



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Cibelle Dias de Carvalho Dantas Maia

Dupla abordagem do Kriging para projeto experimental ótimo Bayesiano

Florianópolis
2024

Cibelle Dias de Carvalho Dantas Maia

Dupla abordagem do Kriging para projeto experimental ótimo Bayesiano

Tese submetida ao Programa de Pós-Graduação
em Engenharia Civil da Universidade Federal de
Santa Catarina como requisito parcial para a obten-
ção do título de Doutora em Engenharia Civil.
Orientador: Prof. Rafael Holdorf Lopez, Dr.
Coorientador: Prof. Leandro Fleck Fadel Miguel, Dr.

Florianópolis
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Maia, Cibelle Dias de Carvalho Dantas

Dupla abordagem do Kriging para projeto experimental
ótimo Bayesiano / Cibelle Dias de Carvalho Dantas Maia ;
orientador, Rafael Holdorf Lopez, coorientador, Leandro
Fleck Fadel Miguel, 2024.

74 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Civil, Florianópolis, 2024.

Inclui referências.

1. Engenharia Civil. 2. Projeto Experimental Ótimo. 3.
Inferência Bayesiana. 4. Ganho de Informação Esperado. 5.
Kriging. I. Lopez, Rafael Holdorf. II. Miguel, Leandro
Fleck Fadel. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia Civil. IV. Título.

Cibelle Dias de Carvalho Dantas Maia

Dupla abordagem do Kriging para projeto experimental ótimo Bayesiano

O presente trabalho em nível de doutorado foi avaliado e aprovado em 09 de dezembro de 2024, pela banca examinadora composta pelos seguintes membros:

Prof^a. Roberta De Queiroz Lima, Dra.
Instituição Pontifícia Universidade Católica do Rio de Janeiro

Prof. Marco Antonio Luersen, Dr.
Instituição Universidade Tecnológica Federal do Paraná

Prof. Wellison José de Santana Gomes, Dr.
Instituição Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutora em Engenharia Civil.

Coordenação do Programa de
Pós-Graduação

Prof. Rafael Holdorf Lopez, Dr.
Orientador

Florianópolis, 2024.

AGRADECIMENTOS

Agradeço, antes de tudo, a Deus, cuja luz me conduz e me dá forças, especialmente nos momentos de desânimo.

Aos meus pais, Rommel e Rosemere, que não apenas me ensinaram o valor do estudo e do trabalho, mas também são exemplos de vida e amor, sendo meus espelhos como pessoas. Sei que sempre estarão ao meu lado, apoiando minhas escolhas e aplaudindo as minhas conquistas.

À minha querida irmã, Isabelle, por todo apoio, companheirismo e cumplicidade, compartilhados comigo ao longo de todos esses anos.

Ao meu namorado, Tiago, por estar sempre ao meu lado, me incentivando, dando forças e contribuindo também com seus conhecimentos.

Ao meu orientador, Prof. Dr. Rafael Holdorf Lopez, que, com dedicação e paciência, compartilhou seus conhecimentos e me ofereceu todo o apoio acadêmico necessário para o desenvolvimento deste projeto.

A todos os professores que fizeram parte da minha formação, pelas contribuições formais e informais, incentivos e exemplos inspiradores.

Aos meus familiares e amigos que, de alguma forma, me incentivaram a seguir em frente. Em especial, às minhas avós, in memoriam, vovó Carminha e vovó Tila, que tanto intercederam ao nosso bom Deus por mim durante toda a minha vida e trajetória acadêmica.

E, finalmente, a todos que, de alguma forma, contribuíram para a realização deste trabalho.

RESUMO

A análise de experimentos é fundamental para a geração de conhecimento científico em diversas áreas da engenharia. Em contextos práticos, onde os experimentos podem ser limitados pela disponibilidade de recursos, o projeto experimental ótimo (OED) é uma importante ferramenta para maximizar a eficiência e a relevância dos dados obtidos. O objetivo do OED é obter a melhor configuração experimental, otimizando uma métrica que quantifica o ganho de informação de um experimento, garantindo que os dados coletados sejam os mais informativos. No contexto Bayesiano, a eficácia de um experimento pode ser quantificada pelo ganho de informação esperado de Shannon (SEIG). No entanto, esse método envolve a avaliação de uma integral dupla que apresenta dois desafios principais. Primeiramente, essas integrais, frequentemente de alta dimensionalidade, dificilmente possuem soluções analíticas. Em segundo lugar, em diversos modelos, a computação do SEIG requer o uso de ferramentas numéricas computacionalmente intensivas, como métodos de elementos finitos. Essas características tornam o processo de otimização do SEIG computacionalmente dispendioso. Neste trabalho, é proposto uma dupla abordagem do Kriging, para resolver de forma eficiente problemas OED Bayesiano. A metodologia é estruturada em duas etapas. Na primeira, o metamodelo do Kriging substitui o modelo original direto, aumentando a eficiência computacional na avaliação do SEIG, estimado por meio do método de Monte Carlo de laço duplo. Na segunda etapa, o processo de otimização emprega o EGO, que requer a construção de um novo metamodelo do Kriging como substituto direto do SEIG. No EGO, é aplicado o critério de melhoria esperada como critério de preenchimento. Outra vantagem dessa metodologia é reduzir o problema da dimensionalidade tipicamente associada aos modelos substitutos do Kriging. Nessa estratégia, a primeira etapa do Kriging foca na modelagem do espaço das variáveis incertas, enquanto a segunda etapa se dedica à modelagem do espaço dos parâmetros de projeto a serem otimizados, evitando a necessidade de construir um modelo do Kriging em ambos os espaços simultaneamente. Essa segmentação permite que a metodologia proposta seja mais eficiente, especialmente em espaços de alta dimensionalidade. O método foi aplicado a três problemas de OED Bayesiano. O primeiro é um modelo genérico não-linear, com dois parâmetros a serem otimizados e duas variáveis incertas. Já o segundo e o terceiro exemplo consistem em posicionar, vertical e horizontalmente, extensômetros com base nas deformações da viga pela teoria de Timoshenko. No segundo exemplo, o objetivo é otimizar um único extensômetro para inferir sobre os módulos de elasticidade transversal e longitudinal da viga, resultando em dois parâmetros a serem otimizados e duas variáveis incertas. E o terceiro exemplo consiste em posicionar dois extensômetros para caracterizar três cargas verticais pontuais e suas respectivas posições, resultando em quatro parâmetros a serem otimizados e seis variáveis incertas. Os resultados demonstram que a dupla abordagem do Kriging (EGO-KR) solucionou eficazmente os problemas analisados, proporcionando boa acurácia e economias computacionais significativas, especialmente no terceiro e mais complexo exemplo.

Palavras-chave: Projeto Experimental Ótimo 1. Inferência Bayesiana 2. Ganho de Informação Esperado 3. Kriging 4. Otimização Global Eficiente 5.

ABSTRACT

The analysis of experiments provides scientific knowledge across various fields of engineering. In practical scenarios where experiments may be constrained by resource availability, optimal experimental design (OED) stands out as an essential tool for maximizing the efficiency and relevance of the data obtained. The goal of OED is to obtain the best experimental configuration by optimizing a metric that quantifies the information gain from an experiment, ensuring that the collected data are the most informative and relevant for the desired application. In the Bayesian context, the effectiveness of an experiment can be quantified by the Shannon's expected information gain (SEIG). However, this method involves evaluating a double integral that presents two main challenges. First, these integrals, which can be high-dimensional, rarely have an analytical solution. Second, in several models, the calculation of SEIG requires the use of computationally intensive numerical tools, such as finite element methods. These characteristics make the SEIG optimization process computationally expensive. In this work, we propose a two-step Kriging approach to efficiently solve Bayesian OED problems. To enhance computational efficiency in evaluating the SEIG, we introduced a Kriging surrogate as a replacement for the original forward model (step 1 Kriging). This surrogate is utilized within the double loop Monte Carlo method for SEIG estimation. We employed the efficient global optimization (EGO) algorithm as the optimizer, which requires the construction of a Kriging surrogate of the SEIG (step 2 Kriging). Within EGO, the expected improvement infill criterion was employed. The underlying rationale of employing a two-step Kriging approach is to alleviate the curse of dimensionality typically associated with Kriging surrogates. In this strategy, the first step of Kriging is focused on surrogating the random parameter space, while the second step is dedicated to modeling the design variable space. By taking this two-step approach, the need to build a global replacement for the model in both spaces is avoided. This segmentation allows for more efficient and accurate surrogate modeling, particularly in high-dimensional spaces, enhancing the overall computational performance of the optimization process. The method was applied to three OED problems. The first is a generic nonlinear model, with two parameters to be optimized and two uncertain variables. The second and third examples involve positioning strain gauges, both vertically and horizontally, based on the beam deformations according to Timoshenko's theory. In the second example, the objective is to optimize a single strain gauge to infer the transverse and longitudinal elasticity moduli of the beam, resulting in two parameters to be optimized and two uncertain design variables. The third example involves positioning two strain gauges to characterize three vertical point loads and their respective positions, resulting in four parameters to be optimized and six uncertain design variables. The results demonstrate that the proposed two-step Kriging approach (EGO-KR) effectively addressed the analyzed problems, offering good precision and significant computational savings, particularly in the third and more complex example.

Keywords: Optimal Experimental Design 1. Bayesian Inference 2. Expected information gain 3. Kriging 4. Efficient Global Optimization 5.

LISTA DE FIGURAS

Figura 1 – Viga bi-apoiada para posicionamento de extensômetro.	22
Figura 2 – PDFs a priori e a posteriori de E para o exemplo da viga de Euler-Bernoulli.	24
Figura 3 – Curva da entropia relativa $D_{KL}(\mathbf{d}, \mathbf{Y})$ do problema da viga de Euler-Bernoulli.	26
Figura 4 – PDFs da priori e da posteriori de E para os casos otimizado e não otimizado do exemplo da viga de Euler-Bernoulli.	27
Figura 5 – Curva da entropia relativa $D_{KL}(\mathbf{d}, \mathbf{Y})$ para os casos otimizado e não otimizado do exemplo da viga de Euler-Bernoulli.	28
Figura 6 – Curva da entropia relativa $\mathcal{I}_{DLMC}$ do exemplo da viga de Euler-Bernoulli.	30
Figura 7 – Representação do metamodelo do Kriging determinístico.	34
Figura 8 – Influência do parâmetro p_k na função de correlação	37
Figura 9 – Influência do parâmetro c_k na função de correlação	38
Figura 10 – Gráfico do modelo e do metamodelo do Kriging para o exemplo da viga de Euler-bernoulli na posição $x = 3,0\text{m}$	47
Figura 11 – Fluxograma para o método de dupla abordagem do Kriging em problemas OED Bayesiano.	53
Figura 12 – Estimativa do SEIG pelo EGO para o exemplo da viga de Euler-Bernoulli.	54
Figura 13 – Representação do modelo original e do metamodelo do Kriging nas posições $d = 0.2$ e $d = 1.0$	57
Figura 14 – Estimativa do SEIG para o projeto de dois experimentos pelo EGO-KR: os círculos laranja representam o plano de amostragem inicial, os círculos azuis indicam os pontos de preenchimento e as estrelas representa o mínimo obtido.	58
Figura 15 – Geometria da viga de Timoshenko e da seção transversal para o problema de caracterização do material.	59
Figura 16 – Boxplots dos valores ótimos obtidos em 10 execuções independentes para o problema de caracterização de propriedades mecânicas da viga.	62
Figura 17 – Geometria da viga de Timoshenko e seção transversal para o problema de caracterização das cargas verticais.	63
Figura 18 – Boxplots dos ótimos obtidos para o problema de caracterização de carga da viga.	66

LISTA DE TABELAS

Tabela 1 – Estatísticas das 10 execuções independentes para o problema não linear.	58
Tabela 2 – Parâmetros da distribuição a priori para o problema de caracterização das propriedades mecânicas da viga.	60
Tabela 3 – Estatísticas das 10 execuções independentes para o problema da viga de caracterização das propriedades mecânicas: os valores entre parênteses indicam o SEIG estimado utilizando DLMC+Kriging no caso do EGO-KR, e utilizando a aproximação PCE+Laplace no caso do EGO-PCE	61
Tabela 4 – Parâmetros da priori para o problema de caracterização de cargas na viga.	64
Tabela 5 – Parâmetros do ruído aditivo para o problema da viga de caracterização da carga.	64
Tabela 6 – Estatísticas para o exemplo caracterização da carga da viga: os valores entre parênteses indicam o SEIG estimado utilizando DLMC+Kriging no caso do EGO-KR, e utilizando a aproximação PCE+Laplace no caso do EGO-PCE.	66

LISTA DE ABREVIATURAS E SIGLAS

BO	<i>Bayesian optimization</i>
DLMC	<i>Double loop Monte Carlo</i>
DLMCIS	<i>Double loop Monte Carlo with importance sampling</i>
EGO	<i>Efficient Global Optimization</i>
EI	<i>Expected Improvement criterion</i>
GPR	<i>Gaussian Process Regression</i>
i.i.d.	<i>Independent and identically distributed</i>
IC	<i>Infill criteria</i>
IPs	<i>Infill points</i>
LHS	<i>Latin hypercube sampling</i>
MC	<i>Monte Carlo</i>
MCI	<i>Monte Carlo integration</i>
MCLA	<i>Monte Carlo with Laplace approximation</i>
MLE	<i>Maximum likelihood estimates</i>
MSE	<i>Mean squared error</i>
NCFM	<i>Number of calls of the forward model</i>
OED	<i>Optimal experimental design</i>
PCE	<i>Polynomial Chaos expansions</i>
PDF	<i>Probability Density Function</i>
PSO	<i>Particle Swarm Optimization</i>
QoI	<i>Quantity of interest</i>
SEIG	<i>Shannon's expected information gain</i>
SGD	<i>Stochastic Gradient Descent</i>

LISTA DE SÍMBOLOS

N_e	Quantidade de experimentos
y_i	Vetor de observações do experimento
$\mathcal{Y}$	Conjunto de observações do experimento
n_y	Quantidade de observações do experimento
$\mathbf{g}$	Vetor do modelo experimental
$\mathbf{d}$	Vetor dos parâmetros do experimento
$\mathcal{S}$	Conjunto dos parâmetros de projeto
n_d	Quantidade de variáveis do problema OED
$\boldsymbol{\theta}$	Vetor da quantidade de Interesse
Θ	Conjunto das variáveis de interesse
n_θ	Quantidade de variáveis de interesse QoI
$\boldsymbol{\varepsilon}$	Vetor de ruído das medições
Σ_ε	Matriz de covariância do ruído das medições
π	Distribuição de probabilidade
E	Módulo de elasticidade
I_n	Momento de Inércia
Ω	Espaço de eventos aleatórios
$\mathcal{Y}$	Conjunto de dados observados do problema OED
D_{KL}	Divergência de Kullback-Leibler
$\mathcal{I}$	Ganho de Informação Esperado de Shannon
N	Quantidade de amostras do laço externo do DLMC
M	Quantidade de amostras do laço interno do DLMC
$\mathbf{d}^*$	Solução ótima para $\mathbf{d}$
$\mathcal{G}$	Vetor de valores do modelo
$\mathbf{x}$	Vetor de variáveis de projeto ou variável independente das funções
n_x	Número de variáveis de projeto
$\mathcal{D}$	Conjunto das variáveis de projeto
$\boldsymbol{\beta}$	Vetor de parâmetros desconhecidos do metamodelo
f	Valor da função determinística
Z	Incerteza extrínseca do metamodelo do Kriging
n_k	Número de pontos amostrais da variável de projeto
$\mathbf{y}$	Vetor de resposta do modelo
$\hat{\mathbf{y}}$	Preditor pelo Kriging
$\hat{\mu}$	Estimador para a média do Kriging
$\hat{\sigma}$	Estimador para o desvio padrão do Kriging
$\mathbf{h}$	Vetor de funções de correlação do Kriging
$\boldsymbol{\Psi}$	Matriz de correlação do Kriging

c_k	Parâmetro do modelo do Kriging
ρ_k	Parâmetro do Kriging
L	Função de verossimilhança
L_{ln}	Função logarítmica negativa da verossimilhança
$\mathcal{L}_{ln}$	Função logarítmica concentrada negativa da verossimilhança
$\tilde{y}$	Vetor aumentado de resposta do Kriging
$\tilde{\Psi}$	Matriz aumentada de correlação do Kriging
s^2	Variância do Kriging
$\hat{g}$	Preditor do modelo OED pelo Kriging
Γ_θ	Vetor de variáveis do problema OED
σ_g	Variância do metamodelo do Kriging para o modelo do problema OED
Σ_g	Matriz de covariância do metamodelo do Kriging para o modelo do problema OED
$\hat{\mu}_g$	Estimador para a média do Kriging para o problema OED
$\hat{\sigma}_g$	Estimador para o desvio padrão Kriging para o problema OED
θ^+	Ponto a ser predito pelo Kriging para o problema OED
G	Módulo de cisalhamento
A_r	Área da seção transversal
K_s	Constante de Timoshenko

SUMÁRIO

1	INTRODUÇÃO	14
1.1	ESTADO DA ARTE	16
1.2	JUSTIFICATIVA E RELEVÂNCIA	18
1.3	OBJETIVOS	19
1.3.1	Objetivo geral	19
1.3.2	Objetivo específico	19
1.4	ORGANIZAÇÃO DO TRABALHO	19
2	PROJETO EXPERIMENTAL ÓTIMO	21
2.1	MODELO EXPERIMENTAL	21
2.2	INFERÊNCIA BAYESIANA	22
2.3	GANHO DE INFORMAÇÃO ESPERADO	25
2.3.1	Divergência de Kullback-Leibler	25
2.3.2	Ganho de Informação Esperado de Shannon	26
2.3.3	Método de Monte Carlo	27
2.4	MAXIMIZAÇÃO DO GANHO DE INFORMAÇÃO ESPERADO	29
3	KRIGING APLICADO AO MODELO DO SEIG	32
3.1	METAMODELO DO KRIGING	32
3.1.1	Conceitos básicos do Kriging	32
3.1.2	Função de correlação	34
3.1.3	Estimativa dos parâmetros do Kriging	36
3.1.4	Preditor do Kriging	40
3.1.5	Erro na predição	44
3.2	CONSTRUÇÃO DO METAMODELO PARA O PROBLEMA OED	44
3.3	PROJETO EXPERIMENTAL ÓTIMO UTILIZANDO O KRIGING	46
4	OTIMIZAÇÃO GLOBAL EFICIENTE	49
4.1	APROXIMAÇÃO DO SEIG PELO KRIGING	49
4.2	CRITÉRIO DE PREENCHIMENTO	51
4.3	DUPLA ABORDAGEM DO KRIGING	52
5	EXEMPLOS NUMÉRICOS	55
5.1	EXEMPLO 01: MODELO NÃO LINEAR	56
5.1.1	Resultados e discussões	56
5.2	EXEMPLO 02: POSICIONAMENTO DE EXTENSÔMETRO PARA CARACTERIZAÇÃO DAS PROPRIEDADES MECÂNICAS DA VIGA	59
5.2.1	Formulação Bayesiana	60
5.2.2	Resultados e discussões	60
5.3	EXEMPLO 03: POSICIONAMENTO DE EXTENSÔMETROS PARA CARACTERIZAÇÃO DE CARGAS EM VIGA.	62

5.3.1	Formulação Bayesiana	64
5.3.2	Resultados e discussões	64
6	CONCLUSÃO	67
6.1	SUGESTÕES PARA TRABALHOS FUTUROS	68
	REFERÊNCIAS	69

1 INTRODUÇÃO

Os experimentos científicos são ferramentas utilizadas em diversas áreas do conhecimento, pois testam teorias, validam modelos e estabelecem relações de causa e efeito, gerando conhecimento científico confiável. Da observação e coleta de dados experimentais, é possível realizar inferências estatísticas e obter informações sobre os fenômenos e quantidades de interesse em estudo. Para tanto, os experimentos precisam fornecer dados informativos e relevantes que possam ser utilizados para fundamentar as tomadas de decisão.

Nesse contexto, o projeto experimental ótimo (OED, do inglês *Optimal experimental design*) surge como uma ferramenta para otimizar o processo de tomada de decisão experimental. Seu objetivo é maximizar a eficiência dos experimentos, buscando a configuração que forneça os dados mais informativos e relevantes para a aplicação de interesse (CHALONER; VERDINELLI, 1995; GALLY *et al.*, 2021). Isso implica em obter a melhor combinação de variáveis de projeto, como geometria, propriedades dos materiais, condições de carga e parâmetros de teste, que sejam as mais informativas para a finalidade do experimento. Como resultado, obtêm-se dados mais precisos com uma significativa economia de recursos.

Essa abordagem é aplicada em cenários nos quais os experimentos são custosos em termos de tempo, recursos financeiros, materiais ou recursos computacionais. Um exemplo prático é a definição da melhor posição para realizar medições em uma ponte para monitoramento de integridade, ou detecção e controle de danos. Os parâmetros de projeto podem incluir a localização de sensores e atuadores ou características da excitação, como amplitude e características da frequência. Já as quantidades de interesse incertas, que se deseja inferir, são variáveis como modos de vibração da estrutura, efeitos das cargas estáticas e dinâmicas, distribuição de tensões, entre outros (PAPADIMITRIOU; HARALAMPIDIS; SOBCZYK, 2005; PAPADIMITRIOU; ARGYRIS, 2017; COSTAS; COSTAS; PANAGIOTIS, 2017).

Outro exemplo de aplicação do OED, abordado nos trabalhos de Ortega *et al.* (2018), Rivas-López *et al.* (2017) e Ganesan, Raj e Shashikala (2013), envolve a realização de testes de fadiga do concreto, que avaliam a capacidade do material de resistir a repetidos ciclos de carga e descarga, aplicados ao longo do tempo. Nesses testes a quantidade de ciclos até a falha pode variar significativamente, de 100 a 100.000, entre os corpos de prova, considerando as mesmas condições de carga. Essa alta dispersão de resultados pode acarretar testes prolongados e sem garantia de atingir o ponto de falha. Neste caso, os parâmetros de projeto a serem otimizados incluem a quantidade adequada de amostras e/ou o número adequado de ciclos de carregamento, para se ter resultados confiáveis com um erro controlado. E as quantidades de interesse incertas, que se deseja inferir, incluem amplitude, frequência, configurações

do corpo de prova (como quantidade, tamanho, forma e características geométricas), além de propriedades do material, como resistência, módulo de elasticidade e taxa de endurecimento.

Para análises experimentais que consideram a existência de ruído nas observações, foram desenvolvidas formulações de OED no contexto Bayesiano, baseadas em métricas de desempenho que quantificam o valor das informações contidas nos dados. A configuração ótima do experimento é obtida pela maximização dessa métrica. Uma opção é o ganho de informação esperado de Shannon (SEIG, do inglês *Shannon's expected information gain*) (CHALONER; VERDINELLI, 1995; SOBCZYK; HOŁOBUT, 2012).

Estimar o SEIG envolve a avaliação de uma integral dupla sobre o espaço dos dados observados e o espaço das quantidades de interesse (QoI, do inglês *Quantity of interest*). Como discutido por Englezou, Waite e Woods (2022), o cálculo do SEIG enfrenta dois principais desafios. Primeiramente, as integrais, que podem ser de alta dimensão e, em geral, são analiticamente intratáveis. Em segundo lugar, em diversos modelos, a computação do SEIG requer o uso de ferramentas numéricas computacionalmente custosas, como o método dos elementos finitos.

Diante do exposto, estudos têm sido desenvolvidos para reduzir o custo computacional associado à avaliação da integral dupla do SEIG dentro de um processo de otimização para maximizar o ganho de informação. Dito isto, é proposta nessa tese a abordagem do metamodelo do Kriging para a solução de problemas OED Bayesiano utilizando o SEIG.

Os metamodelos são superfícies preditoras construídas a partir de respostas conhecidas do domínio de projeto. A utilização de metamodelos ganha destaque em cenários onde há um alto custo computacional associado à avaliação da função objetivo. Essa ferramenta permite realizar previsões e otimizações de maneira mais eficiente, visto que os resultados são previstos sem recorrer à função original.

Entre os metamodelos, o Kriging se destaca por sua capacidade de fornecer boas previsões dos valores da função em pontos não amostrados, ao mesmo tempo em que quantifica a incerteza associada a essas previsões por meio da variância do preditor (FORRESTER; KEANE; BRESSLOFF, 2006). Desenvolvido inicialmente por Krige (1951) no campo da geoestatística, o Kriging foi posteriormente adaptado para o projeto e análise de experimentos computacionais por Sacks *et al.* (1989).

A abordagem do Kriging para otimização global apresenta algumas vantagens principais. A técnica tem alto desempenho na solução de problemas de otimização global com custo computacional limitado. É capaz de construir o modelo a partir de previsões de tendências dos dados ao invés de se mover passo a passo ao longo de alguma trajetória. Além disso, é capaz de fornecer uma regra de parada confiável com base na melhoria esperada de pesquisas adicionais. Essa regra de parada é possível

porque o modelo estatístico fornece intervalos de confiança no valor da função em pontos não amostrados (HUANG *et al.*, 2006).

