

Funções de Aproximação e Modelos Estatísticos Aplicados a Processos de Otimização

Luiz Lebensztajn e Carina Alexandra Rondini Marretto

EPUSP – Escola Politécnica da Universidade de São Paulo

LMAG – Laboratório de Eletromagnetismo Aplicado

Av. Prof. Luciano Gualberto, Trav. 3, n 158, 05508-900, São Paulo, SP, Brasil

Resumo — Este trabalho tem como objetivo mostrar a viabilidade da utilização dos modelos estatísticos, Kriging e Cokriging, para a substituição da função objetivo no âmbito de um processo de otimização. O modelo Kriging utiliza somente os valores da função, enquanto o modelo Cokriging usa os valores da função e seus gradientes. Esta capacidade do Cokriging pode melhorar a qualidade da função de substituição, e como consequência, o resultado da otimização. A robustez de ambas as aproximações é verificada em funções analíticas e posteriormente no Problema 25 do TEAM Workshop.

Palavras-chaves — Modelos por Substituição: Kriging, Cokriging, Otimização.

I. INTRODUÇÃO

No projeto de um equipamento eletromecânico é necessária a determinação das solicitações eletromagnéticas que atuam sobre o dispositivo. Esta tarefa diária para um projetista possui um grau de dificuldade bastante elevado. A complexidade da geometria, a não linearidade do circuito magnético e anisotropia dos materiais são alguns dos itens que dificultam a análise de um dispositivo eletromecânico. Essas particularidades da análise do projeto de um equipamento levaram pesquisadores e projetistas a procurarem técnicas computacionais como, por exemplo, o Método dos Elementos Finitos (MEF) para avaliar campos eletromagnéticos no interior de equipamentos.

Ao se utilizar técnicas de otimização associadas ao cálculo de campos por elementos finitos, o projeto ótimo é alcançado a partir de um processo sistematizado, porém realizado pelo computador. Em geral, as técnicas de otimização, sejam elas determinísticas ou estocásticas, demandam um tempo computacional baixo, ao passo que cada cálculo de campo por elementos finitos demanda um tempo computacional elevado. Nesse trabalho serão utilizadas técnicas de otimização estocásticas, que requerem um número elevado de avaliações da função objetivo, ou seja, diversos cálculos por elementos finitos. Isto pode implicar em um esforço computacional elevado.

Uma estratégia viável para contornar esse problema é a adoção de funções de aproximação para a representação do modelo de elementos finitos. Essas funções devem fornecer uma aproximação do modelo, ou seja, são modelos do modelo, ou metamodelos. Desta forma, no âmbito do

processo de otimização, o modelo de elementos finitos é substituído por uma função de menor custo computacional. Caso a aproximação represente o modelo de elementos finitos com bastante precisão, o ótimo global será alcançado.

Existem diversas metodologias de construção de metamodelos ou funções de aproximação. Dentre elas, pode-se citar as Superfícies de Resposta [1], as Redes Neurais Artificiais [2] e o Método dos Elementos Difusos [3].

As funções de aproximação adotadas no âmbito desse trabalho são de natureza estatística. O primeiro metamodelo é denominado Kriging, que é um termo genérico dado a uma família de métodos de interpolação, cuja principal característica é minimizar o erro da variância da estimação, definida a partir de um modelo de covariância. Sua metodologia explora a correlação espacial dos dados para construir a interpolação.

O segundo metamodelo é uma extensão do modelo Kriging e é denominado Cokriging. Ele pode agregar aos valores da função, nos pontos amostrais, os valores do gradiente para melhorar a precisão do modelo.

Esses valores do gradiente podem ser calculados através de metodologias com custo computacional mais baixo que o MEF, como por exemplo: métodos adjuntos, análise de sensibilidade ou um esquema de diferenças finitas, como proposto por Chung e Alonso [4]. No presente trabalho, os modelos Kriging e Cokriging são apresentados como interpoladores exatos.

