Fourteenth International Conference on Domain Decomposition Methods
Editors: Ismael Herrera , David E. Keyes, Olof B. Widlund, Robert Yates

44. On a selective reuse of Krylov subspaces in Newton-Krylov approaches for nonlinear elasticity

P. Gosselet[1], C. Rey[2]

1. Introduction. We consider the resolution of large-scale nonlinear problems arising from the finite-element discretization of geometrically non-linear structural analysis problems. We use a classical Newton Raphson algorithm to handle the non-linearity which leads to the resolution of a sequence of linear systems with non-invariant matrices and right hand sides. The linear systems are solved using the FETI-2 algorithm. We show how the reuse, as a coarse problem, of a pertinent selection of the information generated during the resolution of previous linear systems, stored inside Krylov subspaces, leads to interesting acceleration of the convergence of the current system.

Nonlinear problems are a category of problems arising from various applications in mathematics, physics or mechanics. Solving these problems very often leads to a succession of linear problems the solution to which converges towards the solution to the considered problem. Within the framework of this study, all linear systems are solved using a conjugate gradient algorithm. It is well known that this algorithm is based on the construction of the so-called Krylov subspaces, on which depends its numerical efficiency and its convergence behaviour.

The purpose of this paper is to accelerate the convergence of linear systems by reusing information arising from previous resolution processes. Such an idea has already led to a classical algorithm for invariant matrices [8] which has been successfully extended to the case of non invariant matrices [6, 7]. We here propose, thanks to a spectral analysis of linear systems, to select the most significant part of the information generated during conjugate gradient iterations to accelerate the convergence via an augmented Krylov conjugate gradient algorithm.

The remainder of this paper is organized as follows: section 2 addresses characteristic properties of preconditioned conjugate gradient, section 3 exposes the acceleration strategies, section 4 gives numerical assessments and section 5 concludes the paper.

2. Basic properties of preconditioned conjugate gradient. We consider the linear system $Ax = b$ solved with a M-preconditioned conjugate gradient (A and M are $N \times N$ real symmetric positive definite matrices). We note x_i the ith estimation to $x = A^{-1}b$, $r_i = b - Ax_i = A(x - x_i)$ the associated residual and $z_i = M^{-1}r_i$ the preconditioned residual. In order to concentrate the notations, we also note with capital letters matrices built from set of vectors, e.g. $R_i = (r_0, \dots, r_{i-1})$. Given initialization x_0, preconditioned conjugate gradient iteration consists in searching

$$
\begin{array}{c}
x_i \in \{x_0\} + \mathcal{K}_i(M^{-1}A, z_0) \text{ with } r_i \perp \mathcal{K}_i(M^{-1}A, z_0) \\
\text{where } \mathcal{K}_i(M^{-1}A, z_0) \text{ is the } i\text{th Krylov subspace} \\
\mathcal{K}_i(M^{-1}A, z_0) = \mathrm{Span}(z_0, \dots, (M^{-1}A)^{i-1}z_0) = \mathrm{Range}(Z_i)
\end{array}
\tag{2.1}
$$

2.1. Augmented conjugate gradient. The augmentation consists in defining full-ranked constraint matrix C and imposing $C^T r_i = 0$. It leads to the definition of a modified Krylov subspaces $\tilde{\mathcal{K}}_i(M^{-1}A, z_0, C)$ [2]:

$$
\begin{array}{c}
\tilde{\mathcal{K}}_i(M^{-1}A, z_0, C) = \mathcal{K}_i(M^{-1}A, z_0) \oplus \mathrm{Range}(C) \\
x_i \in \{x_0\} + \tilde{\mathcal{K}}_i(M^{-1}A, z_0, C) \text{ with } r_i \perp \tilde{\mathcal{K}}_i(M^{-1}A, z_0, C)
\end{array}
\tag{2.2}
$$

