A NOTE ON INDEFINITE MATRIX SPLITTING AND PRECONDITIONING

ANDY WATHEN*

Dedicated to Daniel Szyld on his 70th birthday

Abstract. The solution of systems of linear(ized) equations lies at the heart of many problems in Scientific Computing. In particular for systems of large dimension, iterative methods are a primary approach. Stationary iterative methods are generally based on a matrix splitting, whereas for polynomial iterative methods such as Krylov subspace iteration, the splitting matrix is the preconditioner. The smoother in a multigrid method is generally a stationary or polynomial iteration.

Here we consider real symmetric indefinite and complex Hermitian indefinite coefficient matrices and prove that no splitting matrix can lead to a contractive stationary iteration unless the inertia is exactly preserved. This has consequences for preconditioning for indefinite systems and smoothing for multigrid as we further describe.

Key words. Iterative methods for linear systems, indefinite matrices, matrix splitting, preconditioning, smoothing

AMS subject classifications. 65F08, 65F10, 65M55

Introduction and Main Result. An elementary approach for the iterative solution of an invertible linear system of equations

$$Ax = b$$

is to split

$$A = M - N$$

and from some starting vector x_0 to iterate

$$\text{solve} \quad Mx_{k+1} = Nx_k + b, \quad k = 0, 1, \dots \; . \tag{0.1}$$

M is here called the splitting matrix or the preconditioner and it is required to be invertible. For any x_0 the sequence $\{x_k\}$ converges to the solution if and only if all eigenvalues of $M^{-1}N = I - M^{-1}A$ lie strictly inside the unit disc; in such a situation the iteration is contractive. Equivalently, all eigenvalues of $M^{-1}A$ must lie in $B(1,1)$, the open unit ball centred at 1. In particular, if ever $M^{-1}A$ has an eigenvalue with negative real part then (0.1) certainly can not be contractive. For the situation when A is singular but the equations are consistent, see [26].

More recent use of preconditioning has utilised polynomial iterative methods, either with fixed polynomial sequences as in Chebyshev (semi-)iteration or implicitly defined polynomials as in Krylov subspace iteration (for both, see for example [21],[25]).

Krylov subspace methods are typically applied to the preconditioned system $M^{-1}Ax = M^{-1}b$ so that it is polynomials in $M^{-1}A$ that are relevant in the context of convergence; the polynomials $p(\cdot)$ must satisfy the consistency condition $p(0) = 1$. Convergence of a Krylov subspace iteration will be rapid if $\max |p(\lambda)|$ reduces significantly as the degree of the polynomials p is increased. Here the maximum is over all eigenvalues λ of $M^{-1}A$. Rapid convergence is typically associated with clustering of

*Mathematical Institute, Oxford University, UK (wathen@maths.ox.ac.uk) and Computational Mathematics Group, Rutherford Appleton Laboratory, UK (andrew.wathen@stfc.ac.uk)

these eigenvalues away from the origin [14, Chapter 3]. If all eigenvalues are real and lie only to one side of the origin, then a Conjugate Gradient Krylov subspace method can be utilised. If all eigenvalues of $M^{-1}A$ are real but lie on both sides of the origin, then a less efficient minimum residual method such as MINRES is typically used and convergence can be less rapid (see for example [7]). Whenever there are non-real eigenvalues then the more general (and more expensive) GMRES iterative method [22] is ubiquitously employed as a Krylov subspace iterative solver.

In terms of methods with fixed polynomial sequences, usually it is polynomials in $I - M^{-1}A$ that are employed so that the consisitency condition is $p(1) = 1$. The classical Chebyshev semi-iterative method based on the Chebyshev polynomials on a single interval (see for example [29, Chapter 5]) can be used when all eigenvalues are real and lie to one side of unity. It is more difficult to identify polynomials which lead to an efficient iterative method when all eigenvalues are real, but lie to both sides of 1. For the general situation with complex eigenvalues, see for example [6].

