DELFT UNIVERSITY OF TECHNOLOGY

REPORT 16-02

The Induced Dimension Reduction method applied to convection-diffusion-reaction problems

R. Astudillo and M. B. van Gijzen

ISSN 1389-6520

Reports of the Delft Institute of Applied Mathematics

Delft 2016

The Induced Dimension Reduction method applied to convection-diffusion-reaction problems

Reinaldo Astudillo and Martin B. van Gijzen

Delft Institute of Applied Mathematics, Technical Report 16-02

Abstract. Discretization of (linearized) convection-diffusion-reaction problems yields a large and sparse non symmetric linear system of equations,

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

In this work, we compare the computational behavior of the Induced Dimension Reduction method (IDR(s)) [10], with other short-recurrences Krylov methods, specifically the Bi-Conjugate Gradient Method (Bi-CG) [1], restarted Generalized Minimal Residual (GMRES(m)) [4], and Bi-Conjugate Gradient Stabilized method (Bi-CGSTAB) [11].

1 Introduction

In this paper we consider the following simple convection-diffusion-reaction model problem

$$-\epsilon \triangle u + \mathbf{v}^T \nabla u + \rho u = f, \qquad \text{in } \Omega = [0,\, 1]^d \tag{2}$$

with $d = 2$ or $d = 3$, and Dirichlet boundary conditions $u = 0$ on $\partial\Omega$. In Eq. (2), u represents the concentration of solute, $\mathbf{v} \in \mathbb{R}^d$ is the velocity of the medium or convection vector, $\epsilon > 0$ represents the diffusion coefficient, ρ the reaction coefficient, and f represents the source-term function.

Discretization of the Eq. (2) yields a non-symmetric system of linear equations,

$$A\mathbf{x} = \mathbf{b}, \tag{3}$$

where $\mathbf{x}$ is the unknown vector in $\mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^N$, and $A \in \mathbb{R}^{N\times N}$ is typically large, and sparse. Krylov subspace methods are a popular choice to solve such systems. However, the convergence ratio of these methods are strongly influenced by the numerical properties of the coefficient matrix A, which internally depend on the physical parameters of Eq. (2). For example, in the convection-dominated case, i.e. $\|\mathbf{v}\| >> \epsilon$, the coefficient matrix A has almost purely imaginary eigenvalues and this can slow down the convergence of Krylov methods.

GMRES is an optimal method, it obtains the best approximation in a subspace of dimension j performing j matrix-vector products. Nevertheless,

due the large and ill-conditioned linear systems obtained from the discretization of the convection-diffusion-reaction equation, one can expect that many iterations need to be performed to compute the solution accurately. For this reason and taking into account that the computational cost of GMRES increases per iteration, it is preferable to use a preconditioned short-recurrences Krylov method that keeps the computational work and memory consumption fixed per iteration. Bi-CGSTAB is the most widely used method of this kind. However, IDR(s) outperforms Bi-CGSTAB in the experiments presented in [10] and [2]. In this work we continue this comparison. We compare the numerical behavior of Bi-CG, GMRES(m), Bi-CGSTAB, and IDR(s) to solve the linear systems arising from the discretization of (2).

2 Krylov methods for solving systems of linear equations

A projection method onto m-dimensional subspace $\hat{\mathcal{K}}$ and orthogonal to the m-dimensional subspace $\mathcal{L}$, is an iterative method to solve (3) which finds the approximate solution $\mathbf{x}_m$ in the affine subspace $\mathbf{x}_0 + \hat{\mathcal{K}}$ imposing the Petrov-Galerkin condition, i.e., $\mathbf{r}_m = \mathbf{b} - A\mathbf{x}_m$ orthogonal to $\mathcal{L}$. The subspace $\hat{\mathcal{K}}$ is called search space, while $\mathcal{L}$ is called restriction space.

