



**HAL**  
open science

# Hermitian Preconditioning for a class of Non-Hermitian Linear Systems

Nicole Spillane

► **To cite this version:**

Nicole Spillane. Hermitian Preconditioning for a class of Non-Hermitian Linear Systems. 2023. hal-04028590v1

**HAL Id: hal-04028590**

**<https://hal.science/hal-04028590v1>**

Preprint submitted on 14 Mar 2023 (v1), last revised 27 Oct 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Hermitian Preconditioning for a class of Non-Hermitian Linear Systems

Nicole Spillane \*

March 14, 2023

**Keywords:** GMRES, preconditioning, convergence, Krylov subspace method, GCR, Minimal residual iteration

**MSCodes:** 65F10, 65Y05, 68W40

## Abstract

This work considers the convergence of GMRES for non-singular problems. GMRES is interpreted as the GCR method which allows for simple proofs of the convergence estimates. Preconditioning and weighted norms within GMRES are considered. The objective is to provide a way of choosing the preconditioner and GMRES norm that ensure fast convergence. The main focus of the article is on Hermitian preconditioning (even for non-Hermitian problems). It is proposed to choose a Hermitian preconditioner  $\mathbf{H}$  and to apply GMRES in the inner product induced by  $\mathbf{H}$ . If moreover, the problem matrix  $\mathbf{A}$  is positive definite, then a new convergence bound is proved that depends only on how well  $\mathbf{H}$  preconditions the Hermitian part of  $\mathbf{A}$ , and on how non-Hermitian  $\mathbf{A}$  is. In particular, if a scalable preconditioner is known for the Hermitian part of  $\mathbf{A}$ , then the proposed method is also scalable. This result is illustrated numerically.

## Contents

<b>1 Introduction</b>	<b>2</b>
<b>2 Problem posed, notation and main results</b>	<b>3</b>
<b>3 Convergence of WP-GMRES viewed as WP-GCR</b>	<b>4</b>
3.1 WP-GCR with right preconditioning . . . . .	4
3.2 Restarted and Truncated versions . . . . .	8
3.3 A parenthesis about left preconditioning . . . . .	8
<b>4 Hermitian positive definite preconditioning for positive definite <math>\mathbf{A}</math></b>	<b>9</b>
4.1 Hermitian positive definite preconditioning . . . . .	9
4.2 Positive definite $\mathbf{A}$ . . . . .	11
<b>5 Illustration: Convection-Diffusion-Reaction</b>	<b>13</b>
<b>6 Conclusion</b>	<b>17</b>
<b>A Appendix</b>	<b>17</b>

---

\*CNRS, CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex, France  
(*nicole.spillane@cmmap.polytechnique.fr*)

## 1 Introduction

GMRES, or the Generalized Minimal Residual Method, is a method of choice for solving general linear systems. First introduced by [29], the convergence of GMRES has since been extensively studied [11, 12, 23, 2, 25, 26]. In this manuscript, linear systems

$$\mathbf{A}\mathbf{x}_* = \mathbf{b}, \tag{1}$$

are considered. Initially, the only assumption is that  $\mathbf{A}$  is a general non-singular matrix over the field  $\mathbb{K}$  with  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ . The focus then shifts to Hermitian preconditioning and matrices whose Hermitian part is positive definite. In each of these cases, convergence is examined for the weighted GMRES algorithm, a version of GMRES where the Euclidean inner product has been replaced by  $(\mathbf{x}, \mathbf{y}) \mapsto \langle \mathbf{W}\mathbf{x}, \mathbf{y} \rangle$  with  $\mathbf{W}$  Hermitian positive definite. The objective of the present work is to prove a convergence bound that can then be used to choose the preconditioner  $\mathbf{H}$  and the weight matrix  $\mathbf{W}$  in a smart way.

Convergence of GMRES in the Euclidean inner product when  $\mathbf{A}$  is positive definite is studied in [11]. Interestingly, [11] precedes the introduction of GMRES in [29] because the results are for the GCR algorithm (that produces the same iterates as GMRES).

In his PhD thesis [7], Cai proposes to select for GMRES applied to a matrix  $\mathbf{P}$ , an inner product that ‘is chosen to take advantage of some special properties of  $\mathbf{P}$ ’. In collaboration with Widlund [9], they propose domain decomposition preconditioners for non-symmetric and indefinite second order PDEs. GMRES is considered in the norm induced by the highest order term in the variational form: the energy norm. More will be said about these results below. A detailed presentation of what is now called weighted GMRES, with results both on convergence and implementation strategy, can be found in [30]. In [13], the author reinvents the idea of changing the norm within GMRES in a general framework and coins the term *weighted GMRES*. The inner products considered by [13] are diagonal matrices of weights that changes at each restart of the method. Larger weights are associated to the larger components of the residual at the end of the previous cycle leading to a faster convergence.

The method that is proposed in Section 4.2 assumes that an efficient preconditioner  $\mathbf{H}$  for the Hermitian part of  $\mathbf{A}$  (*i.e.*,  $1/2(\mathbf{A} + \mathbf{A}^*)$ ) is known. The same  $\mathbf{H}$  is applied within GMRES to solve the non-Hermitian problem (for  $\mathbf{A}$ ). As an illustration, the advection-diffusion-reaction discretized by finite elements is solved at the end of this article. The preconditioner  $\mathbf{H}$  is chosen to be a two-level domain decomposition preconditioner [35] with the GenEO coarse space introduced by [32, 33] (see also [34, 16, 22, 10, 27, 31] for a non exhaustive list of extensions and related work). It has been shown that GenEO is a very powerful and scalable solver [21] for symmetric positive definite problems.

The idea of separating the Hermitian and skew-Hermitian part of a matrix in order to approximate the solution of a linear system has been exploited in the Hermitian and skew-Hermitian splitting methods introduced in [1]. The field of domain decomposition for non-spd problems was paved by [7, 9] (see also [35][Chapter 11]). The authors solve convection-diffusion-reaction with a two-level additive Schwarz preconditioner where the coarse space is based on a coarse grid. The one level component in the preconditioner takes two forms: either the local solvers come from the original (non-symmetric and/or indefinite) matrix, or the local solvers come from an spd part of the problem matrix. GMRES is applied in the norm induced by that spd part. In both cases the coarse projector is for the original operator. If the coarse mesh is fine enough, the rate of convergence is shown to be independent of the number of degrees of freedom and the number of local problems (scalability). A simplified explanation for the presence of a condition on the size of the subdomains is that, for the theory to go through, the second order term must dominate the other terms. It is known that the first non-zero eigenvalue of  $-\Delta$  on regular subdomains of diameter  $H$  is of the order of  $1/H$ . The algorithm is

generalized in [39] to any matrix that can be viewed as a perturbation of an spd matrix. Their proposed preconditioner is the combination of a very good preconditioner for the spd part and a coarse solve. A parameter  $\delta_0$  qualifies how efficient the coarse space is at filling the gap between the original problem and the spd one. It enters into the convergence estimate. The algorithm is later called CSPD for Coarse Grid Plus SPD Preconditioning in a numerical comparison with other algorithms [8].

More recently, domain decomposition for Helmholtz has been studied [14, 15]. This case is symmetric indefinite and applying GMRES in the energy norm is a crucial part of the proof. Weighted GMRES was also applied to study the spd GenEO eigenproblem applied to indefinite and non-self-adjoint problems in [5, 6]. It is proved and observed numerically that GenEO performs well also on a family of non-spd second-order problems. Finally, the authors of [3] prove an abstract framework for one-level additive Schwarz for non-Hermitian or indefinite problems. They illustrate their results by solving the convection-diffusion-reaction equation. The present work takes a more algebraic route and proves results that are not restricted to domain decomposition.

The outline for the rest of the article is as follows. In Section 2, some notation is introduced and Theorem 1 gives an overview of some of the convergence results. In Section 3, the GMRES algorithm is studied through the study of an equivalent form that is the GCR algorithm. A convergence bound is proved in Theorem 3 that is connected to field of value, or Elman, estimate [11, 12, 2]. In Section 4.1, Theorem 5, a special case is considered where the preconditioner  $\mathbf{H}$  is Hermitian and GMRES/GCR is applied in the  $\mathbf{H}$  inner product (or  $\mathbf{H}^{-1}$  for left preconditioned GMRES). For positive-definite  $\mathbf{A}$ , a final convergence estimate is proved in Theorem 6 that makes explicit the rate at which the non-Hermitianness of  $\mathbf{A}$  slows down convergence. As an illustration of this result, in Section 5, a solver is proposed for the convection-diffusion-reaction equation. The preconditioner is a domain decomposition preconditioner with a GenEO spectral coarse space. This way, the Hermitian part of the problem is very well preconditioned. It is shown theoretically that the overall convergence does not depend on the number of subdomains (scalability), or on the discretization step. Numerical experiments show that Hermitian preconditioning can be very efficient and scalable for mildly non-Hermitian problems.

## 2 Problem posed, notation and main results

Let  $\mathbf{K} = \mathbb{R}$  or  $\mathbb{C}$  be the field over which the linear system is considered. Let  $\mathbf{A} \in \mathbb{K}^{n \times n}$  be a non-singular matrix. Given any  $\mathbf{b} \in \mathbb{K}^n$ , the problem at hand is to find  $\mathbf{x}_* \in \mathbb{K}^n$  such that:

$$\mathbf{A}\mathbf{x}_* = \mathbf{b}.$$

The chosen methodology is to apply weighted and preconditioned (WP) GMRES. Let  $\mathbf{H} \in \mathbb{K}^{n \times n}$  denote the preconditioner and  $\mathbf{W} \in \mathbb{K}^{n \times n}$  denote the weight matrix. It is assumed that  $\mathbf{H}$  is non-singular and that  $\mathbf{W}$  is Hermitian positive definite (hpd). The inner product and norm induced by  $\mathbf{W}$  are denoted by  $\langle \cdot, \cdot \rangle_{\mathbf{W}}$  and  $\| \cdot \|_{\mathbf{W}}$ , respectively:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{W}} = \langle \mathbf{W}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{W}\mathbf{y} \rangle = \mathbf{y}^* \mathbf{W}\mathbf{x} \text{ and } \|\mathbf{x}\|_{\mathbf{W}} = \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{W}}^{1/2}, \forall \mathbf{x}, \mathbf{y} \in \mathbb{K}^n.$$

Any matrix  $\mathbf{B} \in \mathbb{K}^{n \times n}$  can be split into the sum of its Hermitian part and its skew-Hermitian part. The notation used is

$$\mathbf{M}(\mathbf{B}) = \frac{\mathbf{B} + \mathbf{B}^*}{2} \text{ (Hermitian part) and } \mathbf{N}(\mathbf{B}) = \frac{\mathbf{B} - \mathbf{B}^*}{2} \text{ (skew-Hermitian part)}. \quad (2)$$

An overview of some results in the article is given in the following theorem.