[1]Laboratoire de Modélisation et Mécanique des Structures, gosselet@ccr.jussieu.fr
[2]Laboratoire de Modélisation et Mécanique des Structures, rey@ccr.jussieu.fr

$Ax = b$ with $C^T r_i = 0$	Iterations $i = 0, \ldots, s$
$P_C = Id - C\left(C^T AC\right)^{-1} C^T A$	$z_i = P_C M^{-1} r_i \qquad (w_0 = z_0)$
Initialization (x_{00} is arbitrary)	$w_i = z_i + \sum_{j=0}^{i-1} \beta_i^j w_j \qquad \beta_i^j = -\dfrac{(z_i, Aw_j)}{w_j, Aw_j}$
$x_0 = C\left(C^T AC\right)^{-1} C^T b + P_C x_{00}$	$x_{i+1} = x_i + \alpha_i w_i \qquad \alpha_i = \dfrac{(r_i, z_i)}{(w_i, Aw_i)}$
$r_0 = b - Ax_0$	$r_{i+1} = r_i - \alpha_i A w_i$

Property: $R_i^T Z_i$ and $W_i^T A W_i$ are diagonal

Figure 2.1: Augmented Preconditioned Conjugate Gradient

The augmented preconditioned conjugate gradient can be implemented with a projected algorithm (fig. 2.1): initialization and projector P_C ensure orthogonality conditions.

Remark 2.1 *Although no optimality result holds anymore when matrix A is non-positive, conjugate gradient still proves good convergence behaviour [5].*

Remark 2.2 *As M is definite positive, it can be factorized under Cholevsky's form $M = LL^T$. Following [9] we prove that the M-preconditioned C-augmented conjugate gradient is equivalent to a non-preconditioned $\hat{C}$-augmented conjugate gradient $\hat{A}\hat{x} = \hat{b}$ with :*

$$\begin{array}{ccccc} \hat{A} = L^{-1} A L^{-T} & & \hat{x}_i = L^T x_i & \hat{b} = L^{-1} b & \hat{C} = L^T C \\ \hat{w}_i = L^T w_i & \hat{z}_i = L^T z_i & \hat{r}_i = L^{-1} r_i & \beta_i^j = \hat{\beta}_i^j & \alpha_i = \hat{\alpha}_i \end{array} \tag{2.3}$$

2.2. Ritz's spectral analysis of symmetric system. Ritz's values and vectors $(\theta_i^j, \hat{y}_i^j)_{1 \leqslant j \leqslant i}$ defined in equation (2.4) are the eigenelements of the projection of matrix $\hat{A}$ onto $\tilde{\mathcal{K}}_i(\hat{A}, \hat{r}_0, \hat{C})$, they converge ($i \rightarrow N$) to eigenelements of matrix $\hat{A}$ [5].

$$\left.\begin{array}{c} \hat{V}_i \text{ orthonormal basis of } \tilde{\mathcal{K}}_i(\hat{A}, \hat{r}_0, \hat{C}) \\ B_i = \hat{V}_i^T \hat{A} \hat{V}_i \text{ Rayleigh's matrix} \\ \end{array}\right| \begin{array}{c} \text{Diagonalization } B_i = Q_i^B \Theta_i {Q_i^B}^T \\ \Theta_i = \mathrm{Diag}(\theta_i^j)_{1 \leqslant j \leqslant i} \\ {Q_i^B}^T Q_i^B = Id, \quad \hat{Y}_i = \hat{V}_i Q_i^B \end{array} \tag{2.4}$$

Ritz's representation of conjugate gradient provides meaningful information. Especially, the convergence of Ritz's values is directly linked to the convergence of the conjugate gradient:

$$\hat{x} - \hat{x}_i = \pi(\hat{A})(\hat{x} - \hat{x}_0) \text{ with } \pi(\xi) = \prod_{j=1}^{i} \frac{\theta_i^j - \xi}{\theta_i^j} \tag{2.5}$$