II. METODOLOGIA

O presente trabalho, utiliza a Estimativa de Máxima Verossimilhança (MLE) para construir a interpolação dos modelos Kriging e Cokriging. A formulação básica dos modelos é apresentada a seguir. Detalhes específicos podem ser obtidos no trabalho de Koehler e Owen [5].

A. Formulação do Kriging

O princípio da metodologia Kriging é interpolar um conjunto de dados fornecidos, através da associação de funções locais a uma função global. Desta forma, um modelo Kriging pode ser definido através da expressão:

$$y(x) = f(x) + Z(x) \quad (1)$$

em que $y(x)$ é uma função de interpolação, $f(x)$ é um modelo global, que nesse trabalho foi adotado constante e

L. Lebensztajn, leb@pea.usp.br, C. A. R. Marretto, carinarm@pea.usp.br, Tel +55-11-30915533, Fax +55-11-38144013

Esse trabalho foi financiado pela FAPESP, (Processo 00/08160-7).

igual a β e $Z(x)$ é a realização de um processo estocástico $N(0, \sigma^2)$ com covariância não nula.

A covariância entre $Z(x_i)$ e $Z(x_j)$, ou seja, entre quaisquer dois pontos x_i e x_j do total de N_s pontos amostrais, é obtida pelo produto da variância do processo σ^2 e a função de correlação espacial $\mathfrak{R}(x_i, x_j)$:

$$\text{Cov}[Z(x_i), Z(x_j)] = \sigma^2 \mathfrak{R}(x_i, x_j) \quad (2)$$

A função de correlação espacial determina como o modelo se ajusta aos dados. Existem vários tipos de funções de correlação, porém em experimentos por computador, a função de correlação é normalmente assumida como uma Gaussiana e apresentada da seguinte forma:

$$\mathfrak{R}(x_i, x_j) = \prod_{k=1}^{N_{par}} \exp[-\theta_k (|x_i - x_j|_k)^2] \quad (3)$$

em que N_{par} é o número de variáveis de projeto (ou parâmetros) e $|x_i - x_j|_k$ é a diferença em módulo, entre as coordenadas da amostra x_i e x_j na direção k .

O parâmetro θ_k é uma constante positiva na direção k e mede como os dados estão correlacionados nessa direção. Um valor elevado de θ_k mostra um conjunto de dados pouco correlacionados.

O melhor estimador linear não viesado (BLUP) de $y(x)$ é apresentado como:

$$\hat{y}(x) = \hat{\beta}(\theta) + r^T(x, \theta) \times R^{-1}(\theta) \times (y - f \hat{\beta}(\theta)) \quad (4)$$

em que $\hat{y}(x)$ é o valor estimado em x , y é o vetor dos valores da função de dimensão $(N_s \times 1)$, f é um vetor unitário uma vez que $f(x)$ foi assumido constante também de dimensão $(N_s \times 1)$ e R é a matriz de correlação cujo elemento $\mathfrak{R}(x_i, x_j)$ é definido por (3). A matriz de correlação é simétrica, definida positiva com dimensão $(N_s \times N_s)$ e diagonal de valor unitário.

O vetor $r^T(x)$, calculado na forma:

$$r^T(x) = [\mathfrak{R}(x, x_1) \dots \mathfrak{R}(x, x_{N_s})]^T \quad (5)$$

é o vetor de correlação entre o ponto a ser estimado x e os N_s pontos amostrais.

