



JUBILEO
Prof. Julio Ricaldoni

A SÉRIE POLINOMIAL DE NEUMANN INCOMPLETA COMO PRÉ-CONDICIONADOR EM PROBLEMAS DO MEF

Valério S. Almeida, João B. de Paiva
Departamento de Engenharia de Estruturas
Escola de Engenharia da Universidade de São Paulo
Caixa Postal 359 – CEP 13560-970 – São Carlos, SP, Brasil
e-mail: valerio@sc.usp.br, web page: [http:// www.set.eesc.sc.usp.br](http://www.set.eesc.sc.usp.br)
e-mail: paiva@sc.usp.br, web page: [http:// www.set.eesc.sc.usp.br](http://www.set.eesc.sc.usp.br)

SUMÁRIO

Com a crescente aplicação dos computadores de arquitetura em paralelo para a resolução de problemas do MEF, cada vez mais se tem utilizado métodos iterativos em virtude da sua forma desacopladora de se resolver o sistema, destacando-se os métodos: Gauss-Seidel, Jacobi, Quase-Newton (BFGS) e o Método dos Gradientes Conjugados (MGC). Em particular, o método dos gradientes conjugados tem sido aplicado em virtude da sua boa versatilidade e da sua fácil implementação, tanto em ambiente sequencial quanto em paralelo.

Neste trabalho apresenta-se o método iterativo MGC otimizado com o uso da técnica de pré-condicionamento, sendo aqui desenvolvido uma nova técnica, que nada mais é que a idéia principal da decomposição incompleta de Cholesky aplicada a técnica da série truncada de Neumann.

São apresentados exemplos desta técnica consideradas para *processamento sequencial* aplicados a exemplos estruturais de treliça tridimensional e placa, destacando que esta inovação traz resultados satisfatórios para a utilização deste modelo aperfeiçoado.

1 INTRODUÇÃO

A crescente simulação de modelos estruturais complexos - via métodos dos elementos finitos - tem exigido a manipulação de grande quantidade de dados, bem como a procura da diminuição do tempo de resposta para a resolução do sistema linear gerado.

Mesmo com o impacto que o setor da microeletrônica tem causado no desenvolvimento de componentes computacionais [1], ainda assim essas poderosas máquinas - por si só - nem sempre conseguiram tratar adequadamente diversos modelos estruturais: quer por falta de memória, quer por excessivo tempo de resposta.

Entretanto, o uso de sistemas computacionais interligados para resolver um mesmo problema tem destacado, principalmente nos últimos 20 anos [2,3]. Enfatiza que estes sistemas, denominados mais comumente de análise em processamento paralelo, exigem uma readaptação dos procedimentos numéricos outrora já consagrados como, por exemplo, a resolução do sistema linear.

Assim, em vista deste novo tipo de arquitetura, o meio científico tem buscado métodos de resolução do sistema linear, até então não mais utilizados em virtude de suas deficiências específicas, mas que agora se adequem a este novo modelo computacional. Destaca-se então - para a resolução do sistema linear - o crescente uso dos métodos iterativos frente aos tradicionais métodos diretos.

Desta forma, a aplicação dos métodos iterativos têm se destacado pela sua forma desacopladora de resolver o sistema como, por exemplo, a utilização dos métodos: Gauss-Seidel, Jacobi, Quase-Newton (BFGS) e o Método dos Gradientes Conjugados (MGC). Será dada atenção neste trabalho ao MGC, em virtude de sua boa versatilidade e da sua fácil implementação, tanto em ambiente sequencial quanto em paralelo, e também por ser o método que está sendo usado com mais frequência na análise em processamento paralelo [3, 4, 5, 6, 7].

Para o MGC o grande problema característico desse é a sua instabilidade numérica perante as diferentes propriedades do sistema linear, mais especificamente do condicionamento da matriz. Se o sistema estiver bem condicionado, então o método converge rapidamente. Caso contrário, a convergência é sofrível tornando o método inadequado para a resolução do sistema.

Com isso, para solucionar esta questão, tem-se utilizado as técnicas de pré-condicionadores que nada mais são do que otimizadores que diminuem esta instabilidade própria do método, ver [8, 9, 10, 11, 12, 13].

Então apresenta-se o método iterativo MGC otimizado com o uso da técnica de pré-condicionamento via série truncada de Neumann, sendo que esta técnica é aperfeiçoada aplicando-se uma inovação frente ao método convencional.

