

Uma estratégia para minimização de funções com termos modulares via métodos de pontos interiores

Diego Nunes da Silva

Programa de Pós-Graduação em Engenharia Elétrica

FEB, UNESP, Bauru, SP

E-mail: diegoitapeva996@hotmail.com

Antonio Roberto Balbo

Departamento de Matemática

FC, UNESP, Bauru, SP

E-mail: arbalbo@fc.unesp.br

Ricardo Bento Nogueira Pinheiro

Programa de Pós-Graduação em Engenharia Elétrica

EESC, USP, São Carlos, SP

E-mail: ribenopi@hotmail.com

Resumo: Neste trabalho, apresentamos um método de otimização determinístico híbrido que vincula um método previsor-corretor primal-dual de pontos interiores, o qual emprega a função barreira logarítmica modificada, com uma estratégia baseada na função tangente hiperbólica para minimização de funções objetivo com termos modulares. Para garantir a convergência global do método proposto, uma estratégia variante de Levenberg-Marquardt é inserida no método proposto, a fim de ajustar a matriz dual normal da função lagrangiana barreira modificada, em situações em que a mesma não é definida positiva. Por fim, resultados iniciais obtidos através da abordagem explorada são apresentados.

Palavras-chave: Otimização, Pontos Interiores, Função Barreira Modificada.

1 Introdução e Motivações

A não-diferenciabilidade de funções é um tipo de dificuldade que pode surgir ao resolver problemas de otimização, principalmente quando se utilizam métodos clássicos, os quais empregam informações de vetores gradientes e matrizes hessianas para determinar as direções de busca. Diante desta possibilidade, diversos autores optam por utilizar métodos de otimização estocástica/heurísticas para sua resolução, como algoritmos genéticos ou *simulated annealing* [6]. Tais métodos têm se mostrado bastante eficientes na resolução deste tipo de problema, porém não possuem garantias de convergência.

Diversos problemas reais podem ser modelados como problemas de otimização com função objetivo não-diferenciável. Em particular, em engenharia elétrica, um problema de grande relevância é o Problema de Despacho Econômico com Ponto de Válvula (PDE-PV). O PDE-PV é um problema de otimização em que a função objetivo é não-diferenciável devido à presença de termos modulares, cuja finalidade é alocar a demanda de energia entre as unidades geradoras, ao mesmo tempo em que as restrições operacionais são satisfeitas.

Neste sentido, o objetivo deste trabalho é apresentar uma estratégia para aproximar funções com termos modulares, de modo que a função aproximante seja diferenciável. Isto possibilitará a aplicação de um método clássico de otimização, a saber, o método primal-dual previsor-corretor barreira logarítmica modificada, o qual foi utilizado com sucesso por Sousa et al. [5] e Pinheiro [3] na resolução de problemas de fluxo de potência ótimo, a fim de obter soluções consistentes para um problema de otimização com função objetivo não diferenciável devido a termos modulares.

2 Método Primal-Dual Previsor-Corretor Barreira Logarítmica Modificada

Seja o seguinte modelo de otimização não-linear, em que a função objetivo e as restrições são funções de classe C^2 .

$$\begin{aligned} & \text{Minimizar} && f(\mathbf{x}) \\ & \text{sujeito a} && \mathbf{g}(\mathbf{x}) = \mathbf{0} \\ & && \mathbf{u}_1 \leq \mathbf{h}(\mathbf{x}) \leq \mathbf{u}_2 \\ & && \mathbf{l}_1 \leq \mathbf{x} \leq \mathbf{l}_2 \end{aligned} \quad (1)$$

onde $f(\mathbf{x})$ é a função objetivo, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^r$. O problema (1) é, então, convertido em um problema irrestrito, através da função lagrangiana barreira logarítmica modificada, proposta por Polyak [4], expressa em (2):

$$\begin{aligned} L(\boldsymbol{\omega}) = & f(\mathbf{x}) - \mu \sum_{i=1}^r [(\delta_1)_i \ln (z_1^\diamond)_i + (\delta_2)_i \ln (z_2^\diamond)_i] - \mu \sum_{j=1}^n [(\delta_3)_j \ln (z_3^\diamond)_j + (\delta_4)_j \ln (z_4^\diamond)_j] \\ & + \sum_{t=1}^m (\lambda_0)_t g_t(\mathbf{x}) + \sum_{i=1}^r (\lambda_1)_i [-h_i(\mathbf{x}) + (\mathbf{u}_1)_i + (z_1)_i] + \sum_{i=1}^r (\lambda_2)_i [h_i(\mathbf{x}) - (\mathbf{u}_2)_i + (z_2)_i] \\ & + \sum_{j=1}^n (\lambda_3)_j [-x_j + (\mathbf{l}_1)_j + (z_3)_j] + \sum_{j=1}^n (\lambda_4)_j [x_j - (\mathbf{l}_2)_j + (z_4)_j] \end{aligned} \quad (2)$$