Um algoritmo de otimização que aborda o Kriging é o Otimização Global Eficiente (EGO, do inglês *Efficient Global Optimization*), que explora as informações fornecidas pelo metamodelo para adicionar iterativamente novos pontos de suporte ou pontos de preenchimento (IPs, do inglês *Infill points*), melhorando a precisão do preditor e buscando o mínimo global (JONES, 2001; HUANG *et al.*, 2006).

Diante do exposto, esta tese apresenta uma nova metodologia baseada em uma dupla abordagem do Kriging como uma alternativa para a solução de problemas OED Bayesiano utilizando o SEIG como métrica de desempenho. Essa metodologia tem por objetivo ser uma abordagem robusta e abrangente, reduzindo o custo computacional sem comprometer a acurácia dos resultados.

Na metodologia proposta, a primeira abordagem do Kriging consiste na construção do metamodelo como substituto do modelo original direto, considerando o vetor das QoI incerta, o qual se deseja inferir. Consequentemente, no processo de otimização, a cada nova posição, é construído um modelo substituto que é mais simples do que o originalmente custoso, e o cálculo do SEIG ocorre através deste.

E no processo de maximização do SEIG é empregado o algoritmo EGO, o que requer a construção de um modelo substituto pelo Kriging para o SEIG, considerando os parâmetros do experimento a serem otimizados, sendo essa a segunda abordagem. Essa dupla abordagem do Kriging é denominada EGO-KR.

1.1 ESTADO DA ARTE

O OED é uma metodologia utilizada para otimizar os parâmetros do experimento, visando maximizar sua eficiência e reduzir as dificuldades e custos envolvidos no processo experimental (CHALONER; VERDINELLI, 1995; GALLY *et al.*, 2021). No OED, buscam-se os parâmetros de projeto ótimos que maximizam alguma medida de eficiência de um experimento. Pesquisas vêm sendo desenvolvidas para reduzir os custos relacionados à abordagem do SEIG na solução de problemas OED Bayesiano.

Uma abordagem tradicional para aproximar a integral dupla do SEIG é o método de Monte Carlo de laço duplo (DLMC, do inglês *Double loop Monte Carlo*) (RYAN, 2003; HUAN; MARZOUK, 2013; BARTUSKA; ESPATH; TEMPONE, 2022). Este estimador requer a avaliação de duas amostragens de Monte Carlo (MC) justapostas. A desvantagem dessa abordagem é o alto custo computacional associado à amostragem de MC justaposta, necessitando de um grande número de pontos amostrais. Este custo torna-se ainda mais significativo por ser combinado com algoritmos de otimização e para modelos computacionalmente exigentes.

Uma alternativa envolve empregar o método de Laplace para aproximar a integral interna. Long *et al.* (2013) aplica uma aproximação de Laplace à distribuição

a posteriori em modelos de projeto experimental de equações diferenciais parciais. Essa abordagem é denominada Monte Carlo com aproximação de Laplace (MCLA, do inglês *Monte Carlo with Laplace approximation*). Embora evite a avaliação do laço interno, essa abordagem pode introduzir viés na aproximação, especialmente para posteriores não Gaussianas. Outra abordagem, proposta por Beck *et al.* (2018), é realizar uma amostragem por importância no laço interno da estrutura do DLMC, baseada na aproximação de Laplace, para obter amostras mais informativas, reduzindo o custo da amostragem do DLMC. Este método foi denominado Monte Carlo de laço duplo com amostragem por importância (DLMCIS, do inglês *Double loop Monte Carlo with importance sampling*). Em Beck *et al.* (2020) foi proposto a abordagem do DLMC multinível e métodos de colocação estocástica com amostragem por importância para calcular o SEIG de forma eficiente.

Para lidar com o custo computacional associado ao processo de otimização, Carlon *et al.* (2020a) contribuíram derivando gradientes para DLMC, MCLA e DLMC com amostragem por importância, propondo o uso de algoritmos de descida de gradiente estocástica (SGD, do inglês *Stochastic gradient descent*), resultando na redução do ônus computacional associado ao problema. Outros estudos, como Goda *et al.* (2022), introduziram novos algoritmos para resolver problemas de OED combinando estimadores multivariados não tendenciosos com o método SGD. Outra técnica de otimização abordada foi o método de Troca de Coordenadas Aproximada, explorado por Overstall e Woods (2017), que utiliza uma série de emuladores unidimensionais para o substituto do SEIG. Além disso, Englezou, Waite e Woods (2022) reduziram o custo do MCLA no processo de otimização, eliminando alguns passos de otimização e introduzindo o método chamado amostragem por importância aproximada de Laplace.

Outra abordagem para a redução do custo computacional foi através do uso de modelos substitutos. Huan e Marzouk (2013) construíram um modelo substituto utilizando a expansão polinomial de Caos (PCE, do inglês *Polynomial Chaos expansions*) tanto no espaço da variável de projeto quanto no QoI. Em seguida, aplicaram um algoritmo SGD, baseado em uma variação do algoritmo de Kiefer-Wolfowitz (SPALL, 1988). Em Huan e Marzouk (2014), foi realizada uma análise comparativa entre a aproximação estocástica de Robbins-Monro (ROBBINS; MONRO, 1951) e a aproximação da média amostral combinada com um método determinístico quase-Newton para maximizar o SEIG.

A maioria dos artigos citados abordaram otimizadores locais, pois a otimização global no domínio do OED é escassa na literatura (FOSTER *et al.*, 2019; KLEINGESSE; GUTMANN, 2019). Um exemplo de estudo que aborda otimizador global é o de Foster *et al.* (2019), que se concentra no desenvolvimento de estimadores variacionais para o ganho de informação e emprega um esquema de otimização Bayesiana (BO, do inglês *Bayesian optimization*) (SNOEK; LAROCHELLE; ADAMS, 2012) para

resolver o problema de OED. Outro exemplo é o trabalho de Kleinegesse e Gutmann (2019), que propôs a inferência sem verossimilhança por estimativa de razão para aproximar a razão entre a posteriori e a priori. Mais recentemente, Carlon *et al.* (2023) empregaram o algoritmo EGO baseado no metamodelo Kriging (ANKENMAN; NELSON; STAUM, 2010), utilizando a aproximação de Laplace e PCE para abordar a otimização do SEIG, reduzindo os custos computacionais e permitindo a abordagem de problemas não convexos. No entanto, a precisão da aproximação de Laplace tem algumas limitações, exigindo que a função a posteriori seja unimodal e simétrica. Isso, por sua vez, pode produzir estimativas distorcidas se a distribuição a posteriori for assimétrica ou multimodal.

Diante dos estudos apresentados, fica evidente que um dos principais desafios na otimização no OED Bayesiano é o alto custo computacional.

1.2 JUSTIFICATIVA E RELEVÂNCIA

A resolução de problemas OED Bayesianos utilizando o SEIG como métrica de desempenho, requer a avaliação de uma integral dupla em um problema de otimização, o que resulta em um alto custo computacional, especialmente para modelos que demandam processamento intensivo. O aumento do espaço de busca, por exemplo, implica em um aumento exponencial do custo computacional.

O método clássico de resolução através do DLMC requer uma grande quantidade de pontos amostrais devido a sua estrutura de amostragem justaposta. A redução do número de pontos amostrais pode comprometer a precisão dos resultados, especialmente no laço interno, onde podem ocorrer problemas de *underflow* numérico, caracterizado pelo arredondamento de valores extremamente baixos para zero pelo sistema computacional.

Uma alternativa é utilizar o DLMCIS, no qual a amostragem por importância é aplicada ao laço interno do DLMC. Embora essa abordagem melhore a precisão dos resultados, ela ainda exige um número elevado de pontos amostrais e não reduz significativamente o custo computacional associado à avaliação do modelo.

Outra estratégia é empregar o MCLA para aproximar a integral interna do SEIG por meio de uma expansão de Taylor de segunda ordem seguida de integração Gaussiana, resultando em uma única integral. No entanto, essa técnica apresenta algumas limitações, exigindo que a função seja convexa e tenha uma distribuição a posteriori simétrica. Caso contrário, a aproximação de Laplace não será precisa e pode gerar um viés nas estimativas.

Diante desse cenário e para superar os desafios relacionados a maximização do SEIG, esta tese propõe uma dupla abordagem do Kriging. O metamodelo será utilizado como substituto do modelo direto na estrutura do DLMC, bem como na otimização do SEIG.

Essa metodologia possibilita a solução de problemas OED Bayesiano que envolvam modelos computacionalmente custosos de forma mais robusta, sendo capaz de resolver problemas não convexos ou com distribuições a posteriori assimétricas ou multimodais, sem comprometer a acurácia dos resultados e com menor custo computacional associado ao processo de otimização.

1.3 OBJETIVOS

1.3.1 Objetivo geral

O objetivo desta pesquisa é contribuir para o avanço na resolução de problemas OED Bayesiano utilizando o SEIG como métrica de desempenho por meio do desenvolvimento de uma metodologia que seja eficiente, robusta e reduza o custo computacional, principalmente para modelos computacionalmente custoso.

1.3.2 Objetivo específico

Os principais objetivos específicos são:

- Desenvolver uma metodologia de dupla abordagem do Kriging, estruturada em duas etapas, para solução de problemas OED Bayesiano pelo SEIG;
- Implementar o Kriging da primeira etapa, que consiste em utilizar o metamodelo do Kriging como preditor do modelo direto no espaço das QoI, dentro da estrutura do DLMC, reduzindo o custo computacional associado ao modelo. Além disso, essa abordagem não introduz viés ao cálculo do SEIG;
- Implementar o Kriging da segunda etapa, que utiliza o algoritmo EGO, baseado no metamodelo Kriging, para otimizar o vetor das variáveis de projeto, possibilitando a busca eficiente dos parâmetros experimentais que maximizam o ganho de informação. Esta etapa visa reduzir o número de pontos nos quais o SEIG precisa ser avaliado, diminuindo assim o custo computacional no processo de otimização;
- Testar e validar a abordagem EGO-KR proposta;
- Comparar os resultados obtidos com outras metodologias existentes na literatura para avaliar a eficácia da abordagem proposta em reduzir o custo computacional, mantendo a acurácia dos resultados.

1.4 ORGANIZAÇÃO DO TRABALHO

Esta tese é organizada em seis capítulos. No Capítulo 2 são introduzidos os conceitos e formulações relacionadas ao Projeto Ótimo de Experimentos Bayesiano. É abordado a construção do problema Bayesiano, apresentando o teorema de Bayes

e como aplicá-lo ao modelo experimental. Além disso, é discutido a formulação do SEIG como medida de desempenho empregada para a avaliação do problema OED Bayesiano. Para estimar o SEIG é utilizado o método de DLMC, exposto ao final do capítulo.

O objetivo deste trabalho é explorar a abordagem do metamodelo do Kriging como substituto do modelo original e o algoritmo EGO, também baseado no Kriging, no processo de maximização do SEIG. Desta forma, no Capítulo 3 é apresentada a construção do modelo substituto pelo Kriging para a representação do modelo, abordando detalhes de sua formulação e ajuste dos parâmetros. Em seguida, aborda-se o DLMC para a atualização do cálculo do ganho de informação.

No Capítulo 4, é apresentada a formulação do problema de otimização utilizando o algoritmo EGO. O processo inclui a formulação do problema de otimização por meio do metamodelo do Kriging, a estimativa dos parâmetros, a construção da superfície preditora e a avaliação do erro extrínseco inerente. Em seguida, detalha-se a estrutura do critério de Melhoria Esperada (EI, do inglês *Expected Improvement criterion*), sendo este o critério de preenchimento abordado na busca pelo ponto de projeto que maximiza o SEIG. Por fim, é apresentado um resumo da dupla abordagem do Kriging na resolução de problemas OED Bayesiano.

No Capítulo 5, são avaliados três problemas relacionados ao OED Bayesiano utilizando o SEIG como métrica de desempenho. O procedimento proposto é aplicado com o objetivo de avaliar sua eficiência, precisão e acurácia. Os resultados obtidos são comparados com os de outros métodos disponíveis na literatura.

No Capítulo 6 são apresentadas as conclusões observadas com a aplicação das formulações e procedimentos discutidos neste trabalho. Além disso, são oferecidas sugestões para pesquisas futuras, com o objetivo de aprimorar as técnicas de resolução de problemas OED Bayesiano.

2 PROJETO EXPERIMENTAL ÓTIMO

Neste capítulo é abordada a formulação de problemas de OED Bayesiano, utilizando o ganho de informação esperado de Shannon como medida de desempenho. O SEIG é frequentemente empregado para avaliar a quantidade de informação oferecida pelo experimento, como abordado nos trabalhos de Nesterov (2013), Huan e Marzouk (2013) e Carlon (2019). O conceito central desse método consiste em medir o ganho de informação que uma nova observação ou aquisição de dados proporciona em relação à situação atual (HUAN; MARZOUK, 2013).

A estimativa do SEIG baseia-se na divergência de Kullback-Leibler (D_{KL}) entre a função de densidade de probabilidade (PDF) a posteriori e a PDF a priori das variáveis de interesse (CHALONER; VERDINELLI, 1995; SOBCZYK; HOŁOBUT, 2012).

Uma abordagem direta para a aproximação da integral dupla do SEIG é o método DLMC (RYAN, 2003; HUAN; MARZOUK, 2013), um estimador de amostragem de Monte Carlo que requer a avaliação de duas amostragens MC justapostas.

Assim, neste capítulo, primeiramente é apresentada a formulação do método, seguida de uma introdução à estrutura de inferência Bayesiana para o problema. Ao final, é detalhada a estrutura do SEIG, partindo do D_{KL} , e o cálculo através do DLMC.

2.1 MODELO EXPERIMENTAL

A avaliação do modelo por meio da repetição de N_e experimentos pode ser representada como um modelo computacional determinístico com um componente de ruído observacional aditivo,

$$\mathbf{y}_i(\mathbf{d}, \boldsymbol{\theta}, \varepsilon_i) = \mathbf{g}(\mathbf{d}, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, \dots, N_e \quad (1)$$

em que, $\mathbf{y}_i \in \mathcal{Y} \subset \mathbb{R}^{n_y}$ é o vetor de observações do experimento, $\mathbf{g}(\mathbf{d}, \boldsymbol{\theta}) : \mathbb{R}^{n_d} \times \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_y}$ é o modelo experimental, $\mathbf{d} \in \mathcal{S} \subset \mathbb{R}^{n_d}$ é o vetor com os parâmetros do experimento a serem otimizados, em que $\mathcal{S}$ é o espaço de projeto, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{n_\theta}$ é o vetor QoI a serem recuperadas, em que Θ é o conjunto de eventos aleatórios e ε é o vetor de ruído das medições.

O ruído é amostrado de forma independente e identicamente distribuída (i.i.d., do inglês *Independent and identically distributed*), ou seja, é considerado independente de $\mathbf{g}$, $\mathbf{d}$ e $\boldsymbol{\theta}$. Além disso, segue uma distribuição Gaussiana com média zero e covariância Σ_ε , $\varepsilon_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$, em que a matriz de covariância Σ_ε é simétrica e definida positiva. O vetor de parâmetros $\boldsymbol{\theta}$ é considerado desconhecido, portanto é definido como um vetor de variáveis aleatórias com distribuição a priori $\pi(\boldsymbol{\theta})$.

O projeto ótimo de experimentos consiste em selecionar as condições experimentais $\mathbf{d}$ que levem ao maior ganho de informação sobre $\boldsymbol{\theta}$. O objetivo é minimizar a incerteza nos parâmetros estimados enquanto maximiza as informações obtidas do

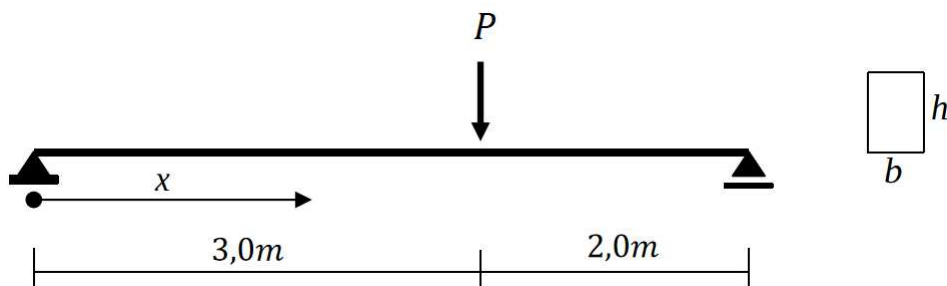
experimento. Um exemplo prático seria realizar um teste de flexão em uma viga de forma a maximizar o ganho de informação sobre o módulo de elasticidade (E) estimado do material. A viga é bi-apoiada submetida somente a uma carga pontual, com seção transversal retangular de base b e altura h , como ilustrado na Figura 1.

Neste exemplo, o objetivo é posicionar um extensômetro de forma a fornecer o máximo ganho de informação sobre E . Nesse contexto, o parâmetro de projeto do experimento que se busca otimizar é a posição, $d = x$, a quantidade de interesse é o E , y é a deflexão observada, e o erro de observação ao registrar os valores de y é representado pelo ε . O modelo adotado para calcular a deflexão da viga é baseado na teoria de Euler-Bernoulli, que relaciona E e x com a deflexão y a partir de $g(x, E) = \frac{2Px}{6EI_nL} (L^2 - x^2 - 4)$, em que P é a carga aplicada, L é o comprimento da viga e I_n é o momento de inércia da seção transversal da viga. Este exemplo simples será empregado no decorrer dos capítulos para ilustrar os conceitos introdutórios.

Neste exemplo, em que a posição do extensômetro mais informativa sobre E está baseada em medições da deflexão, observa-se que o valor de E é diretamente influenciado pela contribuição do momento fletor gerado pela carga concentrada. Desta forma, a demonstração de que o ponto ótimo é próximo a $x^* = 3$ é trivial, pois é o ponto que ocorre o momento fletor máximo.

Neste exemplo, o módulo de elasticidade longitudinal da viga é diretamente influenciado pela contribuição do momento fletor gerado pela carga concentrada. Desta forma, a demonstração de que o ponto ótimo é próximo a $x^* = 3$ é trivial, pois é o ponto que ocorre o momento fletor máximo. Este ponto proporcionará o melhor ajuste entre o modelo e os dados medidos.

Figura 1 – Viga bi-apoiada para posicionamento de extensômetro.



2.2 INFERÊNCIA BAYESIANA

A qualidade dos experimentos é avaliada a partir de uma análise de inferência Bayesiana. Esta estrutura consiste em atualizar informações previamente conhecidas sobre um parâmetro à medida que novos dados são obtidos, expressando os resultados em termos de probabilidades (HOFF, 2009). Ao combinar novas evidências com

conhecimentos prévios através do uso da regra de Bayes, essa abordagem proporciona a melhoria da precisão dos resultados.

Dado que a variável de interesse, θ , é um vetor aleatório com distribuição a priori $\pi(\theta)$, definido como $\theta : \Omega^{\theta} \mapsto \Theta \subset \mathbb{R}^{\theta}$, onde Ω representa o espaço de eventos aleatórios. Assume-se que o modelo $\mathbf{g}(\mathbf{d}, \theta)$ é duas vezes diferenciável, em relação a θ e em relação a $\mathbf{d}$. Para N_e experimentos realizados com a mesma configuração $\mathbf{d}$, o conjunto de dados observados é definido como $\mathbf{Y} = \{y_i\}_{i=1}^{N_e}$. O teorema de Bayes estabelece que:

$$\pi(\theta|\mathbf{Y}, \mathbf{d}) = \frac{p(\mathbf{Y}|\theta, \mathbf{d}) \cdot \pi(\theta)}{p(\mathbf{Y}|\mathbf{d})}, \quad (2)$$

em que $\pi(\theta|\mathbf{Y}, \mathbf{d})$ é a distribuição a posteriori, ou seja, a PDF atualizada do vetor de variáveis aleatórias θ dada as observações $\mathbf{Y}$; $\pi(\theta)$ é a distribuição a priori, ou seja, a PDF de θ antes do experimento; $p(\mathbf{Y}|\theta, \mathbf{d})$ é a verossimilhança, ou seja, a probabilidade de observar $\mathbf{Y}$ dada a distribuição a priori $\pi(\theta)$ conhecida anteriormente; e $p(\mathbf{Y}|\mathbf{d})$ é a evidência, a probabilidade de $\mathbf{Y}$ ser observada.

Portanto, uma vez que os dados são coletados, a PDF a priori, que representa a crença ou informação disponível a respeito do parâmetro antes de realizar o experimento, é atualizada por meio da função de verossimilhança dos eventos, produzindo a PDF a posteriori. A função de verossimilhança fornece uma medida de quão bem os parâmetros do modelo explicam os dados observados. Logo, expressa a probabilidade de obter os dados observados considerando os valores específicos do modelo.

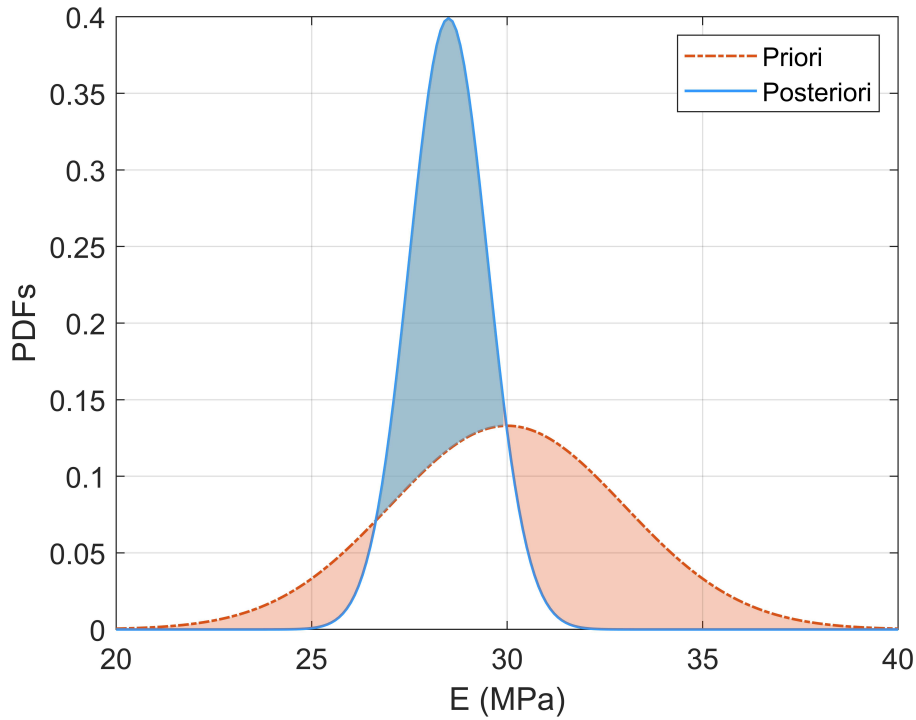
No contexto de experimentos, a inferência Bayesiana é utilizada para estimar uma PDF a posteriori de uma QoI dada uma PDF a priori e algumas observações $\mathbf{Y}$ fornecidas pelo experimento.

Considerando o exemplo da viga mencionado anteriormente, o objetivo é obter a posição ótima que melhor ajuste o valor do módulo de elasticidade do modelo, levando em consideração os valores obtidos para a deflexão e informações prévias de E . Como por exemplo, E deve assumir somente valores positivos e de variância finita.

A Figura 2 apresenta a curva PDF de distribuição a priori, com as informações iniciais sobre E , e a curva a posteriori, obtida após inserir informações do experimento. Pode-se observar que houve uma redução da dispersão dos dados após o experimento, o que significa que as informações obtidas foram relevantes sobre a QoI. A região onde a densidade de probabilidade da distribuição a posteriori supera a da distribuição a priori está destacada em azul, enquanto a área onde a densidade a posteriori é inferior à da priori está destacada em vermelho.

A distribuição de probabilidade mais utilizada na literatura para modelar o conhecimento prévio sobre a QoI é a Gaussiana (KAPIO; SOMERSALO, 2006). No entanto, outras distribuições também podem ser usadas. O único requisito é que seja possível amostrar θ a partir da distribuição a priori.

Figura 2 – PDFs a priori e a posteriori de E para o exemplo da viga de Euler-Bernoulli.



Considerando o modelo do experimento e o ruído aditivo com distribuição Gaussiana, Equação (1), a função de verossimilhança, $p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d})$, ao observar $\mathbf{Y}$ dado $\boldsymbol{\theta}$ e $\mathbf{d}$ é uma Gaussiana multivariada com média $g(\mathbf{d}, \boldsymbol{\theta})$ e matriz de covariância Σ_{ϵ} ,

$$p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) = \det(2\pi\Sigma_{\epsilon})^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} |y_i - g(\mathbf{d}, \boldsymbol{\theta})|_{\Sigma_{\epsilon}^{-1}}^2\right). \quad (3)$$

Como a evidência $p(\mathbf{Y}|\mathbf{d})$ é desconhecida, é introduzida uma nova variável $\boldsymbol{\theta}^*$ e substitui-se a evidência pela marginalização da verossimilhança com respeito a distribuição a priori $\pi(\boldsymbol{\theta}^*)$, isto é,

$$p(\mathbf{Y}|\mathbf{d}) = \int_{\Theta} p(\mathbf{Y}|\boldsymbol{\theta}^*, \mathbf{d}) \pi(\boldsymbol{\theta}^*) d\boldsymbol{\theta}^*. \quad (4)$$

Nessa estrutura $\boldsymbol{\theta}^*$ e $\boldsymbol{\theta}$ são independentes, mas identicamente distribuídos e $\mathbf{Y}$ depende de $\boldsymbol{\theta}$, $\mathbf{d}$ e ϵ , ou seja, o parâmetro $\boldsymbol{\theta}$ usado para gerar $\mathbf{Y}$ é diferente de $\boldsymbol{\theta}^*$ na integral dentro do logaritmo (RYAN, 2003).

