c_d = (\frac{1}{2})^{\frac{4}{3}}(4\pi^2)^{\frac{2}{3}} \approx 4.6012$.

The above analysis can be concluded by the following theorem

**Theorem 1** *For the MTFFD preconditioner with $\Lambda_i$ be an identity matrix, the optimal choice of modification order is $q = \frac{4}{3}$, the optimal relaxation parameter is $c_p = (4\pi^2)^{\frac{2}{3}}$. For $q = \frac{4}{3}$ and fixed c, then asymptotically $(h \to 0)$ the eigenvalues of MTFFD preconditioned matrix $\hat{M}^{-1}A$ are always less than 1, and the condition number of $\hat{M}^{-1}A$ is $\mathcal{O}(h^{-\frac{2}{3}})$.*

**Remarks:** By using the semi-discrete analysis, Y. Achdou and F. Nataf (cf. Reference [1]) obtain an optimal filtering vector that minimizes the condition number of preconditioned matrix by the tangential frequency filtering preconditioner [38]. In this paper, we choose **1** as the filtering vector, but modify the recursion formula of tangential tangential filtering decomposition proposed in [2]. The optimal condition numbers obtained in both papers have the same order. In Reference [13], the same order of the condition number is obtained by using optimized two-frequency filtering decomposition. However, it is not tangential filtering decomposition.

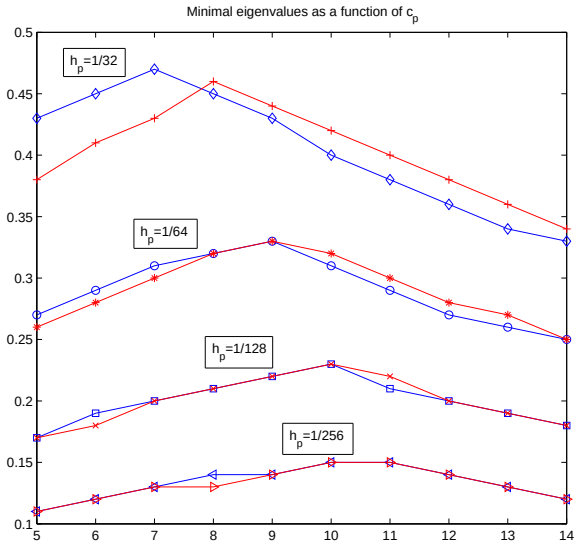


Figure 2: The dependence of minimum eigenvalues on parameter $c_p$.

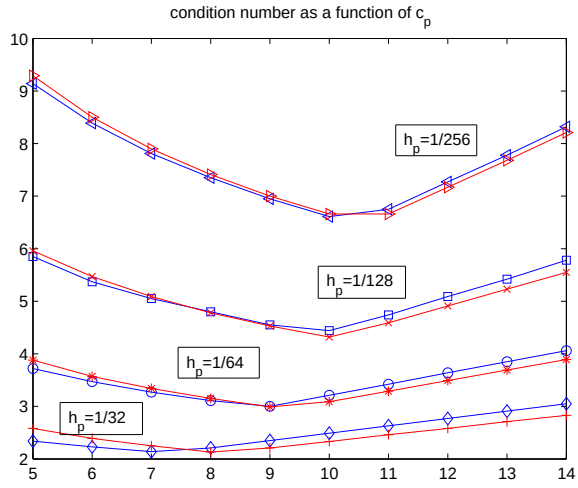


Figure 3: The dependence of condition numbers on parameter $c_p$

Table 1: Dirichlet and periodic results for $c_d = 2.5$.

| $c_d = 2.5$ | $\lambda_{max}$ | | $\lambda_{min}$ | | condition number | |
|---|---|---|---|---|---|---|
| $\frac{1}{h_d}$ | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* |
| 8 | 1.00 | 0.97 | 0.64 | 0.54 | 1.55 | 1.77 |
| 16 | 1.00 | 0.99 | 0.43 | 0.38 | 2.34 | 2.58 |
| 32 | 1.00 | 1.00 | 0.27 | 0.26 | 3.72 | 3.88 |
| 64 | 1.00 | 1.00 | 0.17 | 0.17 | 5.83 | 5.96 |
| 128 | 1.00 | 1.00 | 0.11 | 0.11 | 9.14 | 9.28 |
| 256 | 1.00 | 1.00 | 0.07 | 0.07 | 14.37 | 14.58 |