**Theorem 1** (Summary of main results). *Assume that the operator  $\mathbf{A} \in \mathbb{K}^{n \times n}$  and preconditioner  $\mathbf{H} \in \mathbb{K}^{n \times n}$  are non-singular, and that the weight matrix  $\mathbf{W} \in \mathbb{K}^{n \times n}$  is hpd. The  $i$ -th iterate of right preconditioned WP-GMRES satisfies*

$$\begin{aligned} \frac{\|\mathbf{r}_i\|_{\mathbf{W}}}{\|\mathbf{r}_0\|_{\mathbf{W}}} &\leq \left[ 1 - \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}^2 \|\mathbf{y}\|_{\mathbf{W}}^2} \right]^{i/2} \\ &\leq \left[ 1 - \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \right]^{i/2} \quad \text{if } \mathbf{H} = \mathbf{W} \text{ is hpd} \\ &\leq \left[ 1 - \frac{1/\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))}{1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2} \right]^{i/2} \quad \text{if } \mathbf{H} = \mathbf{W} \text{ is hpd and } \mathbf{A} \text{ positive definite,} \end{aligned}$$

where  $\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))$  is the condition number of  $\mathbf{H}\mathbf{M}(\mathbf{A})$  and  $\rho(\cdot)$  denotes the spectral radius of a matrix. For the last two estimates, it has been assumed that the preconditioner  $\mathbf{H}$  is hpd and that  $\mathbf{W} = \mathbf{H}$ . In this case, the algorithm will be referred to as WHP-GMRES for Weighted with a Hermitian Preconditioner.

*Proof.* The proof proceeds as follows. First, in Theorem 2, the equivalence between WP-GCR (Algorithm 1) and WP-GMRES is established as long as the origin is not in the field of values of the preconditioned operator. In the opposite case, the bounds above all simplify to  $\|\mathbf{r}_i\|_{\mathbf{W}} \leq \|\mathbf{r}_0\|_{\mathbf{W}}$ , a trivial result. The first estimate in the theorem is proved for WP-GCR in Theorem 3 (and is already known [24, 2]). The second estimate in the theorem is proved for WHP-GCR (where again HP stands for Hermitian preconditioning) in Theorem 5 and the last estimate in the theorem is proved in Corollary 2.  $\square$

### 3 Convergence of WP-GMRES viewed as WP-GCR

The Generalized Conjugate Residual (GCR) method [11], also known as Orthomin [37] is equivalent to GMRES in the sense that it generates the same approximate solutions at each iteration. GMRES is usually preferred as it is slightly less computationally expensive and more stable. However, GCR has the advantage of a simpler presentation and the proofs in this article all come from the GCR formulation of GMRES.

#### 3.1 WP-GCR with right preconditioning

Weighted and preconditioned GCR (WP-GCR) with right preconditioning is presented in Algorithm 1. The initial guess  $\mathbf{x}_0$  is assumed to be any vector in  $\mathbb{K}^n$ .

In the following theorem, it is proved that WP-GCR is indeed equivalent to WP-GMRES unless WP-GCR has an unlucky breakdown. This can only happen if 0 is in the  $\mathbf{W}$ -field of values of  $\mathbf{A}\mathbf{H}$ , a set defined by

$$W_{\mathbf{W}}(\mathbf{A}\mathbf{H}) = \left\{ \frac{\langle \mathbf{A}\mathbf{H}\mathbf{u}, \mathbf{u} \rangle_{\mathbf{W}}}{\langle \mathbf{u}, \mathbf{u} \rangle_{\mathbf{W}}}; \mathbf{u} \in \mathbb{C}^n \setminus \{0\} \right\}. \quad (3)$$

Although the proof is not new it has been included. Indeed, some intermediary results in the proof are used in subsequent proofs and to discuss the generalization of all results to truncated and restarted methods.

**Theorem 2.** *Assume that the operator  $\mathbf{A} \in \mathbb{K}^{n \times n}$  and preconditioner  $\mathbf{H} \in \mathbb{K}^{n \times n}$  are non-singular, and that the weight matrix  $\mathbf{W} \in \mathbb{K}^{n \times n}$  is hpd. The  $i$ -th residual of Algorithm 1 satisfies*

$$\|\mathbf{r}_i\|_{\mathbf{W}} = \min \{ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}; \mathbf{x} \in \mathbf{x}_0 + \text{span}\{\mathbf{p}_0, \dots, \mathbf{p}_{i-1}\} \}. \quad (4)$$

**Algorithm 1** WP-GCR with right preconditioning**Require:**  $\mathbf{x}_0 \in \mathbb{R}^n$ 

$$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$$

$$\mathbf{z}_0 = \mathbf{H}\mathbf{r}_0$$

$$\mathbf{p}_0 = \mathbf{z}_0$$

$$\mathbf{q}_0 = \mathbf{A}\mathbf{p}_0$$

**for**  $i = 0, 1, \dots$ , convergence **do**

$$\delta_i = \langle \mathbf{q}_i, \mathbf{q}_i \rangle_{\mathbf{W}}; \quad \gamma_i = \langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}}; \quad \alpha_i = \gamma_i / \delta_i$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i$$

$$\mathbf{z}_{i+1} = \mathbf{H}\mathbf{r}_{i+1}$$

**for**  $j = 0, 1, \dots, i$  **do**

$$\Phi_{i,j} = \langle \mathbf{q}_j, \mathbf{A}\mathbf{z}_{i+1} \rangle_{\mathbf{W}}; \quad \beta_{i,j} = \Phi_{i,j} / \delta_j$$

**end for**

$$\mathbf{p}_{i+1} = \mathbf{z}_{i+1} - \sum_{j=0}^i \beta_{i,j} \mathbf{p}_j$$

$$\mathbf{q}_{i+1} = \mathbf{A}\mathbf{z}_{i+1} - \sum_{j=0}^i \beta_{i,j} \mathbf{q}_j$$

**end for****return** Return  $\mathbf{x}_{i+1}$ 

Moreover, if  $0 \notin W_{\mathbf{W}}(\mathbf{A}\mathbf{H})$ , then  $\text{span}\{\mathbf{p}_0, \dots, \mathbf{p}_{i-1}\} = \mathcal{K}_i$  where

$$\mathcal{K}_i := \{\mathbf{H}\mathbf{r}_0, \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{r}_0, \dots, (\mathbf{H}\mathbf{A})^{i-1}\mathbf{H}\mathbf{r}_0\}$$

is the Krylov subspace, i.e., WP-GCR returns the same approximate solutions as WP-GMRES.

*Proof.* The vectors  $\mathbf{q}_i = \mathbf{A}\mathbf{p}_i$  are pairwise orthogonal, i.e.,

$$\langle \mathbf{q}_j, \mathbf{q}_i \rangle_{\mathbf{W}} = 0 \text{ if } i \neq j. \quad (5)$$

Indeed, by symmetry, it suffices to prove by recursion over  $i \geq 0$  that:  $\langle \mathbf{A}\mathbf{p}_i, \mathbf{A}\mathbf{p}_j \rangle_{\mathbf{W}} = 0$  for every  $j < i$ . This is easy by recalling the definitions of  $\mathbf{p}_i$  and  $\beta_{i,j}$ .

Next, the minimization property (4) is proved. By an immediate recursion, it holds that  $\mathbf{r}_i = \mathbf{r}_0 - \sum_{j=0}^{i-1} \gamma_j / \delta_j \mathbf{A}\mathbf{p}_j$ . We notice that  $\gamma_i = \langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}} = \langle \mathbf{q}_i, \mathbf{r}_0 \rangle_{\mathbf{W}} - \sum_{j=0}^{i-1} \gamma_j / \delta_j \langle \mathbf{q}_i, \mathbf{A}\mathbf{p}_j \rangle_{\mathbf{W}} = \langle \mathbf{q}_i, \mathbf{r}_0 \rangle_{\mathbf{W}}$  by (5). The  $i$ -th residual can now be rewritten as

$$\mathbf{r}_0 - \mathbf{r}_i = \sum_{j=0}^{i-1} \frac{\langle \mathbf{A}\mathbf{p}_j, \mathbf{r}_0 \rangle_{\mathbf{W}}}{\langle \mathbf{A}\mathbf{p}_j, \mathbf{A}\mathbf{p}_j \rangle_{\mathbf{W}}} \mathbf{A}\mathbf{p}_j.$$

This means that  $\mathbf{r}_0 - \mathbf{r}_i$  is the  $\mathbf{W}$ -orthogonal projection of  $\mathbf{r}_0$  onto  $\text{span}\{\mathbf{q}_j; j < i\}$ . Two other equivalent characterizations of the orthogonal projection are

$$\mathbf{r}_i \in \mathbf{r}_0 + \text{span}\{\mathbf{q}_j; j < i\} \text{ with } \mathbf{r}_i = (\mathbf{r}_0 - \mathbf{r}_i) - \mathbf{r}_0 \perp^{\mathbf{W}} \text{span}\{\mathbf{q}_j; j < i\}, \quad (6)$$

and  $\|\mathbf{r}_i\|_{\mathbf{W}} = \min\{\|\mathbf{r}\|_{\mathbf{W}}; \mathbf{r} \in \mathbf{r}_0 + \text{span}\{\mathbf{q}_j; j < i\}\}$ . Then (4) follows by the change of variables  $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$  and  $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$ .