3. Choice of optional constraints. The choice of matrix $\hat{C}$ is a very accurate problem which requires a study of the governing factors of the convergence of the conjugate gradient [11]. The condition number, which is proved to decrease [1] whatever the $\hat{C}$ matrix may be, is not sufficient for a relevant analysis. In the remainder of the paper, we will call "active" eigenelements that are excited by (i.e. non-orthogonal to) the initial residual and "effective" active eigenelements that are not yet properly estimated by Ritz's elements. Only effective condition number influences the convergence rate: when an eigenvalue is sufficiently well approximated inside the Krylov subspace, the conjugate gradient acts as if it had been suppressed from the resolution process. This explains the superconvergent behaviour of the conjugate gradient: when highest eigenvalues are sufficiently well approximated by Ritz's values, the effective condition number is very low and the convergence rate very high. So a good way to ensure a decrease of the effective condition number is to put active eigenvectors of $\hat{A}$ inside matrix $\hat{C}$.

However, computing a priori active eigenvectors of a system is as expensive as solving it, hence in this section we first show, inspiring from [9], how a posteriori computation can be achieved costlessly when reusing information generated during the conjugate gradient process, then we propose within the framework of multiple systems resolution to use approximation of the eigenvectors of previous systems as constraints to accelerate the convergence of current system.

3.1. Efficient computation of Ritz's elements. Hessemberg matrix H_i arising from Lanczos' procedure is a specific tridiagonal Rayleigh matrix the coefficients of which can be recovered from the coefficients of the conjugate gradient:

$$\begin{gathered}\underline{\hat{R}}_i = (\frac{\hat{r}_0}{\|\hat{r}_0\|}, \dots, (-1)^{i-1}\frac{\hat{r}_{i-1}}{\|\hat{r}_{i-1}\|}) \text{ orthonormal basis of } \tilde{\mathcal{K}}_i(\hat{A}, \hat{r}_0, \hat{C}) \\ \underline{Z}_i = (\frac{z_0}{(z_0, r_0)}, \dots, \frac{(-1)^{i-1} z_{i-1}}{\sqrt{(z_{i-1}, r_{i-1})}}) \text{ M-orthonormal basis of } \mathcal{K}_i(M^{-1}A, z_0, C) \\ H_i = \underline{\hat{R}}_i^T \hat{A} \underline{\hat{R}}_i = \underline{Z}_i^T A \underline{Z}_i \\ H_i = \text{Tridiag}(\eta_{j-1}, \delta_j, \eta_j) \text{ with } \eta_j = \frac{\sqrt{\beta_j^{j-1}}}{\alpha_j} \text{ and } \delta_j = \frac{1}{\alpha_j} + \frac{\beta_{j-1}^{j-2}}{\alpha_{j-1}}\end{gathered} \tag{3.1}$$

So a tridiagonal Rayleigh matrix can be computed without vector manipulation, and a specific Lapack procedure can then be used to compute the eigenelements. To have an action on the non-symmetric preconditioned problem, we define "transported Ritz's vectors" $Y_i = L^{-T}\hat{Y}_i = \underline{Z}_i Q_i^H$, they verify the following orthonormalities:

$$Y_i^T A Y_i = \Theta_i \text{ and } Y_i^T M Y_i = Id_i \tag{3.2}$$

3.2. Selective reuse of Krylov subspaces. We focussed on the interest of reusing eigenvectors (or at least good estimations) as constraints. Our strategies are based on the simple equivalence $\hat{C} = (\hat{y}_i^j) \Leftrightarrow C = (y_i^j)$ which means that a spectral action can be achieved acting directly on the preconditioned problem.

We now consider the resolution of a sequence of linear systems $A^k x^k = b^k$ ($k \geqslant 1$ stands for the number of the linear system, matrices and right hand sides are non-invariant) with augmented conjugate gradient. We propose two strategies based on the reuse of spectral information.