O parâmetro estimado de β e a variância estimada de σ^2 entre o modelo global β e y são respectivamente apresentados como:

$$\hat{\beta}(\theta) = (f^T R^{-1}(\theta) f)^{-1} (f^T R^{-1}(\theta) y) \quad (6)$$

$$\hat{\sigma}^2(\theta) = [(y - f \hat{\beta}(\theta))^T R^{-1}(\theta) (y - f \hat{\beta}(\theta))] / N_s \quad (7)$$

O MLE do parâmetro θ pode ser obtido maximizando a expressão:

$$-\frac{N_s (\ln(\hat{\sigma}^2(\theta)) + \ln(|R(\theta)|))}{2} \quad (8)$$

A solução deste sistema não linear sem restrição fornece o valor de θ . É importante observar que qualquer valor de θ produz uma interpolação dos dados, contudo, somente a otimização de (8) e o posterior uso de (4) é que permitem a obtenção de um modelo ótimo.

B. Formulação do Cokriging

No modelo Kriging a interpolação é realizada através do valor da função nos pontos amostrais. Para o modelo Cokriging a interpolação é realizada através do valor da função nos pontos amostrais e do valor da derivada da função nesses pontos. Desta forma, a formulação matemática para o Cokriging segue a formulação apresentada para o Kriging porém, algumas mudanças ocorrem para que os valores das derivadas sejam incorporados.

A metodologia do modelo Cokriging assim como o modelo Kriging, explora a correlação espacial. Então, deve-se calcular a correlação entre os pontos amostrais, tal qual realizada no Kriging. Como o Cokriging incorpora os valores da derivada existe a necessidade de correlacionar esses valores em um ponto com o valor da função em um outro ponto ou mesmo com o valor da derivada da função. Desta forma, apenas o conhecimento da covariância entre $Z(x_i)$ e $Z(x_j)$ não é suficiente. Torna-se necessário conhecer a covariância entre a realização do processo estocástico $Z(x_i)$ e sua derivada $Z'(x_j)$. O cálculo da covariância entre as derivadas das realizações do processo estocástico $Z'(x_i)$ e $Z'(x_j)$ é igualmente necessário. A expressão generalizada para o cálculo dessas covariâncias toma a forma:

$$\text{Cov}[Z^{(a_1, \dots, a_p)}(x_i), Z^{(b_1, \dots, b_p)}(x_j)] = (-1)^a \sigma^2 \prod_{j=1}^p \mathfrak{R}_j^{(a_j, b_j)}(x_i, x_j) \quad (9)$$

em que $a = \sum_{j=1}^p a_j$ e $b = \sum_{j=1}^p b_j$.

Assim, tem-se que a matriz de covariância do modelo é definida como sendo:

$$\text{Cov}[Z(x_i), Z(x_j)] = \sigma^2 \mathfrak{R}(x_i, x_j) \quad (10)$$

$$\text{Cov}[Z(x_i), Z'(x_j)] = \sigma^2 \mathfrak{R}'(x_i, x_j) \quad (11)$$

$$\text{Cov}[Z'(x_i), Z(x_j)] = -\sigma^2 \mathfrak{R}'(x_i, x_j) \quad (12)$$

$$\text{Cov}[Z'(x_i), Z'(x_j)] = -\sigma^2 \mathfrak{R}''(x_i, x_j) \quad (13)$$

O melhor estimador linear não viesado de $y(x)$ agora passa ser:

em que f_{ck} é um vetor preenchido com N_S valores iguais a 1 e $N_S \times N_{par}$ valores iguais a zero. A matriz R_{ck} tem dimensão $((N_{par} + 1)N_S \times (N_{par} + 1)N_S)$.

em que $\mathfrak{R}_{N_{par}}^*(x, x_{N_S})$ é a derivada da função de correlação com respeito a variável de projeto, $x_{N_{par}}$, no ponto (x, x_{N_S}) .