It remains to justify that  $\text{span}\{\mathbf{p}_j; j < i\} = \mathcal{K}_i$ . It is obvious that  $\text{span}\{\mathbf{p}_j; j < i\} \subset \mathcal{K}_i$ . The reverse inclusion is true unless  $\mathbf{A}\mathbf{z}_{i-1} \in \text{span}\{\mathbf{q}_j; j < i-1\}$ . Then, by (6),  $\mathbf{r}_{i-1} \perp^{\mathbf{W}} \mathbf{A}\mathbf{z}_{i-1}$ , i.e.,  $\langle \mathbf{r}_{i-1}, \mathbf{A}\mathbf{H}\mathbf{r}_{i-1} \rangle_{\mathbf{W}} = 0 = \gamma_{i-1}$ . If  $0 \notin W_{\mathbf{W}}(\mathbf{A}\mathbf{H})$ , this implies  $\mathbf{r}_{i-1} = \mathbf{0}$  which is a lucky breakdown of both GCR and GMRES. The proof ends by recalling the characterization of the  $i$ -th iterate of preconditioned GMRES in e.g., [28][Section 9]. Following the idea in [7, 13], the Euclidean product can be changed to the  $\mathbf{W}$ -inner product in order to get the result for WP-GMRES.  $\square$

The following observations can be made:

- Preconditioning modifies the Krylov subspace.
- Weighting modifies which norm of the residual is minimized.
- Weighting does not modify the Krylov subspace.

The speed of convergence of the algorithm is addressed next by comparing two subsequent residuals.

**Theorem 3** (Convergence of WP-GCR). *Assume that the operator  $\mathbf{A} \in \mathbb{K}^{n \times n}$  and preconditioner  $\mathbf{H} \in \mathbb{K}^{n \times n}$  are non-singular, and that the weight matrix  $\mathbf{W} \in \mathbb{K}^{n \times n}$  is hpd. The  $i$ -th iterate of Algorithm 1 satisfies*

$$\frac{\|\mathbf{r}_i\|_{\mathbf{W}}}{\|\mathbf{r}_0\|_{\mathbf{W}}} \leq \left( 1 - \inf_{\mathbf{y} \neq \mathbf{0}} \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}^2 \|\mathbf{y}\|_{\mathbf{W}}^2} \right)^{i/2}. \quad (7)$$

*Proof.* From the residual update formula we get  $\mathbf{r}_i = \mathbf{r}_{i+1} + \alpha_i \mathbf{q}_i$ . By (6),  $\mathbf{r}_{i+1} \perp^{\mathbf{W}} \mathbf{q}_i$  so

$$\|\mathbf{r}_i\|_{\mathbf{W}}^2 = \|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2 + |\alpha_i|^2 \|\mathbf{q}_i\|_{\mathbf{W}}^2 = \|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2 + \frac{|\langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{q}_i\|_{\mathbf{W}}^4} \|\mathbf{q}_i\|_{\mathbf{W}}^2.$$

The relative decrease in residual between two subsequent iterations is

$$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2}{\|\mathbf{r}_i\|_{\mathbf{W}}^2} = 1 - \frac{|\langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{q}_i\|_{\mathbf{W}}^2 \|\mathbf{r}_i\|_{\mathbf{W}}^2}. \quad (8)$$

Taking the  $\mathbf{W}$ -inner product of  $\mathbf{q}_i = \mathbf{A}\mathbf{z}_i - \sum_{j=0}^{i-1} \mathbf{q}_j \beta_{i,j}$  by  $\mathbf{r}_i$  leads to

$$\langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i \rangle_{\mathbf{W}} - \sum_{j=0}^{i-1} \beta_{i,j} \langle \mathbf{q}_j, \mathbf{r}_i \rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i \rangle_{\mathbf{W}},$$

and

$$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2}{\|\mathbf{r}_i\|_{\mathbf{W}}^2} = 1 - \frac{|\langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{q}_i\|_{\mathbf{W}}^2 \|\mathbf{r}_i\|_{\mathbf{W}}^2} = 1 - \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{r}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{q}_i\|_{\mathbf{W}}^2 \|\mathbf{r}_i\|_{\mathbf{W}}^2}. \quad (9)$$

Next, from the orthogonalisation formula and (5), it is deduced that

$$\|\mathbf{A}\mathbf{z}_i\|_{\mathbf{W}}^2 = \|\mathbf{q}_i + \sum_{j=0}^{i-1} \beta_{i,j} \mathbf{q}_j\|_{\mathbf{W}}^2 = \|\mathbf{q}_i\|_{\mathbf{W}}^2 + \sum_{j=0}^{i-1} |\beta_{i,j}|^2 \|\mathbf{q}_j\|_{\mathbf{W}}^2 \geq \|\mathbf{q}_i\|_{\mathbf{W}}^2. \quad (10)$$

Finally, the decrease in residual between two subsequent iterations of Algorithm 1 is bounded by

$$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{W}}}{\|\mathbf{r}_i\|_{\mathbf{W}}} \leq \left[ 1 - \frac{|\langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{z}_i\|_{\mathbf{W}}^2 \|\mathbf{r}_i\|_{\mathbf{W}}^2} \right]^{1/2} = \left[ 1 - \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{r}_i, \mathbf{r}_i \rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{r}_i\|_{\mathbf{W}}^2 \|\mathbf{r}_i\|_{\mathbf{W}}^2} \right]^{1/2}, \quad (11)$$

where the Cauchy-Schwarz inequality ensures that the square root is well defined.  $\square$

Next, this is reformulated to match an often cited result (out of many) in [11].

**Corollary 1** (Field of Value, or Elman, estimate). *Under the assumptions of Theorem 3, the  $i$ -th iterate of Algorithm 1 satisfies*

$$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{W}}}{\|\mathbf{r}_0\|_{\mathbf{W}}} \leq \left[ 1 - \frac{d(0, W_{\mathbf{W}}(\mathbf{A}\mathbf{H}))^2}{\|\mathbf{A}\mathbf{H}\|_{\mathbf{W}}^2} \right]^{i/2},$$

where

- $d(0, W_{\mathbf{W}}(\mathbf{A}\mathbf{H})) = \inf \left\{ \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{u}, \mathbf{u} \rangle_{\mathbf{W}}|}{\langle \mathbf{u}, \mathbf{u} \rangle_{\mathbf{W}}}; \mathbf{u} \in \mathbb{C}^n \setminus \{0\} \right\}$  is the distance to zero of the  $\mathbf{W}$ -field of values of  $\mathbf{A}\mathbf{H}$  defined in (3),
- $\|\mathbf{A}\mathbf{H}\|_{\mathbf{W}}$  denotes the matrix norm of  $\mathbf{A}\mathbf{H}$  induced by the vector norm  $\mathbf{W}$ .

*Proof.* The terms in (7) can be grouped as

$$\left[ \frac{\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}}{\|\mathbf{y}\|_{\mathbf{W}}^2} \times \frac{\|\mathbf{y}\|_{\mathbf{W}}}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}} \right]^2 \geq \left[ \frac{d(0, W_{\mathbf{W}}(\mathbf{A}\mathbf{H}))}{\|\mathbf{A}\mathbf{H}\|_{\mathbf{W}}} \right]^2, \quad (12)$$

where the numerator minimizes the first term in the product and the denominator maximizes the inverse of the second.  $\square$

The result of Theorem 3 is stronger than the field of value estimate in Corollary 1. Indeed, a bound for  $C$  in Theorem 3 can be found without necessarily bounding  $d(0, W_{\mathbf{W}}(\mathbf{A}\mathbf{H}))$  and  $\|\mathbf{A}\mathbf{H}\|_{\mathbf{W}}$  independently. Another way of saying this is that the terms in  $C$  can be grouped differently than in the field of value bound. This is done in the next section and also discussed in [24] and [2]. The proof given here is different from the ones in the references.

**Remark 1** (Breakdown and equivalence with WP-GMRES). *If 0 is in the  $\mathbf{W}$ -field of values of  $\mathbf{H}\mathbf{A}$ , the right hand side of the Elman estimate is 1 rendering it useless. In fact, the proof in [11] makes the assumption that the  $\mathbf{A}$  is positive-definite so the case where 0 is in the field of values is not considered. Still, the formula is not incorrect because it states that  $\mathbf{r}_{i+1} \leq \mathbf{r}_i$ .*

*If 0 is in the  $\mathbf{W}$ -field of values of  $\mathbf{H}\mathbf{A}$ , it can occur that  $\gamma_i = \langle \mathbf{q}_i, \mathbf{r}_i \rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{H}\mathbf{r}_i - \sum_{j=0}^{i-1} \Phi_{ij} / \delta_j \mathbf{q}_j \rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{H}\mathbf{r}_i, \mathbf{r}_i \rangle_{\mathbf{W}} = 0$ . Then the residual does not get updated ( $\mathbf{r}_{i+1} = \mathbf{r}_i$ ) and the next search direction is  $\mathbf{A}\mathbf{H}\mathbf{r}_i$  orthogonalized against all previous ones, including itself. In other words  $\mathbf{p}_{i+1} = \mathbf{q}_{i+1} = \mathbf{0}$ . The algorithm has broken down before zeroing the residual. This is a particularity of the way the search directions are computed in WP-GCR. These unlucky breakdowns do not occur in WP-GMRES if  $\mathbf{A}$  is non-singular (a sufficient but not necessary condition). If an unlucky breakdown occurs in GCR, the algorithm can be restarted by computing the next few search directions as in the weighted and preconditioned Orthodir algorithm.*

*A very simple way of understanding that WP-GCR can breakdown is to consider taking  $\mathbf{r}_0$  such that  $\langle \mathbf{r}_0, \mathbf{A}\mathbf{H}\mathbf{r}_0 \rangle_{\mathbf{W}} = 0$ .*