2 MÉTODO DOS GRADIENTES CONJUGADOS

Dado o sistema linear do tipo

$$[K] \cdot \{U\} = \{F\} \quad (1)$$

onde $[K]$ representa a matriz de rigidez, $\{U\}$ e $\{F\}$ os vetores de deslocamentos e forças respectivamente, o Método dos Gradientes Conjugados (MGC) [3, 4, 9 e 15] é aplicado sobre a matriz $[K]$, sendo esta simétrica e positiva definida e este método é baseado na estratégia do método da *descida íngreme*. Nessa estratégia, é escolhido um ponto $\{U_0\}$ como valor inicial, e mediante uma série de aproximação $\{U_1\}$, $\{U_2\}$, $\{U_3\}$, ..., $\{U_n\}$ encontra-se os pontos numericamente próximos que satisfaçam a equação (1).

Este ponto de mínimo, representado pelo escalar α , é dado por

$$\alpha_{CRITICO} = \frac{\{r_p^T\} \cdot \{r_p\}}{\{r_p^T\} \cdot [K] \cdot \{r_p\}} \quad (2)$$

onde $\{r_p\}$ é o vetor resíduo, dado por:

$$\{r_p\} = \{F\} - [K] \cdot \{U_0\} \quad (3)$$

Então, a aproximação do vetor $\{U\}$ será,

$$\{U_{p+1}\} = \{U_p\} + \alpha_{critico} \cdot \{r_p\} \quad (4)$$

Como avanço no Método dos Gradientes Conjugados (GC) é usada a estratégia de sempre ir buscando direções ortogonais $\{p_0\}, \{p_1\}, \dots, \{p_{n-1}\}$ às direções já calculadas no passo anterior. Para cada uma dessas direções se encontrará uma das coordenadas de $\{U\}$. Com isso, após n passos, que é a dimensão de $\{U\}$, o processo terminará e a solução procurada é encontrada.

A aproximação do vetor $\{U\}$ fica neste processo:

$$\{u_{j+1}\} = \{u_j\} + \alpha_j \{p_j\} \quad (5)$$

e a garantia para atingir a ortogonalidade das direções de $\{p_i\}$ e $\{p_{i+1}\}$ de forma que sejam K- ortogonais é dada por:

$$\{p_i^T\} \cdot [K] \cdot \{p_j\} = \{0\}, \text{ para } i \neq j \quad (6)$$

e o pseudo- algoritmo do Método dos Gradientes Conjugados é apresentado na figura (2).

Os teoremas (1) e (2), que têm suas respectivas demonstrações em [4, 8], complementam os critérios de convergência do método.

Teorema 1: Se as direções de busca $\{p_0\}, \{p_1\}, \dots, \{p_{n-1}\}$ são K-conjugados (K-ortogonais) e o escalar α_j é escolhido da forma que: $\alpha_j = -\frac{\{p_j^T\} \cdot [K] \cdot \{r_j\}}{\{p_j^T\} \cdot [K] \cdot \{p_j\}}$ então o processo termina em, no máximo, n passos, onde n é a dimensão da matriz quadrada $[K]$ do sistema da expressão (1).

Teorema 2: Seja $[K]$ simétrica e definida positiva, então o algoritmo dos gradientes conjugados produz uma seqüência de vetores $\{U_0\}, \{U_1\}, \dots, \{U_j\}, \dots$, com a seguinte propriedade:

$$\|\{U\} - \{U_j\}\|_K \leq 2 \cdot \|\{U\} - \{U_0\}\|_K \cdot \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^\kappa \quad (7)$$

onde $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$ e λ_{\max} e λ_{\min} são, respectivamente, o maior e o menor autovalor de $[K]$ e κ é conhecido como **número de condição**.