A partir da Equação (3), a PDF da verossimilhança da integral da evidência é dada por:

$$p(\mathbf{Y}(\mathbf{d}, \boldsymbol{\theta}, \epsilon) | \boldsymbol{\theta}^*, \mathbf{d}) = \det(2\pi\Sigma_{\epsilon})^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} \|\mathbf{r}_i(\mathbf{d}, \boldsymbol{\theta}, \boldsymbol{\theta}^*, \epsilon)\|_{\Sigma_{\epsilon}^{-1}}^2\right), \quad (5)$$

em que $r_i(\mathbf{d}, \boldsymbol{\theta}, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon}) = \mathbf{g}(\mathbf{d}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon} - \mathbf{g}(\mathbf{d}, \boldsymbol{\theta}^*)$ é o resíduo dos i -ésimos dados experimentais.

2.3 GANHO DE INFORMAÇÃO ESPERADO

2.3.1 Divergência de Kullback-Leibler

A ideia fundamental do OED consiste em encontrar a configuração experimental que produza dados que, em média, maximizem o conhecimento sobre a QoI. Um método para estimar o quanto de informação um experimento fornece é a divergência de Kullback-Leibler (D_{KL}) (KULLBACK; LEIBLER, 1951; KULLBACK, 1997). Este método para experimentos foi utilizado por Long *et al.* (2013), Beck *et al.* (2018), Ben Issaid (2015) e Carlon *et al.* (2023).

O D_{KL} , também conhecido como entropia relativa, é uma medida não simétrica que quantifica a diferença entre duas distribuições de probabilidade sobre a mesma variável ($\boldsymbol{\theta}$). O D_{KL} é utilizado para medir a diferença informacional entre duas distribuições, indicando o quanto uma distribuição difere da outra em termos de informação que elas carregam (KULLBACK, 1997). Portanto, ela pode ser utilizada para avaliar como novos dados alteram a distribuição de probabilidade subjacente, permitindo uma escolha mais informativa e eficaz em relação às ações futuras.

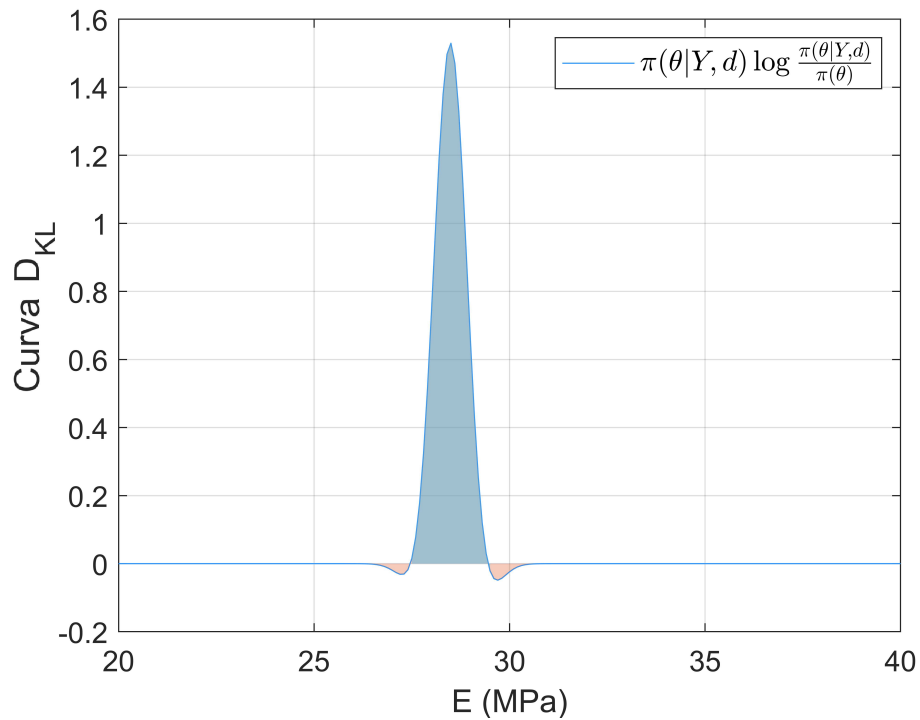
O D_{KL} no contexto do OED é medido entre a PDF a posteriori $\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d})$, que representa a distribuição teórica calculada com maior precisão, e a PDF a priori $\pi(\boldsymbol{\theta})$, que representa uma teoria, descrição ou aproximação sobre o parâmetro. Quanto maior o valor de D_{KL} , mais informativa é $\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d})$ em relação a $\pi(\boldsymbol{\theta})$. Por não ser simétrica, a divergência D_{KL} de $\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d})$ para $\pi(\boldsymbol{\theta})$ geralmente não é a mesma que a divergência se medida da PDF a priori para a posteriori. Para avaliar a qualidade de cada experimento tem-se:

$$D_{KL}(\mathbf{d}, \mathbf{Y}) = \int_{\Theta} \pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d}) \log \frac{\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta}. \quad (6)$$

Em resumo, o D_{KL} pode ser interpretado como uma medida de quão diferente uma distribuição é em relação à outra. Quanto maior o valor da divergência, maior a diferença entre as distribuições. Por outro lado, se a divergência for igual a zero, isso indica que as distribuições são idênticas.

A Figura 3 ilustra a curva integral do $D_{KL}(\mathbf{x}, \mathbf{Y})$ para o exemplo da viga bi-apoiada submetida a uma carga pontual, citado anteriormente, para o caso $x = 3,0m$. As regiões onde o integrando é positivo estão sombreadas em verde, e as regiões onde o integrando é negativo estão sombreadas em vermelho. A entropia relativa é construída entre a PDF a posteriori e a priori obtida para o parâmetro E, ver Figura 2.

Neste exemplo, o ganho máximo de informação sobre E ocorre em uma posição próxima a $x = 3,0m$. Na Figura 4, é apresentada uma comparação entre as curvas PDF de E a priori, a PDF a posteriori otimizada, obtida em $x = 3,0m$, e a PDF a

Figura 3 – Curva da entropia relativa $D_{KL}(\mathbf{d}, \mathbf{Y})$ do problema da viga de Euler-Bernoulli.

posteriori traçada para $x = 1,5m$, que representa o caso não otimizado. Observa-se que a distribuição a posteriori para a configuração otimizada é mais concentrada, em comparação à distribuição a posteriori antes da otimização, indicando que a otimização leva a estimativas mais precisas para E .

O gráfico da Figura 5 apresenta a curva da integral do D_{KL} para as configurações otimizada e não otimizada. Pode-se observar que o D_{KL} para o caso otimizado é maior do que o da configuração não otimizada, como indicado pelas áreas sob as curvas das integrais.

Em suma, para maximizar o ganho de informação de um experimento, é necessário obter $\mathbf{d}^*$ que forneça as observações $\mathbf{Y}^*$ mais informativas, maximizando o valor de D_{KL} .

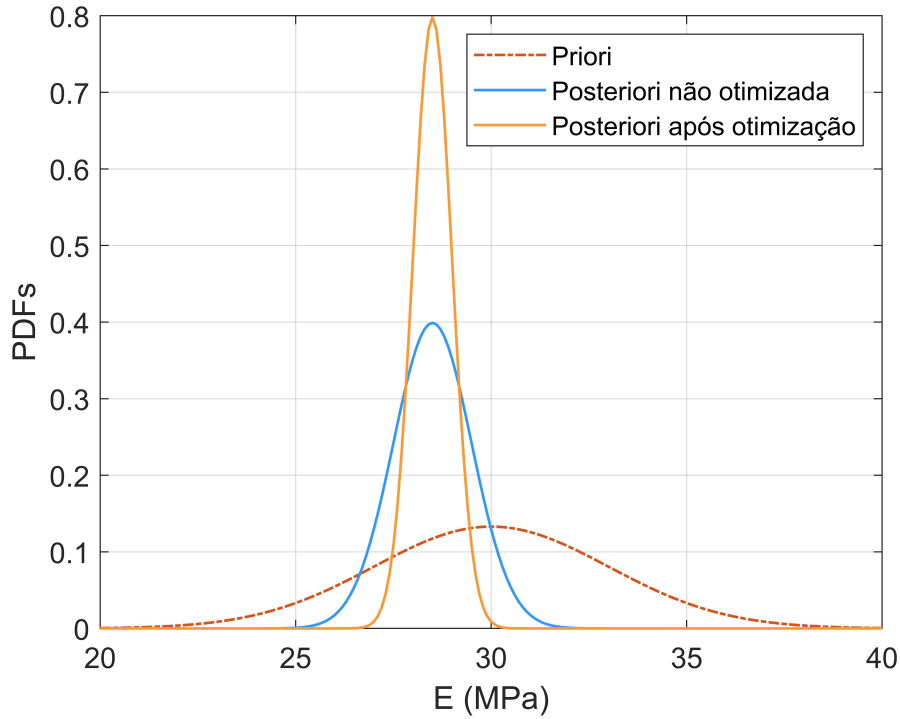
2.3.2 Ganho de Informação Esperado de Shannon

Para estimar o ganho de informação levando em conta o ruído nas observações devido as incertezas nas medições, é necessário marginalizar o D_{KL} em relação a $\mathbf{Y}$, resultando no SEIG (SHANNON, 1948), expresso por:

$$\mathcal{I} = \int_{\mathbf{Y}} \int_{\Theta} \log \left(\frac{\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d})}{\pi(\boldsymbol{\theta})} \right) \pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{d}) d\boldsymbol{\theta} p(\mathbf{Y}|\mathbf{d}) d\mathbf{Y}. \quad (7)$$

Estimar $\mathcal{I}$ requer a solução de uma integral dupla sobre o espaço amostral da

Figura 4 – PDFs da priori e da posteriori de E para os casos otimizado e não otimizado do exemplo da viga de Euler-Bernoulli.



QoI (Θ) e das observações ($\mathcal{Y}$). Para utilizar a função de verossimilhança apresentada na Equação (3), é reescrita a Equação (7) utilizando a equação de Bayes, apresentada na Equação (2) e resultando em:

$$\mathcal{I} = \int_{\Theta} \int_{\mathcal{Y}} \log \left(\frac{p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d})}{p(\mathbf{Y}|\mathbf{d})} \right) p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) d\mathbf{Y} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (8)$$

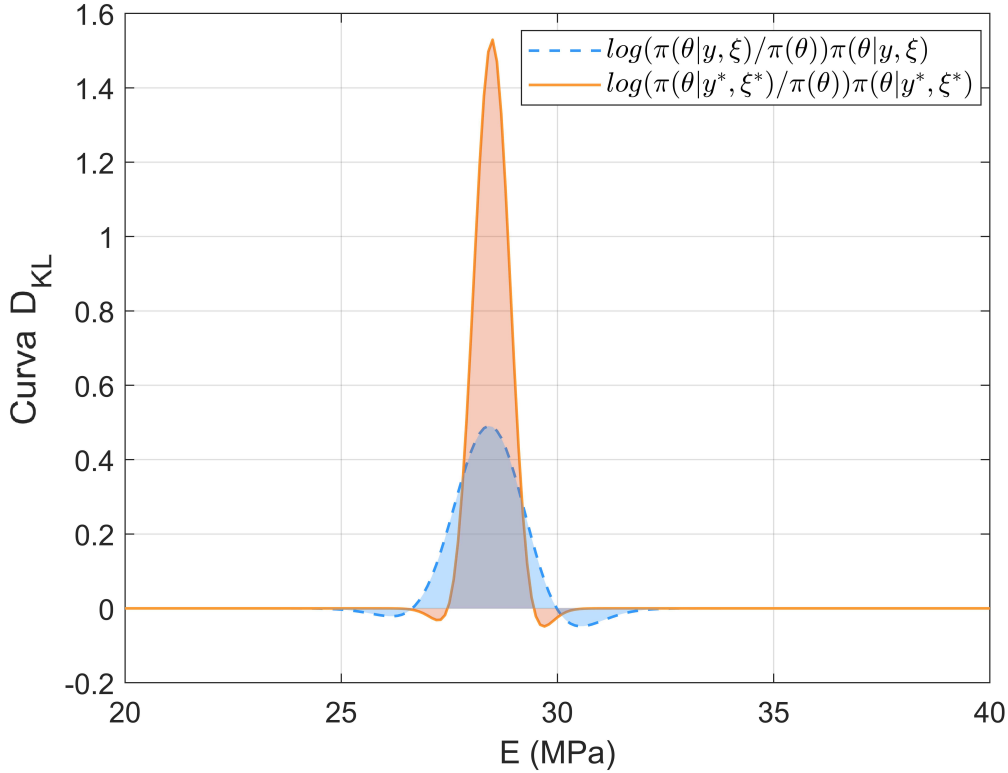
Reescrevendo o ganho de informação esperado substituindo a evidência, Equação (3), tem-se:

$$\mathcal{I} = \int_{\Theta} \int_{\mathcal{Y}} \log \left(\frac{p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d})}{\int_{\Theta} p(\mathbf{Y}|\boldsymbol{\theta}^*, \mathbf{d}) \pi(\boldsymbol{\theta}^*) d\boldsymbol{\theta}^*} \right) p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{d}) d\mathbf{Y} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (9)$$

2.3.3 Método de Monte Carlo

A avaliação de $\mathcal{I}$, na Equação (9), envolve a resolução da integral dupla sobre Θ e $\mathcal{Y}$. Em muitos casos, essas integrais não possuem solução analítica, tornando necessário o uso de métodos numéricos para sua aproximação (ROBERT; CASELLA, 2010; HAMADA *et al.*, 2001, 2001). Nesse contexto, para estimar os valores das integrais será utilizado método de integração de Monte Carlo (MCI, do inglês *Monte Carlo integration*).

Figura 5 – Curva da entropia relativa $D_{KL}(\mathbf{d}, \mathbf{Y})$ para os casos otimizado e não otimizado do exemplo da viga de Euler-Bernoulli.



No MCI, os valores da função são avaliados a partir de uma amostragem dentro da região de integração conhecida. Por exemplo, para $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, e $f(\theta) : \mathbb{R}^{n_\theta}$, a integral de $f(\theta)$ no domínio θ pode ser aproximada como:

$$\int_{\Theta} f(\theta) d\theta \approx \int_{\Theta} d\theta \frac{1}{N} \sum_{n=1}^N f(\theta_n), \quad (10)$$

em que θ_n é amostrado uniformemente a partir de Θ . Se $f(\theta)$ está sendo integrada sobre alguma medida, por exemplo $\pi(\theta)$, sua integral pode ser aproximada como:

$$\int_{\Theta} f(\theta) \pi(\theta) d\theta \approx \int_{\Theta} \pi(\theta) d\theta \frac{1}{N} \sum_{n=1}^N f(\theta_n), \quad (11)$$

em que θ é amostrado de $\pi(\theta)$. Sendo $\pi(\theta)$ uma PDF, a integral $\int_{\Theta} \pi(\theta) d\theta$ resulta no valor um.

O estimador de Monte Carlo converge para o valor real da integral à medida que $N \rightarrow \infty$. Além disso à medida que $N \rightarrow \infty$, o erro na aproximação de uma integral por MCI converge para zero com taxa de $1/\sqrt{N}$ (RUBINSTEIN; KROESE, 2016).

A aproximação da Equação (9) é realizada usando o MCI amostrando θ a partir da distribuição a priori e $\mathbf{Y}$ a partir da verossimilhança. Desta forma, o estimador de Monte Carlo para o problema OED é definido como:

$$\mathcal{I}_{MC} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \log \left(\frac{p(\mathbf{Y}_n | \theta_n, \mathbf{d})}{\int_{\Theta} p(\mathbf{Y}_n | \theta_n^*, \mathbf{d}) \pi(\theta_n^*) d\theta^*} \right), \quad (12)$$

em que θ_n é amostrado uniformemente a partir de Θ . Nessa estrutura, os valores de θ_n são amostrados de forma independente e identicamente distribuída (i.i.d.), e o ruído aditivo ε_n é amostrado i.i.d. a partir dos θ_n . Empregando o MC para aproximar a evidência, resulta no DLMC:

$$\mathcal{I}_{DLMC}(\mathbf{d}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \log \left(\frac{p(\mathbf{Y}_n | \theta_n, \mathbf{d})}{\frac{1}{M} \sum_{m=1}^M p(\mathbf{Y}_n | \theta_{n,m}^*, \mathbf{d})} \right). \quad (13)$$

em que $\mathcal{I}_{MC}$ é a estimativa do problema OED pelo DLMC, enquanto N e M são os respectivos tamanho das amostras nos laços externo e interno. Os valores de $\theta_{n,m}^*$ usados no laço interno são amostrados i.i.d. tanto dos θ_n quanto dos ε_n no laço externo. Uma análise da Equação (13) revela que o estimador DLMC necessita de $N(M+1)$ avaliações do modelo direto para aproximar o SEIG. Em alguns casos, um valor grande para M é necessário para prevenir problemas de *underflow* numérico, isto é, a representação inadequada ou imprecisa de um número muito pequeno em um sistema numérico (BECK *et al.*, 2018).

Considerando o problema da viga discutido anteriormente, Figura 1, a Figura 6 apresenta a curva obtida pelo $\mathcal{I}_{DLMC}$ na posição com o máximo ganho de informação, $x = 3,0m$, considerando $N = 10000$ e $M = 4000$. Nessa metodologia, para o cálculo do SEIG, a função original do modelo foi acessada 10000×4000 , resultando num número de chamadas do modelo direto de $= 4 \times 10^7$ vezes.

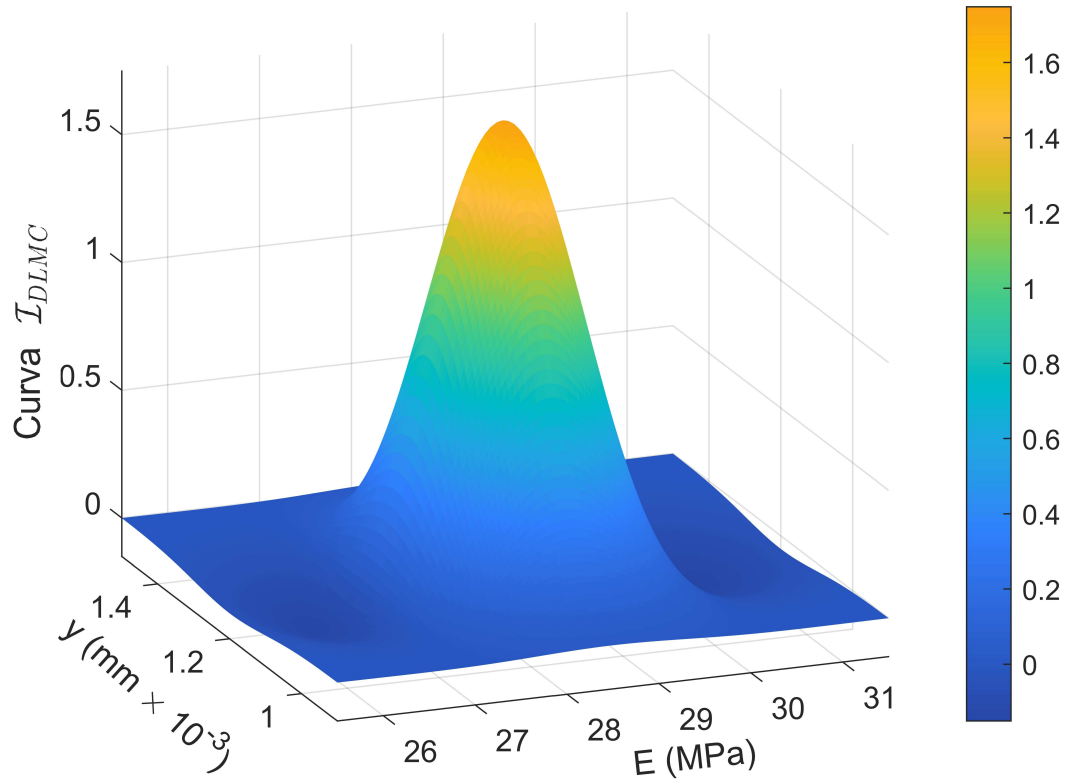
O Algoritmo 1 apresenta a rotina para solução do problema OED pelo DLMC, retornando a estimativa de $\mathcal{I}_{MC}$ para uma determinada posição $\mathbf{d}$.

2.4 MAXIMIZAÇÃO DO GANHO DE INFORMAÇÃO ESPERADO

O objetivo do projeto experimental ótimo Bayesiano é determinar a configuração experimental $\mathbf{d}^*$ mais informativa sobre o vetor de parâmetros θ . Portanto, para esta tese, o objetivo é maximizar a função de utilidade do SEIG, denotado como $\mathcal{I}(\mathbf{d})$, sobre a região viável $\mathbf{d} \in \mathcal{S}$. A formulação do problema de otimização para o OED é dada pela maximização do valor esperado do SEIG:

$$\mathbf{d}^* = \arg \max_{\mathbf{d} \in \mathcal{S}} E[\mathcal{I}(\mathbf{d})]. \quad (14)$$

Figura 6 – Curva da entropia relativa $\mathcal{I}_{DLMC}$ do exemplo da viga de Euler-Bernoulli.



Integrar o DLMC dentro de um procedimento de otimização, como na Equação (14), pode ser computacionalmente inviável, especialmente quando o modelo direto envolve simulações computacionalmente custosas, como alguns casos de códigos de elementos finitos. Portanto, nas seções seguintes, é proposta uma dupla abordagem do Kriging para aliviar os desafios computacionais associados à resolução de problemas de OED Bayesiano.

Algoritmo 1 Pseudocódigo para o estimador DLMC para SEIG.

```

1: funcao DLMC( $\mathbf{d}, N, M$ )
2:   para  $n = 1, 2, \dots, N$  fazer                                Laço externo
3:     Amostrar  $\theta_n$  de  $\pi(\theta)$ 
4:     Atualizar  $g(\mathbf{d}, \theta_n)$ 
5:     para  $i = 1, 2, \dots, N_e$  fazer
6:       Amostrar  $\varepsilon_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$ 
7:        $y_i \leftarrow g(\mathbf{d}, \theta_n) + \varepsilon_i$ 
8:     finalizar para
9:      $\mathbf{Y}_n \leftarrow \{y_i\}_{i=1}^{N_e}$ 
10:     $P(\mathbf{Y}_n | \theta_n, \mathbf{d}) \leftarrow \det(2\pi\Sigma_\varepsilon)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} \|\mathbf{r}_i(\mathbf{d}, \theta, \theta^*, \varepsilon)\|_{\Sigma_\varepsilon^{-1}}^2\right)$ 
11:    para  $m = 1, 2, \dots, M$  fazer                                Laço interno
12:      Amostrar  $\theta_m^*$  para  $\pi(\theta)$ 
13:      Atualizar  $g(\mathbf{d}, \theta_m^*)$ 
14:       $P(\mathbf{Y}_n | \theta_m^*, \mathbf{d}) \leftarrow \det(2\pi\Sigma_\varepsilon)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} \|\mathbf{y}_i^{(n)}(\mathbf{d}) - g(\mathbf{d}, \theta_m^*)\|_{\Sigma_\varepsilon^{-1}}^2\right)$ 
15:    finalizar para
16:     $p(\mathbf{Y}_n | \mathbf{d}) \leftarrow \frac{1}{M} \sum_{m=1}^M p(\mathbf{Y}_n | \theta_m^*, \mathbf{d})$ 
17:  finalizar para
18:   $\mathcal{I}_{DLMC} \leftarrow \frac{1}{N} \sum_{n=1}^N \log\left(\frac{p(\mathbf{Y}_n | \theta_n, \mathbf{d})}{p(\mathbf{Y}_n | \mathbf{d})}\right)$ 
20:  retornar  $\mathcal{I}_{DLMC}(\mathbf{d})$ 
21: finalizar funcao

```

3 KRIGING APLICADO AO MODELO DO SEIG

Neste capítulo é apresentado a construção do metamodelo do Kriging e o cálculo do SEIG, abordando o Kriging como um substituto para o modelo direto. Essa metodologia permite minimizar o custo associado às chamadas do modelo durante o cálculo do SEIG e é denominado Kriging do estágio 1.

3.1 METAMODELO DO KRIGING

O propósito de um metamodelo é construir uma superfície preditora com base em um conjunto finito de observações conhecidas dentro do domínio de projeto. A partir dessa superfície, é possível prever o comportamento de pontos desconhecidos. Em problemas de otimização, os resultados passam então a ser previstos por meio desse preditor, sem a necessidade de consultar diretamente a função objetivo original. Existe uma literatura que aborda diferentes técnicas de metamodelagem, dentre as quais destaca-se o metamodelo do Kriging.