**Remark 2.** *Minimizing  $|\langle \mathbf{H}\mathbf{A}\mathbf{x}, \mathbf{x} \rangle_{\mathbf{W}}| / \langle \mathbf{H}\mathbf{A}\mathbf{x}, \mathbf{H}\mathbf{A}\mathbf{x} \rangle_{\mathbf{W}}$  is equivalent to minimizing*

*$|\langle \mathbf{y}, (\mathbf{H}\mathbf{A})^{-1}\mathbf{y} \rangle_{\mathbf{W}}| / \langle \mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}$  by the change of variables  $\mathbf{y} = \mathbf{H}\mathbf{A}\mathbf{x}$ . This is a way of recovering that the WP-GMRES residual is bounded with respect to the product of the distances to zero of the  $\mathbf{W}$ -fields of value of  $\mathbf{A}\mathbf{H}$  and of  $(\mathbf{A}\mathbf{H})^{-1}$  [2].*



### 3.2 Restarted and Truncated versions

Within WP-GCR as well as WP-GMRES, the new search directions are orthogonalized against all previous ones. The cost of this procedure in terms of computation and storage can become prohibitive if the algorithm takes many iterations to converge. Well established variants of the algorithms have been proposed as early as in [11, 37] where either the orthogonalization is truncated (Orthomin(k)) or the algorithm is restarted every  $k$  iterations (WP-GMRES(k), WP-GCR(k)). If  $k = 0$ , no orthogonalization is performed at all and the algorithm is called the Minimal Residual iteration (MR). A fact that is not so frequently known is that the convergence result given in Theorem 3 holds for all restarted and truncated versions of the algorithms. This was already fully understood by [11].

**Theorem 4** (Convergence of truncated and restarted versions). *The result in Theorem 3 holds for WP-GCR, WP-GMRES as well as all their truncated and restarted versions including WP-MR.*

*Proof.* Looking at the proof of Theorem 3, the only property of the algorithm that is used to establish (8) is that  $\mathbf{r}_{i+1} \perp^{\mathbf{W}} \mathbf{q}_i$ . This remains true even with truncation and restart because of the choice of  $\alpha_i$ . If no orthogonalization is performed at all then  $\mathbf{q}_i = \mathbf{A}\mathbf{z}_i = \mathbf{A}\mathbf{H}\mathbf{r}_i$  and the result has been proved. If partial orthogonalization is performed then the proof is the same except that sums for  $j$  ranging between 0 and  $i - 1$  are replaced by partial sums.  $\square$

Equations (9) and (10) explain how orthogonalization helps: decreasing  $\|\mathbf{q}_i\|_{\mathbf{W}}$  leads to a decrease in the residual and orthogonalization does just that. Full orthogonalization provides the best choice of  $\mathbf{q}_i$  in  $\mathbf{A}\mathcal{K}_i$ . A very important realization is that the field of value convergence bound, as well as all the convergence bounds in this article do not account for the advantages provided by orthogonalization. For this reason they are expected to be over-pessimistic when applied to WP-GMRES and WP-GCR.

Two conclusions can be drawn from this. Either the decrease in residual predicted by the convergence bound is satisfactory, then WP-MR, or another truncated or restarted algorithm can be applied. Or, the decrease guaranteed by the theory is not satisfactory, then full WP-GCR/WP-GMRES can be applied with the hope that the bound is over-pessimistic. In practice this is very likely to be the case as WP-GCR/WP-GMRES often exhibits a superlinear convergence behaviour. The convergence bounds of the form presented in this article allow to check that the algorithm cannot stagnate (or near-stagnate), and the superlinear convergence behaviour should kick in.

All subsequent convergence proofs follow from manipulating the minimized quantity in Theorems 3 and 4 so they hold also for truncated and restarted versions of the algorithms.

### 3.3 A parenthesis about left preconditioning

Left preconditioning, *i.e.* solving  $\mathbf{H}\mathbf{A}\mathbf{x}_* = \mathbf{H}\mathbf{b}$  can be performed instead of right preconditioning. For completeness, the left preconditioned WP-GCR is presented in Algorithm 2.

The  $i$ -th residual of Algorithm 2 satisfies

$$\|\mathbf{z}_i\|_{\mathbf{W}} = \|\mathbf{H}\mathbf{r}_i\|_{\mathbf{W}} = \min \{ \|\mathbf{H}(\mathbf{b} - \mathbf{A}\mathbf{x})\|_{\mathbf{W}}; \mathbf{x} \in \mathbf{x}_0 + \text{span}\{\mathbf{p}_0, \dots, \mathbf{p}_{i-1}\} \}.$$

Moreover, if  $0 \notin W_{\mathbf{W}}(\mathbf{H}\mathbf{A})$ , then  $\text{span}\{\mathbf{p}_0, \dots, \mathbf{p}_{i-1}\} = \mathcal{K}_i$  (the Krylov subspace) so again WP-GCR returns the same approximate solutions as WP-GMRES. Following the same steps as in Section 3.1, it can be proved that

$$\frac{\|\mathbf{z}_{i+1}\|_{\mathbf{W}}^2}{\|\mathbf{z}_i\|_{\mathbf{W}}^2} = 1 - \frac{\langle \mathbf{y}_i, \mathbf{z}_i \rangle_{\mathbf{W}}^2}{\langle \mathbf{y}_i, \mathbf{y}_i \rangle_{\mathbf{W}} \langle \mathbf{z}_i, \mathbf{z}_i \rangle_{\mathbf{W}}} \leq 1 - \frac{\langle \mathbf{H}\mathbf{A}\mathbf{z}_i, \mathbf{z}_i \rangle_{\mathbf{W}}^2}{\langle \mathbf{H}\mathbf{A}\mathbf{z}_i, \mathbf{H}\mathbf{A}\mathbf{z}_i \rangle_{\mathbf{W}} \langle \mathbf{z}_i, \mathbf{z}_i \rangle_{\mathbf{W}}}.$$

The following observations follow:

---

**Algorithm 2** WP-GCR with left preconditioning

---

**Require:**  $\mathbf{x}_0 \in \mathbb{R}^n$ 

$$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$$

$$\mathbf{z}_0 = \mathbf{H}\mathbf{r}_0$$

$$\mathbf{p}_0 = \mathbf{z}_0$$

$$\mathbf{y}_0 = \mathbf{H}\mathbf{A}\mathbf{p}_0$$

**for**  $i = 0, 1, \dots$ , convergence **do**

$$\delta_i = \langle \mathbf{y}_i, \mathbf{y}_i \rangle_{\mathbf{W}}; \quad \gamma_i = \langle \mathbf{y}_i, \mathbf{z}_i \rangle_{\mathbf{W}}; \quad \alpha_i = \gamma_i / \delta_i$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{z}_{i+1} = \mathbf{z}_i - \alpha_i \mathbf{y}_i$$

**for**  $j = 0, \dots, i$  **do**

$$\Phi_{i,j} = \langle \mathbf{y}_j, \mathbf{H}\mathbf{A}\mathbf{z}_{i+1} \rangle_{\mathbf{W}}; \quad \beta_{i,j} = \Phi_{i,j} / \delta_j$$

**end for**

$$\mathbf{p}_{i+1} = \mathbf{z}_{i+1} - \sum_{j=0}^i \beta_{i,j} \mathbf{p}_j$$

$$\mathbf{y}_{i+1} = \mathbf{H}\mathbf{A}\mathbf{z}_{i+1} - \sum_{j=0}^i \beta_{i,j} \mathbf{y}_j$$

**end for****return**  $\mathbf{x}_{i+1}$ 

---

- Left preconditioning produces the same Krylov subspace as right preconditioning.
- Left preconditioning modifies the residual that is considered in the minimization property as well as the Krylov subspace.
- Left preconditioning in the  $\mathbf{H}^{-1}$ -inner product and right-preconditioning in the  $\mathbf{H}$ -inner product are equivalent (as suggested in [28][Problem 9.13]).

## 4 Hermitian positive definite preconditioning for positive definite $\mathbf{A}$

The convergence study now focuses on some not completely general cases.

### 4.1 Hermitian positive definite preconditioning

Two strong assumptions are made:

1. the preconditioner  $\mathbf{H}$  is hpd,
2. the inner product is induced by the preconditioner:  $\mathbf{W} = \mathbf{H}$ .

WP-GCR with right hpd preconditioning and  $\mathbf{W} = \mathbf{H}$  takes the form of Algorithm 3. The name WHP-GCR is adopted where HP stands for Hermitian Preconditioning. The operations have been reorganized so that no additional application of  $\mathbf{H}$  is required compared to unweighted GCR (except in the initialization). The extra cost is the storage of the vectors  $\mathbf{y}_j = \mathbf{H}\mathbf{q}_j$ . Two alternate versions are presented in Algorithms 4 and 5 (of the appendix) that do not require more storage than unweighted GCR. Although no details are given here, the same cost saving measures can be taken in a GMRES algorithm. Note also that if the preconditioner  $\mathbf{H}$  is very cheap to apply, applying it twice per iteration may be entirely feasible. In this case, it suffices to run a right preconditioned GCR or GMRES code with the inner product changed to  $\langle \cdot, \cdot \rangle_{\mathbf{H}}$ .

In application of Theorem 2, WHP-GCR (Algorithm 3) is characterized as a Krylov subspace method by the following properties. First, The vectors  $\mathbf{q}_i = \mathbf{A}\mathbf{p}_i$  are pairwise orthogonal in the  $\mathbf{H}$ -inner product. Second, the residuals satisfy the minimization property:  $\|\mathbf{r}_i\|_{\mathbf{H}} = \min \{\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_{\mathbf{H}}; \mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_i\}$  as long as  $0 \notin W_{\mathbf{H}}(\mathbf{A}\mathbf{H})$ .

WP-GCR with right preconditioning by  $\mathbf{H}$  and weighting by  $\mathbf{W} = \mathbf{H}$  is equivalent to WP-GCR with left preconditioning by  $\mathbf{H}$  and weighting by  $\mathbf{W} = \mathbf{H}^{-1}$ . For this reason the distinction between left and right preconditioning is no longer made. In [30], a similar equivalence is observed for an inner product that arises from a symmetric part of the problem matrix  $\mathbf{A}$ .