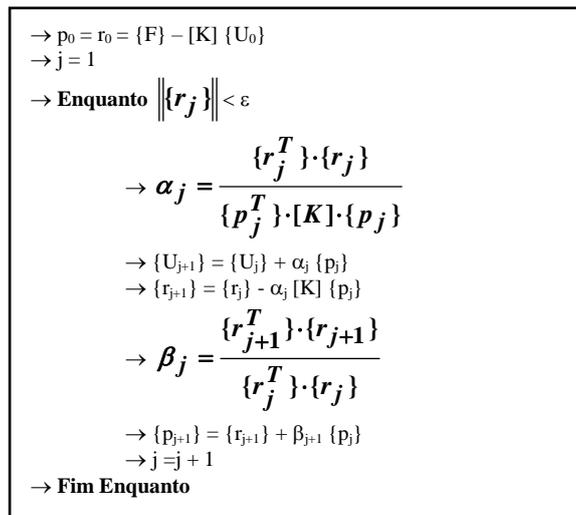


Figura 1 – Pseudo-algoritmo do método dos gradientes conjugados

Pela equação (7) dois fatores influenciam na convergência do método dos gradientes conjugados: a aproximação inicial $\{U_0\}$ e o número de condição (κ).

Em relação a aproximação inicial, na literatura comenta-se que este fator é de pouca importância para a eficiência do método. Entretanto, em [10] é apresentado um estudo de otimização onde se obtém uma aproximação inicial para o vetor $\{U_0\}$. Nos últimos anos, a pesquisa na área de inteligência artificial tem dado atenção a este item, onde a estratégia é criar bancos de dados de deformações e/ou de esforços de determinados modelos estruturais e assim aplicar algoritmos especiais, conhecidos como meta-algoritmos para se otimizar a aplicação deste método.

Nota-se em (7) que a influência da variação do número de condição (κ) na convergência do método é mais significativo do que a aproximação inicial $\{U_0\}$. Assim, quanto maior é a diferença entre os autovalores da matriz $[K]$, mais demorada é a convergência do método. Dessa forma, o principal empecilho para a utilização deste método é justamente a necessidade de se ter um número de condição (κ) de $[K]$ o mais próximo do valor unitário.

Nos últimos cinco anos, extensos estudos têm sido dedicados à busca de técnicas matemáticas para se obter um número de condição (κ) o mais próximo do valor um.

A idéia básica empregada é de ao invés de se resolver o sistema $[K] \cdot \{U\} = \{F\}$, resolve-se um sistema do tipo $[M] \cdot [K] \cdot \{U\} = [M] \cdot \{F\}$, tal que a matriz obtida do produto ($[M] \cdot [K]$) seja o mais próximo possível da matriz identidade, onde sabe-se que $\kappa_{\text{identidade}}=1$. Esta matriz $[M]$ é conhecida como pré-condicionador, que nada mais é do que um acelerador de convergência para o método.

3 PRÉ-CONDICIONADORES

A técnica de pré-condicionamento consiste em transformar a matriz $[K]$ do sistema (1) em uma matriz $[\tilde{K}]$ que seja a mais próxima da matriz identidade $[I]$ e, desta forma, bem condicionada. Ou seja, do sistema inicial dado pela equação (1), pré-multiplica-se pela inversa da matriz pré-condicionadora:

$$([M]^{-1} \cdot [K]) \cdot \{U\} = [M]^{-1} \cdot \{F\} \tag{8}$$

de modo que quanto mais $[M]^{-1}$ se aproxime de $[K]^{-1}$, tem-se a relação

$$\kappa(\tilde{K}) \ll \kappa(K) \text{ e } \kappa(\tilde{K}) \rightarrow 1 \tag{9}$$

fica-se então no limite quando $[M]^{-1} = [K]^{-1}$. A convergência do método se torna direta e a iteração (*ite*) é imediata, ou seja, na representação em forma de limite:

$$\lim_{M^{-1} \rightarrow K^{-1}} \kappa(\tilde{K}) = 1 \text{ ou } \lim_{M^{-1} \rightarrow K^{-1}} \text{ite} = 1 \tag{10}$$

esquemáticamente, resume-se a equação (10) na figura (2).

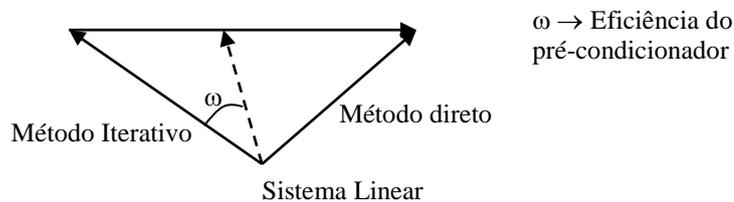


Figura 2 –Eficiência do pré-condicionamento sobre o sistema

Mas para se obter a matriz inversa exata de $[K]$ tem-se um esforço computacional muito maior que o necessário para resolver o sistema sem o uso de pré-condicionador. Assim, a estratégia numérica buscada no estudo destes pré-condicionadores é a obtenção de uma matriz especial que seja a mais próxima possível da inversa de $[K]$, contudo, com um reduzido esforço computacional para a sua montagem e a garantia da eficácia para a convergência do sistema equivalente.