Table 2: Dirichlet and periodic results for $c_d = 5$.

| $c_d = 5$ | $\lambda_{max}$ | | $\lambda_{min}$ | | condition number | |
|---|---|---|---|---|---|---|
| $\frac{1}{h_d}$ | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* |
| 8 | 1.00 | 0.94 | 0.49 | 0.51 | 2.03 | 1.84 |
| 16 | 1.00 | 0.98 | 0.40 | 0.42 | 2.49 | 2.33 |
| 32 | 1.00 | 0.99 | 0.31 | 0.32 | 3.21 | 3.09 |
| 64 | 1.00 | 1.00 | 0.23 | 0.23 | 4.40 | 4.32 |
| 128 | 1.00 | 1.00 | 0.15 | 0.15 | 6.61 | 6.66 |
| 256 | 1.00 | 1.00 | 0.097 | 0.096 | 10.32 | 10.39 |

Table 3: Dirichlet and periodic results for $c_d = 7.5$.

| $c_d = 7.5$ | $\lambda_{max}$ | | $\lambda_{min}$ | | condition number | |
|---|---|---|---|---|---|---|
| $\frac{1}{h_d}$ | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* | *Dirichlet* | *Periodic* |
| 8 | 1.00 | 0.96 | 0.40 | 0.42 | 2.53 | 2.20 |
| 16 | 1.00 | 0.96 | 0.31 | 0.33 | 3.20 | 2.95 |
| 32 | 1.00 | 0.99 | 0.23 | 0.24 | 4.28 | 4.09 |
| 64 | 1.00 | 0.99 | 0.17 | 0.17 | 6.02 | 5.88 |
| 128 | 1.00 | 1.00 | 0.11 | 0.11 | 8.83 | 8.69 |
| 256 | 1.00 | 1.00 | 0.075 | 0.076 | 13.27 | 13.21 |

In Figures 2 -3, the minimum eigenvalues and the condition numbers are plotted as a function of $c$, with $\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}$ (corresponding to $h_p = \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}$). As the maximum eigenvalues are both close to 1 and their plots are not easy to distinguish, so we don't display them. From the figures, it is easy to see that the minimal eigenvalue (and hence the condition numbers) are quite similar. From Figure 3, we can see that the experimental optimal $c_p$ is a little smaller than the theoretical asymptotical optimal value. This is possibly because the mesh size is not sufficiently refined. However, as $h_p \to 0$, the experimental optimal value $c$ indeed tends to $(4\pi^2)^{\frac{2}{3}}$.

As we have shown above, the experimental optimal parameter $c_p$ is slightly smaller than the asymptotically optimal value $(4\pi^2)^{\frac{2}{3}} = 11.59$. In the following test, we compare the numerical results with three different parameter $c_p$ and various mesh sizes. Three parameters $c_p = 5$, $c_p = 10$, and $c_p = 15$ are chosen;

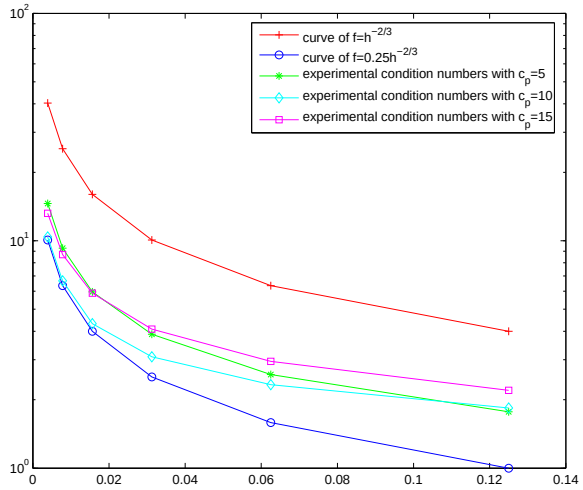


Figure 4: The dependence of condition numbers on $h_p$.