---

**Algorithm 3** WHP-GCR (*i.e.*, WP-GCR with hpd  $\mathbf{H}$  and  $\mathbf{W} = \mathbf{H}$ )

---

**Require:**  $\mathbf{x}_0 \in \mathbb{R}^n$

$$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$$

$$\mathbf{z}_0 = \mathbf{H}\mathbf{r}_0$$

$$\mathbf{p}_0 = \mathbf{z}_0$$

$$\mathbf{q}_0 = \mathbf{A}\mathbf{p}_0$$

$$\mathbf{y}_0 = \mathbf{H}\mathbf{q}_0$$

**for**  $i = 0, 1, \dots$ , convergence **do**

$$\delta_i = \langle \mathbf{y}_i, \mathbf{q}_i \rangle; \quad \gamma_i = \langle \mathbf{q}_i, \mathbf{z}_i \rangle; \quad \alpha_i = \gamma_i / \delta_i$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i$$

$$\mathbf{z}_{i+1} = \mathbf{z}_i - \alpha_i \mathbf{y}_i$$

$$\mathbf{p}_{i+1} = \mathbf{z}_{i+1}$$

$$\mathbf{q}_{i+1} = \mathbf{A}\mathbf{z}_{i+1}$$

**for**  $j = 0, \dots, i$  **do**

$$\Phi_{i,j} = \langle \mathbf{y}_j, \mathbf{q}_{i+1} \rangle; \quad \beta_{i,j} = \Phi_{i,j} / \delta_j^{-1}$$

$$\mathbf{p}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{p}_j$$

$$\mathbf{q}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{q}_j$$

**end for**

$$\mathbf{y}_{i+1} = \mathbf{H}\mathbf{q}_{i+1}^-$$

**end for**

**return**  $\mathbf{x}_{i+1}$

---

**Theorem 5** (Convergence of WHP-GCR). *Assume that the operator  $\mathbf{A} \in \mathbb{K}^{n \times n}$  is non-singular. Assume also that the preconditioner  $\mathbf{H} \in \mathbb{K}^{n \times n}$  is hpd and that right preconditioned GCR is applied in the inner product induced by  $\mathbf{H}$ . The  $i$ -th iterate of Algorithm 3 satisfies*

$$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq \left[ 1 - \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \right]^{i/2}. \quad (13)$$

*The same result holds for all truncated and restarted versions of WHP-GCR and WHP-GMRES, including WHP-MR.*

*Proof.* Applying Theorem 3 to WHP-GCR, the quantity that must be bounded can be

rewritten as

$$\begin{aligned}
\inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle_{\mathbf{H}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{H}}^2 \|\mathbf{y}\|_{\mathbf{H}}^2} &= \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle|^2}{\langle \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{A}\mathbf{H}\mathbf{y} \rangle \langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \\
&= \inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle^2}{\langle \mathbf{H}\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y} \rangle \langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \quad (\text{by } \mathbf{y} \leftarrow \mathbf{H}\mathbf{y}) \\
&\geq \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \\
&= \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}^{-1}\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \quad (\text{by } \mathbf{y} \leftarrow \mathbf{A}\mathbf{y}) \\
&= \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle}
\end{aligned}$$

where, according to the notation (2),  $\mathbf{M}(\mathbf{A})$  and  $\mathbf{M}(\mathbf{A}^{-1})$  are the Hermitian parts of  $\mathbf{A}$  and  $\mathbf{A}^{-1}$ .  $\square$

## 4.2 Positive definite $\mathbf{A}$

Assuming that the problem matrix  $\mathbf{A}$  is positive definite, *i.e.*, that  $\mathbf{M}(\mathbf{A})$  is hpd, the calculations from the previous paragraph are resumed:

$$\begin{aligned}
\inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y} \rangle_{\mathbf{W}}^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}^2 \|\mathbf{y}\|_{\mathbf{W}}^2} &\geq \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle|}{|\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle|} \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \\
&\quad \times \inf_{\mathbf{y} \neq 0} \frac{|\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} \\
&\geq \inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle} \times \inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} \\
&\quad \times \inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle},
\end{aligned}$$

where division by  $\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle$  is licit and removing the absolute values does not change the result. Indeed, by [18][Property (1.6) on page 10],  $\mathbf{M}(\mathbf{A})$  being positive definite implies that  $\mathbf{A}^{-1}$  is well defined and that  $\mathbf{M}(\mathbf{A}^{-1})$  is also positive definite.

Let  $\lambda_{\min}(\mathbf{H}\mathbf{M}(\mathbf{A}))$  and  $\lambda_{\max}(\mathbf{H}\mathbf{M}(\mathbf{A}))$  denote the smallest and largest eigenvalues of  $\mathbf{H}\mathbf{M}(\mathbf{A})$ . The eigenvalues of  $\mathbf{H}\mathbf{M}(\mathbf{A})$ , are also the eigenvalues of the generalized eigenvalue problems  $\mathbf{M}(\mathbf{A})\mathbf{H}\mathbf{M}(\mathbf{A})\mathbf{y} = \lambda\mathbf{M}(\mathbf{A})\mathbf{y}$  and  $\mathbf{H}\mathbf{M}(\mathbf{A})\mathbf{H}\mathbf{y} = \lambda\mathbf{H}\mathbf{y}$ . By the Rayleigh-Ritz characterization of eigenvalues, an interpretation of the last two terms in the product follows:

$$\inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y} \rangle} = \inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{H}\mathbf{M}(\mathbf{A})\mathbf{H}\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} = \lambda_{\min}(\mathbf{H}\mathbf{M}(\mathbf{A})).$$

and

$$\inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{H}\mathbf{y}, \mathbf{y} \rangle} = \left( \sup_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A})\mathbf{H}\mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y} \rangle} \right)^{-1} = \frac{1}{\lambda_{\max}(\mathbf{H}\mathbf{M}(\mathbf{A}))},$$

For the remaining term, recalling from (2) that  $\mathbf{N}(\mathbf{A})$  denotes the skew Hermitian part of  $\mathbf{A}$ , a very useful result is the following.

**Lemma 1** (Corollary 3 in [20], see also [19, 18]). *Let  $\mathbf{A} \in \mathbb{C}^n$  be positive definite and  $c \in \mathbb{R}$ . The matrix  $c\mathbf{M}(\mathbf{A}^{-1}) - \mathbf{M}(\mathbf{A})^{-1}$  is positive definite if and only if  $c > 1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2$ , with*

$$\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})) = \max\{|t_j|; \pm it_j \text{ are the eigenvalues of } \mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})\}.$$

An equivalent reformulation of the lemma is that

$$\inf_{\mathbf{y} \neq 0} \frac{\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y} \rangle} = [1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2]^{-1}. \quad (14)$$

By definition,  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  is the spectral radius of  $\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})$ , *i.e.*, the norm of the eigenvalue of maximal norm. In this particular case, the eigenvalues of  $(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  are conjugate pairs of pure imaginary numbers  $\pm i|t_j|$ , as well as 0 if the order of  $\mathbf{A}$  is odd.

Putting everything together, we get:

**Theorem 6** (Convergence of WHP-GCR for positive definite  $\mathbf{A}$ ). *Assume that the operator  $\mathbf{A} \in \mathbb{K}^{n \times n}$  is positive definite. Assume also that the preconditioner  $\mathbf{H} \in \mathbb{K}^{n \times n}$  is hpd and that right preconditioned GCR is applied in the inner product induced by  $\mathbf{H}$ . The  $i$ -th iterate of Algorithm 3 satisfies*

$$\frac{\|\mathbf{r}_i\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq \left[ 1 - \frac{\lambda_{\min}(\mathbf{H}\mathbf{M}(\mathbf{A}))}{\lambda_{\max}(\mathbf{H}\mathbf{M}(\mathbf{A}))} \times \frac{1}{1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2} \right]^{i/2}. \quad (15)$$

The same result holds for all truncated and restarted versions of WHP-GCR and WHP-GMRES, including WHP-MR.

The estimate in the theorem has split the residual bound into two components:

- the condition number of the Hermitian part of  $\mathbf{A}$  preconditioned by  $\mathbf{H}$ ,
- a measure of the strength of non-Hermitianness of  $\mathbf{A}$  that is independent of  $\mathbf{H}$ .

**Corollary 2.** *Under the assumptions of Theorem 6, if  $\mathbf{H}$  is a preconditioner for  $\mathbf{M}(\mathbf{A})$  such that the condition number of the preconditioned operator is  $\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))$  then*

$$\frac{\|\mathbf{r}_i\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq \left[ 1 - \frac{\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))^{-1}}{1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2} \right]^{i/2}. \quad (16)$$

The same result holds for all truncated and restarted versions of WHP-GCR and WHP-GMRES, including WHP-MR.

In particular, if  $\mathbf{H}$  is a domain decomposition preconditioner such that  $\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))$  can be bounded independently of the number of subdomains, then the convergence bound above does not depend on the number of subdomains either. In other words, if a scalable domain decomposition method is known for the Hermitian part of  $\mathbf{A}$ , the proposed algorithm for the non-Hermitian problem is also scalable. For many problems arising from the discretization of PDEs,  $\mathbf{M}(\mathbf{A})$  is derived from a differential operator of lower order than that producing  $\mathbf{N}(\mathbf{A})$  so  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  is bounded independently of the discretization step (an observation already made in a different context in [38] and [12][Section 10]). If  $\mathbf{H}$  can also be chosen so that the preconditioned Hermitian part is well conditioned independently of  $h$  then the residual bound is  $h$ -independent.

In cases where a factorization of the preconditioner into  $\mathbf{H} = \mathbf{S}\mathbf{S}^*$  is available, split preconditioning by  $\mathbf{S}^*$  on the left and  $\mathbf{S}$  on the right can be applied. It can be shown that this is equivalent to WHP-GMRES. With this remark, there is a connection between WHP-GMRES and the algorithm in Section 10.4 of [12]. However, the bound in [12] is less sharp (the condition number is squared) and split preconditioning is very restrictive.

