Random Shadow Vectors in IDR(s): an Explanation of its GMRES-like Convergence

Peter Sonneveld

The IDR(s) – *Induced Dimension Reduction* – method [Sonneveld and van Gijzen, SISC 31, 1035-1062 (2008)], is a family of short-term recurrent Krylov subspace solvers for large sparse, not necessarily symmetric linear systems. For increasing s, the convergence behaviour shows an increasing similarity with full GMRES [Saad and Schultz, SISC 7, 856-869 (1986)]. This similarity even can be observed in systems arising from non-preconditioned discretizations of Helmholtz' equation, so apparently the classical restrictions on the spectrum of the matrix, such as being located at one side of a straight line through the origin of the complex plane, are not required for IDR(s). We'll give an explanation of this behaviour.

IDR(s) uses s so-called *shadow vectors* $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_s$. In solving the $N \times N$ system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, the method produces iterates $\boldsymbol{x}_n$, for which the residuals $\boldsymbol{r}_n = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n$ are forced to be in spaces $\mathcal{G}_j$ of shrinking dimensions. If $\mathcal{G}_0$ is the complete space, then $\mathcal{G}_j = (\boldsymbol{I} - \omega_j \boldsymbol{A}) \left(\mathcal{G}_{j-1} \bigcap \mathcal{N}(\boldsymbol{P}^H)\right)$, where $\boldsymbol{P}$ is an $N \times s$ matrix of which the shadow vectors are the columns. The scalars ω_j are defined on the basis of a stabilization argument, as is done in BiCGSTAB. Normally $s+1$ residuals in $\mathcal{G}_{j-1}$ are needed before vectors in $\mathcal{G}_j$ can be constructed.

As in other Krylov subspace solvers, the residuals $\boldsymbol{r}_n = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n$ can be written as $\Phi_n(\boldsymbol{A})\boldsymbol{r}_0$; in the case of IDR($s$), we call Φ_n the n-th degree *IDR polynomial*. This polynomial is the product of two other polynomials: $\Phi_n(\boldsymbol{A}) = \Omega_j(\boldsymbol{A})\Psi_{n-j}(\boldsymbol{A})$, where $\Omega_j(\boldsymbol{A}) = (\boldsymbol{I} - \omega_j \boldsymbol{A})\Omega_{j-1}(\boldsymbol{A})$, with $\Omega_0(\boldsymbol{A}) = \boldsymbol{I}$. As a consequence of the algorithm, the degree j satisfies $j = \lfloor \frac{n}{s+1} \rfloor$. The other polynomials Ψ_{n-j} can be identified with the residual polynomials of a two-sided (block-)Lanczos process, with s left side starting vectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_s$. Each step of this virtual Lanczos process can be regarded as a Galerkin approximation in the Krylov subspace $\mathcal{K}_n(\boldsymbol{A}, \boldsymbol{r}_0)$, in which the s test vectors are the first n vectors from the set
$\{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_s, \boldsymbol{A}^H\boldsymbol{p}_1, \boldsymbol{A}^H\boldsymbol{p}_2, \ldots, \boldsymbol{A}^H\boldsymbol{p}_s, (\boldsymbol{A}^H)^2\boldsymbol{p}_1, \ldots\}$

The relation between this Galerkin approximation and the optimal Galerkin method –least squares– depends on the *angle* between $\mathcal{K}_n(\boldsymbol{A}, \boldsymbol{r}_0)$ and the search space: smaller angles produce results closer to the opmtimal process. Least squares in this context is equivalent with the n-th step of full GMRES.

The shadow vectors in IDR(s) are usually chosen randomly. Originally because the authors couldn't find anything better. The random choice now produces the explanation of the GMRES-like convergence of IDR(s). The virtual Lanczos residuals $\widetilde{\boldsymbol{r}}_n$ can be regarded as stochastic vectors. Denote the GMRES residuals by $\widehat{\boldsymbol{r}}_n$, let $\widetilde{\boldsymbol{r}}_n - \widehat{\boldsymbol{r}}_n = \|\widehat{\boldsymbol{r}}_n\| \boldsymbol{z}_n$, then if all test vectors are independent standard normally distributed (meaning $s \geq n$), $\boldsymbol{z}_n$ has the same probability distribution as the solution of a $n \times n$ linear system with random matrix and random righthand side. The derivation of this distribution will be sketched. It turns out that $\|\boldsymbol{z}_n\| = O(\sqrt{n})$ may be expected, implying that the logarithmic convergence history curve is shifted a little bit to the right, compared to GMRES. This is in agreement with the observations.

If s is smaller, then the test vectors contain some stochastic dependence. This ruins the theoretical analysis, but apparently, the method can live very well if the test vectors are only *almost* stochastically independent. Experts in probability theory may try to generalize the analysis to the practial case of moderate s.