A adaptação do pseudoalgoritmo do MGC puro, considerando-se o efeito do pré-condicionamento, pode ser feita de modo que ao invés de se aplicar o pré-condicionador diretamente na matriz de rigidez, realiza-se a contribuição da matriz pré-condicionadora ao vetor resíduo $\{r_p\}$, ou seja:

$$\{\tilde{r}_p\} = [M]^{-1} \cdot \{F\} - [M]^{-1} \cdot [K] \cdot \{U\} \quad (11)$$

4 PRÉ-CONDICIONADOR POLINOMIAL

O pré-condicionador polinomial é uma outra técnica clássica para se obter a aproximação da inversa da matriz $[K]$, mediante uma aproximação para a inversa desta matriz com o uso de uma expansão polinomial em potências de $[K]$.

Assim, como vantagem deste método quando comparado com o método da decomposição incompleta, é que nesta aproximação da inversa de $[K]$ tem-se que realizar apenas um produto matriz-vetor para obter o vetor condicionado $\{z\}$, não tendo que aplicar a substituição e a retrosubstituição como no caso anterior. Em [12] mostram-se todas as hipóteses e provas matemáticas necessárias para aplicar o método nos problemas aqui tratados.

Parte-se inicialmente da hipótese de que a matriz $[K]$ tem que ser simétrica e positiva-definida e, então, pode-se reescrevê-la da seguinte forma:

$$[K] = [D_s] + [\bar{K}] \quad (12)$$

tal que $[D_s]$ é a matriz formada apenas pelos elementos da diagonal principal de $[K]$ e $[\bar{K}]$ é a matriz dos elementos de $[K]$ com valores nulos em toda a diagonal principal.

Colocando em evidência a matriz diagonal $[D_s]$ na equação (12), tem-se:

$$[K] = [D_s] \cdot ([I] + [D_s]^{-1} \cdot [\bar{K}]) \quad (13)$$

E elevando os dois membros a (-1) na equação (13) e aplicando as propriedades matriciais, chega-se a

$$[K]^{-1} = ([I] + [D_s]^{-1} \cdot [\bar{K}])^{-1} \cdot [D_s]^{-1} \quad (14)$$

E conforme [12, 13], se a matriz $[K]$ for *estritamente diagonal dominante*, pode se escrever a primeira parcela do segundo membro da equação (14) como uma série polinomial da seguinte maneira:

$$([I] + [D_s]^{-1} \cdot [\bar{K}])^{-1} = \sum_{J=0}^{\infty} \{(-1)^J \cdot ([D_s]^{-1} [\bar{K}])^J\} \quad (15)$$

onde $[I]$ é a matriz identidade.

Ou seja, substituindo a equação (15) em (16), a aproximação da inversa de $[K]$ é dada por:

$$[K]^{-1} = \left\{ \sum_{J=0}^{\infty} \{(-1)^J \cdot ([D_s]^{-1} [\bar{K}])^J\} \right\} [D_s]^{-1} \quad (16)$$

O fato da matriz $[K]$ ter que ser estritamente diagonal dominante é porque sabendo-se que a aproximação em série da expressão

$$\frac{1}{1+x} \quad \text{tal que } (0 \leq x < 1) \quad (17)$$

é dada por

$$\frac{1}{1+x} \cong 1 - x + x^2 - x^3 + \dots \quad (18)$$

Analogamente esta idéia, considerando as devidas diferenças, pode ser estendida para a aplicação em matrizes, fazendo:

$$[\tilde{X}] = \frac{[K]}{[D_s]} \quad (19)$$

onde $[\tilde{X}]$ é a matriz resultante do quociente entre $[K]$ e $[D_s]$. Resulta-se então na aproximação polinomial

$$\frac{[I]}{[I] + [\tilde{X}]} \cong [I] - [\tilde{X}] + [\tilde{X}]^2 - [\tilde{X}]^3 + \dots \quad (20)$$

que é equivalente a equação (15).