in the sense that $c_p = 5$ is less than the optimal value; $c_p = 10$ is close to the optimal value; and $c_p = 15$ is larger than the optimal value. The test results are shown in Table 1 - 3, where we use *Dirichlet* and *Periodic* to denote the results for Dirichlet case and periodic case, respectively. In order to approximate the extremal eigenvalues of the preconditioned Dirichlet system, we use restarted harmonic Arnoldi method [31] when mesh size $h_d \leq \frac{1}{64}$. The computed approximate eigenpairs $(\lambda_i, \hat{\varphi}_i)$ satisfy $||A\hat{\varphi}_i - \hat{\lambda}_i\hat{\varphi}_i|| < 10^{-2}$. From the three tables, we can see that the periodic values are very close to the Dirichlet values. The condition number of the preconditioned Dirichlet system can be captured by the periodic results. By comparing Table 1, Table 3 with Table 2 respectively, we can see that $c_d = 5$ produces the best condition number as $h_d$ is refined. The results are consistent with the theoretical results.

To illustrate that the condition number of $\hat{M}^{-1}A$ is $\mathcal{O}(h^{-\frac{2}{3}})$, we display in Figure 4 the experimental periodical condition numbers of Tables 1- 3. The x-axis denotes the values of $h_p$; the y-axis is the logarithmic scale of the experimental condition numbers and the function values of $h_p^{-\frac{2}{3}}$ and $\frac{1}{4}h_p^{-\frac{2}{3}}$. The results at the points $h_p = \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}$ are plotted. From the Figures we can see that the periodical experimental condition numbers depend linearly on $h_p^{-\frac{2}{3}}$. As $h_p$ tends to zero, the plot of experimental results with $c_p = 10$ (c.f. Table 2) becomes very close to the curve of function $\frac{1}{4}h_p^{-\frac{2}{3}}$.

To compare the range and clustering of the Fourier eigenvalues with that of preconditioned Dirichlet system, we display the spectrum distributions when $c_p = 5$ and $c_p = 10$, see Figures 5 and 6 respectively. The test results of mesh size $h_d = \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$, (corresponding to $h_p = \frac{1}{16}, \frac{1}{32}, \frac{1}{64}$) are plotted. From the Figures, we see that the range and clustering of the preconditioned Dirichlet system and the Fourier eigenvalue distribution are extremely close. As mesh

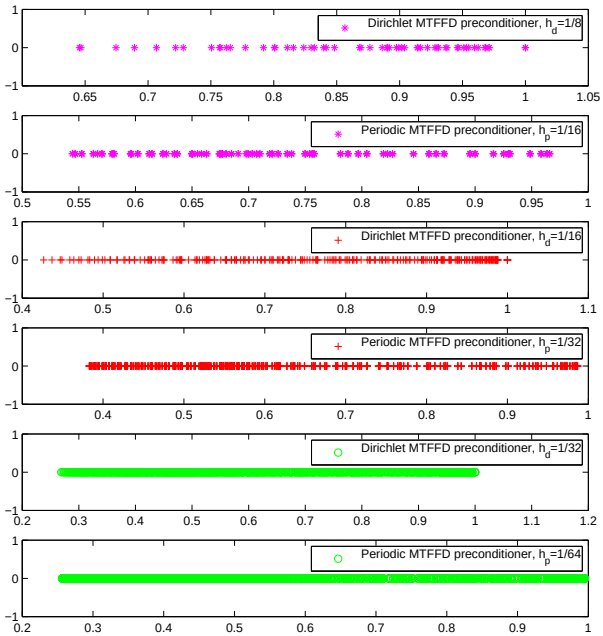


Figure 5: Spectrum distribution of the preconditioned matrices, with $c_p = 5$.