The Krylov subspace methods are projection methods for which the search space is the Krylov subspace $\mathcal{K}_m(A, \mathbf{r}_0) = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, \ldots, A^{m-1}\mathbf{r}_0\}$, where $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$ with $\mathbf{x}_0$ as initial guess in $\mathbb{C}^N$. The different Krylov methods are obtained from the different choices of the restriction space. For a comprehensive description of the Bi-CG, GMRES(m) and Bi-CGSTAB, we refer the reader to [3].

2.1 The Induction Dimension Reduction method (IDR(s))

IDR(s) was introduced in 2008 [10]. This method can also be described as a Krylov projection method (see [5]), however, the original formulation and implementation of IDR(s) is based on the following theorem.

Theorem 1 (IDR theorem). *Let A be any matrix in $\mathbb{C}^{N\times N}$, and let $P = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_s]$ be an $N \times s$ matrix with s linear independent columns. Let $\{\mu_j\}$ be a sequence in $\mathbb{C}$. With $\mathcal{G}_0 \equiv \mathbb{C}^N$, define*

$$\mathcal{G}_{j+1} \equiv (A - \mu_{j+1}I)(\mathcal{G}_j \cap P^{\perp}) \quad j = 0, 1, 2 \ldots,$$

where $P^{\perp}$ represents the orthogonal complement of P. If $P^{\perp}$ does not contain an eigenvector of A, then, for all $j = 0, 1, 2 \ldots$, the following hold

1. *$\mathcal{G}_{j+1} \subset \mathcal{G}_j$, and*
2. *dimension($\mathcal{G}_{j+1}$) $<$ dimension($\mathcal{G}_j$) unless $\mathcal{G}_j = \{\mathbf{0}\}$.*

Proof. See [10].

Assuming that $s+1$ approximations are available with their corresponding residuals belonging to $\mathcal{G}_{j-1}$, IDR(s) constructs the new approximation $\mathbf{x}_k$ at the iteration k, imposing the condition that the vector $\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$ should be in the subspace $\mathcal{G}_j$. Moreover, using the fact that $\mathcal{G}_j \subset \mathcal{G}_{j-1}$, IDR($s$) creates inductively $s+1$ residuals in the subspace $\mathcal{G}_j$. After this, it is possible to create new residuals in $\mathcal{G}_{j+1}$.

IDR(s) has three attractive numerical properties. First, IDR(s) uses recurrences of size $s+1$, and the parameter s is normally selected between 2 and 8. Second, the subspaces $\mathcal{G}_j$ with $j = 1, 2, , \ldots$ are nested and shrinking, and for this reason, IDR(s) has guarantee convergence in at most $N + \frac{N}{s}$ matrix-vector multiplication in exact arithmetic (see Corollary 3.2 in [10]). Third, IDR(1) and Bi-CGSTAB are mathematically equivalent (see [8]); and IDR(s) is commonly faster than Bi-CGSTAB for $s > 1$. The details of the implementation of IDR(s) can be found in [2].

3 Numerical experiments

All the experiment presented in this section are the discretization of Eq. (2) with homogeneous Dirichlet boundary conditions over the unit cube, The right-hand-side function f is defined by the solution $u(x, y, z) = x(1-x)y(1-y)z(1-z)$. We use as stopping criteria that,

$$\frac{\|\mathbf{b} - A\mathbf{x}_k\|_2}{\|\mathbf{b}\|_2} < 10^{-8}.$$

The discretization of Eq. (2) using central finite differences may produce unphysical oscillations in the numerical solution of convection or reaction dominated problems. This problem can be solved discretizing the convection term using upwind schemes. However, we use central finite differences rather than upwind dicretization in this set of problems, in order to illustrate the effect of unfavorable numerical conditions over the Krylov subspace solvers.