Se a matriz $[K]$ é *estritamente diagonal dominante*, na matriz resultante $[\tilde{X}]$ pode ser mostrado que seu determinante é menor que 1. Assim, vale a aproximação e a série é convergente.

Na literatura para a aplicação da série é comum truncar o polinômio em, no máximo, potências de cinco. Comenta-se que valores acima destes é comum começar a ocorrer instabilidade no sistema tratado, conforme o aumento na ordem da potência.

Em [10], por exemplo, faz-se aplicações com expansões de potências de ordem um e dois e, em geral, esta técnica de pré-condicionamento é denominada de POLY (n_T), tal que n_T é o número de potências retidas no truncamento da série.

Um pré-condicionador que é fácil de ser implementado tanto em ambiente em série como em paralelo é o POLY(0), que também é denominado de pré-condicionador Jacobi. Ele é dado pela matriz formada apenas pela diagonal principal onde os termos são calculados pelo inverso do valor da referida posição da matriz $[K]$. Ou seja:

$$[K]^{-1} \cong \text{diag}(1/k_{ii}) \quad i = 1, n \quad (21)$$

4.1 PRÉ-CONDICIONADOR POLINOMIAL INCOMPLETO - POLY (n_T, m_I)

Para cada aproximação da série pelo truncamento da potência, realiza-se também um *truncamento sobre o número e colunas* da matriz que está sendo multiplicada por ela mesma, ou seja, $[K]^J$. Monta-se então a aproximação pela série, considerando apenas as colunas que estejam mais próximas da diagonal principal, já que são estes elementos que tem valores mais relevantes para obtenção da inversa de $[K]$.

Em função da possibilidade de se variar o espectro de influência sobre o produto da matriz por ela mesma, foi necessário desenvolver aqui uma outra notação para especificar o tipo de série polinomial e o número de diagonais (m_I) a partir da diagonal principal, que serão consideradas. Desta forma, utilizando a notação apresentada anteriormente incluiu-se apenas o número de diagonais, ou seja, a notação fica POLY (n_T, m_I).

5 APLICAÇÕES

A seguir são apresentados os resultados com a aplicação do pré-condicionador polinomial incompleto comparando-os com a técnica convencional em que se considera a influência de todo da espectro a matriz $[K]$. O termo POLY(n_T, M) significa que foi adotado a série truncada na potência n_T , e realizado a contribuição de toda matriz até o comprimento de banda (M) para a parcela da potência da série.

Os resultados aqui apresentados foram obtidos em um microcomputador do tipo PENTIUM III 450. Os valores de tempo têm apenas o intuito de comparar a eficiência relativa entre as técnicas consideradas.

5.1 TRELIÇA TRIDIMENSIONAL

Apresenta-se resultados de convergência em termos de tempo e número de iterações para o exemplo de treliça tridimensional em regime elastolinar. Este exemplo, extraído em [14], bem como as características adotadas, sendo que aplicou-se uma carga concentrada no nó central da estrutura, no banzo superior, de valor indicado na tabela da

figura 3. A precisão estipulada para o critério de convergência do método foi de 0,01 e para o vetor inicial atribuiu-se zero.