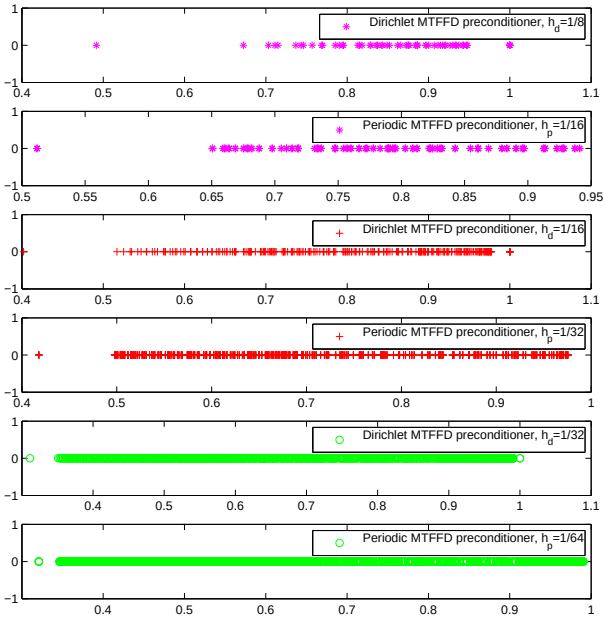


Figure 6: Spectrum distribution of the preconditioned matrices, with $c_p = 10$.

size $h$ decreases, the extremal eigenvalues (and hence the condition number) of both cases become closer.

# 4 Numerical Examples

The performance of the MTFFD preconditioner, the TFFD preconditioner [2], and the ILU(0) [36] preconditioner are compared on several problems arising from the discretization of partial differential equations. All the tests are run on an Intel Pentium IV Dual-Core with main memory 1G and the machine precision $eps = 2.22 \times 10^{-16}$ using Matlab 7.5 on a Linux-based system.

In the tests, we stop the algorithm when the relative norm $\frac{||b-Ax_k||}{||b||}$ is less than $10^{-12}$. Both the exact solution and the initial approximate solutions are chosen randomly. In the following discussions, the restarted GMRES [36] is used with maximum subspace dimension 200. The filtering vector is always chosen as $\mathbf{1} = [1, \ldots, 1]^T$. In the following tables, *iter* denotes the number of iterations, *error* denotes the infinite norm of the difference between the final approximate solution and the exact solution. We use "†" to denote that the method fails to converge within 200 iteration steps.

**Example 1.** We consider the boundary value problem as in [2]

$$\begin{aligned} \eta(x)u + div(\mathbf{a}(x)u) - div(\kappa(x)\nabla u) = f \ \ in \ \ \Omega \\ u = 0 \ \ on \ \ \partial\Omega_D \\ \tfrac{\partial u}{\partial n} = 0 \ \ on \ \ \partial\Omega_N \end{aligned} \tag{31}$$

where $\Omega = [0,1]^k$ with $k = 2$, $\partial\Omega_N = \partial\Omega \setminus \partial\Omega_D$, $\partial\Omega_D = [0,1] \times \{0,1\}$.

We consider the following five different cases. As these problems are no longer constant coefficient, we choose the additional term as $c\Lambda_i h^{\frac{4}{3}}$, where $\Lambda_i = diag(D_i)$, i.e. the diagonal matrix of the $i$th diagonal block of the coefficient matrix.

`Case I:` *The advection-diffusion problem with a rotating velocity in two dimensions:*
The tensor $\kappa$ is the identity, and the velocity is $\mathbf{a}(x) = (2\pi(x_2 - 0.5), 2\pi(x_1 - 0.5))^T$. The function $\eta(x)$ is zero. The uniform grid with $n \times n$ nodes, $n = 100, 200, 300, 400$ nodes are tested respectively. The diagonal elements of $A$ are close to 4. We set parameter $c$ to be 2.5 in the numerical test. Table 4 displays the results obtained by using three different preconditioners.

Table 4: Test results for advection-diffusion problems, nonsymmetric matrix

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| $1/h$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| 100 | 26 | 2.4e-13 | 57 | 8.3e-12 | 108 | 7.3e-10 |
| 200 | 32 | 5.6e-9 | 82 | 1.3e-11 | 191 | 3.2e-9 |
| 300 | 37 | 9.7e-10 | 101 | 1.6e-11 | † | 2.4e-6 |
| 400 | 40 | 8.1e-12 | 117 | 1.4e-11 | † | 2.0e-5 |