The first strategy is a simple total reuse of Ritz's vectors which is equivalent, since $\text{Range}(Y_i) = \text{Range}(W_i) = \mathcal{K}_i(M^{-1}A, z_0, C)$, to a total reuse of Krylov subspaces: matrix C^k is built concatenating all previous Krylov subspaces $C^k = (W_1, \dots, W_{k-1})$ ($C^1 = 0$). As all the information is reused without selection, this strategy gives the best decrease of the number of iterations of the conjugate gradient expectable from the reuse of Krylov subspaces. Of course it quickly leads to huge C^k matrices and expensive computations to handle the augmented algorithm. Note that when A^k is invariant ($\forall k,\ A^k = A$), ${C^k}^T A C^k$ is a diagonal matrix and this algorithm is equivalent to a multiple right hand side conjugate gradient [8].

The second strategy aims at reducing the dimension of matrices C^k concentrating the information stored inside Krylov subspaces into few vectors. It is managed through the spectral analysis exposed above and the selection of Ritz's vectors associated to converged

Ritz's values. The convergence of the values is estimated computing the values for the last two iterations and comparing them.

$$\text{for } j \leqslant (i+1),\ \theta_i^j \text{ is converged if } \left| \frac{\theta_i^j - \theta_{i-1}^{j-1}}{\theta_i^j} \right| \leqslant \varepsilon \tag{3.3}$$

4. Numerical assessment. We now assess the reuse of Krylov subspaces on the computation of the buckling of a clamped-free beam (fig. 4.1). The beam is a composite structure made up of Saint-Venant–Kirchoff materials, fibers are 1000 times stiffer than the matrix. It is decomposed into 32 substructures. We use Newton Raphson's algorithm [10] to linearize the problem, the resolution is then conducted in 28 linear systems with non-invariant matrices $K^k u^k = f^k$ (k is the linear system number). The linear systems are solved with FETI-2 method equipped with Dirichlet's preconditioner and superlumped projector.

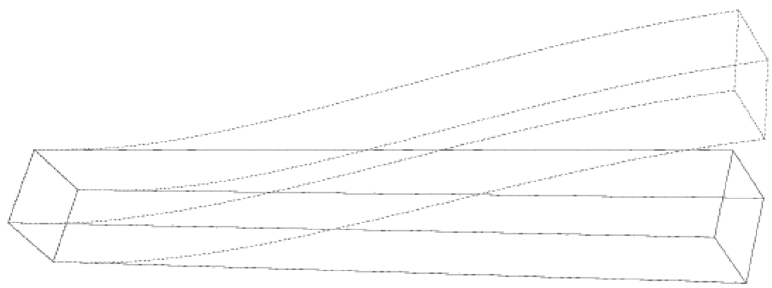


Figure 4.1: Buckling of the beam

4.1. Application of the reuse of Krylov subspaces to FETI-2. The Finite Elements Tearing and Interconnecting (FETI) method was first introduced by Farhat and Roux [4]. It consists in solving with a projected conjugate gradient the system arising from dual domain decomposition method. FETI-2 [3] solves the same problem with augmented conjugate gradient. Readers should refer to referenced papers for a complete description, we only show here the specificity of our strategies applied to FETI-2 method (fig. 4.2). With notations from [3], the system arising from the condensation writes:

$$\begin{pmatrix} F_I & -G_I \\ -G_I^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} d \\ -e \end{pmatrix} \tag{4.1}$$

The first level projection P and initialization λ_{01} handle floating substructures, second level projector P_C and initialization λ_{02} handle the augmentation associated to matrix C. Note that constraints have to be made compatible with the first level projector $(PC)^T r_i = 0$. In the case of FETI algorithm, the augmentation possesses a mechanical interpretation: $P^T r_i$ represents the jump of the displacement field between substructures. Constraints matrix C then ensures a weak continuity of the displacement field. Forming and factorizing the so-called coarse problem matrix $\left((PC)^T F(PC)\right)$ is a complex operation requiring all-to-all exchanges between substructures, in a parallel processing context these operations are penalizing then matrix C has to be chosen as small as possible.