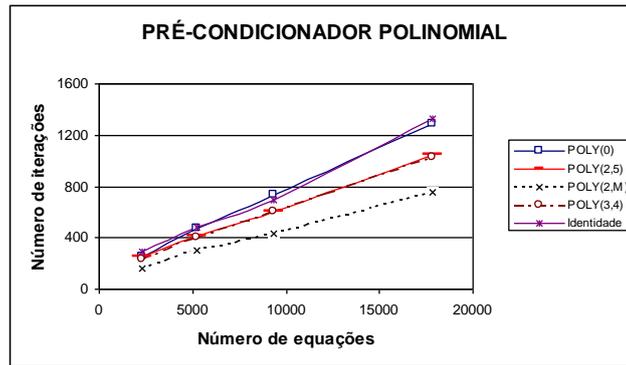


Figura 4 –Curva (Nº de iterações x Nº de equações) para a treliça

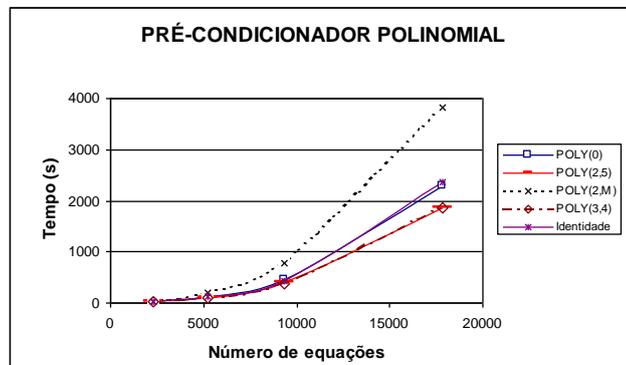


Figura 5 –Curva (Tempo x Nº de equações) para a treliça

5.2 PLACA QUADRADA SIMPLESMENTE APOIADA

O modelo aqui utilizado para verificar os dois métodos é o advindo de um problema do MEF aplicado a uma estrutura bidimensional de placa com elementos DKT, sendo que as características geométricas do material e o tipo de carregamento é apresentado a seguir. Também tem-se os resultados para o sistema sem o uso de pré-condicionadores, considerando a matriz identidade como pré-condicionador.

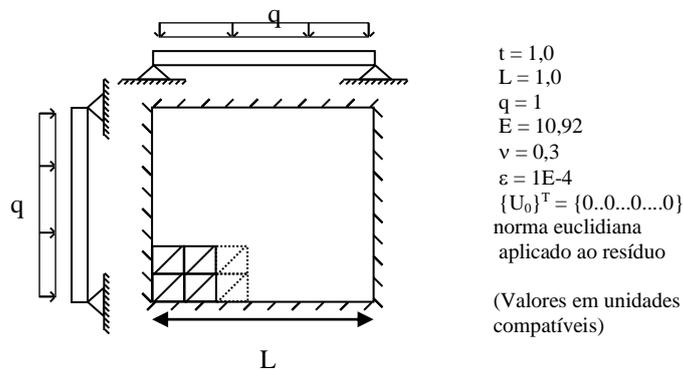


Figura 4 – Esquema da placa, sua orientação e parâmetros de geometria e do material

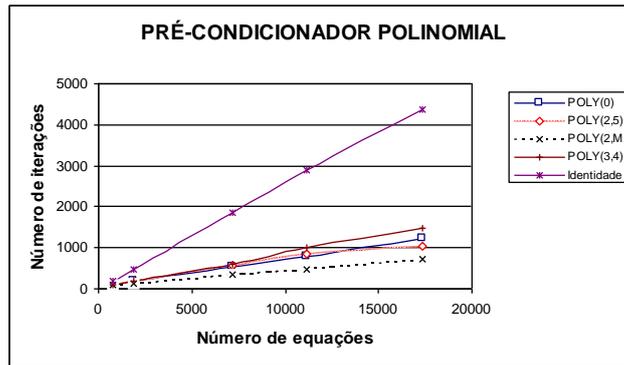


Figura 6 –Curva (Nº de iterações x Nº de equações) para a placa

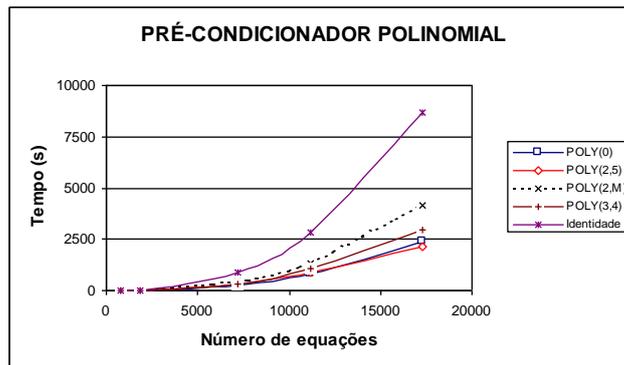


Figura 7 –Curva (Tempo x Nº de equações) para a placa

6 CONCLUSÕES

O trabalho pretendia avaliar o grau de convergência em termos de número de iterações e de tempo para a resolução do sistema linear advindo de problemas elastostáticos do MEF usando-se o método dos gradientes conjugados com pré-condicionamento. Pelo fato de que este método empregado é dependente do número de condição da matriz de rigidez, é comum – e muitas vezes necessário – empregar as técnicas de pré-condicionamento para reduzir o mau-condicionamento do sistema.

Neste contexto, este artigo propôs uma adaptação para a clássica forma de otimização de pré-condicionamento, sendo que esta técnica clássica é conhecida como série truncada de Neumann. A adaptação realizada incorpora além da técnica de truncamento da série, fato este já comumente empregado, como também realiza um truncamento na contribuição das potências da matriz de rigidez para a formação da série.