It is thus possible to have a convergent polynomial iterative method for all eigenvalue distributions. Nevertheless, when eigenvalues are all real, convergence is usually slower when $M^{-1}A$ has eigenvalues with different signs rather than just of one sign.

If A is real and symmetric, it's *inertia* (the triple of number of positive, zero and negative eigenvalues) is (p, z, n) for some integers $p, z, n \geq 0$. $z = 0$ for any invertible matrix. A matrix with inertia $(p, 0, 0)$ is *positive definite* and a matrix with inertia $(0, 0, n)$ is *negative definite*; in either case the matrix is said to be *definite*.

LEMMA. *If matrices A and M with the same dimension are real symmetric and invertible with different inertia, then $M^{-1}A$ has at least one negative real eigenvalue and thus the stationary iteration (0.1) can not be contractive.*

Proof. For any real parameter θ, the real symmetric matrix

$$T(\theta) = (1 - \theta)A + \theta M \tag{0.2}$$

satisfies $T(0) = A, T(1) = M$ and so it's eigenvalues are real and depend continuously on the entries, therefore continuously on θ. Since the inertia of A and M are different, at least one eigenvalue of T must change sign as θ varies between 0 and 1, hence there must exist $\widehat{\theta} \in (0, 1)$ such that $T(\widehat{\theta})$ is singular (by the Intermediate Value Theorem). Thus

$$A + \frac{\widehat{\theta}}{(1 - \widehat{\theta})} M$$

is singular, so

$$\frac{\widehat{\theta}}{\widehat{\theta} - 1} < 0$$

is an eigenvalue of $M^{-1}A$ and the result follows. ∎

Remark. An identical result and proof holds also for Complex Hermitian matrices.

Remark. Having identical inertia is of course not sufficient for contraction of (0.1) as trivial examples like $A = diag(1, -1), M = -A$ show.

Remark. This lemma certainly does not rule out the existence of complex eigenvalues for $M^{-1}A$, equivalently complex generalised eigenvalues of the 'pencil' $A - \lambda M$ even in the case of real, symmetric matrices A, M. Indeed, unless at least one of A, M is definite, it is generic that complex conjugate pairs of eigenvalues exist. When either A or M are definite, all eigenvalues of the pencil $A - \lambda M$ must be real.

There is much analysis of matrix splittings when A is positive definite (see for example the classic book by Richard Varga [29]). In this case the above lemma indicates that M should also be positive definite; this is no surprise!

However, positive definite preconditioning is commonly applied to indefinite symmetric matrices, notably in block diagonal preconditioning of saddlepoint systems ([2],[24],[20]). In this situation, the eigenvalues of $M^{-1}A$ must all be real because of the similarity transform

$$(M^{\frac{1}{2}})^T(M^{-1}A)M^{-\frac{1}{2}} = M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$$

which is a real symmetric matrix (and so must have all real eigenvalues). Sylvester's Law of Inertia (see for example [12]) ensures in this case that the number of positive, negative and zero eigenvalues of $M^{-1}A$ and of A are the same because $M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$ is congruent to A since M and so $M^{\frac{1}{2}}$ are symmetric. (Note the unambiguous definition of $M^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}Q^T$ in terms of the diagonalisation $M = Q\Lambda Q$ when positive square roots are taken).

For this reason, an iterative method for indefinite systems such as MINRES is required when definite preconditioning is applied to an indefinite system. Convergence will depend on how small $|p(\lambda)|$ can be for all positive and negative eigenvalues for polynomials, p that satisfy the consistency condition $p(0) = 1$ as the degree of the polynomial is increased (see [7, Section 4.1] or for a more comprehensive description [11]). For example for block diagonal preconditioning of the Stokes saddle point system, see the convergence analysis in [30].