Table 5: Test results for non-Homogeneous problems, symmetric matrix

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| $1/h$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| 100 | 26 | 8.7e-13 | 57 | 7.5e-12 | 108 | 3.0e-10 |
| 200 | 32 | 3.0e-12 | 82 | 9.2e-12 | 186 | 4.3e-9 |
| 300 | 37 | 5.8e-10 | 101 | 1.4e-11 | † | 2.6e-6 |
| 400 | 41 | 1.1e-12 | 116 | 1.7e-11 | † | 4.9e-5 |

`Case II:` *Non-Homogenous problems with large jumps in the coefficients in two dimensions:*

The coefficient $\eta(x)$ and $\mathbf{a}(x)$ are both zero. The tensor $\kappa$ is isotropic and discontinuous. It jumps from the constant value $10^3$ in the ring $\frac{1}{2\sqrt{2}} \leq |x - \mathbf{c}| \leq \frac{1}{2}$, $\mathbf{c} = (\frac{1}{2}, \frac{1}{2})^T$, to 1 outside. We tested uniform grids with $n \times n$ nodes, $n = 100, 200, 300, 400$. The choice of the parameter $c$ is the same with `Case I`. Table 5 displays the results obtained by using three different preconditioners. The results are quite similar to the advection-diffusion problem.

From Table 4 - 5, we can see that MTFFD is much more efficient; it only needs less than half of the iteration numbers that TFFD needs.

`Case III:` *Skyscraper problems:*
The tensor $\kappa$ is isotropic and discontinuous. The domain contains many zones of high permeability which are isolated from each other. Let $[x]$ denote the integer value of $x$. In 2D, we have

$$\kappa(x) = \begin{cases} 10^3 * ([10 * x_2] + 1), & if \ \ [10 * x_i] = 0 \ mod(2) \ , \ i = 1, 2, \\ 1, & otherwise. \end{cases}$$

The diagonal elements of $A$ jump between 4 and 36000. The parameter $c$ is chosen as 10 in the test. The numerical results are shown in table 6.

`Case IV:` *Convective skyscraper problems:*
The same with the Skyscraper problems except that the velocity field is changed to be $\mathbf{a} = (1000, 1000, 1000)^T$. The diagonal elements of $A$ jump between 24 and 36020. The parameter is chosen as 1. The tested results are displayed in Table 7.

From Table 6 - 7 we can see that Skyscraper and Convective skyscraper problems are quite difficult. The TFFD and ILU(0) preconditioned GMRES fail to converge for both problems. The MTFFD preconditioned GMRES has much better performance. For Skyscraper problem with $h = \frac{1}{400}$, we note that MTFFD needs 222 iterations to converge.

`Case V:` *Anisotropic layers:*
The domain is made of 10 anisotropic layers with jumps of up to four orders of magnitude and an anisotropy ratio of up to $10^3$ in each layer. The diagonal elements jump between 22 and 220000. The parameter $c$ is chosen as $\frac{2}{5}$. The test results are displayed in Table 8. From the table we can see that MTFFD preconditioner is much more efficient; as $h$ decreases, it needs only $\frac{1}{3}$ of the number of iterations that TFFD preconditioner needs.

Table 6: Test results for skyscraper problems, nonsymmetric matrix

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| $1/h$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| 100 | 151 | 8.8e-7 | † | 1.3e-1 | † | 1.1e-3 |
| 200 | 185 | 2.9e-6 | † | 2.6e-1 | † | 6.6e-3 |
| 300 | 159 | 6.2e-6 | † | 3.9e-1 | † | 8.6e-3 |
| 400 | † | 3.6e-4 | † | 4.8e-1 | † | 3.6e-2 |

In Figure 7, the eigenvalue distributions of the preconditioned matrices by TFFD and MTFFD preconditioners are displayed. The test matrices are generated from discretization of (31) with the above five different conditions and mesh size $h = \frac{1}{50}$. From the figures we can see that the MTFFD preconditioner can improve the eigenvalue distributions considerably. Particularly, the smallest eigenvalues are shifted in the positive direction, which makes the smallest eigenvalues to be well separated from the origin. The largest eigenvalues are remain very close to 1.