em que $\boldsymbol{\omega} = (\mathbf{x}, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4)^t$, μ é o parâmetro de barreira, $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \in \mathbb{R}^r$, $\boldsymbol{\delta}_3, \boldsymbol{\delta}_4 \in \mathbb{R}^n$ são denominados *estimadores dos multiplicadores de Lagrange* relativos às restrições de desigualdade, $\boldsymbol{\lambda}_0 \in \mathbb{R}^m$ é o vetor de multiplicadores de Lagrange das restrições de igualdade, $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}^r$, $\boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4 \in \mathbb{R}^n$ são os vetores de multiplicadores de Lagrange das restrições de desigualdade, $(z_1^\diamond)_i = \frac{\mu + (z_1)_i}{\mu}$, $(z_2^\diamond)_i = \frac{\mu + (z_2)_i}{\mu}$, $(z_3^\diamond)_j = \frac{\mu + (z_3)_j}{\mu}$ e $(z_4^\diamond)_j = \frac{\mu + (z_4)_j}{\mu}$, onde $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^r$, $\mathbf{z}_3, \mathbf{z}_4 \in \mathbb{R}^n$ são as variáveis de folga, as quais foram relaxadas através da introdução da função barreira modificada, de modo que $\mathbf{z}_1 > -\mu \mathbf{e}_r$, $\mathbf{z}_2 > -\mu \mathbf{e}_r$, $\mathbf{z}_3 > -\mu \mathbf{e}_n$ e $\mathbf{z}_4 > -\mu \mathbf{e}_n$, com $\mathbf{e}_r = (1, 1, \dots, 1)^t \in \mathbb{R}^r$ e $\mathbf{e}_n = (1, 1, \dots, 1)^t \in \mathbb{R}^n$.

Sobre (2) são aplicadas as condições necessárias de KKT, impondo que $\nabla L(\boldsymbol{\omega}) = \mathbf{0}$. O sistema não-linear obtido é, então, linearizado utilizando um aproximante de Taylor de primeira ordem. O sistema linear obtido é da forma:

$$A_k \mathbf{d}_{\boldsymbol{\omega}}^k = \mathbf{b}^k \quad (3)$$

em que

$$A_k = \begin{pmatrix} K & 0 & 0 & 0 & 0 & \mathbf{Jg}(\mathbf{x}^k)^t & -\mathbf{Jh}(\mathbf{x}^k)^t & \mathbf{Jh}(\mathbf{x}^k)^t & -I_n & I_n \\ \mathbf{Jg}(\mathbf{x}^k) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\mathbf{Jh}(\mathbf{x}^k) & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{Jh}(\mathbf{x}^k) & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -I_n & 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ I_n & 0 & 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 \\ 0 & \Lambda_{1_k} & 0 & 0 & 0 & 0 & \bar{Z}_{1_k} & 0 & 0 & 0 \\ 0 & 0 & \Lambda_{2_k} & 0 & 0 & 0 & 0 & \bar{Z}_{2_k} & 0 & 0 \\ 0 & 0 & 0 & \Lambda_{3_k} & 0 & 0 & 0 & 0 & \bar{Z}_{3_k} & 0 \\ 0 & 0 & 0 & 0 & \Lambda_{4_k} & 0 & 0 & 0 & 0 & \bar{Z}_{4_k} \end{pmatrix} \quad (4)$$

$$\begin{aligned} \bar{Z}_1^{-1} &= \begin{pmatrix} \frac{1}{\mu + (z_1)_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\mu + (z_1)_r} \end{pmatrix}, \bar{Z}_2^{-1} = \begin{pmatrix} \frac{1}{\mu + (z_2)_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\mu + (z_2)_r} \end{pmatrix} \\ \bar{Z}_3^{-1} &= \begin{pmatrix} \frac{1}{\mu + (z_3)_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\mu + (z_3)_n} \end{pmatrix}, \bar{Z}_4^{-1} = \begin{pmatrix} \frac{1}{\mu + (z_4)_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\mu + (z_4)_n} \end{pmatrix} \end{aligned} \quad (5)$$

com $\mathbf{Jg}(\mathbf{x}) \in \mathbb{R}^{n \times m}$ e $\mathbf{Jh}(\mathbf{x}) \in \mathbb{R}^{n \times r}$ as matrizes jacobianas dos funcionais \mathbf{g} e \mathbf{h} , respectivamente. Os