Next, it is illustrated that WHP-GCR is efficient for mildly non-Hermitian problems. If the preconditioner is well chosen, the algorithm is scalable and optimal (in the sense that convergence does not depend on the discretization step).

## 5 Illustration: Convection-Diffusion-Reaction

In this section, the problem considered is the convection-diffusion-reaction problem posed in  $\Omega = [0, 1]^2$ . It is a real problem ( $\mathbb{K} = \mathbb{R}$ ), so Hermitian means symmetric. The presentation, notation and test case are inspired by [3, 4]. The strong formulation of the problem is:

$$\begin{aligned} c_0 u + \operatorname{div}(\mathbf{a}u) - \operatorname{div}(\nu \nabla u) &= f \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned}$$

The variational formulation is: Find  $\mathbf{u} \in H_0^1(\Omega)$  such that

$$\underbrace{\int_{\Omega} \left( \left( c_0 + \frac{1}{2} \operatorname{div} \mathbf{a} \right) uv + \nu \nabla u \cdot \nabla v \right)}_{\text{symmetric part}} + \int_{\Omega} \left( \frac{1}{2} \mathbf{a} \cdot \nabla uv - \frac{1}{2} \mathbf{a} \cdot \nabla vu \right) = \int_{\Omega} f v,$$

for all  $v \in H_0^1(\Omega)$ .

The right hand side is chosen as

$$f(x, y) = \exp(-10((x - 0.5)^2 + (y - 0.1)^2))$$

and the convection field as

$$\mathbf{a}(x, y) = 2\pi[-(y - 0.1), x - 0.5].$$

The reaction coefficient  $c_0 > 0$  and viscosity  $\nu > 0$  are chosen to be constant over  $\Omega$ . Varying  $c_0$  and  $\nu$  would not cause any additional difficulty since the proposed preconditioner handles heterogeneous  $c_0$  and  $\nu$  (in the Hermitian part). The problem is discretized by Lagrange  $\mathbb{P}_1$  finite elements. The WHP-GCR algorithm is implemented in Freefem++ [17] with the ffdm library [36]. Except in the last case (where it is specified otherwise), all iteration counts correspond to the number of iterations needed to reach  $\|\mathbf{r}_i\|_{\mathbf{H}} < 10^{-6} \|\mathbf{b}\|_{\mathbf{H}}$  starting from a zero initial guess.

Let  $\mathbf{M}(\mathbf{A})$  be the matrix corresponding to the discretization of the symmetric part of the problem.  $\mathbf{M}(\mathbf{A})$  is preconditioned by the Additive Schwarz domain decomposition method with the GenEO coarse space [32, 33]. The partition of  $\Omega$  into  $N$  subdomains  $\Omega^s$  is computed automatically by Metis. One layer of overlap is added to each  $\Omega^s$ . Letting  $\mathbf{R}^{s\top}$  ( $s = 1, \dots, N$ ) denote the prolongation by zero of local finite element functions (in  $\Omega^s$ ) to the whole of  $\Omega$ , the preconditioner can be written as:

$$\mathbf{H} = \mathbf{\Pi} \sum_{s=1}^N \mathbf{R}^{s\top} \underbrace{(\mathbf{R}^s \mathbf{M}(\mathbf{A}) \mathbf{R}^{s\top})^{-1}}_{\text{local solves}} \mathbf{R}^s \mathbf{\Pi}^{\top} + \mathbf{R}^{0\top} \underbrace{(\mathbf{R}^0 \mathbf{M}(\mathbf{A}) \mathbf{R}^{0\top})^{-1}}_{\text{coarse solve}} \mathbf{R}^0,$$

where  $\mathbf{\Pi} = \mathbf{I} - \mathbf{R}^{0\top} (\mathbf{R}^0 \mathbf{M}(\mathbf{A}) \mathbf{R}^{0\top})^{-1} \mathbf{R}^0 \mathbf{M}(\mathbf{A})$  is the coarse projector (also known as a deflation operator) and the vectors in  $\mathbf{R}^{0\top}$  span the coarse space (or deflation space). The particularity of GenEO is that the coarse vectors are constructed by solving the low frequency eigenmodes for a generalized eigenvalue problem in each subdomain. The user provides a threshold  $\tau > 0$ , e.g.,  $\tau = 0.15$ . The corresponding ffdm options are

```
-ffdm_schwarz_method asm
-ffdm_geneo_threshold 0.15
-ffdm_schwarz_coarse_correction BNN.
```

The condition number of the resulting preconditioned operator is bounded by

$$\kappa(\mathbf{H}\mathbf{M}(\mathbf{A})) \leq (k_0 + 1)(1 + k_1\tau),$$

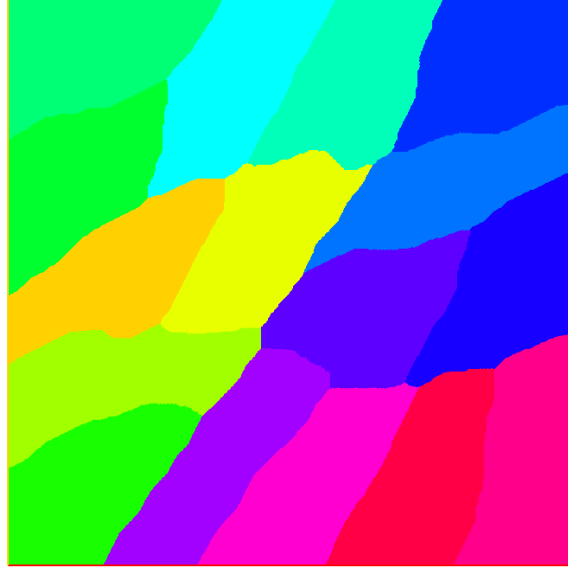


Figure 1: Partition into  $N = 16$  subdomains computed by Metis

where  $k_0$  denotes the maximal number of neighbours of a subdomain, and  $k_1$  denotes the maximal number of subdomains that a mesh element belongs to. These constants do not depend on the total number  $N$  of subdomains. In all examples  $\tau$  has been set to 0.15. As an illustration, the partition into  $N = 16$  subdomains provided by Metis is shown in Figure 1. For this case, it holds that  $k_0 = 6$  and  $k_1 = 3$ . Consequently, the condition number of the preconditioned symmetric part of the problem is bounded from above by

$$\kappa(\mathbf{H}\mathbf{M}(\mathbf{A})) \leq (6 + 1)(1 + 3 \times 0.15) = 10.15.$$

Injecting this into the bound from Theorem 6, it is obtained that the residuals produced by WHP-GCR or WHP-GMRES will be bounded by

$$\frac{\|\mathbf{r}_i\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq \left[ 1 - \frac{0.0985}{1 + \rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})^2)} \right]^{i/2}.$$

If, for example  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})^2) \leq 1$ , the bound gives

$$\frac{\|\mathbf{r}_i\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq \sqrt{\left[ 1 - \frac{0.0985}{1 + 1} \right]}^i = 0.975^i, \quad \frac{\|\mathbf{r}_{100}\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq 0.080 \quad \text{and} \quad \frac{\|\mathbf{r}_{500}\|_{\mathbf{H}}}{\|\mathbf{r}_0\|_{\mathbf{H}}} \leq 3.3 \cdot 10^{-6}.$$

The residual is guaranteed to decrease only by 2.5% at each iteration. As was previously explained, the bound is pessimistic for fully orthogonalized WP-GCR and WP-GMRES and we expect in practice to observe much faster decrease in residual.

**Value of  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$**  It remains to consider the value of  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$ . In the proof of [3][Lemma 4.5], by Cauchy Schwarz and some arithmetic identities, it is proved that

$$\langle \mathbf{y}, \mathbf{N}(\mathbf{A})\mathbf{x} \rangle \leq \alpha \|\mathbf{x}\|_{\mathbf{M}(\mathbf{A})} \|\mathbf{y}\|_{\mathbf{M}(\mathbf{A})}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Discretization step $h$	1/200	1/200	1/30	1/10
$\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$	0.3391	0.3389	0.3380	0.3136

Table 1: Computation of  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  when the discretization step  $h$  varies. Case  $\nu = c_0 = 1$ .

201 x 201 dofs				
Number of subdomains	4	8	16	32
Iteration count	59	59	58	58

501 x 501 dofs				
Number of subdomains	4	8	16	32
Iteration count	111	113	114	114

Table 2: Scalability. In each table the same problem (with  $201^2$  or  $501^2$  dofs) is solved for an increasing number of subdomains. The iteration count remains constant. Case  $\nu = c_0 = 1$ .

with  $\alpha = \frac{1}{2} \frac{\|\mathbf{a}\|_{L^\infty(\Omega)}}{\sqrt{\inf(\nu) \inf(c_0 + \frac{1}{2} \operatorname{div}(\mathbf{a}))}}$ . Consequently, for  $\mathbf{x} \neq \mathbf{0}$ , it holds that

$$\begin{aligned} \|\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})\mathbf{x}\|_{\mathbf{M}(\mathbf{A})}^2 &\leq \alpha \langle \mathbf{x}, \mathbf{M}(\mathbf{A})\mathbf{x} \rangle^{1/2} \langle \mathbf{N}(\mathbf{A})\mathbf{x}, \mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})\mathbf{x} \rangle^{1/2} \\ &= \alpha \|\mathbf{x}\|_{\mathbf{M}(\mathbf{A})} \|\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})\mathbf{x}\|_{\mathbf{M}(\mathbf{A})}. \end{aligned}$$

The spectral radius of a matrix is bounded from above by any of its natural norms, and in particular by the norm induced by  $\mathbf{M}(\mathbf{A})$ , from which it immediately follows that

$$\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})) \leq \|\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})\|_{\mathbf{M}(\mathbf{A})} \leq \frac{1}{2} \frac{\|\mathbf{a}\|_{L^\infty(\Omega)}}{\sqrt{\inf(\nu) \inf(c_0 + \frac{1}{2} \operatorname{div}(\mathbf{a}))}}. \quad (17)$$

Bound (17) for  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  does not depend on the discretization step  $h$  so neither does the overall convergence bound.