Experiment 1: In this example, we consider the parameters $\epsilon = 1.0$ and $\mathbf{v} = (1.0, 1.0, 1.0)^T/\sqrt{3}$. We want to illustrate the effect of non-negative reaction parameter over the Krylov solver, then, we select $\rho \in \{0, 50, \ldots, 300\}$. Figure 1 (a) shows the number of matrix-vector multiplication required for each Krylov method as a function of the reaction parameter ρ. In these problems, the increment of the reaction parameter produces a reduction in the number of matrix-vector products required for each Krylov method. All the methods perform very efficiently for these examples. Figure 1 (b) shows the evolution of the residual norm for $\rho = 0.0$. The execution times are: IDR(4) 0.62s, Bi-CGSTAB 0.64s, Bi-CG 0.92s, and GMRES 2.83s.

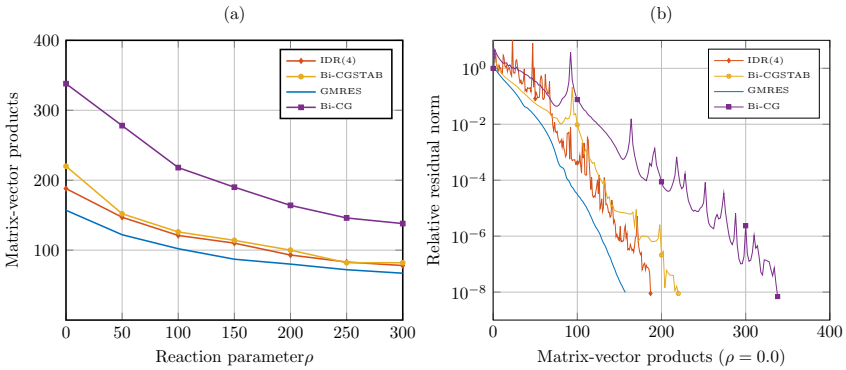

Fig. 1. *Example 1:* (a) Number of matrix-vector products required to converge as a function of the parameter ρ for a diffusion-dominated problem. (b) Comparison of the residual norms, the physical parameters are $\epsilon = 1.0$, $\mathbf{v} = (1.0, 1.0, 1.0)^T/\sqrt{3}$, and $\rho = 0.0$.

Experiment 2: In order to illustrate the effect of the magnitude of the convection velocity, we consider $\epsilon = 1.0$, $\rho = -50.0$, and $\mathbf{v} = \beta(1.0, 1.0, 1.0)^T/\sqrt{3}$ with $\beta \in \{100.0, 200.0, \ldots, 800.0\}$. As the parameter β grows we obtain a more convection-dominated problem. Figure 2 (a) shows how many matrix-vector products are required for each Krylov method as function of the convection speed. The problem is more convection-dominated as $\|\mathbf{v}\|_2$ grows. It is interesting to remark the linear of the number of matrix-vector product for Bi-CGSTAB. Figure 1 (b) shows the evolution of the residual norm for $\beta = 800.0$. Execution time IDR(4) 1.24s, Bi-CGSTAB 5.64s, Bi-CG 1.01s, and GMRES 3.26s.

Experiment 3: Here we use the same set of problems presented in experiment 1, but selecting negative reaction parameters, we consider $\rho \in \{-300, 250, \ldots, -50\}$. In Figure 3 (a), one can see how the negative of the reaction parameter generates a considerable increment of the matrix-vector needed for solving the corresponding linear system. Bi-CGSTAB perform poorly for large negative reaction parameter. Figure 1 (b) shows the evolution of the residual norm for $\epsilon = 1$ and $\rho = 300.0$. The execution time are: IDR(4) 4.02s, Bi-CGSTAB 15.38s, Bi-CG 3.52 s, and GMRES 28.57s.