At another extreme, *constraint preconditioning* of saddle point systems ensures that the inertia of the preconditioner

$$M = \begin{bmatrix} W & B^T \\ B & 0 \end{bmatrix}$$

and the original saddle point matrix

$$A = \begin{bmatrix} H & B^T \\ B & 0 \end{bmatrix} \tag{0.3}$$

are identical at least when W, H are positive definite. Indeed, in the paper [16] that introduced constraint preconditioning, it is proved that all eigenvalues of the pencil $A-\lambda M$ are real and positive when W, H are positive definite. Clearly this is consistent with the Lemma above. A consequence is that a specialised and rapidly converging Conjugate Gradient method can be employed with constraint preconditioning [13]. For inexact constraint preconditioning, there are important eigenvalue bounds in [3], [4]. We comment that in the multigrid literature for flow problems, relevant references are [5], [31].

In the common situation where $H \in \mathbb{R}^{m\times m}$ is invertible in (0.3), the congruence transform of the saddle point matrix

$$A = \begin{bmatrix} H & B^T \\ B & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ BH^{-1} & 0 \end{bmatrix} \begin{bmatrix} H & 0 \\ 0 & -BH^{-1}B^T \end{bmatrix} \begin{bmatrix} I & H^{-1}B^T \\ 0 & I \end{bmatrix}$$

is well known. A consequence is that if H is positive definite and $B \in \mathbb{R}^{n\times m}$ is of full rank, then the inertia of A must be $(m, 0, n)$ by Sylvester's Law of Inertia. For such problems preservation of this known inertia in a preconditioner is possibly simpler.

This observation is implicitly employed in Vanka iteration for the indefinite saddle point system of Stokes flow [28] where an indefinite splitting matrix is used. It may not be immediately clear whether it shares the same inertia as the original Stokes matrix, however by careful construction, associating local problems to small numbers of pressure variables (Lagrange multipliers) often in a 'patch', the number of negative eigenvalues can be guaranteed to equal the number of Lagrange multipliers, m, and hence the inertia can be preserved [10],[9].

Unless the inertia is identical, the above lemma implies that the simple iteration (0.1) could not be contractive. Further, even standard Chebyshev polynomial acceleration (based on eigenvalues in one interval to the right of the origin) can not be convergent for any chosen interval if $M^{-1}A$ has negative as well as positive eigenvalues. This follows because Chebyshev polynomials grow exponentially in magnitude away from the chosen interval; in turn this implies that $|p(\lambda)|$ must be large (certainly greater than 1) for any negative eigenvalue λ. The most successful Vanka-like iterations must not only satisfy the criterion in the lemma above, they should cluster eigenvalues in an appropriate way. There are now many examples in the literature where Vanka-like iteration is used for smoothing in a multigrid context—for some recent examples see [1], [27], [15]—even though there remains limited theoretical literature on the convergence of Vanka-type iterations. Twenty years ago, Manservisi [19] declared that there was "nothing at all on Vanka-type smoothers" and Larin and Reusken [17] say that "As far as we know there is no convergence analysis of a multigrid method with a Vanka smoother applied...to Stokes problems". Even in 2022, Saberi et al. [23] can only refer to the paper of Manservisi for a proof of convergence of "some Vanka-type smoothers for the Stokes...equations", though Larin and Reusken noted that "smoothing properties are not, however, considered in that paper". Smoothing does not strictly require contraction, but non-contractive smoothers are not so common. Our theory here only relates to the symmetric Stokes flow situation with block Jacobi-like smoothing (additive Vanka relaxation), that is a situation where a symmetric preconditioner is used for a symmetric matrix problem. Nevertheless, if the iteration (0.1) is not contractive for the symmetric Stokes problem, it would seem risky to employ a related iteration for more complicated Navier-Stokes problems since these problems give rise to matrices that have additional nonsymmetric terms (see [7, Chapters 8 & 9]).

Indefinite systems necessarily arise for wave problems like the Helmholtz and Maxwell equations. For the Helmholtz problem for example, inertia will depend on the wave frequency and discretisation. Finding a preconditioner with identical inertia in such a situation can be expected to be challenging (see [8]).

Avoidance of crossing. Because $T(\theta)$ in (0.2) is symmetric for all θ, the concept of *avoidance of crossing* or *eigenvalue avoidance* due to Lax [18] applies. Though this may be of less importance than the Lemma above, it allows some slightly more precise statements the be made about the number of negative eigenvalues of $M^{-1}A$.