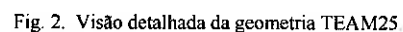
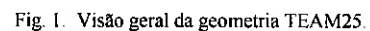
Para validação dos modelos Kriging e Cokriging desenvolvidos, utilizou-se duas funções analíticas ambas a 3 variáveis. A primeira (f_1) proposta por Alotto et al. [6] apresenta-se como:

A segunda (f_2) proposta por Vasconcelos et al. [7] apresenta-se como:

Para essas funções analíticas, duas aproximações foram usadas para o cálculo das derivadas das funções para o modelo Cokriging. A primeira é analítica e fornece o valor exato da derivada. A segunda é numérica, baseada num esquema de diferenças finitas (EDF) através de dois pontos. Um passo de 0.5% do intervalo do parâmetro foi adotado.

No problema 25, cuja geometria é apresentada nas Fig. 1 e 2, tem-se um magnetizador em que o objetivo é encontrar uma forma geométrica que possibilite uma distribuição de indução magnética extremamente homogênea e radial ao longo da linha $e - f$. Para isto, deve-se minimizar a função objetivo (20), calculada ao longo do arco $e - f$ representado em detalhe na Fig. 2.

Na expressão (20), faz-se a diferença entre induções magnéticas especificadas e calculadas. O índice p indica um valor de indução calculado e o índice o um valor especificado.



Os parâmetros geométricos para a otimização são R1, L2 e L4 cujas variações em (mm) são definidas por: $R1 = [5, 9.4]$,

L2 = [12.6, 18] e L4 = [4, 19]. O parâmetro L3 foi fixado em 14 mm.

O mesmo EDF descrito para as funções analíticas foi usado para a aplicação do Cokriging.

É importante observar que o EDF tem um custo computacional mais elevado do que uma análise de sensibilidade contudo, quando N_{par} é baixo, esse custo computacional não é muito alto.

Para mostrar a performance dos modelos Kriging e Cokriging, todos os cálculos foram realizados sobre uma grade regular, porém, para o modelo Cokriging, em cada ponto amostral, foi calculado o valor da função e três valores da derivada. Desta forma, para o modelo Kriging, o tamanho amostral deve ser maior do que o tamanho amostral para o Cokriging.

Para o Cokriging, não utilizou-se a informação do gradiente da função para a determinação do MLE do parâmetro θ . Assim, manipula-se matrizes de dimensão bastante reduzida o que facilita o processo de otimização descrito por (8).

IV. RESULTADOS

Os resultados obtidos nos casos descritos anteriormente estão dispostos nas Tabelas subsequentes.

Nas Tabelas I e II apresentadas a seguir, r^2 é o coeficiente de determinação e f_{min} é o valor do mínimo global obtido. O símbolo CK foi reservado para abreviar o termo Cokriging, que foi calculado com derivadas exatas ou através do EDF descrito anteriormente.

Todos os resultados de otimização foram obtidos através do método Simulated Annealing.

TABELA I MINIMIZAÇÃO DA FUNÇÃO f_1

Método	N_s	$1-r^2$	Mínimo	f_{min}
Análítico	----	----	[-4.45 -4.45 -4.45]	-7.85
Kriging	10^3	7×10^{-3}	[-4.55 -4.54 -4.55]	-7.98
CK (Exata)	216	2×10^{-6}	[-4.46 -4.46 -4.47]	-7.88
CK(EDF)	216	2×10^{-5}	[-4.44 -4.33 -4.54]	-7.90

TABELA II MINIMIZAÇÃO DA FUNÇÃO f_2

Método	N_s	$1-r^2$	Mínimo	f_{min}
Análítico	----	----	[3.59 3.59 3.59]	17.31
Kriging	10^3	0,057	[3.49 3.49 3.49]	17.36
CK(Exata)	216	0,0341	[3.65 3.64 3.64]	17.71
CK(EDF)	216	0,042	[3.64 3.65 3.64]	17.71

Ao se analisar os resultados apresentados nas Tabelas I e II vê-se que os valores obtidos através dos modelos estão muito próximos dos valores analíticos. Este aspecto é reforçado com o alto fator de determinação obtido pelos modelos o que corresponde à qualidade da aproximação obtida.