Table 7: Test results for convective skyscraper problems, nonsymmetric matrix

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| $1/h$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| 100 | 66 | 1.1e-8 | † | 1.0e-4 | 173 | 3.7e-8 |
| 200 | 94 | 9.2e-8 | † | 4.3e-2 | † | 1.1e-3 |
| 300 | 82 | 5.0e-8 | † | 8.4e-2 | † | 8.6e-3 |
| 400 | 133 | 4.1e-8 | † | 2.9e-1 | † | 4.5e-2 |

Table 8: Test results for anisotropic layers problems, nonsymmetric matrix

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| $1/h$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| 100 | 29 | 3.1e-11 | 68 | 1.4e-8 | 190 | 3.6e-7 |
| 200 | 36 | 6.8e-7 | 97 | 1.8e-8 | † | 2.1e-4 |
| 300 | 40 | 2.3e-6 | 120 | 1.2e-8 | † | 2.8e-4 |
| 400 | 42 | 5.0e-6 | 139 | 2.6e-8 | † | 6.5e-3 |

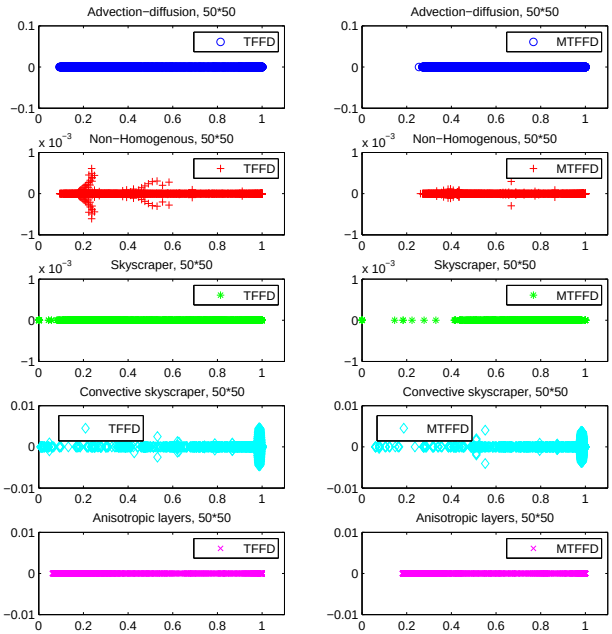


Figure 7: Spectrum distribution the preconditioned matrices.

**Example 2.** We consider the constant-coefficient convection diffusion equation

$$\begin{array}{ll} -\Delta u + 2p_1 u_x + 2p_2 u_y - p_3 u = f & in \quad [0,1]^2, \\ u = g & on \quad \partial[0,1]^2, \end{array} \tag{32}$$

where $p_1$, $p_2$ and $p_3$ are positive constants. Discretization of the equation by the standard second order 5 -point stencil on a uniform $n \times n$ mesh gives rise to a sparse linear system

$$Ax = b$$

where

$$A = Btrid_n(-(\beta+1)I, T, (\beta+1)I)$$

and

$$T = trid_n(-\gamma - 1, 4 - \sigma, \gamma - 1),$$

with $\beta = p_1 h_d$, $\gamma = p_2 h_d$ and $h_d = \frac{1}{n+1}$. The matrices series cdde1-cdde6 are based on the above equation with different parameters. We have tested all of the matrices, and the results are shown in Table 9. In the tests, the parameter $c_d$ is set to be 8 for cdde3 and cdde5, and 1 for other matrices.

Table 9: Test results for the cdde series matrices, nonsymmetric matrices

| | MTFFD | | TFFD | | ILU(0) | |
|---|---|---|---|---|---|---|
| matrix$(p_1, p_2, p_3)$ | iter | *error* | *iter* | *error* | *iter* | *error* |
| cdde1 (1,2,30) | 32 | 6.4e-11 | 197 | 3.4e-11 | 50 | 5.0e-11 |
| cdde2 (25,50,30) | 10 | 2.6e-11 | 10 | 7.3e-12 | 15 | 2.3e-12 |
| cdde3 (1,2,80) | 42 | 5.9e-10 | † | 3.4e-2 | 62 | 3.0e-10 |
| cdde4 (25,50,80) | 10 | 2.4e-11 | 10 | 3.0e-12 | 18 | 7.2e-12 |
| cdde5 (1,2,250) | 68 | 1.5e-10 | † | 1.8e-1 | 96 | 5.6e-10 |
| cdde6 (25,50,250) | 12 | 4.2e-12 | 11 | 1.2e-11 | 19 | 5.0e-12 |

From Table 9 we can see that the MTFFD preconditioner produces nearly the same results as that of the TFFD preconditioner for cdde2, cdde4 and cdde6. For the relatively difficult problems cdde1, cdde3 and cdde5, we can see that the MTFFD preconditioner is more efficient.