In Table 1, the actual value of  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  computed by Octave's *eigs* is given in the case where  $c_0 = \nu = 1$ . The discretization step varies between  $h = 1/10$  and  $h = 1/200$  and  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))$  varies only between 0.31 and 0.34. In comparison, bound (17) gives  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})) \leq 2.29$ . The bound is approximately 10 times larger than the computed value of the spectral radius.

**Scalability (Table 2)** Since Additive Schwarz with the GenEO coarse space is scalable, it has been proved that the overall algorithm is scalable. This is checked by solving the same problem for different partitions into subdomains (all computed by Metis). Two discretizations are considered:  $h = 1/200$  and  $h = 1/500$ . For this test  $\nu = c_0 = 1$ . It is observed that the method is indeed scalable: the iteration counts reported in Table 2 do not depend on the number of subdomains. A dependency on  $h$  is observed and this is studied next.

**Dependency on  $h$  (Table 3)** The influence of  $h$  on the iteration count is studied in Table 3. The partition is set to  $N = 8$  subdomains. First the problem is solved with  $c_0 = \nu = 1$ . It has been shown that  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})) \leq 2.29$  so the symmetric part and skew-symmetric part of the problem matrix are of comparable magnitude. When the discretization step decreases, the iteration count increases almost linearly. This is not



	$1/h$	2000	1000	500	200	100
	# dofs ( $n$ )	4 004 001	1 002 001	251 001	40 401	10 201
$c_0 = \nu = 1$	it. count	403	217	113	59	37
$c_0 = \nu = 10$	it. count	61	36	23	16	13
For the test below (and only for this test) $\mathbf{a} = [100, 0]$ and $c_0 = \nu = 1$						
$\mathbf{a} = [100, 0]$	it. count	9	10	10	11	12

Table 3: Dependency on  $h$ . Each line corresponds to a value of  $(\nu, c_0)$ . Case  $N = 8$  subdomains. In the last line, and only for this test, the convection field is  $\mathbf{a} = [100, 0]$  and  $c_0 = \nu = 1$

$c_0 = \nu$	0.1	1	5	10	only symmetric part
Iteration count	> 200 (0.020)	113	36	23	10

Table 4: Influence of the relative importance of the symmetric term and the skew-symmetric term. Case  $N = 8$  subdomains and  $h = 1/500$ . The value in parenthesis corresponds to the relative residual when the algorithm stopped after 200 iterations.

what was expected from the theory. However, as  $h \rightarrow 0$ , spurious modes also pollute the solution (independently of any preconditioner). Stabilization would fix this issue, *e.g.*, by the SUPG method as in [3][Section 11.8]. When  $c_0 = \nu = 10$ ,  $\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})) \leq 0.72$  so the symmetric part of the problem dominates slightly. The iteration count does not vary as much with  $h$  (it gets multiplied by 5 when  $h$  gets divided by 20). In the last line of Table 3, the case where the convection field  $\mathbf{a}$  is the constant vector  $[100, 0]$  is considered. The iteration count does not vary with  $h$  as predicted by the theory.

**Dependency on strength of non-symmetry (Table 4)** For this test,  $h = 1/500$  and  $N = 8$ . The value of  $\nu$  and  $c_0$  are varied and the corresponding iteration counts are reported in Table 4. As expected, the problem converges very fast when the symmetric part dominates and not so fast otherwise. This is expected by design of the algorithm.

**GMRES in the Euclidean inner product (Table 5)** It has already been observed that changing the inner product in GMRES does not influence its convergence very much [30, 6]. As a final test, we run a scalability test for  $h = 1/500$  in two settings  $\nu = c_0 = 1$  and  $\nu = c_0 = 10$ . The number of subdomains varies between 4 and 32. Two algorithms are applied with the same preconditioner as previously: right preconditioned GMRES and WHP-GCR. The difference is that GMRES works in the Euclidean inner product while WHP-GCR (which produces the same iterates as WHP-GMRES) works in the  $\mathbf{H}$ -inner product. For both, the stopping criterion has been set to  $\|\mathbf{r}_i\| < 10^{-6}\|\mathbf{b}\|$  where the norm is the Euclidean norm. It is absolutely remarkable (and not a typo) that the iteration counts reported in Table 5 are almost identical. This means that in practice using the hpd preconditioner and the Euclidean norm will most likely give results that are in agreement with the developed theory. The advantage is to save the effort of implementing WHP-GCR if GMRES is already available. The extra cost of running WHP-GCR is small compared to GCR (or GMRES) since it is only the cost of storing one (or two) extra vectors per iteration.

**Case  $\nu = c_0 = 1$**

Number of Subdomains	4	8	16	32
GMRES	118	122	124	124
WHP-GCR (Euclidean stopping criterion)	118	121	124	124

**Case  $\nu = c_0 = 10$**

Number of Subdomains	4	8	16	32
GMRES	31	32	33	33
WHP-GCR (Euclidean stopping criterion)	31	32	33	33

Table 5: Influence of the inner product: GMRES in Euclidean norm compared to WHP-GCR. The stopping criterion is in Euclidean norm for both algorithms. Case  $h = 1/500$ . The number of subdomains  $N$  varies and two cases are considered :  $\nu = c_0 = 1$  and  $\nu = c_0 = 10$ .

## 6 Conclusion

In this article, the convergence of GMRES, GCR and their truncated and restarted versions has been studied. The influence of the preconditioner and the inner product have been made explicit. It has been proposed, even for non-Hermitian problems to apply a hpd preconditioner  $\mathbf{H}$ . Then, GMRES or GCR can be applied in the  $\mathbf{H}$  inner product. This is referred to as WHP-GMRES. A new convergence result is proved for cases where  $\mathbf{A}$  is positive definite. The two terms in the convergence bound are the condition number of the Hermitian part of  $\mathbf{A}$  once preconditioned by  $\mathbf{H}$  and the spectral radius of  $\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A})$ . This last term can be seen as a measure of the strength of non-Hermitianness. It does not depend on the choice of the preconditioner. A particular application is the case where  $\mathbf{H}$  is a domain decomposition preconditioner. If the preconditioner applied to the Hermitian part of  $\mathbf{A}$  leads to a scalable method, then WHP-GCR will be scalable too. Numerical results have confirmed these findings. It remains to improve the algorithm in cases where the problem is strongly non-Hermitian or indefinite.

## A Appendix

Algorithms 4 and 5 propose two alternate implementations of WHP-GCR (Algorithm 3) that require no more storage than the usual GCR algorithm.

The notation  $\tilde{\cdot}$  has been used to emphasize that vectors with a tilde do not get orthogonalized and saved. In exact arithmetic all three versions produce the same iterates. The Euclidean residual  $\mathbf{r}_i$  is not updated in Algorithm 4 which may be a drawback. In finite precision, computing  $\alpha_i$  from vectors that have not been explicitly orthogonalized could lead to inaccuracy. This is why Algorithm 3 is emphasized and implemented in the numerical result section.

**Algorithm 4** Alternate WHP-GCR

---

**Require:**  $\mathbf{x}_0 \in \mathbb{R}^n$   
 $\mathbf{z}_0 = \mathbf{H}(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$   
 $\mathbf{p}_0 = \mathbf{z}_0$   
 $\tilde{\mathbf{q}}_0 = \mathbf{A}\mathbf{p}_0$   
 $\mathbf{y}_0 = \mathbf{H}\tilde{\mathbf{q}}_0$   
**for**  $i = 0, 1, \dots$ , convergence **do**  
 $\delta_i = \langle \mathbf{y}_i, \tilde{\mathbf{q}}_i \rangle; \quad \gamma_i = \langle \tilde{\mathbf{q}}_i, \mathbf{z}_i \rangle$   
 $\alpha_i = \gamma_i / \delta_i$   
 $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$   
  
 $\mathbf{z}_{i+1} = \mathbf{z}_i - \alpha_i \mathbf{y}_i$   
 $\mathbf{p}_{i+1} = \mathbf{z}_{i+1}$   
 $\tilde{\mathbf{q}}_{i+1} = \mathbf{A}\mathbf{z}_{i+1}$   
 $\mathbf{y}_{i+1} = \mathbf{H}\tilde{\mathbf{q}}_{i+1}$   
**for**  $j = 0, \dots, i$  **do**  
 $\Phi_{i,j} = \langle \mathbf{y}_j, \tilde{\mathbf{q}}_{i+1} \rangle$   
 $\beta_{i,j} = \Phi_{i,j} / \delta_j^{-1}$   
**end for**  
 $\mathbf{p}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{p}_j$   
  
 $\mathbf{y}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{y}_j$   
**end for**  
**return**  $\mathbf{x}_{i+1}$

---

**Algorithm 5** Alternate WHP-GCR

---

**Require:**  $\mathbf{x}_0 \in \mathbb{R}^n$   
 $\mathbf{z}_0 = \mathbf{H}(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$   
 $\mathbf{p}_0 = \mathbf{z}_0$   
 $\mathbf{q}_0 = \mathbf{A}\mathbf{p}_0$   
 $\tilde{\mathbf{y}}_0 = \mathbf{H}\mathbf{q}_0$   
**for**  $i = 0, 1, \dots$ , convergence **do**  
 $\delta_i = \langle \tilde{\mathbf{y}}_i, \mathbf{q}_i \rangle; \quad \gamma_i = \langle \mathbf{q}_i, \mathbf{z}_i \rangle$   
 $\alpha_i = \gamma_i / \delta_i$   
 $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$   
 $\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i$   
 $\mathbf{z}_{i+1} = \mathbf{z}_i - \alpha_i \mathbf{y}_i$   
 $\mathbf{p}_{i+1} = \mathbf{z}_{i+1}$   
 $\mathbf{q}_{i+1} = \mathbf{A}\mathbf{z}_{i+1}$   
 $\tilde{\mathbf{y}}_{i+1} = \mathbf{H}\mathbf{q}_{i+1}$   
**for**  $j = 0, \dots, i$  **do**  
 $\Phi_{i,j} = \langle \tilde{\mathbf{y}}_j, \mathbf{q}_{i+1} \rangle$   
 $\beta_{i,j} = \Phi_{i,j} / \delta_j^{-1}$   
**end for**  
 $\mathbf{p}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{p}_j$   
  