PROPOSITION. *If A has inertia $(p, 0, n)$ and M has inertia $(p + r, 0, n - r)$ where the integer r necessarily satisfies $-p \leq r \leq n$, then $M^{-1}A$ has $|r| + 2s$ real and negative eigenvalues for some $s \in \{0, 1, 2, \ldots, \lfloor \frac{p+n-r}{2} \rfloor\}$.*

Proof. Because $T(\theta)$ is real symmetric, eigenvalue avoidance implies that the trajectories of the eigenvalues $\lambda(\theta)$ of $T(\theta)$ do not intersect. Thus there must be $|r|$ distinct values $\widehat{\theta}_j \in (0, 1)$, $j = 1, 2, \ldots, |r|$ for which $T(\widehat{\theta})$ is singular. As in the lemma above, the real value

$$\frac{\widehat{\theta}_j}{\widehat{\theta}_j - 1} < 0$$

must be an eigenvalue of $M^{-1}A$ for $j = 1, 2, \ldots, |r|$.

There may be more values of $\theta \in (0,1)$ where $T(\theta)$ is singular; if such occurs, then the trajectory $\lambda(\theta)$ must intersect with the line $\lambda = 0$ an even number of times. The upper bound on s arises simply because there can be no more than $m + n$ eigenvalues of $M^{-1}A$ overall.

Example. The symmetric matrices

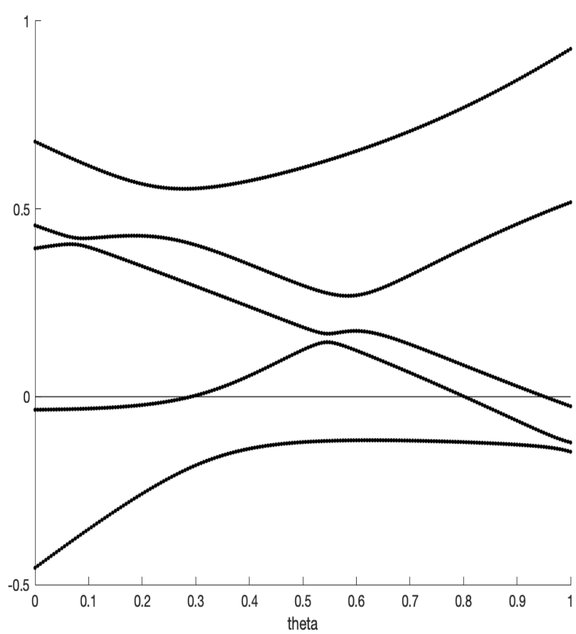
$$A = \begin{bmatrix} 0.33 & -.05 & -.29 & 0.01 & 0.01 \\ -.05 & 0.36 & -.11 & -.22 & -.19 \\ -.29 & -.11 & -.32 & 0.11 & -.01 \\ 0.01 & -.22 & 0.11 & 0.49 & -.12 \\ 0.01 & -.19 & -.01 & -.12 & 0.18 \end{bmatrix}, \quad M = \begin{bmatrix} 0.14 & 0.10 & 0.25 & 0.09 & -.28 \\ 0.10 & -.07 & 0.02 & 0.08 & -.11 \\ 0.25 & 0.02 & 0.49 & -.11 & -.23 \\ 0.09 & 0.08 & -.11 & 0.24 & -.34 \\ -.28 & -.11 & -.23 & -.34 & 0.35 \end{bmatrix}$$

have eigenvalues respectively

$$\lambda_A = -0.4553, -0.0346, 0.3949, 0.4560, 0.6791,$$

$$\lambda_M = -0.1464, -0.1216, -0.0252, 0.5174, 0.9258,$$

hence in the notation of the above Proposition $p = 3, n = 2, r = -1$. The eigenvalues of $T(\theta)$ are plotted as θ varies from 0 to 1; it is seen that $s = 1$ in this example.