O Kriging é um metamodelo de base Gaussiana, que em sua forma padrão é definido como um método de interpolação estocástica (LATANIOTIS; MARELLI; SUDRET, 2017). Algumas extensões foram propostas para lidar com observações ruidosas, ficando o Kriging também conhecido como modelo de Regressão de Processo Gaussiano (GPR, do inglês *Gaussian Process Regression*) (RASMUSSEN, 2003).

O Kriging é uma ferramenta altamente bem sucedida no projeto e análise de experimentos em simulações computacionais e tem sido bastante aplicado no campo da Engenharia (SANTNER *et al.*, 2003). A principal vantagem desse metamodelo é a quantificação da incerteza da superfície de resposta através do erro quadrático médio (MSE, do inglês *Mean squared error*), também conhecido como variância do Kriging (VAN BEERS; KLEIJNEN, 2003).

Quando aplicado no campo da otimização, este metamodelo tem demonstrado um ótimo desempenho para lidar com funções que apresentam não convexidade, multimodalidade e não linearidade, se tornando uma alternativa aos métodos tradicionais (PICHENY; WAGNER; GINSBOURGER, 2013; JALALI; VAN NIEUWENHUYSE; PICHENY, 2017).

3.1.1 Conceitos básicos do Kriging

Seja $\mathcal{G}$ a representação de um modelo computacional, onde $\mathbf{x} = (x_1, x_2, \dots, x_{n_x})^T$ é o vetor de variáveis de projeto de dimensão n_x , que produz um resultado escalar uni-dimensional y , definido por $\mathcal{G} : \mathbf{x} \in \mathcal{D}_x \subset \mathbb{R}^{n_x} \mapsto y \in \mathbb{R}$, em que $\mathcal{D}_x$ representa o domínio de projeto. Pelo metamodelo do Kriging, a superfície de resposta, obtida a partir da k -ésima replicação da simulação em $\mathbf{x}$, pode ser representada como (ANKENMAN;

NELSON; STAUM, 2010):

$$\mathcal{G}_k(\mathbf{x}) = \beta^T \mathbf{f}(\mathbf{x}) + Z(\mathbf{x}), \quad (15)$$

em que $\beta^T \mathbf{f}(\mathbf{x})$ é a tendência constante de resposta do modelo, ou seja, o valor médio do processo Gaussiano, sendo $\mathbf{f}(\mathbf{x})$ o vetor de valores conhecidos da função e β o vetor de parâmetros desconhecidos e de mesma dimensão. Enquanto $Z(\mathbf{x})$ é definida como um processo estocástico Gaussiano estacionário com média zero e variância e covariância não nulas entre os pontos, sendo esta responsável por determinar como o metamodelo interpola os dados e aproximar a função (FORRESTER; KEANE, 2009). Dito isto, a parcela $Z(\mathbf{x})$ representa a incerteza da predição, referido como incerteza extrínseca, pois é imposta ao problema para o desenvolvimento do metamodelo.

Ao assumir que a variável de saída segue um modelo Gaussiano, é possível quantificar a incerteza associada às predições em pontos não amostrados, abordando conceitos estatísticos como o MSE, que mede a qualidade da aproximação (JONES; SCHONLAU; WELCH, 1998).

A Figura 7 ilustra uma superfície de resposta obtida através do metamodelo do Kriging determinístico para a função $y(\mathbf{x}) = (6x - 2)^2 \sin(12x - 4)$. O metamodelo foi construído com base em uma amostragem de seis pontos no domínio do projeto, assumindo que a resposta do modelo nos pontos conhecidos é considerada livre de ruído. Para o ponto $x = 0,9$, que não pertence ao conjunto de pontos de suporte, existe um erro ε associado ao processo estocástico Gaussiano subjacente ao Kriging. Esse erro quantifica a incerteza da predição do metamodelo nesse ponto, decorrente da natureza probabilística do método.

O objetivo do metamodelo é obter uma previsão $\mathcal{G}_k$ para um novo ponto arbitrário $\mathbf{x}$, dado uma amostragem inicial $\mathbf{F} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_k)}\}^T$, de tamanho n_k , e das respostas $\mathbf{y}$ do modelo nesses pontos, dado por $\mathbf{y} = \{y^{(1)} = \mathcal{G}(\mathbf{x}^{(1)}), \dots, y^{(n_k)} = \mathcal{G}(\mathbf{x}^{(n_k)})\}^T$. Essas previsões são baseadas nas propriedades do processo Gaussiano posterior, condicionado aos dados observados. Assim, a predição do Kriging em $\mathbf{x}$ é uma variável aleatória de distribuição normal com média $\hat{\mu}$ e desvio padrão $\hat{\sigma}$, isto é, $\hat{y} \sim \mathcal{N}(\hat{\mu}, \hat{\sigma})$. De acordo com a abordagem Gaussiana, o vetor composto pelas respostas verdadeiras do modelo, $\mathbf{y}$, e pela previsão, $\hat{y}$, segue uma distribuição conjunta normal, conforme descrito em Sudret e Der Kiureghian (2000), definida como:

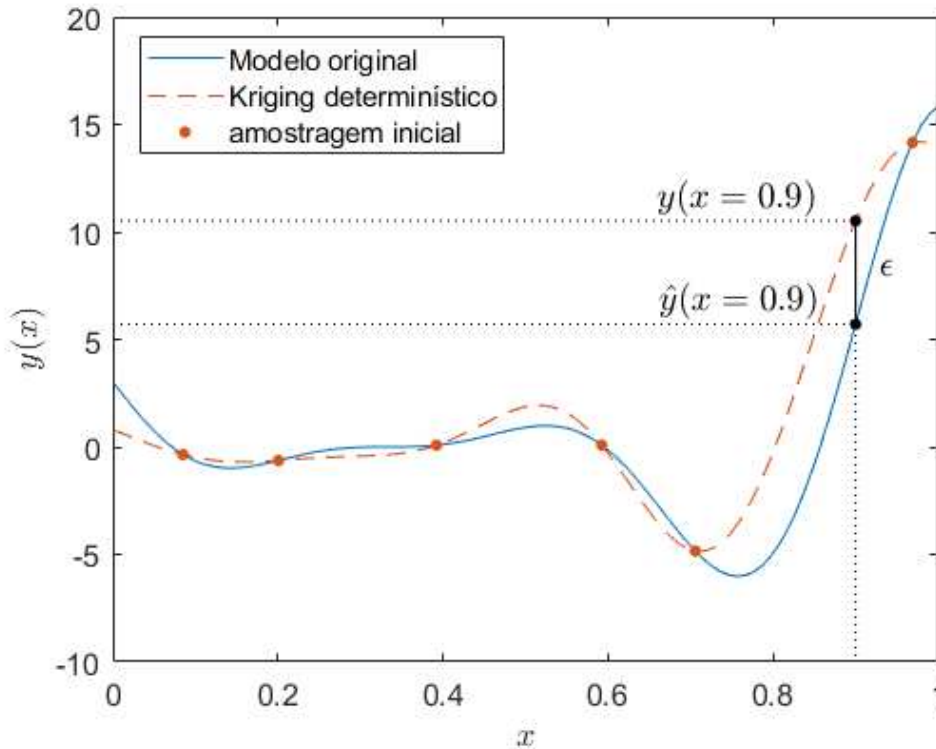
$$\begin{bmatrix} \hat{y}(\mathbf{x}) \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}_{k+1} \left(\begin{bmatrix} \beta^T \mathbf{f}(\mathbf{x}) \\ \beta^T \mathbf{F} \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \mathbf{h}^T \\ \mathbf{h} & \boldsymbol{\Psi} \end{bmatrix} \right), \quad (16)$$

em que,

$\mathbf{F}$ é uma matriz de valores conhecidos da função $f(\mathbf{x})$;

$\mathbf{h}$ é o vetor de correlação entre a nova previsão e os outros pontos amostrados;

Figura 7 – Representação do metamodelo do Kriging determinístico.



Ψ é a matriz de correlação do metamodelo.

3.1.2 Função de correlação

A função de correlação do Kriging, também conhecida como *kernel*, contém as suposições sobre a função que está sendo aproximada, sendo responsável por determinar como o metamodelo irá interpolar os dados e aproximar a função. A escolha da função de correlação influencia diretamente a capacidade do metamodelo do Kriging em capturar as características da função de resposta (SUDRET; DER KIUREGHIAN, 2000). Ao incorporar informações da variância do processo Gaussiano, σ^2 , essa função é denominada de função de covariância e é definida como:

$$\text{Cov} [Z(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})] = \sigma^2 h(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \quad (17)$$

em que, $h(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ é o vetor de correlação (ou função de base) entre os pontos $\mathbf{x}^{(i)}$ e $\mathbf{x}^{(j)}$, e σ^2 representa a variância de $Z(\mathbf{x})$ para todo $\mathbf{x}$, ou seja, representa a variabilidade de incerteza ocasionada pela construção do modelo substituto medida por $Z(\mathbf{x})$.

O valor de σ^2 é ajustado de acordo com os pontos de suporte $\mathbf{x}$ maximizando a função de verossimilhança (MLE, do inglês *Maximum likelihood estimation*), juntamente com os parâmetros da função de correlação, conforme será explicado na Seção 3.1.3.

A função de correlação a ser adotada deve ser positiva definida, garantindo a existência de um processo Gaussiano associado e a positividade da matriz de covariância, sendo esses aspectos importantes para a estimativa dos parâmetros do metamodelo (RASMUSSEN, 2003). Além disso, as funções de correlação utilizadas em processos Gaussianos são estacionárias, dependendo apenas da distância euclidiana entre dois pontos, $r = \|x^{(i)} - x^{(j)}\|$. Essa propriedade implica que a correlação entre duas observações depende apenas de sua separação espacial e não de suas localizações absolutas (WILLIAMS; RASMUSSEN, 2006; SUDRET; DER KIUREGHIAN, 2000). As funções de correlação mais comumente empregadas no campo do processo Gaussiano que apresentam essas características são:

- Linear:

$$h(x^{(i)}, x^{(j)}) = \sum_{k=1}^{n_x} x_k^{(i)} \cdot x_k^{(j)}, \quad (18)$$

- Exponencial:

$$h(x^{(i)}, x^{(j)}) = \exp \left[- \sum_{k=1}^{n_x} c_k |x_k^{(i)} - x_k^{(j)}| \right], \quad (19)$$

- Matêrn 3/2:

$$h(x^{(i)}, x^{(j)}) = \left[1 + \sqrt{3} \sum_{k=1}^{n_x} \frac{|x_k^{(i)} - x_k^{(j)}|}{c_k} \right] \exp \left[- \sum_{k=1}^{n_x} \frac{|x_k^{(i)} - x_k^{(j)}|}{c_k} \right], \quad (20)$$

- Matêrn 5/2:

$$h(x^{(i)}, x^{(j)}) = \left(1 + \sqrt{5}r + \frac{5r^2}{3} \right) \exp(-\sqrt{5}r), \quad r = \sum_{k=1}^{n_x} \frac{|x_k^{(i)} - x_k^{(j)}|}{c_k}, \quad (21)$$

- Gaussiana:

$$h(x^{(i)}, x^{(j)}) = \exp \left[- \sum_{k=1}^{n_x} c_k |x_k^{(i)} - x_k^{(j)}|^{p_k} \right], \quad (22)$$

em que,

n_x é o número de variáveis de projeto;

c_k, p_k são parâmetros do metamodelo.

Os parâmetros c_k e p_k refletem a variância do processo e devem ser ajustados para cada dimensão k do problema nos pontos de suporte. Em suma, a função de correlação h irá quantificar o grau de dependência entre os n_k pontos de Γ , e dependem apenas da distância euclidiana entre os pontos amostrais $\|x_k^{(i)} - x_k^{(j)}\|$ e os parâmetros c_k e p_k .

Cada tipo de função de correlação representa diferentes níveis de suavidade para os processos Gaussianos associados. Neste trabalho, adotou-se a covariância Gaussiana, também conhecida por covariância exponencial quadrática estacionária.

Por ser uma função de covariância infinitamente diferenciável, ela proporciona um comportamento suave entre os pontos amostrados, que resulta em uma solução mais robusta e flexível (RASMUSSEN, 2003). Além disso, a simplicidade analítica do *kernel* Gaussiano facilita sua implementação e a interpretação dos resultados. Para modelos com descontinuidades e variações abruptas, faz-se importante testar outros *kernels* como o exponencial ou o de Matern, que permitem modelar diferentes graus de suavidade (SUDRET; DER KIUREGHIAN, 2000).

Analisando a estruturas das funções de correlação, observa-se que quando $\|\mathbf{x}_k^{(i)} - \mathbf{x}_k^{(j)}\| \rightarrow 0$, então a correlação entre eles tende para 1, ou seja, o preditor do Kriging é exatamente o valor da função em qualquer ponto de suporte. Por outro lado, quando $\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| \rightarrow \infty$, a correlação tende a zero, indicando que não há influência de um ponto de suporte sobre o outro e os pontos são considerados independentes entre si (FORRESTER; KEANE, 2009).

Para os n_k pontos de suporte do metamodelo, a matriz de correlação Ψ , presente na formulação do Kriging, Equação (16), pode ser construída a partir da função de correlação, Equação (22), como:

$$\Psi = \begin{bmatrix} h(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & h(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & h(\mathbf{x}^{(1)}, \mathbf{x}^{(n_k)}) \\ h(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & h(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & h(\mathbf{x}^{(2)}, \mathbf{x}^{(n_k)}) \\ \dots & \dots & \ddots & \dots \\ h(\mathbf{x}^{(n_k)}, \mathbf{x}^{(1)}) & h(\mathbf{x}^{(n_k)}, \mathbf{x}^{(2)}) & \dots & h(\mathbf{x}^{(n_k)}, \mathbf{x}^{(n_k)}) \end{bmatrix}. \quad (23)$$

A partir da análise da matriz de correlação é possível quantificar a importância relativa de cada variável e identificar interações entre elas. Na Equação (23), é possível observar que a predição interpola os dados, pois em um dado ponto de suporte, o vetor de correlação entre a nova previsão do ponto e os outros n_k pontos amostrados, $\mathbf{h}$, é uma coluna da matriz de correlação Ψ . Consequentemente, o erro é nulo nesses pontos.

A próxima fase é estimar os parâmetros da função de correlação, Equação (22), para realizar o ajuste do metamodelo do Kriging.

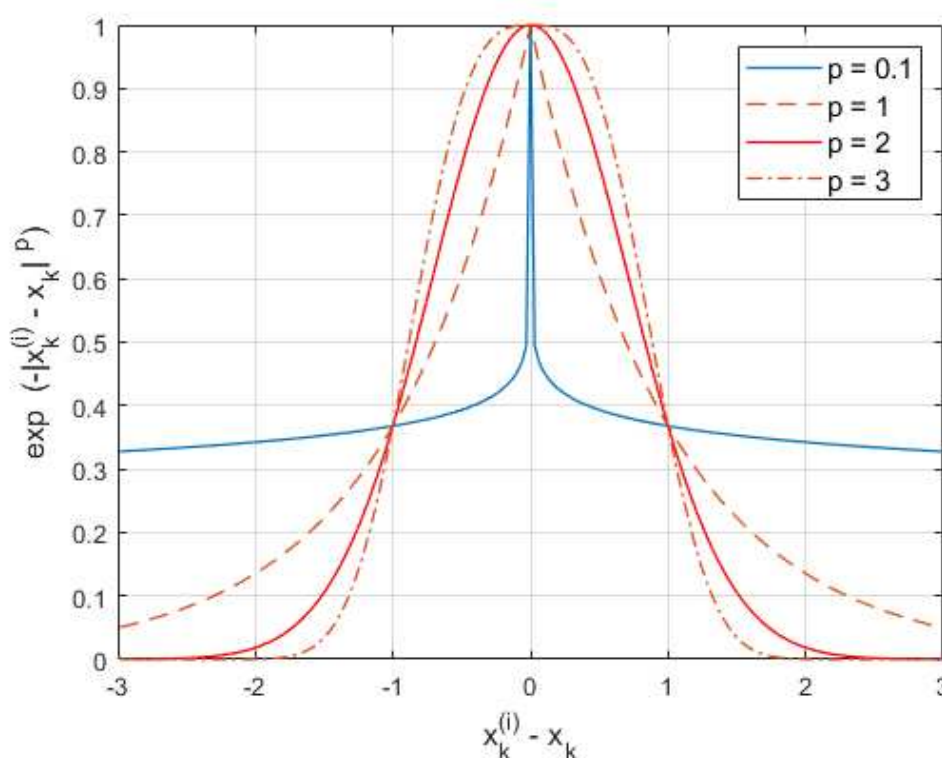
3.1.3 Estimativa dos parâmetros do Kriging

Os parâmetros c_k e p_k na função de correlação, Equação (22), desempenham papéis distintos no comportamento da correlação. O parâmetro p_k pode ser interpretado como um parâmetro de suavidade, que influencia a taxa de decaimento da correlação à medida que a distância entre os pontos aumenta, Figura 8. Por exemplo,

quando $p_k = 2$, a correlação apresenta um comportamento suave com um gradiente contínuo. Enquanto que para valores menores, como $p_k = 0,1$, resultam em uma queda acentuada da correlação. Assim, um valor muito baixo de p_k sugere uma correlação quase nula entre os pontos, o que pode ser interpretado como uma descontinuidade na relação entre $Z(x^{(i)})$ e $Z(x^{(j)})$.

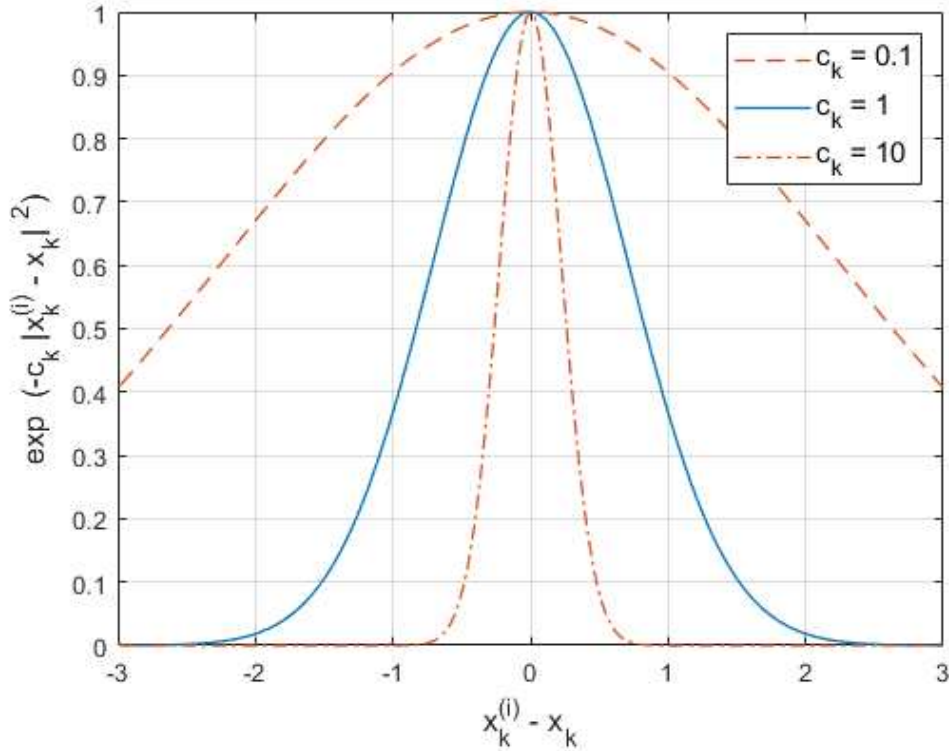
Por outro lado, para valores alto de p_k , a função de correlação decairá de forma muito mais lenta à medida que a distância entre os pontos aumenta. Consequentemente, os valores das variáveis preditas pelo Kriging serão muito semelhantes, mesmo para pontos distantes, resultando em previsões quase constantes, sem variações significativas entre as amostras. Esse comportamento pode resultar em um metamodelo excessivamente suave, com pouca capacidade de capturar variações de comportamento do problema.

Figura 8 – Influência do parâmetro p_k na função de correlação



Já o parâmetro c_k determina a extensão da influência de um ponto amostral sobre outros, onde um valor baixo implica em uma alta correlação entre todos os pontos e um valor elevado resulta em diferenças significativas entre os valores das variáveis correlacionadas, conforme apresentado na Figura 9 (FORRESTER; KEANE, 2009).

Em suma, para a abordagem do metamodelo do Kriging, Equação (16), faz-se necessário obter a estimativa dos valores da média e da variância do Kriging, μ e σ^2 , e

Figura 9 – Influência do parâmetro c_k na função de correlação

os parâmetros c_k e p_k , que minimize o erro de generalização do modelo. Na maioria das aplicações práticas que abordam o Kriging, os parâmetros são estimados a partir de informações do domínio de projeto $\mathbf{I}$ e das respostas do modelo y . Os métodos mais comumente utilizados para essa estimativa são a maximização da função de verossimilhança e o método da validação cruzada (SUDRET; DER KIUREGHIAN, 2000). Neste trabalho foi abordado o método de maximização da função de verossimilhança dos dados observados y .

Como y segue uma distribuição normal multivariada, a função de verossimilhança é dada por:

$$\begin{aligned} L(y \mid \mu, \sigma^2, c_k, p_k) &= (2\pi)^{-n/2} |\sigma^2 \Psi|^{-1/2} \exp \left[-\frac{1}{2} (y - 1\mu)^T (\sigma^2 \Psi)^{-1} (y - 1\mu) \right], \\ &= (2\pi\sigma^2)^{-n/2} |\Psi|^{-1/2} \exp \left[-\frac{1}{2\sigma^2} (y - 1\mu)^T \Psi^{-1} (y - 1\mu) \right]. \end{aligned} \quad (24)$$

Para simplificar os cálculos, é comum maximizar o logaritmo natural da função de verossimilhança, uma vez que este é uma função monotônica da verossimilhança. Temos que os parâmetros μ , σ^2 , c_k e p_k , que são capazes de maximizar a Equação (24), também são minimizadores para a função negativa $\ln(L(y \mid \mu, \sigma^2, c_k, p_k))$. Desconsiderando os termos constantes, a função logarítmica negativa da verossimi-

lhança é dada por:

$$L_{\ln}(y | \mu, \sigma^2, c_k, p_k) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \ln|\Psi| - \frac{(y - 1\mu)^T \Psi^{-1} (y - 1\mu)}{2\sigma^2}. \quad (25)$$

Para obter os valores da média μ e da variância σ^2 , presentes na função de verossimilhança, é utilizado o conceito do gradiente, que indica a direção de menor crescimento da função Nascentes *et al.* (2019). Para tanto, é calculado as derivadas parciais da Equação (25) em relação a essas variáveis:

$$\frac{\partial L_{\ln}(y | \mu, \sigma^2, c_k, p_k)}{\partial \mu} = 0, \quad (26)$$

$$\frac{\partial L_{\ln}(y | \mu, \sigma^2, c_k, p_k)}{\partial \sigma^2} = 0. \quad (27)$$

Substituindo a Equação (25) na Equação (26), e resolvendo a derivada parcial, obtém-se a estimativa da média μ da seguinte forma:

$$\begin{aligned} -\frac{1}{2\sigma^2} [1^T \Psi^{-1} (y - 1\mu) - (y - 1\mu)^T \Psi^{-1} 1] &= 0, \\ \frac{1}{2\sigma^2} 1^T \Psi^{-1} (y - 1\mu) &= 0, \\ 1^T \Psi^{-1} 1\mu &= 1^T \Psi^{-1} y, \\ \mu &= \frac{1^T \Psi^{-1} y}{1^T \Psi^{-1} 1}. \end{aligned} \quad (28)$$

O mesmo é realizado para obter a estimativa do valor da variância σ^2 . Resolvendo agora a derivada parcial da Equação (27), substituindo novamente a Equação (25), tem-se:

$$\begin{aligned} -\frac{n}{\sigma^2} + \frac{(y - 1\mu)^T \Psi^{-1} (y - 1\mu)}{(\sigma^2)^2} &= 0, \\ \sigma^2 &= \frac{(y - 1\mu)^T \Psi^{-1} (y - 1\mu)}{n}. \end{aligned} \quad (29)$$

Uma vez que ambas as derivadas devem ser satisfeitas simultaneamente, o estimador da média e da variância do metamodelo é dado por:

$$\hat{\mu} = \beta^T y = \frac{1^T \Psi^{-1} y}{1^T \Psi^{-1} 1}, \quad (30)$$

$$\hat{\sigma}^2 = \frac{(y - 1\hat{\mu})^T \Psi^{-1} (y - 1\hat{\mu})}{n}. \quad (31)$$

Substituindo o valor de $\hat{\mu}$ e $\hat{\sigma}^2$ na Equação (25), tem-se:

$$L_{\ln}(c_k, p_k) = -\frac{n}{2} \ln \hat{\sigma}^2 - \frac{1}{2} \ln |\boldsymbol{\Psi}| - \frac{n}{2} [\ln(2\pi) + 1], \quad (32)$$

Ignorando os termos constantes, a equação da logarítmica da função negativa da verossimilhança resulta em:

$$\mathcal{L}_{\ln}(c_k, p_k) \approx -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln |\boldsymbol{\Psi}|. \quad (33)$$

Na Equação (33), o parâmetro c_k está implícito na matriz de correlação $\boldsymbol{\Psi}$. Devido a natureza não convexa e multimodal da $\mathcal{L}_{\ln}$, uma abordagem de otimização global é através de algoritmos metaheurísticos, como, por exemplo, o algoritmo Otimização por Enxame de Partículas (PSO, do inglês Particle Swarm Optimization) (FORRESTER; KEANE; BRESSLOFF, 2006).