 $\mathbf{q}_{i+1}^- = \sum_{j=0}^i \beta_{i,j} \mathbf{q}_j$   
**end for**  
**return**  $\mathbf{x}_{i+1}$

---

## References

- [1] Z.-Z. BAI, G. H. GOLUB, AND M. K. NG, *Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 603–626, <https://doi.org/10.1137/S0895479801395458>.
- [2] B. BECKERMANN, S. A. GOREINOV, AND E. E. TYRTYSHNIKOV, *Some remarks on the Elman estimate for GMRES*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 772–778, <https://doi.org/10.1137/040618849>.
- [3] M. BONAZZOLI, X. CLAEYS, F. NATAF, AND P.-H. TOURNIER, *Analysis of the SO-RAS domain decomposition preconditioner for non-self-adjoint or indefinite problems*, J. Sci. Comput., 89 (2021), p. 27, <https://doi.org/10.1007/s10915-021-01631-8>. Id/No 19.
- [4] M. BONAZZOLI, X. CLAEYS, F. NATAF, AND P.-H. TOURNIER, *How does the partition of unity influence SORAS preconditioner?* working paper or preprint, Dec. 2022, <https://hal.science/hal-03882577>.
- [5] N. BOOTLAND, V. DOLEAN, I. G. GRAHAM, C. MA, AND R. SCHEICHL, *GenEO coarse spaces for heterogeneous indefinite elliptic problems*, arXiv preprint arXiv:2103.16703, (2021), <https://doi.org/10.48550/ARXIV.2103.16703>.
- [6] N. BOOTLAND, V. DOLEAN, I. G. GRAHAM, C. MA, AND R. SCHEICHL, *Overlapping schwarz methods with GenEO coarse spaces for indefinite and non-self-adjoint problems*, arXiv preprint arXiv:2110.13537, (2021), <https://doi.org/10.48550/ARXIV.2110.13537>.
- [7] X.-C. CAI, *Some domain decomposition algorithms for nonselfadjoint elliptic and parabolic partial differential equations*, PhD thesis, New York University, 1989, <https://archive.org/details/somedomaindecomp00caix/page/70/mode/2up>.
- [8] X.-C. CAI, W. D. GROPP, AND D. E. KEYES, *A comparison of some domain decomposition algorithms for nonsymmetric elliptic problems*, in Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations, Philadelphia, PA, 1992.
- [9] X.-C. CAI AND O. B. WIDLUND, *Domain decomposition algorithms for indefinite elliptic problems*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 243–258, <https://doi.org/10.1137/0913013>, [semanticscholar.org/paper/106a429bba2ed3c761256c704edacd40770b8e1b](https://www.semanticscholar.org/paper/106a429bba2ed3c761256c704edacd40770b8e1b).
- [10] J. G. CALVO AND O. B. WIDLUND, *An adaptive choice of primal constraints for BDDC domain decomposition algorithms*, Electron. Trans. Numer. Anal, 45 (2016), pp. 524–544.
- [11] S. C. EISENSTAT, H. C. ELMAN, AND M. H. SCHULTZ, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357, <https://doi.org/10.1137/0720023>, [semanticscholar.org/paper/1fc0ce44410db9fa9767a8e2661638c9269e3d10](https://www.semanticscholar.org/paper/1fc0ce44410db9fa9767a8e2661638c9269e3d10).
- [12] H. C. ELMAN, *Iterative methods for large, sparse, nonsymmetric systems of linear equations*, PhD thesis, Yale University, 1982.
- [13] A. ESSAI, *Weighted FOM and GMRES for solving nonsymmetric linear systems*, Numer. Algorithms, 18 (1998), pp. 277–292, <https://doi.org/10.1023/A:1019177600806>.
- [14] I. G. GRAHAM, E. A. SPENCE, AND E. VAINIKKO, *Domain decomposition preconditioning for high-frequency Helmholtz problems with absorption*, Math. Comput., 86 (2017), pp. 2089–2127, <https://doi.org/10.1090/mcom/3190>.

- [15] I. G. GRAHAM, E. A. SPENCE, AND J. ZOU, *Domain decomposition with local impedance conditions for the Helmholtz equation with absorption*, SIAM J. Numer. Anal., 58 (2020), pp. 2515–2543, <https://doi.org/10.1137/19M1272512>.
- [16] R. HAFERSSAS, P. JOLIVET, AND F. NATAF, *An additive Schwarz method type theory for Lions’s algorithm and a symmetrized optimized restricted additive Schwarz method*, SIAM Journal on Scientific Computing, 39 (2017), pp. A1345–A1365.
- [17] F. HECHT, *New development in freefem++*, J. Numer. Math., 20 (2012), pp. 251–265, <https://freefem.org/>.
- [18] C. R. JOHNSON, *Matrices whose hermitian part is positive definite*, PhD thesis, California Institute of Technology, 1972.
- [19] C. R. JOHNSON, *An inequality for matrices whose symmetric part is positive definite*, Linear Algebra Appl., 6 (1973), pp. 13–18, [https://doi.org/10.1016/0024-3795\(73\)90003-7](https://doi.org/10.1016/0024-3795(73)90003-7).
- [20] C. R. JOHNSON, *Inequalities for a complex matrix whose real part is positive definite*, Trans. Am. Math. Soc., 212 (1975), pp. 149–154, <https://doi.org/10.2307/1998618>.
- [21] P. JOLIVET, F. HECHT, F. NATAF, AND C. PRUD’HOMME, *Scalable domain decomposition preconditioners for heterogeneous elliptic problems*, in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC ’13, New York, NY, USA, 2013, Association for Computing Machinery, <https://doi.org/10.1145/2503210.2503212>, <https://doi.org/10.1145/2503210.2503212>.
- [22] A. KLAWONN, M. KUHN, AND O. RHEINBACH, *Adaptive coarse spaces for FETI-DP in three dimensions*, SIAM Journal on Scientific Computing, 38 (2016), pp. A2880–A2911.
- [23] J. LIESEN AND Z. STRAKOŠ, *Krylov subspace methods. Principles and analysis*, Numer. Math. Sci. Comput., Oxford: Oxford University Press, reprint of the 2013 hardback edition ed., 2015, <https://doi.org/10.1093/acprof:oso/9780199655410.001.0001>.
- [24] J. LIESEN AND P. TICHÝ, *The field of values bound on ideal gmres*, arXiv preprint arXiv:1211.5969, (2012), <https://doi.org/10.48550/ARXIV.1211.5969>.
- [25] G. MEURANT AND J. D. TEBBENS, *The role eigenvalues play in forming GMRES residual norms with non-normal matrices*, Numer. Algorithms, 68 (2015), pp. 143–165, <https://doi.org/10.1007/s11075-014-9891-3>.
- [26] N. M. NACHTIGAL, S. C. REDDY, AND L. N. TREFETHEN, *How fast are non-symmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795, <https://doi.org/10.1137/0613049>.
- [27] C. PECHSTEIN AND C. R. DOHRMANN, *A unified framework for adaptive BDDC*, Electron. Trans. Numer. Anal, 46 (2017), p. 3.
- [28] Y. SAAD, *Iterative methods for sparse linear systems.*, Philadelphia, PA: SIAM Society for Industrial and Applied Mathematics, 2nd ed. ed., 2003.
- [29] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869, <https://doi.org/10.1137/0907058>.
- [30] M. SARKIS AND D. B. SZYLD, *Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms*, Comput. Methods Appl. Mech. Eng., 196 (2007), pp. 1612–1621, <https://doi.org/10.1016/j.cma.2006.03.027>.

- [31] N. SPILLANE, *An abstract theory of domain decomposition methods with coarse spaces of the genEO family*, arXiv preprint arXiv:2104.00280, (2021), <https://doi.org/10.48550/ARXIV.2104.00280>.
- [32] N. SPILLANE, V. DOLEAN, P. HAURET, F. NATAF, C. PECHSTEIN, AND R. SCHEICHL, *A robust two-level domain decomposition preconditioner for systems of PDEs*, C. R. Math. Acad. Sci. Paris, 349 (2011), pp. 1255–1259, <https://doi.org/10.1016/j.crma.2011.10.021>.
- [33] N. SPILLANE, V. DOLEAN, P. HAURET, F. NATAF, C. PECHSTEIN, AND R. SCHEICHL, *Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps*, Numer. Math., 126 (2014), pp. 741–770, <https://doi.org/10.1007/s00211-013-0576-y>.
- [34] N. SPILLANE AND D. J. RIXEN, *Automatic spectral coarse spaces for robust FETI and BDD algorithms*, Int. J. Numer. Meth. Engng., 95 (2013), pp. 953–990.
- [35] A. TOSELLI AND O. WIDLUND, *Domain decomposition methods – algorithms and theory.*, vol. 34 of Springer Ser. Comput. Math., Berlin: Springer, 2005.
- [36] P.-H. TOURNIER, P. JOLIVET, AND F. NATAF, *FFDDM: Freefem domain decomposition method*. <https://doc.freefem.org/documentation/ffddm/index.html>, 2019.
- [37] P. K. VINSOME, *Orthomin, an iterative method for solving sparse sets of simultaneous linear equations*, in SPE Symposium on Numerical Simulation of Reservoir Performance, OnePetro, 1976.
- [38] O. WIDLUND, *A Lanczos method for a class of nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 15 (1978), pp. 801–812, <https://doi.org/10.1137/0715053>.
- [39] J. XU AND X.-C. CAI, *A preconditioned GMRES method for nonsymmetric or indefinite problems*, Math. Comput., 59 (1992), pp. 311–319, <https://doi.org/10.2307/2153059>.