We checked that for this class of problem Dirichlet's preconditioner is positive for all the systems. We also verified the imbrication of the kernel of local matrices $\forall$ (substructure s, system k) $\mathrm{Ker}({K^{(s)}}^{k+1}) \subset \mathrm{Ker}({K^{(s)}}^{k})$ which implies that $\forall k\ \ \mathrm{Range}(G_I^{k+1}) \subset \mathrm{Range}(G_I^k)$.

$$P = Id - QG_I \left(G_I^T QG_I\right)^{-1} G_I^T$$
$$P_C = Id - (PC)\left((PC)^T F_I (PC)\right)^{-1} (PC)^T F_I$$

Initialization (λ_{00} is arbitrary)

$$\lambda_{01} = QG_I \left(G_I^T QG_I\right)^{-1} e \qquad \lambda_{02} = (PC)\left((PC)^T F_I (PC)\right)^{-1} (PC)^T d$$
$$\lambda_0 = P_C(P\lambda_{00} + \lambda_{01}) + \lambda_{02}$$
$$r_0 = d - F_I \lambda_0$$

Iterations $i = 0, \ldots, s$

$$z_i = P_C P \tilde{F}_I^{-1} P^T r_i$$
$$w_i = z_i + \sum_{j=0}^{i-1} \beta_{ij} w_j \qquad (w_0 = z_0)$$
$$\lambda_{i+1} = \lambda_i + \alpha_i w_i \qquad \beta_{ij} = -\frac{(w_j, F_I z_i)}{(w_j, F_I w_j)}$$
$$r_{i+1} = r_i - \alpha_i F_I w_i \qquad \alpha_i = \frac{(w_i, r_i)}{(w_i, F_I w_i)}$$

Figure 4.2: Two-level FETI algorithm

So all previous Krylov subspaces are built orthogonally to the G_I^k matrix, hence when using vectors from Krylov subspaces as constraints we already have $P^k C^k = C^k$. Then the two projectors are decoupled which suppresses time consuming step of making constraints admissible.

4.2. Performance results. The first point concerns the choice of the ε parameter introduced in section 3.2 to determine whether Ritz's values are converged or not. Experiments (e.g. fig. 4.3) showed that the criterion is either very low ($> 10^{-14}$) or very high ($> 10^{-8}$), value ε can then be chosen inside a wide range without modifying the selection, typically we chose $\varepsilon = 10^{-13}$.

Figures 4.4, 4.5 and 4.6 summarize the action of the reuse of Krylov subspaces through the resolution of the linear systems. First figure 4.4 shows how effective the selection is: the number of constraints is quickly divided by a factor 2. Figure 4.5 presents the evolution of the number of iterations per linear system, the total reuse corresponds to the best result expectable from the reuse of Krylov subspaces, the number of iterations is divided by a factor 10, which proves the interest of the information stored inside Krylov subspaces. The selective reuse also proves interesting: with a two-time smaller constraints space, its performance results are quite near the total reuse. Figure 4.6 shows the performance results in terms of CPU time: the total reuse is already relevant, the selective reuse since its performance results in terms of iterations are almost equivalent with a lower number of constraints leads to impressive gain, it is 60% faster than the non accelerated method.