3.1 IDR(s) and Bi-CG

Despite being a method that is not drastically affected by the increment of the reaction parameter or the convection speed, Bi-CG is not the faster method in terms matrix-vector products required. Bi-CG requires two matrix-vector multiplications to produce one new approximation. IDR(4) in most of the experiments requires less matrix-vector multiplication to get the desired

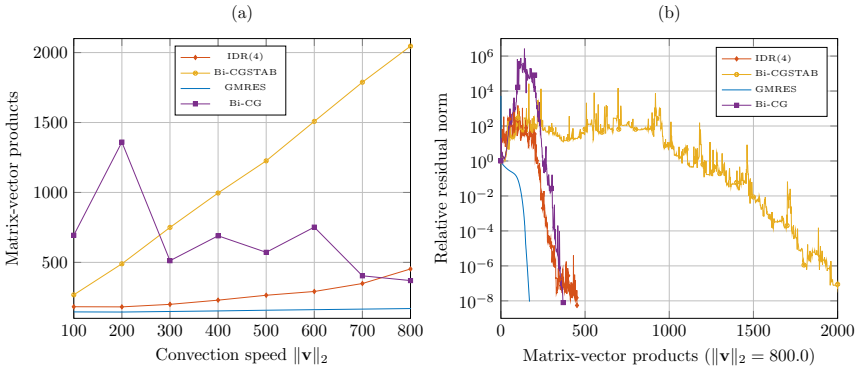

Fig. 2. *Example 2:* (a) Number of matrix-vector products required to converge as a function of the convection speed. (b) Comparison of the residual norms, the physical parameters are $\epsilon = 1.0$, $\mathbf{v} = 800.0 \times (1.0, 1.0, 1.0)^T/\sqrt{3}$, and $\rho = -50.0$.

residual tolerance. Only in the highly convection-dominated examples presented in the experiment 2, Bi-CG presents a similar behavior as IDR(4). A discussion of the phenomena is presented in section 3.3.

3.2 IDR(s), GMRES, and restarted GMRES

In the numerical experiments presented in the previous section, Full GMRES is the methods that uses less matrix-vector products to obtain the desired residual reduction. This result is expected due the optimal residual condition of GMRES. Nevertheless, the computational requirements of full GMRES grow in every iteration. Restarting GMRES or GMRES(m) is an option to overcome this issue. The idea of GMRES(m) is to limit to a maximum of m matrix-vector products, and then restart the process using the last approximation as initial vector. The optimal residual property is lost in this restarting scheme.

In terms of memory consumption, GMRES(m) is equivalent to IDR(s) when $m = 3(s + 1)$. In order to compare the behavior of GMRES(m) and IDR(4), we consider the discretization of Eq. (2) with this parameters: $\epsilon = 1$, $\mathbf{v} = (1.0, 1.0, 1.0)^T/\sqrt{3}$ and $\rho = 40.0$, and we take as restarting parameter $m = 15, 16, \ldots, 170$. Figure 4 shows the number of matrix-vector multiplication required for GMRES(m) for different values of m. GMRES(160) and IDR(4) solve this system using the same number of matrix-vector products (262), however, GMRES(160) consumes approximately ten times more memory than IDR(4). Moreover, CPU time for GMRES(160) is 4.60s while IDR(4) runs in only 0.79s.