A etapa mais computacionalmente custosa na estimativa dos parâmetros do metamodelo é a determinação dos elementos da matriz de correlação $\boldsymbol{\Psi}$. O cálculo da inversa e do determinante dessa matriz pode tornar o processo computacionalmente oneroso para matrizes de grande porte. Entretanto, a utilização da base Gaussiana é vantajosa neste aspecto, pois resulta em uma matriz de correlação definida positiva simétrica. Essa característica possibilita empregar a decomposição de Cholesky para a matriz $\boldsymbol{\Psi}$, o que aumenta significativamente a eficiência do cálculo (FORRESTER; KEANE; BRESSLOFF, 2006).

Outra estratégia para redução do custo, adotada também neste trabalho, é definir o valor do parâmetro p_k igual a 2, conforme abordado por Forrester e Keane (2009) e Jalali, Van Nieuwenhuyse e Picheny (2017). Essa abordagem proporciona suavidade à função de correlação e uma boa representação das características do modelo na maioria dos casos práticos.

Na próxima seção será discutido, por fim, a construção do preditor pelo Kriging.

3.1.4 Preditor do Kriging

Após a determinação dos parâmetros na Seção 3.1.3, o objetivo é obter o valor da predição $\hat{y}$ no novo ponto $\mathbf{x}^+$. Para garantir a consistência dos resultados o mesmo método será adotado, ou seja, a minimização do negativo da log-verossimilhança concentrada. Partindo da definição do metamodelo apresentada na Equação (16), o vetor aumentado de respostas é dado por:

$$\tilde{\mathbf{y}} = \{\hat{\mathbf{y}}, \mathbf{y}\}^T, \quad (34)$$

o vetor de correlações entre os dados observados e a nova previsão:

$$\mathbf{h} = \begin{bmatrix} h(x^1, x^+) \\ h(x^2, x^+) \\ \vdots \\ h(x^{n_k}, x^+) \end{bmatrix}, \quad (35)$$

e a matriz de correlação aumentada $\tilde{\Psi}$ da ordem $(n_k + 1) \times (n_k + 1)$ definida por:

$$\tilde{\Psi} = \begin{bmatrix} \boldsymbol{\psi} & \mathbf{h} \\ \mathbf{h}^T & 1 \end{bmatrix}, \quad (36)$$

em que $[\tilde{\Psi}]_{(n+1)(n+1)} = 1$ representa a correlação entre x^+ e ele mesmo. Substituindo a Equação (34) e a Equação (36) na função da log-verossimilhança concentrada, Equação (25), tem-se:

$$L_{ln}(\hat{y}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln |\tilde{\Psi}| - \frac{(\tilde{y} - 1\hat{\mu})^T \tilde{\Psi}^{-1} (\tilde{y} - 1\hat{\mu})}{2\hat{\sigma}^2}. \quad (37)$$

Como apenas o último termo da Equação (37) é dependente de $\hat{y}$, podemos simplificar a expressão da seguinte forma:

$$L_{ln}(\hat{y}) \approx -\frac{1}{2\hat{\sigma}^2} \begin{bmatrix} \mathbf{y} - 1\hat{\mu} \\ \hat{y} - \hat{\mu} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\psi} & \mathbf{h} \\ \mathbf{h}^T & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} - 1\hat{\mu} \\ \hat{y} - \hat{\mu} \end{bmatrix}. \quad (38)$$

Para obter o valor da predição é necessário minimizar a Equação (38) em relação a $\hat{y}$. O primeiro passo é obter a inversa $\tilde{\Psi}$. Considerando a hipótese de não singularidade de $\tilde{\Psi}$, garante-se a existência de uma única inversa $\tilde{\Psi}^{-1}$ tal que $\tilde{\Psi}^{-1} \tilde{\Psi} = \mathbf{I}$. Dito isto, para simplificar o cálculo, a inversa pode ser expressa da seguinte forma:

$$\Psi^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & C \end{pmatrix}^{-1}, \quad (39)$$

em que $\mathbf{A}$ é uma matriz não singular, $\mathbf{b}$ é um vetor qualquer com $\mathbf{b}^T$ sendo seu transposto e, por fim, C é um escalar real. Logo, tem-se que:

$$\tilde{\Psi}^{-1} \tilde{\Psi} = \mathbf{I} \implies \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & C \end{pmatrix} \begin{pmatrix} \boldsymbol{\psi} & \mathbf{h} \\ \mathbf{h}^T & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}. \quad (40)$$

Resolvendo o sistema matricial, tem-se que:

$$\mathbf{A}\boldsymbol{\psi} + \mathbf{b}\mathbf{h}^T = \mathbf{I} \quad (41)$$

$$\mathbf{A}\mathbf{h} + \mathbf{b} = \mathbf{0} \quad (42)$$

$$\mathbf{b}^T \boldsymbol{\psi} + C\mathbf{h}^T = 0 \quad (43)$$

$$\mathbf{b}^T \mathbf{h} + C = 1. \quad (44)$$

Isolando b^T da Equação (43):

$$b^T = -Ch^T\psi^{-1} \quad (45)$$

Substituindo na Equação (44) e isolando b^T novamente:

$$\begin{aligned} b^T &= -Ch^T\psi^{-1} \\ C(1 - h^T\psi^{-1}h) &= 1 \\ C &= (1 - h^T\psi^{-1}h)^{-1} \\ b^T &= -(1 - h^T\psi^{-1}h)^{-1}h^T\psi^{-1}. \end{aligned} \quad (46)$$

Dado que o termo $(1 - h^T\psi^{-1}h)^{-1}h^T\psi^{-1}$ é de um escalar e que $\psi^T = \psi$, pois ψ é uma matriz simétrica, então aplicando as propriedades da transposição matricial $(A_1, A_2)^T = A_2^T A_1^T$ e $[A_1^{-1}]^T = [A_1^T]^{-1}$, tem-se que:

$$\begin{aligned} b &= - \left[(1 - h^T\psi^{-1}h)^{-1}h^T\psi^{-1} \right]^T \\ &= - \left[h^T\psi^{-1} \right]^T \left[(1 - h^T\psi^{-1}h)^{-1} \right]^T \\ &= - \left[\psi^{-1} \right]^T h \left[(1 - h^T\psi^{-1}h)^T \right]^{-1} \\ &= - \left[\psi^T \right]^{-1} h \left(1 - h^T\psi^{-1}h \right)^{-1} \\ &= -\psi^{-1}h \left(1 - h^T\psi^{-1}h \right)^{-1}. \end{aligned} \quad (47)$$

Substituindo o valor de b na Equação (41), tem-se que:

$$\begin{aligned} A\psi - \psi^{-1}h \left(1 - h^T\psi^{-1}h \right)^{-1} h^T &= I \\ A\psi &= I + \psi^{-1}h \left(1 - h^T\psi^{-1}h \right)^{-1} h^T \\ A &= \left[I + \psi^{-1}h \left(1 - h^T\psi^{-1}h \right)^{-1} h^T \right] \psi^{-1} \\ A &= \psi^{-1} + \psi^{-1}h \left(1 - h^T\psi^{-1}h \right)^{-1} h^T\psi^{-1}. \end{aligned} \quad (48)$$

Reorganizando a Equação (38):

$$\begin{aligned}
L_n(\hat{Y}) &\approx -\frac{1}{2\hat{\sigma}^2} \left[\begin{pmatrix} y-1\hat{\mu} \\ \hat{y}-\hat{\mu} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{h} \\ \mathbf{h}^T & 1 \end{pmatrix}^{-1} \begin{pmatrix} y-1\hat{\mu} \\ \hat{y}-\hat{\mu} \end{pmatrix} \right] \\
&= \frac{1}{2\hat{\sigma}^2} \left[\begin{pmatrix} y-1\hat{\mu} \\ \hat{y}-\hat{\mu} \end{pmatrix}^T \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & C \end{pmatrix} \begin{pmatrix} y-1\hat{\mu} \\ \hat{y}-\hat{\mu} \end{pmatrix} \right] \\
&= -\frac{1}{2\hat{\sigma}^2} \left[\mathbf{A}(y-1\hat{\mu})^2 + (\mathbf{b} + \mathbf{b}^T)(y-1\hat{\mu})(\hat{y}-\hat{\mu}) + C(\hat{y}-\hat{\mu})^2 \right]. \quad (49)
\end{aligned}$$

Dessa última equação, considerando que a matriz $\boldsymbol{\Psi}^{-1}$ é simétrica, o segundo termo será dado por:

$$\begin{aligned}
\mathbf{b} + \mathbf{b}^T &= -\boldsymbol{\Psi}^{-1}\mathbf{h} \left(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h}\right)^{-1} - \left(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h}\right)^{-1} \mathbf{h}^T\boldsymbol{\Psi}^{-1} \\
&= -\frac{\boldsymbol{\Psi}^{-1}\mathbf{h}}{(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} - \frac{\mathbf{h}^T\boldsymbol{\Psi}^{-1}}{(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} \\
&= -2\frac{\mathbf{h}^T\boldsymbol{\Psi}^{-1}}{(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})}. \quad (50)
\end{aligned}$$

Eliminando o termo $\mathbf{A}(y-1\hat{\mu})^2$ por não conter $\hat{y}$, tem-se que:

$$L_n(\hat{Y}) = \left[\frac{1}{2\hat{\sigma}^2(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} \right] (\hat{y}-\hat{\mu})^2 + \left[\frac{\mathbf{h}^T\boldsymbol{\Psi}^{-1}(y-1\hat{\mu})}{\hat{\sigma}^2(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} \right] (\hat{y}-\hat{\mu}). \quad (51)$$

A Equação (51) expressa L_n em função de $\hat{y}$. Para determinar o valor do preditor $\hat{y}$, será novamente utilizado o conceito de gradiente. O objetivo é encontrar o minimizador da função negativa da log-verossimilhança. Para isso, realiza-se a derivada parcial em relação a $\hat{y}$ e iguala-se a zero, resultando em:

$$\frac{dL_n(\hat{y})}{d\hat{y}} = 0, \quad (52)$$

$$\left[-\frac{1}{\hat{\sigma}^2(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} \right] (\hat{y}-\hat{\mu}) + \left[\frac{\mathbf{h}^T\boldsymbol{\Psi}^{-1}(y-1\hat{\mu})}{\hat{\sigma}^2(1 - \mathbf{h}^T\boldsymbol{\Psi}^{-1}\mathbf{h})} \right] = 0, \quad (53)$$

$$\hat{y}(\mathbf{x}^+) = \hat{\mu} + \mathbf{h}^T\boldsymbol{\Psi}^{-1}(\mathbf{y}-1\hat{\mu}). \quad (54)$$

em que,

$\hat{\mu}$ é o estimador da média, Equação (30);

$\mathbf{h}$ é o vetor de correlação entre a nova previsão, $\mathbf{x}^+$, e os outros n_k pontos amostrados, Equação (22);

$\boldsymbol{\Psi}$ é a matriz de correlação do metamodelo, Equação (23);

$\mathbf{y}$ é o vetor de pontos amostrados.

3.1.5 Erro na predição

O preditor obtido na Seção 3.1.4 representa a média da variável aleatória Gaussiana. A partir da variância da variável aleatória Gaussiana fornecida pelo Kriging, é possível obter uma estimativa para o erro das respostas do metamodelo. O MSE, s^2 , é dado por (SANTNER *et al.*, 2003):

$$s^2(\mathbf{x}^+) = \sigma^2 \left[1 - \mathbf{h}^T \boldsymbol{\Psi}^{-1} \mathbf{h} + \frac{1 - \mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{h}}{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{1}} \right]. \quad (55)$$

Como discutido anteriormente durante a construção da matriz de correlação $\boldsymbol{\Psi}$, o valor do preditor é igual ao valor da função objetivo nos pontos amostrados, portanto, o erro é nulo nesses pontos. Isso é esperado, pois não há incerteza sobre os pontos conhecidos no caso determinístico.

Após apresentado a estrutura geral do Kriging, nas próximas subseções serão discutidas a abordagem do Kriging para o problema OED.

3.2 CONSTRUÇÃO DO METAMODELO PARA O PROBLEMA OED

Para a construção do metamodelo do Kriging como um substituto do modelo direto $g(\mathbf{d}^{(k)}, \boldsymbol{\theta})$, a variável do problema será $\mathbf{x} = \boldsymbol{\theta}$, sendo $\boldsymbol{\theta}$ considerado como a única variável independente. A variável $\boldsymbol{\theta}$ representa o vetor dos parâmetros incertos de interesse e $\mathbf{d}^{(k)}$ o vetor das variáveis de projeto. Desta forma, para um dado vetor fixo $\mathbf{d}^{(k)}$, é construída uma superfície preditora pelo Kriging $\hat{g}(\boldsymbol{\theta}) = \hat{g}(\mathbf{d}^{(k)}, \boldsymbol{\theta})$ para aproximar $g(\mathbf{d}, \boldsymbol{\theta})$. Então, em vez de executar o modelo direto durante o DLMC, será empregado o metamodelo criado, que é computacionalmente barato, reduzindo assim o custo computacional associado às avaliações do modelo direto durante o cálculo do SEIG.

Esse processo resulta na construção de um modelo substituto, denotado como $\hat{g}_j(\boldsymbol{\theta})$, para cada resultado do experimento, onde $j = 1, \dots, n_y$. No processo de otimização, a cada novo ponto de projeto $\mathbf{d}^{(k)}$, j novos metamodelos são gerados. Para simplificar a notação, será considerado o caso em que $n_y = 1$, ou seja, considerando apenas um resultado de experimento g . O substituto do modelo direto é construído com base em um plano de amostragem de n_k pontos:

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = [\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n_k)}]^T, \quad (56)$$

em que a amostragem é definida utilizando o método de Hipercubo Latino (LHS, do inglês *Latin hypercube sampling*) (FORRESTER; SOBESTER; KEANE, 2008). Então, o valor da função objetivo $g(\boldsymbol{\theta}^{(n_k)})$ de cada ponto do plano amostral é avaliado e armazenado no vetor:

$$\mathcal{G} = [g(\boldsymbol{\theta}^{(1)}), g(\boldsymbol{\theta}^{(2)}), \dots, g(\boldsymbol{\theta}^{(n_k)})]^T. \quad (57)$$

Em linhas gerais, o metamodelo do Kriging coleta valores de saída $\mathcal{G}$, resultado de um conjunto de dados de entrada $\mathbf{I}_\theta$, para realizar um mapeamento das demais regiões de $g(\mathbf{d}, \theta)$ baseado nessas observações conhecidas.

O Kriging assume que a resposta do modelo, através da realização de um processo Gaussiano, apresenta média constante μ , variância σ_g e matriz de covariância Σ_g . O que relaciona uma observação a outra é a função de covariância, Σ_g . Considerando um plano de amostragem inicial, a covariância entre quaisquer dois pontos de entrada $\theta_k^{(i)}$ e $\theta_k^{(j)}$ será dada por:

$$[\Sigma_g]_{ij} = \sigma_g^2 h(\theta_k^{(i)}, \theta_k^{(j)}), \quad (58)$$

em que,

σ_g^2 é a variância da incerteza ocasionada pela construção do Kriging;

h é o vetor de correlação, definido como:

$$h(\theta_k^{(i)}, \theta_k^{(j)}) = \exp \left[- \sum_{k=1}^{n_\theta} c_k |\theta_k^{(i)} - \theta_k^{(j)}|^2 \right], \quad (59)$$

em que n_k é o tamanho do vetor da variável de projeto θ e c_k é um parâmetro que reflete a variância do processo e deve ser ajustado para cada dimensão do problema dos k pontos de suporte. O c_k é estimado pelo MLE, minimizando o erro de generalização. Os limites de busca para o hiperparâmetro é de $-3 \leq c_k \leq 2$, utilizado em base logarítmica de dez (FORRESTER; KEANE; BRESSLOFF, 2006). Para os n_k pontos de suporte do metamodelo, a partir da Equação (59), a matriz de correlação Ψ é dada por:

$$\Psi = \begin{bmatrix} h(\theta^{(1)}, \theta^{(1)}) & h(\theta^{(1)}, \theta^{(2)}) & \dots & h(\theta^{(1)}, \theta^{(n_k)}) \\ h(\theta^{(2)}, \theta^{(1)}) & h(\theta^{(2)}, \theta^{(2)}) & \dots & h(\theta^{(2)}, \theta^{(n_k)}) \\ \dots & \dots & \ddots & \dots \\ h(\theta^{(n_k)}, \theta^{(1)}) & h(\theta^{(n_k)}, \theta^{(2)}) & \dots & h(\theta^{(n_k)}, \theta^{(n_k)}) \end{bmatrix}. \quad (60)$$

A função logaritmo concentrada ($\mathcal{L}_{ln}$) negativa utilizada para obter os parâmetros do metamodelo é dada por:

$$\mathcal{L}_{ln}(\sigma_g, \mathbf{c}_k) \approx -\frac{n_\theta}{2} \ln(\hat{\sigma}_g^2) - \frac{1}{2} |\Psi|. \quad (61)$$

E o valor médio e a variância do processo Gaussiano para o modelo é dado por:

$$\hat{\mu}_g = \frac{\mathbf{1}^T \Psi^{-1} \mathcal{G}}{\mathbf{1}^T \Psi^{-1} \mathbf{1}}, \quad (62)$$

$$\hat{\sigma}_g^2 = \frac{(\mathcal{G} - \mathbf{1}\hat{\mu})^T \Psi^{-1} (\mathcal{G} - \mathbf{1}\hat{\mu})}{n_\theta}. \quad (63)$$

A minimização da Equação (61) será realizada através do algoritmo Otimização por Enxame de Partículas. Uma vez conhecido os hiperparâmetros, a previsão do Kriging para um ponto específico θ^+ no problema OED é dada por:

$$\hat{g}(\theta^+) = \overbrace{\hat{\mu}_g}^{\text{Média}} + \overbrace{\mathbf{h}^T \Psi^{-1} (\mathcal{G} - \mathbf{1}\hat{\mu}_g)}^{\text{Incerteza do metamodelo}}, \quad (64)$$

em que,

$\hat{\mu}_g$ é o estimador da média, Equação (62);

$\mathbf{h}$ é o vetor de correlação entre a nova previsão θ^+ e os demais n_θ pontos amostrados;

Ψ é a matriz de correlação do metamodelo, Equação (60);

$\mathcal{G}$ é o vetor de valores calculado de $g(\theta)$ nos n_k pontos amostrados.

O segundo termo no lado direito da Equação (64) é tido como a incerteza do modelo substituto, uma vez que seu valor é inferido com base no valor da função calculada nos pontos do plano amostral.

3.3 PROJETO EXPERIMENTAL ÓTIMO UTILIZANDO O KRIGING

O modelo, definido na Equação (1), agora é representado como:

$$y_i(\mathbf{d}) = \hat{\mathbf{g}}(\mathbf{d}^{(k)}, \theta) + \varepsilon_i, \quad i = 1, \dots, N_e, \quad (65)$$

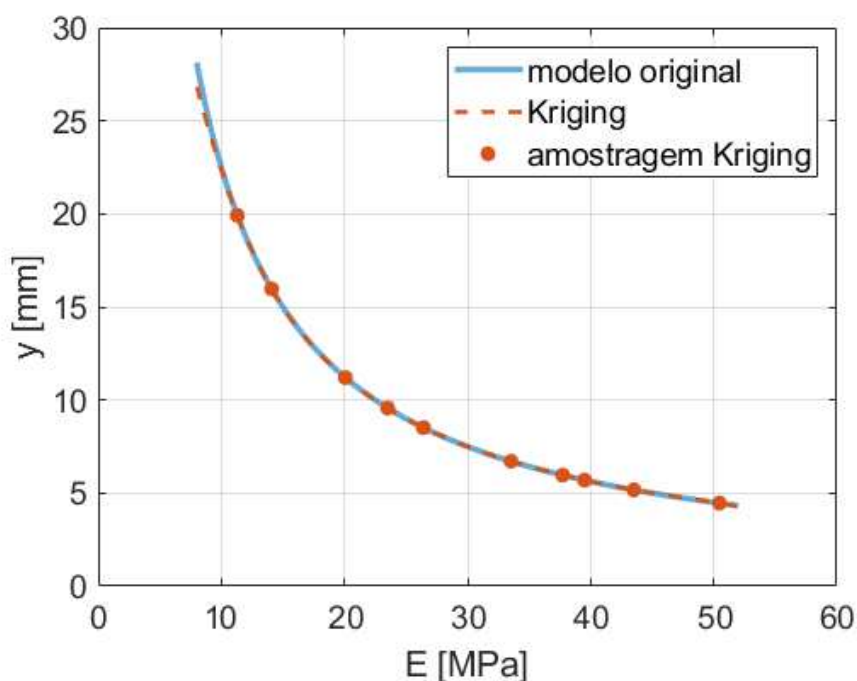
em que o modelo do Kriging $\hat{\mathbf{g}}$ gerado substitui o procedimento original dispendioso empregado para avaliar o modelo direto. Assim, a verossimilhança da inferência Bayesiana, conforme definida na Equação (3), é aproximada como:

$$\hat{p}(\mathbf{Y}|\theta, \mathbf{d}^{(k)}) = \det(2\pi\Sigma_\varepsilon)^{-\frac{N_e}{2}} \exp \left(-\frac{1}{2} \sum_{i=1}^{N_e} \left\| y_i(\mathbf{d}^{(k)}) - \hat{\mathbf{g}}(\mathbf{d}^{(k)}, \theta) \right\|_{\Sigma_\varepsilon^{-1}}^2 \right). \quad (66)$$

A Figura 10 apresenta o gráfico do modelo para o exemplo da viga de Euler-Bernoulli submetida a uma carga pontual, Figura 1, traçado para o modelo original e através do metamodelo do Kriging, com uma amostragem de 10 pontos. A equação que define o modelo é $g(x, E) = \frac{2P}{6EIL}(L^2x - 4x^2 - x^3)$, e os gráficos foram traçados para a posição $x = 3m$, ou seja, $g(3, E)$, onde E é a variável incerta a ser recuperada.

No Algoritmo 2 é apresentado a rotina de cálculo do SEIG para uma dada posição $\mathbf{d}^{(k)}$, utilizando o Kriging como substituto do modelo direto. Assim, a atualização

Figura 10 – Gráfico do modelo e do metamodelo do Kriging para o exemplo da viga de Euler-bernoulli na posição $x = 3,0\text{m}$.



do valor do modelo através do DLMC é realizada através do metamodelo, $\hat{\mathbf{g}}(\mathbf{d}, \boldsymbol{\theta})$, linhas 13 e 17 do Algoritmo 2. Ao empregar a superfície preditora do Kriging em vez do modelo originalmente dispendioso, espera-se que o custo relacionado ao modelo seja insignificante. Além disso, este método permite superar as instabilidades numéricas do DLMC, além de lidar com problemas envolvendo variáveis discretas, descontinuidades no domínio e distribuições de probabilidade não Gaussianas.

Algoritmo 2 Pseudocódigo para o estimador SEIG abordando o Kriging.

```

1: funcao  $\mathcal{I}_{KR}(\mathbf{d})$ 
2:   Amostrar  $\theta^{(n_k)}$ , obtendo  $\Gamma_\theta$ 
3:   Calcular  $g(d, \theta^{n_k})$  nos pontos amostrados, obtendo  $\mathcal{G}$ 
4:   Construir a superfície preditora  $\hat{\mathbf{g}}(d, \theta)$ 
5:   para  $n = 1, 2, \dots, N$  fazer Laço externo
6:     Amostrar  $\theta_n$  de  $\pi(\theta)$ 
7:     Atualizar  $\hat{\mathbf{g}}(\mathbf{d}, \theta_n)$ 
8:     para  $i = 1, 2, \dots, N_e$  fazer
9:       Amostrar  $\varepsilon_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$ 
10:       $y_i \leftarrow \hat{\mathbf{g}}(\mathbf{d}, \theta_n) + \varepsilon_i$ 
11:    finalizar para
12:     $\mathbf{Y}_n \leftarrow \{y_i\}_{i=1}^{N_e}$ 
13:     $P(\mathbf{Y}_n | \theta_n, \mathbf{d}) \leftarrow \det(2\pi\Sigma_\varepsilon)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} \|\mathbf{r}_i(\mathbf{d}, \theta, \theta^*, \varepsilon)\|_{\Sigma_\varepsilon^{-1}}^2\right)$ 
14:    para  $m = 1, 2, \dots, M$  fazer Laço interno
15:      Amostrar  $\theta_m^*$  para  $\pi(\theta)$ 
16:      Atualizar  $\mathbf{g}(\mathbf{d}, \theta_m^*)$ 
17:       $P(\mathbf{Y}_n | \theta_m^*, \mathbf{d}) \leftarrow \det(2\pi\Sigma_\varepsilon)^{-\frac{N_e}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N_e} \|\mathbf{y}_i^{(n)}(\mathbf{d}) - \hat{\mathbf{g}}(\mathbf{d}, \theta_m^*)\|_{\Sigma_\varepsilon^{-1}}^2\right)$ 
18:    finalizar para
19:     $p(\mathbf{Y}_n | \mathbf{d}) \leftarrow \frac{1}{M} \sum_{m=1}^M p(\mathbf{Y}_n | \theta_m^*, \mathbf{d})$ 
20:  finalizar para
21:   $\mathcal{I}_{DLMC} \leftarrow \frac{1}{N} \sum_{n=1}^N \log\left(\frac{p(\mathbf{Y}_n | \theta_n, \mathbf{d})}{p(\mathbf{Y}_n | \mathbf{d})}\right)$ 
22:  retornar  $\mathcal{I}_{DLMC}(\mathbf{d})$ 
23: finalizar funcao

```

4 OTIMIZAÇÃO GLOBAL EFICIENTE

Algoritmos baseados no metamodelo do Kriging são amplamente utilizados em problemas de otimização devido à sua capacidade de gerar previsões em cenários de alto custo computacional, ao mesmo tempo que fornece estimativa dos erros dessas previsões. A incerteza na superfície de resposta é quantificada por meio do MSE, possibilitando reduzir o número de avaliações da função objetivo.