Os maiores fatores de determinação foram alcançados quando do uso do Cokriging com os valores exatos da derivada, o que mostra a potencialidade do método.

A Tabela III lista os resultados obtidos para o problema 25 do TEAM Workshop. Esse problema também foi resolvido pelo Método dos Elementos Difusos (MED) e pelo Método Multiquadrics, descritos por Kobetski et al. [9]

Na Tabela III, f_{min} é o mínimo global e h_{min} é o valor da função objetivo calculado pelo MEF no ponto de mínimo.

TABELA III MINIMIZAÇÃO DO PROBLEMA 25 DO TEAM

Método	R1 (mm)	L2 (mm)	L4 (mm)	f_{min} $\times 10^{-3}$	h_{min} $\times 10^{-3}$
Multiquadrics	6.76	13.03	14.81	-4.4	4.4
MED	7.17	14.06	14.35	-0.43	0.11
Kriging	7.20	14.40	14.00	0.278	0.280
CK(EDF)	6.80	13.48	13.92	-2.9	0.392

A análise da Tabela III mostra que os métodos propostos têm bom desempenho, quando comparados com métodos mais tradicionais, como as multiquadrics. É importante observar que a função objetivo analisada é estritamente positiva. Deste ponto de vista a melhor das aproximações é o Kriging. O desempenho do Cokriging neste caso foi inferior, porque as derivadas não foram calculadas com a devida precisão. Entende-se que o esquema de diferenças finitas adotado seja viável em alguns casos, ou seja, é difícil determinar a priori o valor do passo de variação do parâmetro, a fim de calcular as derivadas. O uso de métodos adjuntos tornaria provavelmente o Cokriging mais efetivo.

V. CONCLUSÕES

Os modelos estatísticos propostos permitiram construir funções de substituição com alto fator de determinação o que comprova a validade da ferramenta desenvolvida. Ambos os modelos: Kriging e Cokriging se mostraram uma metodologia alternativa na construção de funções de substituição através da solução do problema 25 do TEAM Workshop mostrando a viabilidade dos modelos em problemas reais.

REFERÊNCIAS

- [1] R. H. Myers, and D. C. Montgomery, Response Surface Methodology Using Designed Experiments, John Wiley & Sons, New York, 1995.
- [2] T. Ishikawa, Y. Tsukui, M. Matsunami, "Optimization of electromagnetic devices using artificial neural network with quasi-Newton algorithm", IEEE Transactions on Magnetics, Volume: 32, Issue: 3, Pages: 1226 - 1229, May 1996.
- [3] M. Caldora Costa, J.-L. Coulomb, Y. Maréchal, A. B. Dietrich, S. I. Nabeta, "Diffuse element method and quadrees: two ingredients for an adaptive response surface" IEEE Transactions on Magnetics, vol. 38, Issue: 2, pp. 1085 -1088, March 2002
- [4] H. S. Chung, and J. J. Alonso, "Using gradients to construct Cokriging approximation models for high-dimensional design optimization problems", AIAA2002-0317, pp. 1-15, Janeiro 2002.
- [5] J. R. Koehler and A. B. Owen, Computer Experiments In: Handbook of Statistics, vol. 13. Elsevier-Science, New York, 1996, pp. 261-308.
- [6] P. Alotto, et al., "A multiquadrics based algorithm for the acceleration of simulated annealing optimization procedures", IEEE Transactions on Magnetics, vol. 32, nº 3, pp. 1198-1201, 1996.
- [7] J. A. Vasconcelos, et al., "Simulated Annealing coupled with the Tabu Search Method for continuum optimization in electromagnetics", IEEE Transactions on Magnetics, vol. 32, nº 3, pp. 1206-1209, 1996.
- [8] <http://ics.ecs-lyon.fr/team.html>.
- [9] A. Kobetski, et al., "Comparison of radial basis function approximation techniques", In: COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol. 22, nº 3, pp. 616-629, 2003.