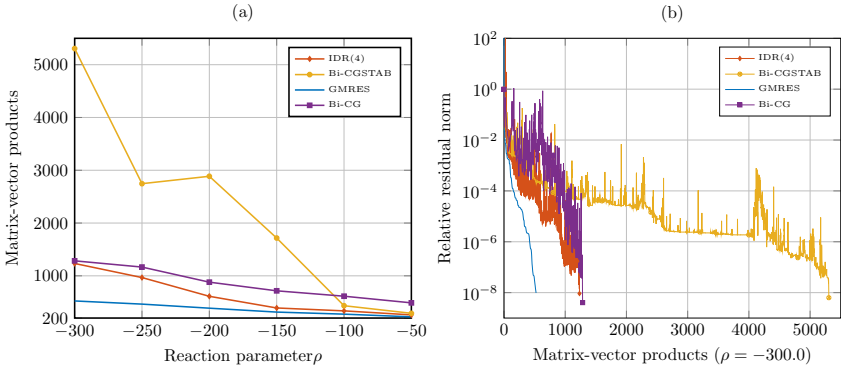

Fig. 3. *Example 3:* (a) Number of matrix-vector products required to converge as a function of the parameter ρ. (b) Comparison of the residual norms. The physical parameters $\epsilon = 1$, $\mathbf{v} = (1.0, 1.0, 1.0)^T/\sqrt{3}$, and $\rho = -300.0$.

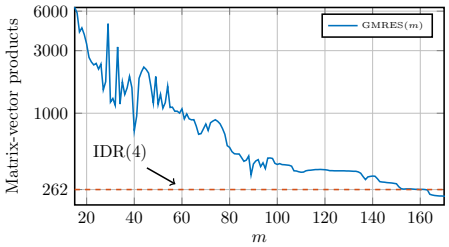

Fig. 4. (GMRES(m) and IDR(s) comparison) Number of matrix-vector products required for GMRES(m) as a function of the parameter m.

3.3 IDR(s) and Bi-CGSTAB

One can see in the experiments that Bi-CGSTAB performs poorly for convection-dominated problems. This can be explained throughout the study of the residual formulas for Bi-CGSTAB. The residual vector in Bi-CGSTAB can be written in the form,

$$\mathbf{r}_k^{(B)} = \Omega_k(A)\phi_k(A)\mathbf{r}_0,$$

where $\phi_k(t)$ is residual associated with Bi-CG and $\Omega_k(t)$ is the Minimal Residual (MR) polynomial defined as,

$$\Omega_k(t) = (1 - \omega_k t)\Omega_{k-1}(t).$$

The parameter ω_k are selected such that $\|\mathbf{r}_k^{(B)}\|_2$ is minimized. However, for indefinite matrices or real matrices that have non-real eigenvalues with an imaginary part that is large relative to the real part, the parameter ω_k is close to zero (see [9]), and the MR-polynomial suffers from slow convergence

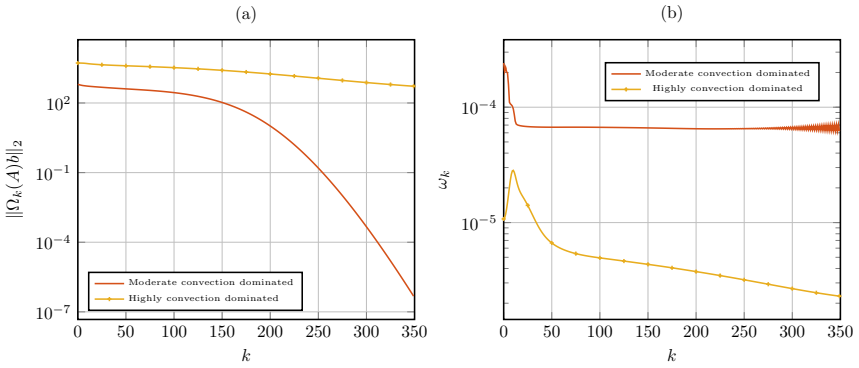

Fig. 5. (a) Behavior of the norm of the MR-polynomial $\Omega_k(A)$. (b) Values of the parameter ω_k.

or numerical instability. To illustrate this we show the behavior of the polynomial $\Omega_k(A)$ applied to two different matrices from the second set of experiments. We consider $\beta = 100.0$ and $\beta = 800.0$ labeled in Figure 5 as moderate convection-dominated and highly convection-dominated respectively.