# 5 Conclusions

A modified tangential frequency filtering preconditioner is proposed and analyzed in this paper. The optimal order of modification and the optimal parameter are determined by the Fourier analysis. With the optimal order of modification, the results show that the preconditioned matrix has the condition number $\mathcal{O}(h^{-\frac{2}{3}})$, which is much better than the BILU and MBILU precodnitioner. All the theoretical results are illustrated by the numerical tests. Finally, the performance of the new preconditioner is examined by some problems arising from discretization of PDEs with discontinuous coefficient. With the optimal order of modification, the major inconvenience of the present preconditioner is the choice of the relaxation parameter $c$, whose value is problem dependent. For future work, it may be worthwhile to investigate the idea of dynamically relaxed methods [25, 26, 32]. This would hopefully further improve the robustness of the current preconditioner.

## References

[1] Y. Achdou and F. Nataf, *An iterated tangential filtering decomposition*, Numer. Linear Algebra Appl., 10, (2003), pp.511-539.

[2] Y. Achdou, F. Nataf, *Low frequency tangential filtering decomposition*, Numer. Linear Algebra Appl., 14, (2007), pp.129-147

[3] J. R. Appleyard and I. M. Cheshire, *Nested Factorization*, SPE 12264, presented at the Seventh SPE Symposium on Reservoir Simulation, San Francisco, 1983.

[4] O. Axelsson, *A generalized SSOR method*, BIT., 13, (1972), pp. 443-467.

[5] O. Axelsson and L. Kolotilina, *Diagonally compensated reduction and related preconditioning methods*, Numer. Linear Algebra Appl.,1, (1994), 155-177.

[6] O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems*, Theory and Computation., New York: Academic Press, 1984.

[7] O. Axelsson and H. Lu, *On the eigenvalue estimates for block incomplete factorization methods*, SIAM. J. Matrix Anal. Appl., 16, (1995), pp. 1074-1085.

[8] O. Axelsson and G. Lindskog, *On the eigenvalue distribution of a class of preconditioning methods,* Numer. Math., 48, (1986), pp. 479-498.

[9] O. Axelsson, *Iterative solution methods*, Cambridge University Press, New York, 1994.

[10] T. Boonen, J. Van Lent and S. Vandewalle, *Local Fourier analysis of multigrid for the curl curl equation*, SIAM J. Sci. Comput., 30, (2008), pp.1730-1755.

[11] A. Brandt, *Rigorous local model analysis of Multigrid*, Math. Comp., 31, (1977), pp. 333-390.

[12] A. Buzdin, *Tangential decomposition*, Computing., 61, (1998), pp.257-276.

[13] A. Buzdin and G. Wittum, *Two-frequency decomposition,* Numer. Math., 97, (2004), pp.269-295.

[14] A. Buzdin, D. Logashenko, and G. Wittum, *IBLU decompositions based on Pade approximants*, Numer. Linear Algebra Appl.,15, (2008), pp.717-746.

[15] T. F. Chan and H. C. Elman, *Fourier analysis of iterative methods for elliptic problems*, SIAM Review., 31, (1989), pp.20-49.

[16] T. F. Chan, *Fourier analysis of relaxed incomplete factorization preconditioners*, SIAM J. Sci. Comput., 12, (1991), pp.668-690.

[17] T. F. Chan and J. M. Donato, *Fourier analysis of incomplete factorization preconditioners for three-dimensional anisotropic problems*, SIAM J. Sci. Stat. Comput., 13, (1992), pp.319-338.