resíduos do método são expressos por (6):

$$\mathbf{b}^k = \begin{pmatrix} \mathbf{m}^k \\ \mathbf{t}_0^k \\ \mathbf{t}_1^k \\ \mathbf{t}_2^k \\ \mathbf{t}_3^k \\ \mathbf{t}_4^k \\ \boldsymbol{\pi}_1^k \\ \boldsymbol{\pi}_2^k \\ \boldsymbol{\pi}_3^k \\ \boldsymbol{\pi}_4^k \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}^k) - \mathbf{Jg}(\mathbf{x}^k)^t \boldsymbol{\lambda}_0^k - \mathbf{Jh}(\mathbf{x}^k)^t (\boldsymbol{\lambda}_2^k - \boldsymbol{\lambda}_1^k) + \boldsymbol{\lambda}_3^k - \boldsymbol{\lambda}_4^k \\ -\mathbf{g}(\mathbf{x}^k) \\ \mathbf{h}(\mathbf{x}^k) - \mathbf{z}_1^k - \mathbf{u}_1 \\ -\mathbf{h}(\mathbf{x}^k) - \mathbf{z}_2^k + \mathbf{u}_2 \\ \mathbf{x}^k - \mathbf{z}_3^k - \mathbf{l}_1 \\ -\mathbf{x}^k - \mathbf{z}_4^k + \mathbf{l}_2 \\ -\bar{\mathbf{Z}}_{1k} \boldsymbol{\lambda}_1^k + \mu_k \boldsymbol{\delta}_1^k - \mathbf{D}_{z_1^k} \mathbf{d}_{\lambda_1}^k \\ -\bar{\mathbf{Z}}_{2k} \boldsymbol{\lambda}_2^k + \mu_k \boldsymbol{\delta}_2^k - \mathbf{D}_{z_2^k} \mathbf{d}_{\lambda_2}^k \\ -\bar{\mathbf{Z}}_{3k} \boldsymbol{\lambda}_3^k + \mu_k \boldsymbol{\delta}_3^k - \mathbf{D}_{z_3^k} \mathbf{d}_{\lambda_3}^k \\ -\bar{\mathbf{Z}}_{4k} \boldsymbol{\lambda}_4^k + \mu_k \boldsymbol{\delta}_4^k - \mathbf{D}_{z_4^k} \mathbf{d}_{\lambda_4}^k \end{pmatrix} \quad (6)$$

onde $\mathbf{K} = \nabla_{\mathbf{xx}}^2 f(\mathbf{x}^k) + \sum_{t=1}^m [(\boldsymbol{\lambda}_0^k)_t \nabla_{\mathbf{xx}}^2 \mathbf{g}_t(\mathbf{x}^k)] + \sum_{i=1}^r \{[(\boldsymbol{\lambda}_2^k)_i - (\boldsymbol{\lambda}_1^k)_i] \nabla_{\mathbf{xx}}^2 \mathbf{h}_i(\mathbf{x}^k)\}$, \mathbf{I}_n é a matriz identidade de ordem n , \mathbf{I}_r a matriz identidade de ordem r , $\mathbf{D}_{z_v^k} = \text{diag}(\mathbf{d}_{z_v^k}^k)$, $\boldsymbol{\Lambda}_{v_k} = \text{diag}(\boldsymbol{\lambda}_v^k)$, com $v = 1, \dots, 4$ e $\mathbf{d}_{\boldsymbol{\omega}}^k = (\mathbf{d}_{\mathbf{x}}^k, \mathbf{d}_{z_1}^k, \mathbf{d}_{z_2}^k, \mathbf{d}_{z_3}^k, \mathbf{d}_{z_4}^k, \mathbf{d}_{\lambda_0}^k, \mathbf{d}_{\lambda_1}^k, \mathbf{d}_{\lambda_2}^k, \mathbf{d}_{\lambda_3}^k, \mathbf{d}_{\lambda_4}^k)^t$ é o vetor de direções.