The eigenvalues of $M^{-1}A$ are:

$$\lambda_{M^{-1}A} = -2.4405, -0.2468, -0.0506, 1.7245 \pm 0.8315i$$

with the three negative real eigenvalues corresponding respectively to the values

$$\widehat{\theta} = 0.2907, 0.8021, 0.9518$$

as seen in the plot.

Rather than $T(\theta)$, by considering the alternative homotopy

$$S(\theta) = (1-\theta)A + \theta(-M)$$

one can say something about positive real eigenvalues.

COROLLARY. *If A has inertia $(p,0,n)$ and M has inertia $(p+r,0,n-r)$ where the integer r necessarily satisfies $-p \le r \le n$, then $M^{-1}A$ has $|p+r-n|+2t$ real and positive eigenvalues for some $t \in \{0,1,2,\ldots,\min\left(\lfloor\frac{2p+r}{2}\rfloor, \lfloor\frac{2n-r}{2}\rfloor\right)\}$.*

Proof. Simply note that $S(0) = A$ has inertia $(p,0,n)$ and $S(1) = -M$ has inertia $(n-r,0,p+r)$ so that if $p \ne n-r$ then $S(\theta)$ must be singular for at least $|p+r-n|$ values $\widehat{\theta} \in (0,1)$ implying $|p+r-n|$ real positive eigenvalues $\frac{\widehat{\theta}}{1-\widehat{\theta}}$. Any further crossings of the axis must appear in pairs as in the Proposition above. ∎

In the example above we have $p = n-r$, (ie. the inertia of $A =$ the inertia of $-M$), hence the absence of any real positive eigenvalues; here $t = 0$.

Conclusions. The elementary lemma of this paper has implications for preconditioned iterative solution of symmetric linear systems. Unless the inertia of the coefficient matrix is preserved in a preconditioner/splitting matrix then simple iteration will generally be non-contractive. Correspondingly when a Krylov subspace iteration method is used, slower convergence might be expected if the inertia of the preconditioner does not match the inertia of the coefficient matrix.

Where smoothing for multigrid is based on an indefinite matrix splitting, the smoother can not be a contractive iteration if the inertia is not exactly preserved.

Acknowledgments. I am grateful to Hussam Al Daas for useful conversations about this work.

REFERENCES

[1] M. Anselmann and M. Bause, *Efficiency of local Vanka smoother geometric multigrid preconditioning for space-time finite element methods to the Navier–Stokes equations*, Proc. Appl. Math. Mech., 23 (2023), p. e202200088, https://doi.org/10.1002/pamm.202200088.

[2] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica, 14 (2005), pp. 1–137, https://doi.org/10.1017/S0962492904000212.

[3] L. Bergamaschi, J. Gondzio, M. Venturin, and G. Zilli, *Inexact constraint preconditioners for linear systems arising in interior point methods*, Comput. Optim. Appl., 36 (2007), pp. 137–147, https://doi.org/10.1007/s10589-006-9001-0.

[4] L. Bergamaschi, J. Gondzio, M. Venturin, and G. Zilli, *erratum to: Inexact constraint preconditioners for linear systems arising in interior point methods*, Comput. Optim. Appl., 49 (2011), pp. 401–406, https://doi.org/10.1007/s10589-009-9298-6.

[5] D. Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Appl. Numer. Math., 23 (1997), pp. 3–19, https://doi.org/10.1016/s0168-9274(96)00059-1.

[6] M. Eiermann, W. Niethammer, and R. Varga, *A study of semiiterative methods for nonsymmetric systems of linear equations*, Numerische Mathematik, 47 (1985), pp. 505–534, https://doi.org/doi.org/10.1007/BF01389454.

[7] H. Elman, D. Silvester, and A. Wathen, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, United Kingdom, second ed., 2014.

[8] O. Ernst and M. Gander, *Why it is difficult to solve Helmholtz problems with classical iterative methods*, Springer, 2012, pp. 325–363, https://doi.org/10.1007/978-3-642-22061-6_10.