Uma abordagem tradicional para otimização através do Kriging é o EGO (JONES, 2001). Este método tem sido amplamente aplicado e se destacando por sua eficiência na busca pelo ótimo global de funções complexas e caras de avaliar (NASCENTES *et al.*, 2019; CARRARO *et al.*, 2019; MIGUEL *et al.*, 2022; GU, 2021; CARLON *et al.*, 2023).

O algoritmo EGO começa simulando um conjunto restrito de combinações de entrada, denominado plano de amostragem inicial, para construir um modelo substituto pelo Kriging para $\mathcal{I}_{KR}(\mathbf{d})$. Subsequentemente, são selecionados iterativamente novas combinações de entrada, utilizando um critério de preenchimento (IC, do inglês *Infill criteria*) que captura informações do metamodelo. Nesta pesquisa é empregado o critério EI para este propósito. A superfície de resposta é então atualizada sequencialmente com as informações adquiridas dos novos pontos simulados, chamados IPs. Este processo é repetido até que o nível de desempenho desejado seja alcançado, que, neste trabalho, será determinado pelo número máximo de chamadas do modelo direto.

Essa segunda abordagem do Kriging, aplicada ao processo de otimização do SEIG utilizando o EGO, é denominada de etapa 2. Neste capítulo, é apresentado a construção do metamodelo de Kriging para aproximar o SEIG e o processo de maximização utilizando o algoritmo EGO, abordando o EI para a seleção de novos pontos. Por fim, é apresentado um esquema completo da abordagem em duas etapas do Kriging para problemas de OED Bayesiano através do SEIG.

4.1 APROXIMAÇÃO DO SEIG PELO KRIGING

No processo de otimização, primeiramente é criado um plano de amostragem do Kriging para o SEIG $\mathbf{r}$ contendo n_{is} pontos, ou seja,

$$\mathbf{r} = \{\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(n_{is})}\}. \quad (67)$$

Para a obtenção das amostras é utilizado o método LHS (FORRESTER; SOBSTER; KEANE, 2008). Em seguida, é avaliado o valor da função objetivo $\mathcal{I}_{KR}(\mathbf{d})$ em cada um desses pontos, obtendo-se:

$$\mathcal{I} = \{\mathcal{I}_{KR}^{(1)}, \mathcal{I}_{KR}^{(2)}, \dots, \mathcal{I}_{KR}^{(n_{is})}\}. \quad (68)$$

Na etapa seguinte, o modelo preditivo $\hat{\mathcal{I}}_{KR}$ é construído com base nas informações existentes do plano de amostragem inicial ($\mathbf{f}$ e $\mathcal{I}$) por meio do Kriging (SACKS *et al.*, 1989), conforme detalhado na Seção 3.2. É importante observar que tal substituto é construído com base no vetor de variáveis de projeto $\mathbf{d}$. A covariância entre dois pontos de entrada $\mathbf{d}^{(i)}$ e $\mathbf{d}^{(j)}$ é expressa da seguinte forma:

$$[\Sigma_\theta]_{ij} = \sigma_\theta^2 \exp \left[- \sum_{r=1}^{n_d} c_r |d_r^{(i)} - d_r^{(j)}|^2 \right], \quad (69)$$

onde σ_θ^2 é a variância do processo Gaussiano. O parâmetro desconhecido c_r é obtido pelo MLE, Equação (70), com limite de $-3 \leq c_k \leq 2$, em base logarítmica de dez.

$$L_{\ln}(\sigma_\theta, c_r) \approx -\frac{n_{is}}{2} \ln(\hat{\sigma}_\theta^2) - \frac{1}{2} |\boldsymbol{\Psi}| \quad (70)$$

Resultando na seguinte média e variância do processo Gaussiano:

$$\hat{\mu}_\theta = \frac{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathcal{I}}{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{1}} \quad (71)$$

$$\hat{\sigma}_\theta^2 = \frac{(\mathcal{I} - \mathbf{1} \hat{\mu}_\theta)^T \boldsymbol{\Psi}^{-1} (\mathcal{I} - \mathbf{1} \hat{\mu}_\theta)}{n_{is}} \quad (72)$$

Para otimizar a Equação (70) será utilizado o algoritmo metaheurístico PSO (FORRESTER; KEANE, 2009). Uma vez conhecido os hiperparâmetros, a superfície preditora e a variância do Kriging para o SEIG em um determinado ponto $\mathbf{d}^*$ é dada por:

$$\hat{\mathcal{I}}_{KR}(\mathbf{d}^*) = \hat{\mu}_\theta + \mathbf{h}^T \boldsymbol{\Psi}^{-1} (\mathcal{I} - \mathbf{1} \hat{\mu}_\theta) \quad (73)$$

$$s_{\mathcal{I}}^2(\mathbf{d}^*) = \hat{\sigma}_\theta^2 \left[1 - \mathbf{h}^T \boldsymbol{\Psi}^{-1} \mathbf{h} + \frac{(1 - \mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{h})^2}{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{1}} \right]. \quad (74)$$

Em que,

$\hat{\mu}_\theta$ é o estimador da média, Equação (71);

$\mathbf{h}$ é o vetor de correlação entre a nova previsão $\mathbf{d}^*$ e os outros n_{is} pontos amostrados;

$\boldsymbol{\Psi}$ é a matriz de correlação do metamodelo, Equação (60);

$\mathcal{I}$ é o vetor de valores calculado de $g(\boldsymbol{\theta})$ nos n_{is} pontos amostrados.

Como discutido na Seção 3.2, na matriz de correlação $\boldsymbol{\Psi}$, o valor do preditor é igual ao valor da função objetivo nos pontos amostrados, portanto, o erro é nulo nesses pontos. Isso é esperado pois supõe-se um problema determinístico, portanto não há incerteza sobre pontos conhecidos.

4.2 CRITÉRIO DE PREENCHIMENTO

As estimativas do erro associadas à construção do Kriging permitem fazer compensações entre novas posições de amostragem onde haja uma boa previsão (exploração local) e amostragem onde há alta incerteza do preditor (exploração global) (FORRESTER; KEANE; BRESSLOFF, 2006). Ou seja, equilibrar a necessidade de explorar o espaço de projeto através com a necessidade de melhorar a aproximação local.

Neste trabalho é empregado o algoritmo EGO combinado com o critério de melhoria esperada (EI, do inglês *Expected Improvement*) como critério de preenchimento. Este método estima a quantidade de melhoria esperada em um dado ponto não simulado do domínio. A melhoria no ponto de projeto $\mathbf{d}$ é avaliada usando as informações do preditor do Kriging $\hat{\mathcal{I}}_{KR}(\mathbf{d})$ e sua variância $s^2\mathcal{I}(\mathbf{d})$. Se a melhor solução encontrada até a iteração atual é $\mathcal{I}_{\min} = \min \{ \mathcal{I}_{KR}^{(1)}, \mathcal{I}_{KR}^{(2)}, \dots, \mathcal{I}_{KR}^{(n_{is})} \}$, então, conforme mostrado em (FORRESTER; KEANE; BRESSLOFF, 2006), o EI é analiticamente tratável e dado por:

$$EI(\mathbf{d}) = \left[\mathcal{I}_{\min} - \hat{\mathcal{I}}_{KR}(\mathbf{d}) \right] \Phi \left(\frac{\mathcal{I}_{\min} - \hat{\mathcal{I}}_{KR}(\mathbf{d})}{s_{\mathcal{I}}(\mathbf{d})} \right) + s_{\mathcal{I}}(\mathbf{d}) \varphi \left(\frac{\mathcal{I}_{\min} - \hat{\mathcal{I}}_{KR}(\mathbf{d})}{s_{\mathcal{I}}(\mathbf{d})} \right), \quad (75)$$

onde $\Phi(\cdot)$ e $\varphi(\cdot)$ são, respectivamente, a função de distribuição cumulativa e a função de densidade de probabilidade Gaussianas. Assim, a cada iteração do EGO, o seguinte problema de otimização é resolvido para escolher o próximo IP:

$$\mathbf{d}^{**} = \arg \max_{\mathbf{d} \in S} (EI(\mathbf{d})). \quad (76)$$

Em seguida, $\mathbf{d}^{**}$ é adicionado ao plano de amostragem $\mathbf{I}$, e um novo substituto para o SEIG é construído. Esse processo é repetido até que um critério de convergência seja alcançado, no caso trabalhado nessa pesquisa, até que o máximo número de IPs seja atingido.

A rotina da abordagem do EGO para a maximização do SEIG é representada no Algoritmo 3. O processo consiste em obter a posição $\mathbf{d}_{\min}$ mais promissora maximizando o EI. No algoritmo, observa-se que, dada a amostragem inicial de $\mathbf{d}^{(n_{is})}$ e seus respectivos valores $\mathcal{I}_{KR}^{(n_{is})}$, é traçada a superfície preditora do Kriging para o SEIG, $\hat{\mathcal{I}}_{KR}$. Os pontos amostrados compõe o vetor inicial dos valores de projeto, $\mathbf{d}_{EGO}$ e $\mathcal{I}_{EGO}$. A partir do IP simulado, é calculado $\mathcal{I}_{KR}(\mathbf{d}^*)$ na nova posição. Esses valores são adicionados ao conjunto de soluções viáveis e utilizados para atualizar o metamodelo $\hat{\mathcal{I}}_{KR}(\mathbf{d}^*)$. O processo se repete até que o número de pontos adicionados atinja o número máximo de IPs desejado, e o valor do mínimo seja obtido.

Algoritmo 3 Pseudocódigo do algoritmo EGO.

```

1: funcao EGO
2:   Amostrar  $\mathbf{d}^{(n_{is})}$ , obtendo  $\Gamma$ 
3:   Calcular  $\mathcal{I}_{KR}^{(n_{is})}$  nos pontos amostrados, obtendo  $\mathcal{I}$ 
4:   Construir a superfície preditora  $\hat{\mathcal{I}}_{KR}$ 
5:    $\mathbf{d}_{EGO} \leftarrow \Gamma$ 
6:    $\mathcal{I}_{EGO} \leftarrow \mathcal{I}$ 
7:    $\mathcal{I}_{min} \leftarrow \min \mathcal{I}_{EGO}$ 
8:   enquanto  $n < nIPs$ 
9:      $\mathbf{d}^* \leftarrow \max \left[ \mathcal{I}_{min} - \hat{\mathcal{I}}_{KR}(\mathbf{d}) \right] \Phi \left( \frac{\mathcal{I}_{min} - \hat{\mathcal{I}}_{KR}(\mathbf{d})}{s_{\mathcal{I}}(\mathbf{d})} \right) + s_{\mathcal{I}}(\mathbf{d}) \varphi \left( \frac{\mathcal{I}_{min} - \hat{\mathcal{I}}_{KR}(\mathbf{d})}{s_{\mathcal{I}}(\mathbf{d})} \right)$ 
10:     $\mathcal{I}_{KR} \leftarrow \text{funcao } \mathcal{I}_{KR}(\mathbf{d}^*) \text{ (Algoritmo 2)}$ 
11:     $\mathbf{d}_{EGO} \leftarrow [\mathbf{d}_{EGO} + \mathbf{d}^*]$ 
12:     $\mathcal{I}_{EGO} \leftarrow [\mathcal{I}_{EGO} + \mathcal{I}_{KR}]$ 
13:    Atualizar  $\hat{\mathcal{I}}_{KR}$  a partir de  $\mathbf{d}_{EGO}$  e  $\mathcal{I}_{EGO}$ 
14:    se  $\mathcal{I}_{KR} < \mathcal{I}_{min}$ 
15:       $\mathcal{I}_{min} \leftarrow \mathcal{I}_{KR}$ 
16:       $\mathbf{d}_{min} \leftarrow \mathbf{d}^*$ 
17:    finalizar
18:  finalizar
19:  retornar  $\mathbf{d}_{min}$  e  $\mathcal{I}_{min}$ 
20: finalizar funcao

```

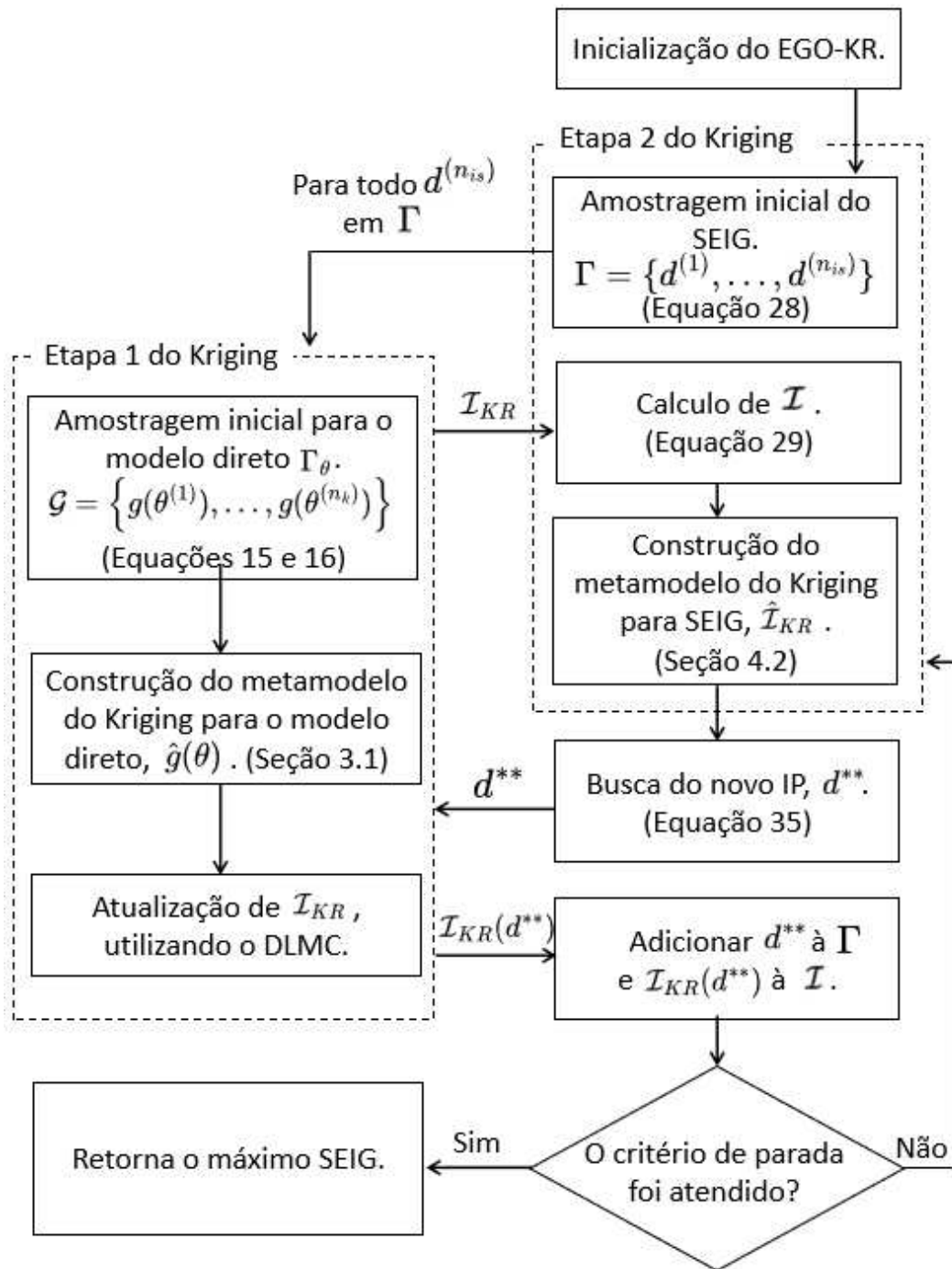
4.3 DUPLA ABORDAGEM DO KRIGING

A metodologia de dupla abordagem do Kriging proposta para a solução eficiente de problemas OED Bayesiano é ilustrada no fluxograma apresentado na Figura 11 e será denominada EGO-KR.

Como apresentado no fluxograma, o método proposto começa construindo a amostragem inicial Γ , composta dos $d^{(n_{is})}$ pontos, referente ao parâmetro do experimento a ser otimizado. Para cada ponto, é realizado o cálculo do $\mathcal{I}_{KR}$, sendo este a primeira etapa da dupla abordagem do Kriging. O processo consiste em construir uma superfície preditora para o modelo direto do problema utilizando o Kriging, tendo o parâmetro da quantidade de interesse desconhecida (θ) como variável de projeto. Em seguida, será utilizado essa superfície preditora para obter o $\mathcal{I}_{KR}$ através do DLMC. O custo desse processo será irrisório, visto que será utilizado a superfície preditora ao invés da função objetivo originalmente custosa.

Na segunda etapa, os $\mathcal{I}_{KR}$ pontos irão compor o vetor de amostragem inicial $\mathcal{I}$, que é utilizado para construir a superfície de resposta do SEIG, $\hat{\mathcal{I}}_{KR}$, utilizada no processo de otimização. Subsequentemente, é selecionado o novo IP, d^{**} , que apresente a maior melhoria, utilizando o EI como métrica de preenchimento. A cada novo IP, repete-se a etapa 1 do Kriging atualizando o modelo $\hat{\mathcal{I}}_{KR}$ com as novas

Figura 11 – Fluxograma para o método de dupla abordagem do Kriging em problemas OED Bayesiano.

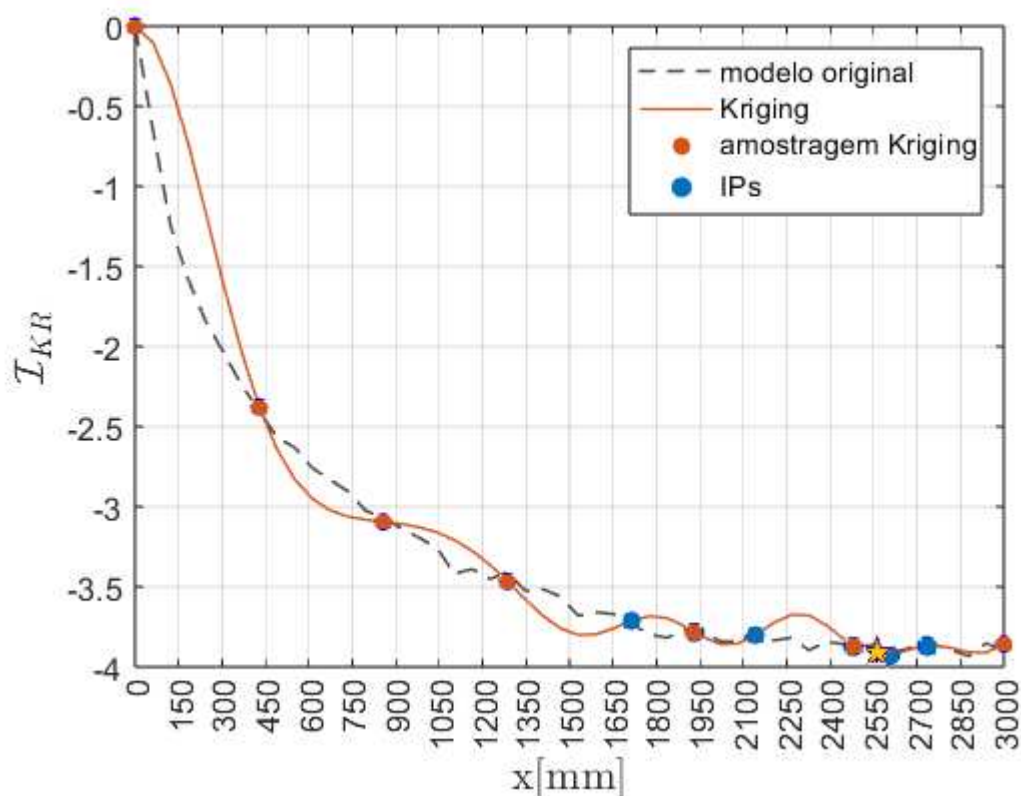


informações até que o critério de parada seja atendido.

Retomando o exemplo da viga de Euler Bernoulli submetida a uma carga pontual, a Figura 12 apresenta graficamente o processo de maximização do SEIG utilizando o metamodelo de Kriging e o EI como métrica de preenchimento. Nesse caso, a otimização do SEIG é realizada como um projeto de minimização, invertendo o valor do

SEIG ao multiplicar por -1 . O metamodelo foi construído com base em 8 pontos de suporte d , representados pelos marcadores circulares laranjas, e o critério de parada foi definido como a quantidade máxima de IPs igual a 4, representados pelos marcadores laranja. O valor ótimo obtido é indicado pelo marcador em forma de estrela amarela.

Figura 12 – Estimativa do SEIG pelo EGO para o exemplo da viga de Euler-Bernoulli.



Observa-se que a métrica de preenchimento adotada aloca pontos ao longo do domínio de busca, abrangendo regiões de maior incerteza do metamodelo e uma maior concentração na área do mínimo global. Isso indica um equilíbrio eficaz entre a exploração global do metamodelo e a exploração local, visando obter uma estimativa mais precisa do valor mínimo da função. A superfície preditora ajustada passa pelos pontos de suporte, mantendo uma boa representação do problema. Além disso, o metamodelo apresenta uma boa aproximação da curva da função original, representado pela linha pontilhada.

5 EXEMPLOS NUMÉRICOS

Nesta seção, a eficiência e robustez da dupla abordagem Kriging proposta (EGO-KR) é avaliada na resolução de problemas OED Bayesiano. Os resultados obtidos pelo EGO-KR são comparados com os obtidos por outros algoritmos da literatura, bem como com o EGO-DLMC, que utiliza o EGO como ferramenta de otimização sem empregar o modelo substituto para avaliar o SEIG. Esta comparação visa verificar se o uso da abordagem do modelo substituto nas duas etapas apresenta algum impacto negativo na precisão geral do algoritmo proposto. Além disso, o desempenho da abordagem proposta é avaliado comparando os resultados com os de algoritmos recentes e eficientes, como EGO-PCE e EGO-BCS (CARLON *et al.*, 2023). Esses algoritmos foram escolhidos para comparação pois demonstraram ser mais eficientes do que métodos de gradiente na resolução de problemas de OED Bayesiano, conforme demonstrado em Carlon *et al.* (2020b, 2023).

O desempenho do método proposto será avaliado usando três exemplos. O primeiro exemplo envolve a resolução de um problema OED bidimensional, que apresenta não-linearidade nos parâmetros e nas variáveis de projeto. O segundo e o terceiro exemplos consistem em problemas de vigas modelados de acordo com a teoria de Timoshenko. No segundo exemplo, o objetivo é otimizar o posicionamento de um extensômetro, definindo sua posição horizontal e vertical na viga, a fim de inferir sobre propriedades mecânicas incertas, como os módulos de elasticidade transversal e longitudinal. Por fim, no terceiro exemplo, o objetivo é posicionar dois extensômetros, ou seja, quatro variáveis de projeto, a fim de caracterizar a posição e magnitude de três cargas verticais aplicadas na viga, totalizando seis parâmetros incertos.

O custo computacional é avaliado através do número de chamadas do modelo direto (NCFM, do inglês *Number of calls of the forward model*) e é empregado como critério de parada em todos os exemplos. O NCFM representa a frequência com que o modelo direto dispendioso foi acessado durante o processo de maximização do SEIG. Os resultados foram analisados considerando tanto o custo computacional associado ao NCFM, quanto a precisão dos resultados obtidos. O custo computacional para estimar o máximo SEIG é calculado multiplicando o número necessário de chamadas do modelo direto para construir o metamodelo, $\hat{g}$, pelo número de vezes que o SEIG foi calculado durante o processo de otimização EGO-KR, ou seja, $n_k \times (n_{is} + IPs)$.