Figures 4.7 and 4.8 enable us to check the spectral action announced above, they represent the Ritz's spectrum for 4 linear systems (the 1st, 5th, 10th and 28th). The selective reuse filters the highest and the negative values, and suppresses part of mid-range values, giving better spectral properties for the resolution. Figures 4.9 and 4.10 show how the resolution process is improved by the selective reuse: two actions are combined, first a better

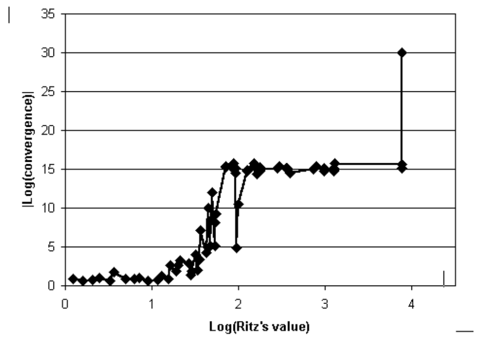


Figure 4.3: Convergence of Ritz's values

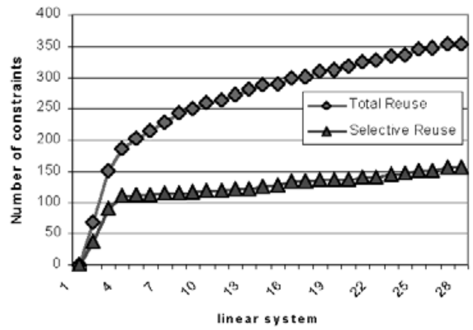


Figure 4.4: Action of Selective Reuse: number of constraints

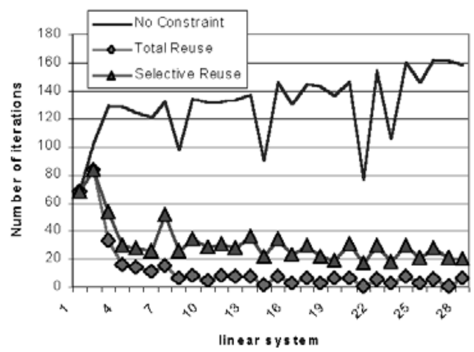


Figure 4.5: Action of Selective Reuse: number of iterations

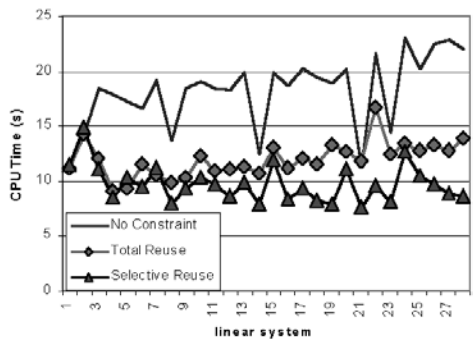


Figure 4.6: Action of Selective Reuse: CPU time

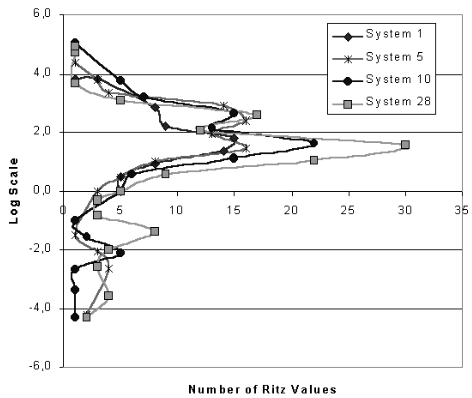
initialization is found, second the superconvergence is achieved from the beginning of the resolution.

5. Conclusion. In this paper we considered the resolution of a sequence of linear systems arising from geometrically nonlinear structural analysis, with a FETI-2 method. We proposed an algorithm to realize a spectral analysis of linear systems solved with a conjugate gradient algorithm with positive preconditioner. We showed that the complete reuse of former Krylov subspaces has already led to good performance results and that a selective reuse of Ritz vectors associated to Ritz's values giving good estimates of eigenvalues gave even better computational performance (up to 60% CPU time gain). Next studies will focuss on additional selection criteria for the Ritz vectors based on the activity of former vectors for the resolution of current system, the aim is to be even more selective and to suppress vectors containing information which is non-relevant for the current system.

We authors acknowledge support from the Centre Informatique National Enseignement Superieur (CINES) for computational resources.