[18] I. Chihiro, *Fast solver for large systems of linear equations for finite element analysis on unstructured meshes*, Ph.D thesis, Swinburne University of Technology, 2004.

[19] P. Concus, G. H. Golub and G. Meurant, *Block Preconditioning for the Conjugate Gradient method*, SIAM J. Sci. Statist. Comput., 6, (1985), pp.220-252.

[20] I. Gustafsson, *A class of first order factorization methods*, BIT., 18, (1978), pp.142-156.

[21] W. Hackbusch, *Iterative solution of large sparse systems of equations*, Springer, New York, 1994.

[22] B. N. Khoromskij, G. E. Mazurkevich and G. Wittum, *Frequency filtering for elliptic interface probems with lagrange multipliers*, SIAM J. Sci Comput., 21, (1999), pp. 421-440.

[23] R J. Le Veque and L. N. Trefethen, *Fourier anlaysis of the SOR iterations*, IMA J. Numer. Anal., 8, (1988), pp.273-279.

[24] H. Lu, *Matrix compensation and diagonal compensation*, J. Comput. Math. Appl., 63, (1995), 237-244.

[25] M. Magolu Monga-Made, *Taking Advantage of the potentialities of dynamically modified block incomplete factorizations*, SIAM J. Sci Comput., 19, (1998), pp. 1083-1108.

[26] M. Magolu Monga-Made, *Dynamically relaxed block incomplete factorizations for solving two- and three-dimensional problems*, SIAM J. Sci Comput., 21, (2000), pp. 2008-2028.

[27] The MathWorks, Inc. *MATLAB 7*, September 2004.

[28] J. A. Meijerink and H. A. van der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is symmetric M-matrix*, Math. Comput., 137, (1977), pp.148-162.

[29] C. Mense and R. Nabben, *On algebraic multilevel methods for non-symmetric systems - comparison results* Linear Algebra Appl. (2008), doi:10.1016/j.laa.2008.04.045.

[30] C. Mense and R. Nabben, *On algebraic multilevel methods for non-symmetric systems - convergence results*, to appear in ETNA.

[31] R.B. Morgan and M. Zeng, *Harmonic Projection Methods for Large Non-symmetric Eigenvalue Problems*, Numer. Linear Algebra Appl., 5, (1998) pp. 33-55.

[32] Y. Notay, *DRIC: A dynamic version of the RIC method*, Numer. Linear Algebra Appl.,1, (1994), 511-532.

[33] Y. Notay, *Algebraic multigrid and algebraic multilevel methods: a theoretical comparison*, Numer. Linear Algebra Appl. 2005; 12:419-451.

[34] K. Otto, *Analysis of preconditioners for hyperbolic partial differential equations*, SIAM J. Numer. Anal. 33, (1996), pp.2131-2165.

[35] J.W. Ruge and K. Stüben, *Algebraic Multigrid* (*AMG*), In Multigrid Methods, Frontiers in Applied Mathematics, Vol 3, SIAM: Philadephia, PA, 1987, 73-130.

[36] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company: Boston, MA, 1996.

[37] P. S. Vassilevski, *Multilevel Block Factorization Preconditioners: Matrix-based Analysis and Algorithms for Solving Finite Element Equations*, Springer, 2008.

[38] C. Wagner, *Tangential frequency filtering decompositions for symmetric matrices*, Numer. Math., 78, (1997), pp.119-142.

[39] C. Wagner, *Tangential frequency filtering decompositions for unsymmetric matrices* Numer. Math., 78, (1997), pp.143-163.

[40] C. Wagner and G. Wittum, *Adaptive filtering,* Numer. Math., 78, (1997), pp.305-382.

[41] R. Wienands, C. W. Oosterlee and T. Washio, *Fourier analysis of GMRES(M) preconditioned by multigrid*, SIAM J. Sci. Comput., 22, (2000), pp.582-603.

[42] G. Wittum, *Filternde Zerlegungen*, Schnelle Löser für große Gleichungssysteme. Teubner Skripten zur Numerik, Band 1, Teubner-Verlag, Stuttgart, 1992.

[43] G. Wittum, *Shifted iterations*, Numer. Math. 76, (1997), pp.265-278.

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399