Devido à natureza estocástica do SEIG, os resultados obtidos não são determinísticos e podem mudar quando o algoritmo é executado várias vezes. Por essa razão, serão apresentados a média, o desvio padrão (std) e o melhor valor obtido em um conjunto de pelo menos 10 execuções independentes para cada problema. Todos os valores de SEIG (e suas médias e desvios padrão) apresentados nesta seção foram avaliados utilizando DLMC e o modelo direto original (g). As únicas exceções são os

valores entre parênteses nas Tabelas 2 e 6, que fornecem a estimativa do SEIG utilizando DLMC+Kriging ou PCE+Laplace, respectivamente. Além disso, todos os valores de SEIG foram calculados utilizando as mesmas realizações das variáveis aleatórias θ_n , ε_n e $\theta_{n,m}^*$. O número de pontos amostrais nos laços interno (M) e externo (N) é detalhado no texto de cada exemplo.

As rotinas computacionais foram desenvolvidas utilizando o software comercial *Matlab*® (MATHWORKS, 2021).

5.1 EXEMPLO 01: MODELO NÃO LINEAR

Nesta seção é avaliado o modelo algébrico apresentado por Huan e Marzouk (2013), o qual apresenta não-linearidade nos parâmetros e nas variáveis de projeto. O experimento visa inferir a posição $\mathbf{d}$ que resulta no máximo ganho de informação sobre o parâmetro de incerteza θ , com base em duas medições de y , ou seja, $N_e = 2$. O problema é dado pelo seguinte modelo:

$$\mathbf{y}(\mathbf{d}, \theta) = g(\mathbf{d}, \theta) + \varepsilon, \quad (77)$$

$$= \theta^3 \mathbf{d}^2 + \theta \exp(-|0.2 - \mathbf{d}|) + \varepsilon, \quad (78)$$

em que $y_{1,2}$ é a variável escalar observável, $n_y = 1$, $g(\mathbf{d}, \theta)$ é o modelo, θ é a variável de interesse incerta a ser recuperada, $n_\theta = 1$, que apresenta a distribuição a priori $\pi(\theta) \sim U(0,1)$, $\mathbf{d}$ é o domínio de projeto $d_1, d_2 \in [0,1]$, $n_d = 2$, e ε representa o erro de medição com distribuição Gaussiana $\varepsilon_{1,2} \sim N(0, 10^{-4})$.

5.1.1 Resultados e discussões

Neste exemplo, ambos os experimentos são descritos pelo mesmo modelo, entretanto, os resultados de um experimento não podem ser usados para projetar o outro. A Figura 13 apresenta a curva de comportamento do modelo g_i , $i = 1,2$, nas posições $d=0$ e $d=1$ traçados a partir do modelo original, curva azul, e através do metamodelo do Kriging, linha tracejada laranja, usando uma amostragem de cinco pontos para a construção do preditor, representados pelas marcações redondas.

O desempenho do EGO-KR é comparado com os resultados obtidos pelo EGO-DLMC e com os resultados obtidos por Huan e Marzouk (2013). Em ambos os algoritmos, EGO-KR e EGO-DLMC, o número de pontos amostrais para o laço externo foi fixado em $N=15000$, enquanto $M=1000$ pontos amostrais foram usadas para o laço interno. No processo de otimização, uma amostragem inicial $n_{is} = 10$ foi empregada para construir o substituto Kriging da fase 2, $\mathcal{I}_{KG}(\mathbf{d})$, e o critério de parada foi estabelecido para 5 IPs.

Para construir o substituto Kriging da fase 1, $\hat{g}(\theta)$, dois orçamentos computacionais foram analisados. No primeiro cenário, um plano de amostragem de $n_k = 20$ foi

Figura 13 – Representação do modelo original e do metamodelo do Kriging nas posições $d = 0.2$ e $d = 1.0$.

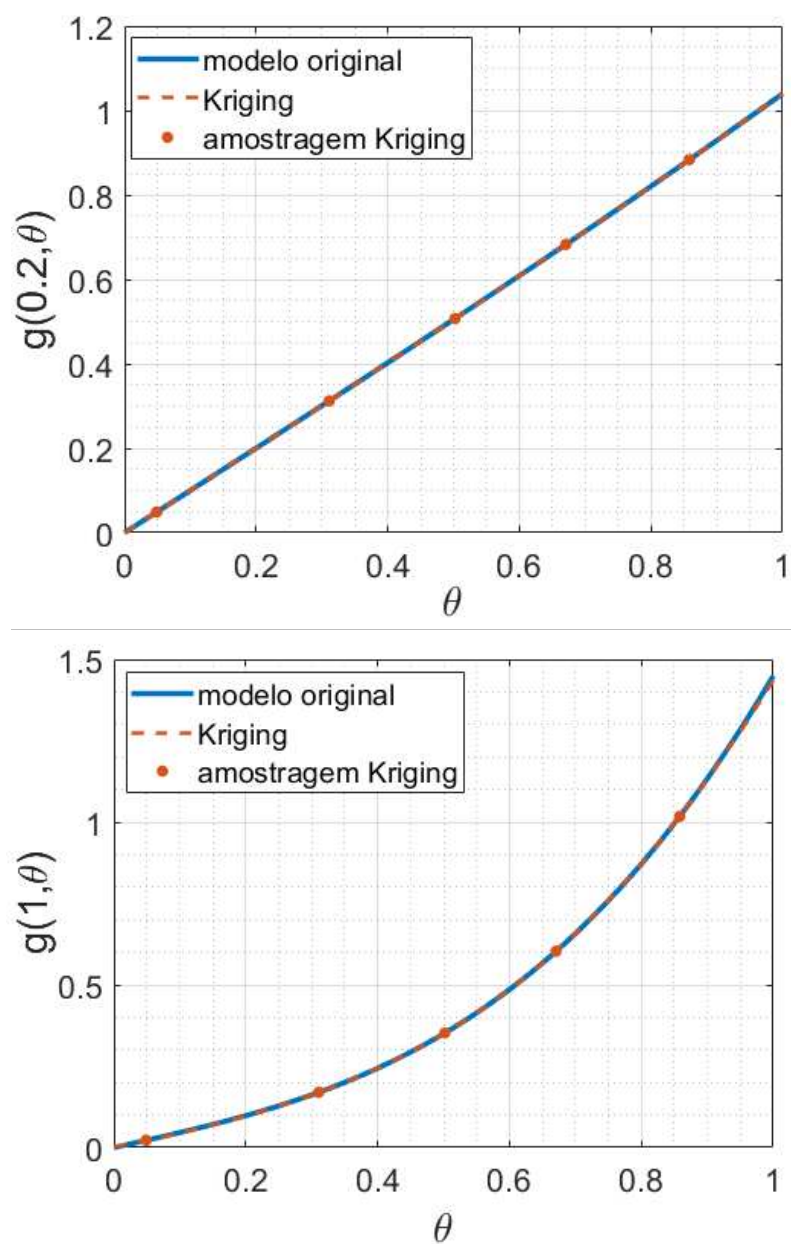
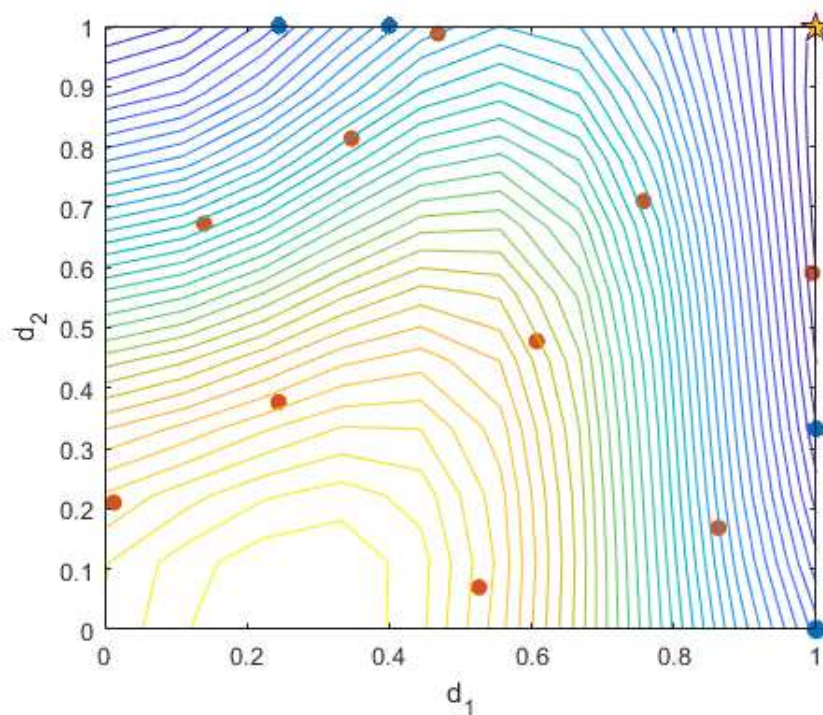


Tabela 1 – Estatísticas das 10 execuções independentes para o problema não linear.

	NCFM	média	std
EGO-DLMC	150×10^3	4.1035	0.0229
EGO-KR	300	4.0969	0.0188
EGO-KR	150	4.0906	0.0226

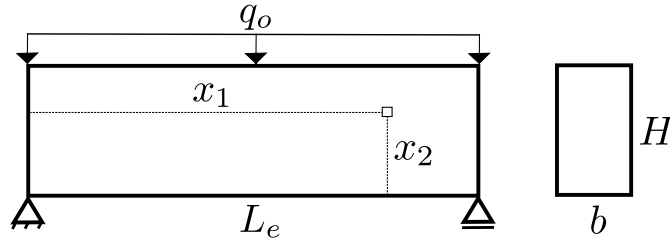
Figura 14 – Estimativa do SEIG para o projeto de dois experimentos pelo EGO-KR: os círculos laranja representam o plano de amostragem inicial, os círculos azuis indicam os pontos de preenchimento e as estrelas representa o mínimo obtido.



utilizado, resultando em $NCFM = 20 \times (10 + 5) = 300$. No segundo cenário, foi empregado $n_k = 10$, totalizando $NCFM = 10 \times (10 + 5) = 150$. Cada algoritmo passou por dez execuções independentes, e as médias e desvios resultantes para o máximo SEIG são apresentadas na Tabela 1.

A partir dos resultados apresentados na Tabela 1, o algoritmo EGO-KR demonstrou uma consistência em seus resultados para os diferentes orçamentos computacionais analisados, mantendo um desvio padrão baixo. Um destaque significativo ocorre no caso de $n_k = 10$, que exigiu apenas $NCFM = 150$, e forneceu resultados promissores. O melhor resultado obtido pelo EGO-KR com $n_k = 10$ ocorreu em $d = [1, 1]$, resultando em $\mathcal{I} = 4.11560$. De acordo com os resultados obtidos pelo EGO-DLMC e apresentados por Huan e Marzouk (2013), o ganho máximo de informação é encontrado nas regiões de $d = [0.2, 1]$, $d = [1, 0.2]$ e $d = [1, 1]$. A Figura 14 ilustra uma execução típica do algoritmo EGO-KR.

Figura 15 – Geometria da viga de Timoshenko e da seção transversal para o problema de caracterização do material.



A análise desses resultados revela que o EGO-KR alcança um valor médio de SEIG muito próximo ao obtido pelo EGO-DLMC, mas com um custo computacional menor. Portanto, o algoritmo EGO-KR identifica com sucesso o ponto de máximo ganho de informação, apresentando precisão satisfatória.

5.2 EXEMPLO 02: POSICIONAMENTO DE EXTENSÔMETRO PARA CARACTERIZAÇÃO DAS PROPRIEDADES MECÂNICAS DA VIGA

O exemplo apresentado por Carlon *et al.* (2023) envolve posicionar um extensômetro na viga de forma a maximizar o ganho de informação sobre os módulos de elasticidade longitudinal e de cisalhamento (E e G). Neste contexto, há duas variáveis de projeto, $n_d = 2$, que representam as posições horizontal e vertical do extensômetro, que serão obtidas fazendo inferência sobre as duas propriedades incertas do material, $n_\theta = 2$. A viga é simplesmente apoiada e submetida a uma carga uniformemente distribuída, conforme ilustrado na Figura 15. Apresenta comprimento de 10 metros, e seção transversal com altura de 2 metros e largura da base de 0,1 metro. O sistema de coordenadas é definido com origem no apoio esquerdo da viga e na metade da altura da seção transversal, $H/2$. Uma carga transversal uniformemente distribuída de 1000 kN/m é aplicada ao longo do eixo longitudinal da viga.

A viga é modelada utilizando a teoria de Timoshenko (TIMOSHENKO, 1921), que é um modelo mecânico que considera as deformações resultantes tanto das tensões normais quanto das tensões de cisalhamento. O modelo é dado por:

$$\begin{cases} K_s G A_r \varepsilon_{12} = \frac{q_o L_e}{2} - q_o x_1, \\ E I_n \varepsilon_{11} = \frac{q_o x_1 (L_e - x_1)}{2} x_2, \end{cases} \quad (79)$$

em que ε_{11} é a deformação normal, ε_{12} é a deformação de cisalhamento, x_1 e x_2 são as posições do extensômetro nos eixos horizontal e vertical, respectivamente, q_o é a carga uniforme, L_e é o comprimento da viga, I_n é o momento de inércia da seção transversal, K_s é a constante de Timoshenko ($K_s = 5/6$ em todos os casos considerados neste trabalho) e A_r é a área da seção transversal.

Tabela 2 – Parâmetros da distribuição a priori para o problema de caracterização das propriedades mecânicas da viga.

Parameters	θ_1 (GPa)	θ_2 (GPa)	σ_θ^1 (GPa)	σ_θ^2 (GPa)	σ_ε^1 (mm)	σ_ε^2 (mm)
Valores	30	11.538	6	2.310	6.25×10^{-4}	1.30×10^{-4}

5.2.1 Formulação Bayesiana

A posição ótima para o extensômetro, que fornece o máximo ganho de informação sobre os parâmetros desconhecidos E e G , é denotada por $\mathbf{d}^* = (x_1^*, x_2^*)$. Portanto, com base na Equação (79), o modelo de Timoshenko é representado como

$$y(\mathbf{d}, \boldsymbol{\theta}) = (\varepsilon_{11}(\mathbf{d}, \boldsymbol{\theta}), \varepsilon_{12}(\mathbf{d}, \boldsymbol{\theta})) + \varepsilon_j \quad i = 1, 2, \dots, N_e$$

$$= \left(\frac{d_2 (q_o L_e d_1 - q_o d_1^2)}{2\theta_1 I_n}, \frac{\frac{L_e}{2} q_o - q_o d_1}{K_s \theta_2 A_r} \right) + \varepsilon_j, \quad (80)$$

em que (x_1, x_2) e (E, G) são substituídos por (d_1, d_2) e (θ_1, θ_2) , respectivamente. A distribuição a priori da QoI é $\pi(\boldsymbol{\theta}) \sim \mathcal{N} \left((\theta_1, \theta_2)^T, \text{diag}[(\sigma_\theta^1)^2, (\sigma_\theta^2)^2] \right)$ e o erro aditivo da medição é $\varepsilon_j \sim \mathcal{N}(0, \Sigma_\varepsilon)$, onde a matriz de covariância é $\Sigma_\varepsilon = \text{diag}\{(\sigma_\varepsilon^1)^2, (\sigma_\varepsilon^2)^2\}$. O valor numérico dessas quantidades usadas neste experimento está detalhado na Tabela 2.

Devido à simetria da estrutura, o espaço de projeto é $d_1 \in [5, 10]$ m e $d_2 \in [-1, 0]$ m. E por último, para este caso, o número de experimentos para cada configuração é $N_e = 3$.

5.2.2 Resultados e discussões

Neste exemplo, será comparado o desempenho do EGO-KR com os algoritmos EGO-PCE (CARLON *et al.*, 2023) e EGO-DLMC. Para os algoritmos EGO-KR e EGO-PCE, o SEIG máximo foi recalculado nos pontos ótimos utilizando DLMC, garantindo uma comparação justa entre os algoritmos. O EGO-DLMC foi exclusivamente empregado para servir de referência, pois apresenta um custo computacional consideravelmente alto. Cada algoritmo foi executado independentemente 10 vezes. Analisamos o EGO-KR e o EGO-PCE fixando o orçamento computacional como NCFM=150. Para todos os algoritmos, o número de pontos amostrais para o laço externo foi fixado em $N = 20000$, e para o laço interno foram utilizadas $M = 2000$ pontos amostrais. Para o EGO-KR, o número de pontos amostrais para construir o modelo do Kriging da etapa 1, $\hat{g}(\boldsymbol{\theta})$, foi analisado em dois cenários, para $n_k = 10$ e $n_k = 15$. Para o modelo Kriging da etapa 2 requerido pelo EGO, $\hat{\mathcal{I}}_{KG}(\mathbf{d})$, o plano de amostragem inicial nos dois cenários foi $n_{is} = 5$, e o critério de parada foi mantido como 5 IPs. O que resulta em um custo

Tabela 3 – Estatísticas das 10 execuções independentes para o problema da viga de caracterização das propriedades mecânicas: os valores entre parênteses indicam o SEIG estimado utilizando DLMC+Kriging no caso do EGO-KR, e utilizando a aproximação PCE+Laplace no caso do EGO-PCE

	NCFM	média	std
EGO-DLMC	4×10^8	2.58 (-)	0.035 (-)
EGO-PCE	162	2.50 (2.34)	0.14 (0.13)
EGO-KR ($n_k = 15$)	150	2.57 (2.56)	0.016 (0.024)
EGO-KR ($n_k = 10$)	100	2.52 (2.61)	0.050 (0.045)

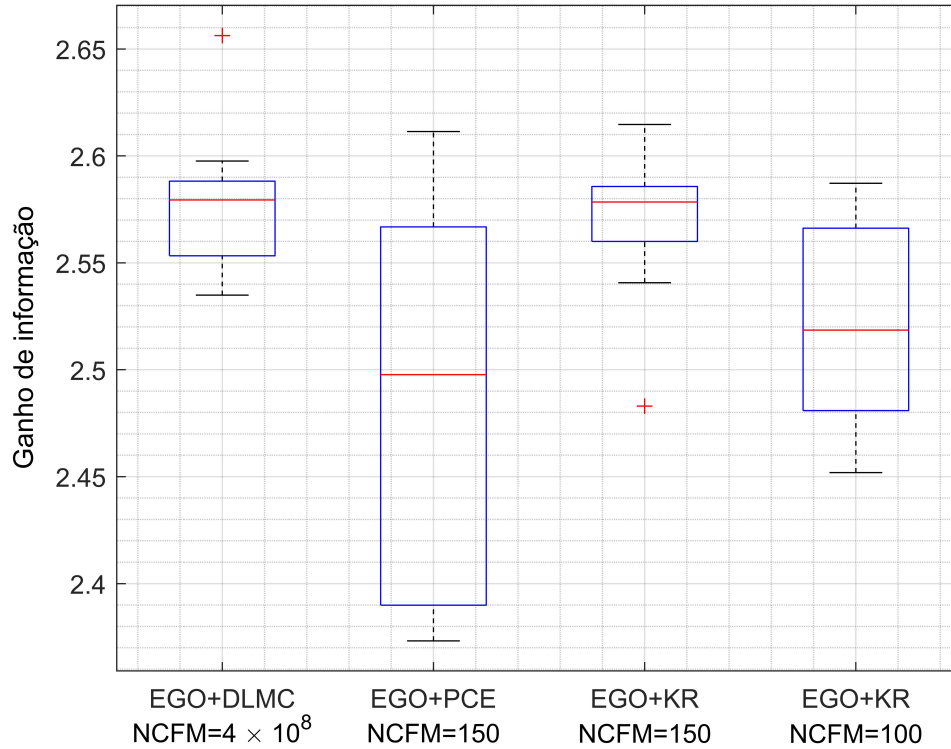
total para cada cenário do EGO+KR de $NCFM=10(5+5)=100$ e $NCFM=15(5+5)=150$, respectivamente. Para o algoritmo EGO-PCE, foi definido $n_{is} = 3$ e o critério de parada igual a 3 IPs, totalizando $NCFM = 162$.

Na Tabela 3, é observado que a média e o desvio padrão do SEIG para EGO-DLMC e EGO-KR foram muito próximos, com a diferença entre eles sendo inferior a 0,5% para $NCFM = 150$ e inferior a 2,3% para $NCFM = 100$. Por outro lado, o EGO-PCE alcança um valor médio quase 3,1% menor do que o valor fornecido pelo EGO-DLMC. Além disso, o desvio padrão dos resultados do EGO-PCE é consideravelmente maior do que os outros métodos.

A Tabela 3 também apresenta, entre parênteses, o valor médio do SEIG obtido diretamente pelo EGO-KR e EGO-PCE, sem recalcular o valor do SEIG nos pontos ótimos. Estes resultados mostram que a qualidade da aproximação aumenta à medida que n_k é aumentado. Por exemplo, a diferença percentual entre o valor do SEIG calculado pelo DLMC usando o modelo original e o Kriging da etapa 1 diminuiu de 3,6%, no caso em que $n_k = 10$ é empregado, para 0,8%, quando $n_k = 15$ foi aplicado. Além disso, a qualidade dos ótimos encontrados também é influenciada pela qualidade do modelo substituto, uma vez que os resultados obtidos pelo EGO-KR com $n_k = 15$ são consideravelmente melhores do que os resultados quando $n_k = 10$ é utilizado. Por exemplo, o valor médio é mais alto e a dispersão é menor nos resultados do EGO-KR quando $n_k = 15$ é empregado, conforme ilustrado na Figura 16.

O melhor projeto experimental alcançado pelo EGO-KR foi em $\mathbf{d} = [7991,7; -1000,0]$ mm, o que resultou em $\mathcal{I} = 2.6147$. Considerando que o módulo de elasticidade longitudinal é influenciado pelo momento fletor (com o valor máximo ocorrendo no meio do vão) e o módulo de cisalhamento é afetado pelo esforço cortante (com o valor máximo ocorrendo nos apoios), a sugestão resultante de OED para posicionar o sensor entre o meio do vão e o apoio direito é consistente. Essa recomendação implica uma compensação no ganho de informação relacionado aos módulos de elasticidade longitudinal e cisalhamento.

Figura 16 – Boxplots dos valores ótimos obtidos em 10 execuções independentes para o problema de caracterização de propriedades mecânicas da viga.

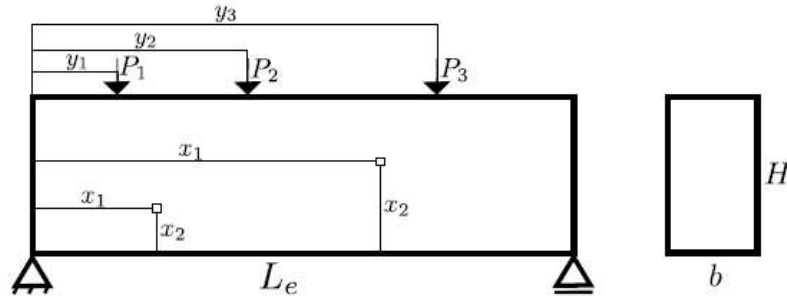


5.3 EXEMPLO 3: POSICIONAMENTO DE EXTENSÔMETROS PARA CARACTERIZAÇÃO DE CARGAS EM VIGA.

Neste exemplo, apresentado em Carlon *et al.* (2023), o problema OED envolve a localização de dois extensômetros na viga de forma a maximizar o ganho de informação sobre a magnitude e a posição de três cargas pontuais verticais. Neste cenário, tanto a magnitude quanto o posicionamento das cargas são incertos, resultando em seis variáveis de interesse ($n_{\theta} = 6$). Este problema OED envolve quatro variáveis de projeto ($n_{\mathbf{d}} = 4$), que são as posições horizontal e vertical dos dois sensores de extensômetro. Isso implica que o experimento é projetado para fazer quatro medições de deformação para inferir seis parâmetros desconhecidos, θ , ou seja, determinar quatro variáveis de projeto, $\mathbf{d}$, através de um processo de otimização destinado a maximizar a informação sobre θ . Assim, a representação do modelo direto de seis dimensões é feita por meio do metamodelo de Kriging. Esta abordagem minimizará o custo associado às chamadas do modelo durante o cálculo do SEIG.

A viga é simplesmente apoiada e sujeita a três cargas pontuais verticais com magnitudes P_1 , P_2 e P_3 , juntamente com posições y_1 , y_2 e y_3 , conforme ilustrado na Figura 17. Suas dimensões são as seguintes: um comprimento de 10 metros, e a seção transversal tem uma altura de 20 cm e uma largura de base de 10 cm. O sistema

Figura 17 – Geometria da viga de Timoshenko e seção transversal para o problema de caracterização das cargas verticais.



de coordenadas de referência é definido com origem no apoio esquerdo da viga e na base da seção transversal. Assume-se que o material da viga é isotrópico e elástico linear, com módulo de elasticidade longitudinal $E = 200$ GPa e coeficiente de Poisson 0.3.