[9] P. Farrell, Y. He, and S. MacLachlan, *A local Fourier analysis of additive Vanka relaxation for the Stokes equations*, Numer. Linear Algebra Appl., 28 (2021), p. e2306, https://doi.org/10.1002/nla.2306.
[10] P. Farrell, M. Knepley, L. Mitchell, and F. Wechsung, *Pcpatch: Software for the topological construction of multigrid relaxation methods*, ACM TOMS, 47 (2021), pp. 1–22, https://doi.org/10.1145/3445791.
[11] B. Fischer, *Polynomial based iteration methods for symmetric linear systems*, Wiley-Teubner, 1996.
[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, third ed., 1996.
[13] N. Gould, M. Hribar, and J. Nocedal, *On the solution of equality constrained quadratic programming problems arising in optimization*, SIAM J. Sci. Comput., 23 (2001), pp. 1376–1395, 10.1137/s1064827598345667.
[14] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, 1997.
[15] Y. He, *A Vanka-based parameter-robust multigrid relaxation for the Stokes–Darcy Brinkman problems*, Numer. Linear Algebra Appl., 30 (2023), p. e2514, https://doi.org/10.1002/nla.2514.
[16] C. Keller, N. I. M. Gould, and A. J. Wathen, *Constraint preconditioning for indefinite linear systems*, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1300–1318, https://doi.org/10.1137/S0895479899351805.
[17] M. Larin and A. Reusken, *A comparative study of efficient iterative solvers for generalized Stokes equations*, Numer. Linear Algebra Appl., 15 (2008), pp. 13–34, https://doi.org/10.1002/nla.561.
[18] P. D. Lax, *Linear Algebra*, Wiley-Interscience, New York, 1997.
[19] S. Manservisi, *Numerical analysis of Vanka-type solvers for steady Stokes and Navier-Stokes flows*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 2025–2056, https://doi.org/10.1137/060655407.
[20] J. Pestana and A. J. Wathen, *Natural preconditioning and iterative methods for saddle point systems*, SIAM Review, 57 (2015), pp. 71–91, http://www.jstor.org/stable/24248520.
[21] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, second ed., 2003, https://doi.org/10.1137/1.9780898718003.
[22] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 856–869, https://doi.org/10.1137/0907058.
[23] S. Saberi, G. Meschke, and A. Vogel, *A restricted additive Vanka smoother for geometric multigrid*, Journal of Computational Physics, 459 (2022), p. 111123, https://doi.org/10.1016/j.jcp.2022.111123.
[24] D. Silvester and A. Wathen, *Fast iterative solution of stabilised Stokes systems part ii: Using general block preconditioners*, SIAM Journal on Numerical Analysis, 31 (1994), pp. 1352–1367, http://www.jstor.org/stable/2158225.
[25] V. Simoncini and D. Szyld, *Recent computational developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59, https://doi.org/https://doi.org/10.1002/nla.499.
[26] D. Szyld, *Equivalence of conditions for convergence of iterative methods for singular equations*, Numer. Linear Algebra Appl., 1 (1994), pp. 151–154, https://doi.org/https://doi.org/10.1002/nla.1680010206.
[27] Y. Tao and E. Sifakis, *A symmetric multigrid-preconditioned Krylov subspace solver for Stokes equations*, Computers and Mathematics with Applications, 172 (2024), pp. 168–180, https://doi.org/10.1016/j.camwa.2024.08.018.
[28] S. Vanka, *Block-implicit multigrid solution of Navier-Stokes equations in primitive variables*, Journal of Computational Physics, 65 (1986), pp. 138–158, https://doi.org/10.1016/0021-9991(86)90008-2.
[29] R. Varga, *Matrix Iterative Analysis*, Prentice-Hall, 1962, http://catalog.hathitrust.org/Record/006763844.
[30] A. Wathen, B. Fischer, and D. Silvester, *The convergence rate of the minimal residual method for the Stokes problem*, Numer. Math., 71 (1995), pp. 121–134, https://doi.org/10.1007/s002110050138.
[31] W. Zulehner, *A class of smoothers for saddle point problems*, Computing, 65 (2000), pp. 227–246, https://doi.org/10.1007/s006070070008.