REFERENCES

[1] Z. Dostal. Conjugate gradient method with preconditioning by projector. *INT. J. Comput.. Math.*, 23:315–323, 1988.



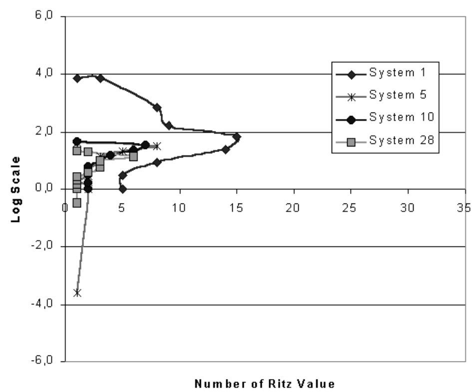
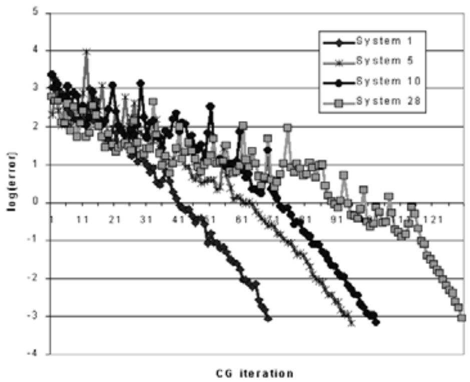


Figure 4.7: Ritz's spectrum, no constraint

Figure 4.8: Ritz's spectrum, selective reuse



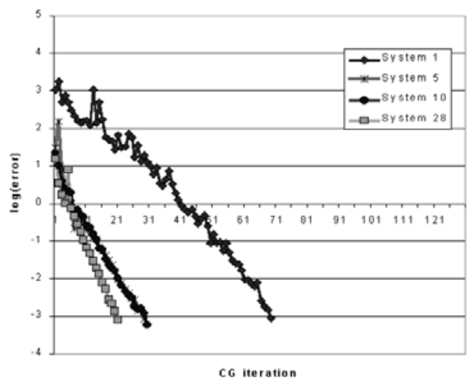


Figure 4.9: Evolution of the error, no constraint

Figure 4.10: Evolution of the error, selective reuse

[2] J. Erhel and F. Guyomarc'h. An augmented conjugate gradient method for solving consecutive symmetric positive definite linear systems. *SIAM J. Matrix Anal. Appl.*, 21(4):1279–1299, 2000.

[3] C. Farhat, K. H. Pierson, and M. Lesoinne. The second generation of feti methods and their application to the parallel solution of large-scale linear and geometrically nonlinear structural analysis problems. *Computer Methods in Applied Mechanics and Engineering*, 184:333–374, 2000.

[4] C. Farhat and F.-X. Roux. A Method of Finite Element Tearing and Interconnecting and its Parallel Solution Algorithm. *Int. J. Numer. Meth. Engrg.*, 32:1205–1227, 1991.

[5] C. Paige, B. Parlett, and H. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2(2):115–133, 1995.

[6] C. Rey and F. Risler. A Rayleigh-Ritz preconditioner for the iterative solution to large scale nonlinear problems. *Numerical Algorithms*, 17:279–311, 1998.

[7] F. Risler and C. Rey. Iterative accelerating algorithms with Krylov subspaces for the solution to large-scale nonlinear problems. *Numerical Algorithms*, 23:1–30, 2000.

[8] Y. Saad. On the lanczos method for solving symmetric linear systems with several right hand sides. *Math. Comp.*, 48:651–662, 1987.

[9] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, 1996.

[10] P. L. Tallec. Numerical methods for non-linear three-dimensional elasticity. In J. L. Lions and P. Ciarlet, editors, *Handbook of numerical analysis vol.3*, pages 465–622. Elsevier, 1994.

[11] A. van der Sluis and H. van der Vorst. The rate of convergence of conjugate gradients. *Numer. Math.*, 48:543–560, 1986.