A viga é modelada utilizando a teoria de Timoshenko (TIMOSHENKO, 1921), pela qual podemos aproximar as deformações normal e de cisalhamento em qualquer ponto x_1, x_2 do domínio por:

$$\begin{cases} K_s G A_r \varepsilon_{12} = Q(x_1), \\ E I_n \varepsilon_{11} = M(x_1) x_2, \end{cases} \quad (81)$$

onde ε_{11} é a deformação normal, ε_{12} é a deformação de cisalhamento, x_1 e x_2 são as posições do extensômetro nos eixos horizontal e vertical, respectivamente, K_s é a constante de Timoshenko ($K_s = 5/6$ em todos os casos de teste), G e E são o módulo de elasticidade longitudinal e o módulo de cisalhamento da viga, A_r é a área da seção transversal, I_n é o momento de inércia da seção transversal, Q é o esforço cortante e M é o momento fletor.

O esforço cortante e o momento fletor na viga, em um ponto com uma distância x_1 para o apoio esquerdo, são dados na Equação (82) e Equação (83), respectivamente.

$$Q(x_1) = \frac{\sum_{i=1}^3 P_i (L_e - y_i)}{L} - \sum_{i=1}^3 P_i \mathbf{1}_{y_i < x_1}, \quad (82)$$

$$M(x_1) = \frac{\sum_{i=1}^3 P_i (L_e - y_i)}{L} x_1 - \sum_{i=1}^3 P_i (x_1 - y_i) \mathbf{1}_{y_i < x_1} \quad (83)$$

onde P_i são as magnitudes das cargas pontuais verticais, y_i são as posições das P_i , L_e é o comprimento da viga e $\mathbf{1}_{(.)}$, especificamente nesta equação, representa a função indicadora.

Tabela 4 – Parâmetros da priori para o problema de caracterização de cargas na viga.

	Carga 1		Carga 2		Carga 3	
	P_1 (kN)	y_1 (mm)	P_2 (kN)	y_2 (mm)	P_3 (kN)	y_3 (mm)
μ_θ	4000	1000	6000	3500	1000	7000
σ_θ	1000	100	500	500	300	800

Tabela 5 – Parâmetros do ruído aditivo para o problema da viga de caracterização da carga.

Parâmetros	σ_ϵ^1 (mm)	σ_ϵ^2 (mm)	σ_ϵ^3 (mm)	σ_ϵ^4 (mm)
Valores	6.25×10^{-4}	1.3×10^{-4}	6.25×10^{-4}	1.3×10^{-4}

5.3.1 Formulação Bayesiana

Vamos representar os parâmetros desconhecidos como $\theta = [P_i, y_i]$, onde $i = 1, 2, 3$ corresponde às cargas pontuais verticais. Esses parâmetros são modelados como uma distribuição prévia Gaussiana $\pi(\theta) \sim \mathcal{N}(\mu_\theta, \text{diag}[(\sigma_\theta)^2])$, conforme especificado na Tabela 4.

O erro aditivo da medição é dado por $\epsilon_i \sim \mathcal{N}(0, \Sigma_\epsilon)$, onde a matriz de covariância é $\Sigma_\epsilon = \text{diag}[(\sigma_\epsilon^1)^2, (\sigma_\epsilon^2)^2, (\sigma_\epsilon^3)^2, (\sigma_\epsilon^4)^2]$, conforme fornecido na Tabela 5.

A posição ótima para cada extensômetro é denotada como x^j , com $j = 1, 2$. O vetor de projeto é $\mathbf{d} = (x_1^1, x_2^1, x_1^2, x_2^2)$, e o espaço de projeto é definido da seguinte forma: $x_1^1 \in [0, 10000]$ mm, $x_2^1 \in [0, 100]$ mm, $x_1^2 \in [0, 10000]$ mm e $x_2^2 \in [0, 100]$ mm. Neste caso, o número de experimentos para cada configuração é $N_e = 1$.

5.3.2 Resultados e discussões

Neste exemplo, comparamos o desempenho do EGO-KR com o EGO-BCS (CARLON *et al.*, 2023). O método EGO-BCS utiliza o PCE, avaliado usando a técnica de Sensoriamento Compressivo Bayesiano (BCS) (MARELLI; SUDRET, 2015). Cada algoritmo foi executado independentemente 10 vezes. Posteriormente, avaliamos a precisão dos resultados do EGO-KR e EGO-BCS recalculando o SEIG nos dez pontos ótimos, desta vez utilizando o método DLMLC, e obtendo a média e o desvio padrão do ganho máximo de informação.

No cálculo do SEIG usando DLMLC, o número de pontos amostrais foi fixado como $N = 20000$ e $M = 4000$. Para a construção do modelo substituto do Kriging da etapa 1, $\hat{g}(\theta, \mathbf{d})$, foram analisados dois cenários, para $n_k = 30$ e $n_k = 40$. No processo de maximização do SEIG usando EGO-KR, foram utilizados $n_{is} = 10$ pontos para a construção do modelo de substituição do Kriging da etapa 2, juntamente com 5 IPs no processo de otimização, resultando em um orçamento computacional total

de $NCFM = 30 \times (10 + 5) = 450$ e $NCFM = 40 \times (10 + 5) = 600$, respectivamente. Para a abordagem EGO-BCS, foram empregados $n_{IS} = 8$ para o plano de amostragem inicial em todos os casos, com 12 IPs para o BCS de primeira ordem e 3 IPs para o BCS de segunda ordem, resultando em $NCFM = 1400$ e 1925 , respectivamente. Os orçamentos computacionais alocados para PCE-BCS, embora consideravelmente maiores do que os do EGO-KR, foram determinados como o mínimo necessário para alcançar resultados razoáveis.

As estatísticas das 10 execuções independentes de cada algoritmo são apresentadas na Tabela 6. Comparando os resultados dos três algoritmos de otimização, observamos que o método EGO-KR alcançou o maior valor médio de SEIG, com um número significativamente menor de avaliações do modelo direto em comparação com ambas as abordagens EGO-BCS. De fato, EGO-BCE de 1ª ordem superou o EGO-BCS de 2ª ordem exigindo menos de 40% do orçamento computacional. O desempenho superior da abordagem proposta é ainda mais evidente ao analisar os diagramas de caixa dos ótimos encontrados, como mostrado na Figura 18. Esses resultados destacam a eficiência computacional superior do EGO-KR, pois alcançou melhores resultados com menos avaliações do modelo direto. Vale ressaltar que o EGO-BCS não utiliza uma abordagem PCE comum; em vez disso, emprega uma abordagem de regressão eficiente, indicando que seu custo computacional não poderia ser reduzido ainda mais usando técnicas mais avançadas. Isso enfatiza a eficiência da abordagem proposta para resolver este problema de OED. É importante observar que a qualidade dos ótimos encontrados é influenciada pela precisão do modelo substituto. Isso é evidenciado pelos resultados obtidos utilizando o EGO-KR com $n_k = 40$, que foram notavelmente melhores do que os resultados com $n_k = 30$. Por exemplo, o valor médio foi mais alto e os resultados apresentaram menor dispersão quando $n_k = 40$ foi utilizado.

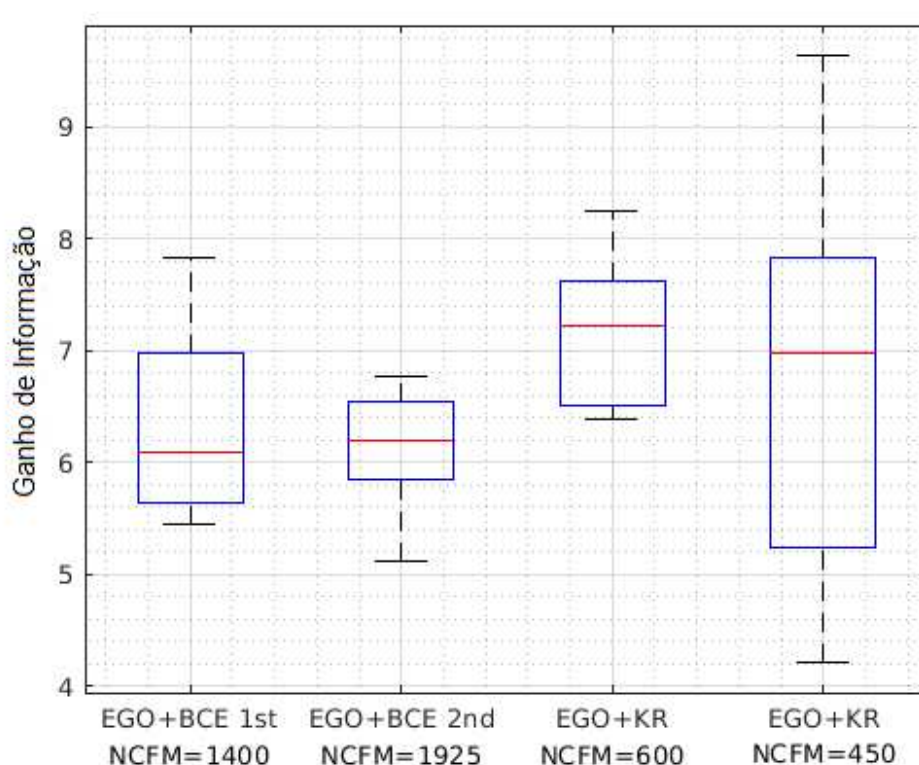
A Tabela 6 também apresenta os valores de SEIG estimados por DLMC + Kriging no caso do EGO-KR, e pela aproximação PCE + Laplace no caso do EGO-PCE. Pode-se observar que, embora a diferença desses estimadores em comparação com a solução de referência (DLMC usando o modelo original) seja maior, os algoritmos de otimização conseguem buscar pela solução ótima.

O melhor ponto encontrado pelo algoritmo EGO-KR para $n_k = 30$ foi $\mathbf{d} = [2501,7; 100; 5067,3; 100]$ mm, resultando em $\mathcal{I} = 8.2545$. Considerando que as forças aplicadas à viga estão posicionadas próximas às coordenadas $y = [1000, 3500, 7000]$ mm, pode-se inferir que o ganho máximo de informação será alcançado posicionando os dois sensores entre as cargas aplicadas e no topo da viga. Portanto, o projeto ótimo para este problema de OED, com o objetivo de localizar os dois sensores, requer que o primeiro sensor seja colocado mais à esquerda (em 2501 mm), enquanto o segundo sensor deve ser posicionado entre as cargas centrais e da direita (em 5067 mm).

Tabela 6 – Estatísticas para o exemplo caracterização da carga da viga: os valores entre parênteses indicam o SEIG estimado utilizando DLMC+Kriging no caso do EGO-KR, e utilizando a aproximação PCE+Laplace no caso do EGO-PCE.

	NCFM	média	std
EGO-BCS (1ª ordem)	1400	6.37 (5.75)	0.87 (0.60)
EGO-BCS (2ª ordem)	1925	6.13 (5.45)	0.47 (0.39)
EGO-KR	600	7.05 (8.32)	0.96 (1.02)
EGO-KR	450	6.68 (8.07)	1.65 (1.24)

Figura 18 – Boxplots dos ótimos obtidos para o problema de caracterização de carga da viga.



6 CONCLUSÃO

Nesta tese foi proposta uma metodologia de dupla abordagem do Kriging para resolver problemas de projeto experimental ótimo Bayesiano, denominada EGO-KR. O SEIG foi empregado como medida de informação. Para aumentar a eficiência computacional na avaliação do SEIG, foi introduzido uma superfície preditora do Kriging como substituto do modelo direto original (Kriging da etapa 1) e empregado o método de Monte Carlo de laço duplo para aproximar o ganho de informação utilizando o metamodelo. No processo de maximização do ganho de informação, foi utilizado o algoritmo EGO, que requer a construção de um modelo substituto do Kriging para o SEIG (Kriging da etapa 2). A otimização através do EGO empregou o critério de preenchimento EI para guiar a busca pela solução global do problema OED.

O método foi aplicado e avaliado em três problemas de OED com ruído nas observações e parâmetros estocásticos. O primeiro caso consistiu em um problema não linear com dois parâmetros incertos e duas variáveis de projeto. Os dois problemas subsequentes foram baseados em vigas, utilizando o modelo de Timoshenko. No segundo problema, o objetivo foi obter a posição horizontal e vertical de um extensômetro de forma a maximizar o ganho de informação de duas propriedades desconhecidas dos materiais da estrutura, resultando em dois parâmetros incertos e duas variáveis de projeto. O terceiro problema, consistiu em determinar a posição vertical e horizontal de dois extensômetros dada a intensidade e posição incertas de três cargas pontuais atuantes na estrutura, resultando em seis parâmetros aleatórios e quatro variáveis de projeto.

O custo computacional foi avaliado com base no NCFM, enquanto a precisão e eficácia, através da comparação dos resultados com a abordagem do DLMLC sem o uso do preditor como substituto do modelo e a abordagem EGO-PCE e EGO-BCE (CARLON *et al.*, 2023), que empregam a aproximação de Laplace.

Os resultados indicam que a abordagem proposta EGO-KR resolveu efetivamente os problemas analisados, com boa precisão e fornecendo economias computacionais significativas em comparação com o EGO usando o modelo direto original ou o método EGO-PCE, representando um método competitivo em problemas de OED Bayesiano. No terceiro e mais complexo exemplo, foi observada uma redução significativa no NCFM, obtendo resultados com precisão superior.

Em resumo, a estrutura proposta permite lidar com problemas computacionalmente caros sem comprometer a precisão. Também tem o potencial de lidar com casos de distribuições posteriores assimétricas ou multimodais e funções não convexas, já que o EGO-KR não depende da aproximação de Laplace.

6.1 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestão para o aprimoramento de métodos de solução de problemas OED bayesiano, propõe-se para trabalhos futuros:

- Aperfeiçoar o método de dupla abordagem do Kriging explorando o aprendizado ativo para otimizar amostras no Kriging da etapa 1. Os resultados dessa pesquisa evidenciam que a representação do modelo impacta diretamente na precisão no cálculo do SEIG. Desta forma, sugere-se o aprimoramento dessa etapa através do projeto de experimentos (DoE), ou seja, utilizando uma seleção inteligente de entrada de projeto experimental;
- Determinar automaticamente o tamanho das amostras do Kriging da etapa 1. Aproveitar o potencial do aprendizado ativo para realizar esse processo de forma adaptativa e eficiente;
- Investigar a eficiência de outros metamodelo, com as redes neurais artificiais, para a representação do modelo;
- Explorar uma abordagem alternativa à dupla abordagem do Kriging, considerando tanto os parâmetros incertos quanto as variáveis de interesse na construção do metamodelo durante o processo de maximização do SEIG;

REFERÊNCIAS

- ANKENMAN, Bruce; NELSON, Barry L; STAUM, Jeremy. Stochastic kriging for simulation metamodeling. **Operations research**, INFORMS, v. 58, n. 2, p. 371–382, 2010.
- BARTUSKA, Arved; ESPATH, Luis; TEMPONE, Raúl. Small-noise approximation for Bayesian optimal experimental design with nuisance uncertainty. **Computer Methods in Applied Mechanics and Engineering**, v. 399, p. 115320, 2022.
- BECK, Joakim *et al.* Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. **Computer Methods in Applied Mechanics and Engineering**, v. 334, p. 523–553, 2018.
- BECK, Joakim *et al.* Multilevel double loop Monte Carlo and stochastic collocation methods with importance sampling for Bayesian optimal experimental design. **International Journal for Numerical Methods in Engineering**, v. 121, n. 15, p. 3482–3503, 2020.
- BEN ISSAID, Chaouki. **Bayesian Optimal Experimental Design Using Multilevel Monte Carlo**. 2015. Tese (Doutorado) – King Abdullah University of Science e Technology.
- CARLON, André Gustavo. **Bayesian-based optimization of experiments using stochastic gradient methods**. 2019. Tese (Doutorado) – Federal university of Florianopolis.
- CARLON, André Gustavo *et al.* A polynomial chaos efficient global optimization approach for Bayesian optimal experimental design. **Probabilistic Engineering Mechanics**, v. 72, p. 103454, 2023.
- CARLON, André Gustavo *et al.* Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. **Computer Methods in Applied Mechanics and Engineering**, v. 363, p. 112909, 2020.
- CARLON, André Gustavo *et al.* Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. **Computer Methods in Applied Mechanics and Engineering**, v. 363, p. 112909, 2020.

- CARRARO, Felipe *et al.* Monte Carlo integration with adaptive variance selection for improved stochastic efficient global optimization. **Structural and Multidisciplinary Optimization**, v. 60, p. 245–268, 2019.
- CHALONER, Kathryn; VERDINELLI, Isabella. Bayesian experimental design: A review. **Statistical Science**, JSTOR, p. 273–304, 1995.
- COSTAS, Argyris; COSTAS, Papadimitriou; PANAGIOTIS, Panetsos. Bayesian optimal sensor placement for modal identification of civil infrastructures. **J. Smart Cities**, v. 2, n. 2, p. 69–86, 2017.
- ENGLEZOU, Yiolanda; WAITE, Timothy W; WOODS, David C. Approximate Laplace importance sampling for the estimation of expected Shannon information gain in high-dimensional Bayesian design for nonlinear models. **Statistics and Computing**, v. 32, n. 5, p. 82, 2022.
- FORRESTER, Alexander; KEANE, Andy; BRESSLOFF, Neil. Design and analysis of "Noisy" computer experiments. **AIAA journal**, v. 44, n. 10, p. 2331–2339, 2006.
- FORRESTER, Alexander I.J.; SOBESTER, Andras; KEANE, Andy J. **Engineering design via surrogate modelling**. [S.l.: s.n.], 2008.
- FORRESTER, Alexander IJ; KEANE, Andy J. Recent advances in surrogate-based optimization. **Progress in aerospace sciences**, v. 45, n. 1-3, p. 50–79, 2009.
- FOSTER, Adam *et al.* Variational Bayesian Optimal Experimental Design. In _____. **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2019. v. 32.
- GALLY, Tristan *et al.* Identification of model uncertainty via optimal design of experiments applied to a mechanical press. **Optimization and Engineering**, p. 1–28, 2021.
- GANESAN, N; RAJ, J Bharati; SHASHIKALA, AP. Flexural fatigue behavior of self compacting rubberized concrete. **Construction and Building Materials**, v. 44, p. 7–14, 2013.

GODA, Takashi *et al.* Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. **SIAM Journal on Scientific Computing**, SIAM, v. 44, n. 1, a286–a311, 2022.

GU, Jichao. An efficient multiple meta-model-based global optimization method for computationally intensive problems. **Advances in Engineering Software**, v. 152, p. 102958, 2021.

HAMADA, M *et al.* Finding near-optimal Bayesian experimental designs via genetic algorithms. **The American Statistician**, v. 55, n. 3, p. 175–181, 2001.

HOFF, Peter D. **A first course in Bayesian statistical methods**. [S.l.]: Springer, 2009. v. 580.

HUAN, Xun; MARZOUK. Gradient-based stochastic optimization methods in Bayesian experimental design. **International Journal for Uncertainty Quantification**, Begel House Inc., v. 4, n. 6, p. 1–41, 2014.

_____. Simulation-based optimal Bayesian experimental design for nonlinear systems. **Journal of Computational Physics**, v. 232, n. 1, p. 288–317, 2013.

HUANG, Deng *et al.* Global optimization of stochastic black-box systems via sequential kriging meta-models. **Journal of global optimization**, v. 34, n. 3, p. 441–466, 2006.

JALALI, Hamed; VAN NIEUWENHUYSE, Inneke; PICHENY, Victor. Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. **European Journal of Operational Research**, v. 261, n. 1, p. 279–301, 2017.

JONES, Donald R. A taxonomy of global optimization methods based on response surfaces. **Journal of global optimization**, v. 21, n. 4, p. 345–383, 2001.

JONES, Donald R.; SCHONLAU, Matthias; WELCH, William J. Efficient global optimization of expensive black-box functions. **Journal of Global optimization**, v. 13, n. 4, p. 455–492, 1998.

KAPIO, Jari; SOMERSALO, Erkki. **Statistical and computational inverse problems**. [S.l.]: Springer Science & Business Media, 2006. v. 160.

KLEINEGESSE, Steven; GUTMANN, Michael U. Efficient Bayesian Experimental Design for Implicit Models. *In*: CHAUDHURI, Kamalika; SUGIYAMA, Masashi (Ed.). **Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics**. [S.l.]: PMLR, 2019. v. 89. (Proceedings of Machine Learning Research), p. 476–485.

KRIGE, Daniel G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Southern African Institute of Mining and Metallurgy**, Southern African Institute of Mining e Metallurgy, v. 52, n. 6, p. 119–139, 1951.

KULLBACK, Solomon. **Information theory and statistics**. [S.l.]: Courier Corporation, 1997.

KULLBACK, Solomon; LEIBLER, Richard A. On information and sufficiency. **The annals of mathematical statistics**, v. 22, n. 1, p. 79–86, 1951.

LATANIOTIS, Christos; MARELLI, Stefano; SUDRET, Bruno. The Gaussian process modelling module in UQLab. **arXiv:1709.09382**, 2017.

LONG, Quan *et al.* Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. **Computer Methods in Applied Mechanics and Engineering**, v. 259, p. 24–39, 2013.

MARELLI, Stefano; SUDRET, Bruno. UQLab user manual - Polynomial chaos expansions. **Chair of risk, safety & uncertainty quantification**, ETH Zürich, v. 2, p. 97–110, 2015.

MATHWORKS. **I. T. MatLab**. [S.l.: s.n.], 2021.

MIGUEL, Leandro F Fadel *et al.* Reliability-based optimization of multiple Folded Pendulum TMDs through Efficient Global Optimization. **Engineering Structures**, v. 266, p. 114524, 2022.

NASCENTES, Fábio Felipe dos Santos *et al.* **Contribuições à eficiência da otimização global estocástica adaptativa**. 2019. Dissertação – Universidade Federal de Santa Catarina, Florianópolis, Brasil.

NESTEROV, Yurii. **Introductory lectures on convex optimization: A basic course**. [S.l.]: Springer Science & Business Media, 2013. v. 87.

ORTEGA, José J *et al.* Number of tests and corresponding error in concrete fatigue. **International journal of fatigue**, v. 116, p. 210–219, 2018.

OVERSTALL, Antony M; WOODS, David C. Bayesian design of experiments using approximate coordinate exchange. **Technometrics**, v. 59, n. 4, p. 458–470, 2017.

PAPADIMITRIOU, C; ARGYRIS, C. Bayesian optimal experimental design for parameter estimation and response predictions in complex dynamical systems. **Procedia engineering**, v. 199, p. 972–977, 2017.

PAPADIMITRIOU, C; HARALAMPIDIS, Y; SOBCZYK, K. Optimal experimental design in stochastic structural dynamics. **Probabilistic engineering mechanics**, v. 20, n. 1, p. 67–78, 2005.

PICHENY, Victor; WAGNER, Tobias; GINSBOURGER, David. A benchmark of kriging-based infill criteria for noisy optimization. **Structural and multidisciplinary optimization**, v. 48, n. 3, p. 607–626, 2013.

RASMUSSEN, Carl Edward. **Gaussian processes in machine learning**. [S.l.]: Springer, 2003. P. 63–71.

RIVAS-LÓPEZ, MJ *et al.* Optimal experimental design on the loading frequency for a probabilistic fatigue model for plain and fibre-reinforced concrete. **Computational Statistics & Data Analysis**, v. 113, p. 363–374, 2017.

ROBBINS, Herbert; MONRO, Sutton. A stochastic approximation method. **The annals of mathematical statistics**, p. 400–407, 1951.

ROBERT, Christian P; CASELLA, George. **Monte Carlo statistical methods**. [S.l.]: Springer, 2010. v. 2.

RUBINSTEIN, Reuven Y; KROESE, Dirk P. **Simulation and the Monte Carlo method**. [S.l.]: John Wiley & Sons, 2016.

RYAN, Kenneth J. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. **Journal of Computational and Graphical Statistics**, v. 12, n. 3, p. 585–603, 2003.

SACKS, Jerome *et al.* Design and analysis of computer experiments. **Statistical science**, v. 4, n. 4, p. 409–423, 1989.

SANTNER, Thomas J *et al.* **The design and analysis of computer experiments**. [S.l.]: Springer, 2003. v. 1.

SHANNON, Claude E. A mathematical theory of communication. **Bell Syst. Tech. J.**, v. 27, p. 623–656, 1948.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. Practical Bayesian Optimization of Machine Learning Algorithms. *In* _____. **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2012. v. 25.

SOBCZYK, K; HOŁOBUT, P. Information-theoretic approach to dynamics of stochastic systems. **Probabilistic engineering mechanics**, v. 27, n. 1, p. 47–56, 2012.

SPALL, James C. A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. *In*: IEEE. PROCEEDINGS of the 27th IEEE Conference on Decision and Control. [S.l.: s.n.], 1988. P. 1544–1548.

SUDRET, Bruno; DER KIUREGHIAN, Armen. **Stochastic finite element methods and reliability: a state-of-the-art report**. [S.l.]: Department of Civil e Environmental Engineering, University of California, 2000.

TIMOSHENKO, Stephen P. On the correction for shear of the differential equation for transverse vibrations of prismatic bars. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 41, n. 245, p. 744–746, 1921.

VAN BEERS, Wim CM; KLEIJNEN, Jack PC. Kriging for interpolation in random simulation. **Journal of the Operational Research Society**, v. 54, n. 3, p. 255–262, 2003.

WILLIAMS, Christopher KI; RASMUSSEN, Carl Edward. **Gaussian processes for machine learning**. [S.l.]: MIT press Cambridge, MA, 2006. v. 2.