IDR(s) and Bi-CGSTAB are closely related, in fact, Bi-CGSTAB and IDR(1) are mathematically equivalent for the same parameter choice (see [8]). The convergence of IDR is also affected by the convection speed for a similar reason. The IDR(s) residual vector $\mathbf{r}_k$ in the subspace $\mathcal{G}_j$ can be written as,

$$\mathbf{r}_k^{(I)} = \Omega_j(A)\psi_{k-j}(A)\mathbf{r}_0,$$

where $\psi_{k-j}(t)$ is a block Lanczos polynomial. For IDR(s) the degree of the polynomial $\Omega_k(t)$ increases by one every $s+1$ matrix-vector products, while in Bi-CGSTAB this degree grows by one every two matrix vector products. For this reason, IDR(s) controls the negative effects of the MR-polynomial when A has complex spectrum or is an indefinite matrix.

The bad convergence for strongly convection-dominated problems of Bi-CGSTAB has been observed by several authors and has given rise to BiCGstab(ℓ) [6]. This method uses polynomial factors of degree ℓ, instead of MR-polynomial. A similar strategy has been implemented in IDR(s) which led to the method IDRstab [7]. For the comparison of the convergence of BiCGstab(ℓ) and IDRstab with IDR(s) we refer the reader to [7].

4 Conclusions

Throughout the numerical experiment, we have shown that IDR(s) is a competitive option to solve system of linear equation arising in the discretization of the convection-diffusion-reaction equation.

GMRES, Bi-CG, and IDR(s) exhibit a stable behavior in the most numerically difficult examples conducted in this work. Despite performing more matrix-vector products to obtain convergence, IDR(s) consumes less CPU time than GMRES. We show that for diffusion-dominated problems with a positive reaction term the convergence of the Bi-CGSTAB and IDR(s) are very similar, and for this kind of problems it is often preferable to simply choose $s = 1$. However, for the more difficult to solve convection dominated problems, or problems with a negative reaction term, IDR(s), with $s > 1$ greatly outperform Bi-CGSTAB.

References

1. R. Fletcher, *Conjugate gradient methods for indefinite systems*, Proceedings of the Dundee Conference on Numerical Analysis, 1976, pp. 73–89.
2. M. B. van Gijzen and P. Sonneveld, *Algorithm 913: An Elegant IDR(s) Variant that Efficiently Exploits Bi-orthogonality Properties*, ACM Trans. Math. Software **38**:1 (2011), 5:1–5:19.
3. Y. Saad, *Iterative methods for sparse linear systems*, 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003.
4. Y. Saad and M. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymetric linear systems*, SIAM J. Sci. Stat. Comput. **7** (1986), 856–869.
5. V. Simoncini and D. B. Szyld, *Interpreting IDR as a Petrov-Galerkin method*, SIAM J. Sci. Comput. **32**:4 (2010), 1898–1912.
6. G. L. G. Sleijpen and D. R. Fokkema, *BiCGstab(ℓ) for Linear Equations involving Unsymmetric Matrices with Complex Spectrum*, Electron. Trans. Numer. Anal. **1** (1993), 11–32.
7. G. L. G. Sleijpen and M. B. van Gijzen, *Exploiting BiCGstab(ℓ) Strategies to Induce Dimension Reduction*, SIAM J. Sci. Comput. **32**:5 (2010), 2687–2709.
8. G. L. G. Sleijpen, P. Sonneveld, and M. B. van Gijzen, *Bi-CGSTAB as an induced dimension reduction method*, Appl. Numer. Math. **60** (2010), 1100–1114.
9. G. L. G. Sleijpen and H. A. van der Vorst, *Maintaining convergence properties of bicgstab methods in finite precision arithmetic*, Numer. Algorithms **10**:2 (1995), 203–223.
10. P. Sonneveld and M. B. van Gijzen, *IDR(s): a family of simple and fast algorithms for solving large nonsymmetric linear systems.*, SIAM J. Sci. Comput. **31**:2 (2008), 1035–1062.
11. H. A. van der Vorst, *Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems*, SIAM J. Sci. Stat. Comput. **13**:2 (1992), 631–644.