Esta adaptação, mostrada nos exemplos anteriores, apresentaram excelentes resultados à diminuição do tempo de convergência. O pré-condicionador POLY(2,m), como era de se esperar, apresentou uma alta taxa de convergência em termos de número de iterações, pois o erro estava apenas relacionado a escolha da potência a ser relevada na série. Por outro lado, em função da grande quantidade de operações que é necessário para se realizar em cada iteração, o tempo consumido foi alto quando comparado aos outros pré-condicionadores.

No emprego da série POLY(2,5) e POLY(3,4) percebe que ao realizar o truncamento também nos valores da matriz de rigidez, o número de iterações cresce, já que estamos diminuindo a precisão do método. Por outro lado a quantidade de operações a serem realizadas em cada iteração é menor quando comparada com o POLY(2,M). Assim, nas figuras (5) e (7) nota-se que a estratégia de truncar o espectro de influência para a aproximada da inversa de [K] mostrou ser uma ótima opção frente a tradicional técnica polinomial, sendo que o POLY(2,5) é a melhor escolha a ser adotada.

Com relação ao uso desta inovadora técnica polinomial em processamento paralelo, é esperado que ela venha a se sobressair também, pois o grande “gargalo” para o aumento de tempo desta técnica é para a realização do produto matriz-vetor várias vezes. Entretanto, este produto possui vantagens para se realizar neste ambiente computacional.

7 AGRADECIMENTOS

O trabalho presente foi financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo). Desta forma, os autores agradecem o apoio da instituição para a confecção do trabalho.

8 BIBLIOGRAFIA

- [1] A. K Noor, "New Computing systems and future high-performance computing environment and their impact on structural analysis and design". *Computers & Structures*, 64, n. 1-4, p. 1-30 (1997).
- [2] M. N Rezende, *Processamento paralelo em análise estrutural*, São Carlos. 112p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo (1995).
- [3] Valério S Almeida, *Uma adaptação do MEF para análise em multicomputadores: aplicações em alguns modelos estruturais*, São Carlos. 126p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo (1999).
- [4] J. A Cuminato e M. Jr Meneguette, *Discretização de equações diferenciais parciais: Técnicas de Diferenças finitas*. Instituto de Ciência e Matemática de São Carlos, Universidade de São Paulo (1998).
- [5] A. del Grosso e G. Righetti, "Finite element techniques and artificial intelligence on parallel machines", *Computers & Structures*, v. 30, n.04, p. 999-1007 (1988).
- [6] K. H. Law, "A parallel finite element solution method", *Computers & Structures*, v. 23, n. 6, p. 845-858 (1986).
- [7] S. Bitzarakis, M. Papadrakakis e A. Kotsopoulos, "Parallel solution techniques in computational structural mechanics", *Comput. Methods Appl. Mech. Engr.*, v. 148, p. 75-104 (1997).
- [8] J. A Meijerink e H. van der Vorst, "An iterative solution method for linear systems of wich the coefficient matrix is a symmetric M-matrix", *Mat. Comp.*, v.31 (137), p.148-162 (1977).
- [9] M. Cimerman, *Resolução de Sistemas Lineares Via Métodos Iterativos com Pré-Condicionadores – Aplicação em Problemas de Engenharia de estruturas*. São Paulo. 111p. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo (1996).
- [10] L. A. Jr Schmt e Y. C. Lai, "Structural optimization based on preconditioned conjugate gradient analysis methods". *International Journal for numerical methods in engineering*, v. 37, p. 943-964 (1994).
- [11] T. A Manteuffel, "An incomplete factorization technique for positive definitive linear systems", *Mat. Comp.*, v.34 (150), p.473-497 (1980).
- [12] P. F Dubois, A. Greenbaum e G. H. Rodrigue, "Aproximating the Inverse of a Matrix for Use in Iterative Algorithms on Vector Processors". *Computing*, v.22, p. 257-268 (1977).
- [13] O. G. Johnson, C. A. Micchelli e G. Paul, "Polynomial preconditioners for conjugate gradient calculations", *SIAM J. Num. Anal.*, v.20, p. 362-376 (1983).
- [14] A. S. C. Souza, "*Contribuição ao estudo das estruturas metálicas espaciais*". São Carlos. 147p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo (1998).