2.1 Procedimentos Previsor e Corretor

No passo *previsor*, desconsideram-se os termos não-lineares da forma $\mathbf{D}_{z_v^k} \mathbf{d}_{\lambda_v}^k$ presentes nos resíduos $\boldsymbol{\pi}_v^k$. Em Pinheiro [3], a esparsidade da matriz \mathbf{A}_k é explorada para calcular as direções explicitadas a seguir:

$$\mathbf{d}_{\lambda_0}^k = -\boldsymbol{\lambda}_0^k - [\mathbf{Jg}(\mathbf{x}^k) \boldsymbol{\theta}_k^{-1} \mathbf{Jg}(\mathbf{x}^k)^t]^{-1} \cdot \{\mathbf{Jg}(\mathbf{x}^k) \boldsymbol{\theta}_k^{-1} [\nabla f(\mathbf{x}^k) + \mathbf{c}^k + \mu_k \boldsymbol{\varphi}^k] + \mathbf{t}_0^k\} \quad (7)$$

$$\mathbf{d}_{\mathbf{x}}^k = -\boldsymbol{\theta}_k^{-1} [\nabla f(\mathbf{x}^k) + \mathbf{c}^k + \mu_k \boldsymbol{\varphi}^k + \mathbf{Jg}(\mathbf{x}^k)^t \boldsymbol{\lambda}_0^k + \mathbf{Jg}(\mathbf{x}^k)^t \mathbf{d}_{\lambda_0}^k] \quad (8)$$

$$\mathbf{d}_{z_1}^k = \mathbf{Jh}(\mathbf{x}^k) \mathbf{d}_{\mathbf{x}}^k + \mathbf{t}_1^k; \mathbf{d}_{z_2}^k = -\mathbf{Jh}(\mathbf{x}^k) \mathbf{d}_{\mathbf{x}}^k + \mathbf{t}_2^k; \mathbf{d}_{z_3}^k = \mathbf{d}_{\mathbf{x}}^k + \mathbf{t}_3^k; \mathbf{d}_{z_4}^k = -\mathbf{d}_{\mathbf{x}}^k + \mathbf{t}_4^k \quad (9)$$

$$\mathbf{d}_{\lambda_v}^k = -\bar{\mathbf{Z}}_{v_k}^{-1} \boldsymbol{\Lambda}_{v_k} \mathbf{d}_{z_v^k} + \mu_k \bar{\mathbf{Z}}_{v_k}^{-1} \boldsymbol{\delta}_v^k - \boldsymbol{\lambda}_v^k, v = 1, \dots, 4 \quad (10)$$

com

$$\boldsymbol{\theta}_k = \mathbf{K} + \mathbf{Jh}(\mathbf{x}^k)^t (\bar{\mathbf{Z}}_{1k}^{-1} \boldsymbol{\Lambda}_{1k} + \bar{\mathbf{Z}}_{2k}^{-1} \boldsymbol{\Lambda}_{2k}) \mathbf{Jh}(\mathbf{x}^k) + \bar{\mathbf{Z}}_{3k}^{-1} \boldsymbol{\Lambda}_{3k} + \bar{\mathbf{Z}}_{4k}^{-1} \boldsymbol{\Lambda}_{4k} \quad (11)$$

$$\mathbf{c}^k = \mathbf{Jh}(\mathbf{x}^k)^t (\bar{\mathbf{Z}}_{1k}^{-1} \boldsymbol{\Lambda}_{1k} \mathbf{t}_1^k - \bar{\mathbf{Z}}_{2k}^{-1} \boldsymbol{\Lambda}_{2k} \mathbf{t}_2^k) + \bar{\mathbf{Z}}_{3k}^{-1} \boldsymbol{\Lambda}_{3k} \mathbf{t}_3^k - \bar{\mathbf{Z}}_{4k}^{-1} \boldsymbol{\Lambda}_{4k} \mathbf{t}_4^k \quad (12)$$

$$\boldsymbol{\varphi}^k = \mathbf{Jh}(\mathbf{x}^k)^t (-\bar{\mathbf{Z}}_{1k}^{-1} \boldsymbol{\delta}_1^k + \bar{\mathbf{Z}}_{2k}^{-1} \boldsymbol{\delta}_2^k) - \bar{\mathbf{Z}}_{3k}^{-1} \boldsymbol{\delta}_3^k + \bar{\mathbf{Z}}_{4k}^{-1} \boldsymbol{\delta}_4^k \quad (13)$$

Uma vez que as direções do procedimento previsor são calculadas, efetua-se o procedimento *corretor*. Nesta etapa, os resíduos não-lineares desconsiderados no passo corretor são utilizados, empregando as direções do passo corretor. Isto permite calcular as seguintes direções corrigidas:

$$\tilde{\mathbf{d}}_{\lambda_0}^k = -\boldsymbol{\lambda}_0^k - [\mathbf{Jg}(\mathbf{x}^k) \boldsymbol{\theta}_k^{-1} \mathbf{Jg}(\mathbf{x}^k)^t]^{-1} \cdot \{\mathbf{Jg}(\mathbf{x}^k) \boldsymbol{\theta}_k^{-1} [\nabla f(\mathbf{x}^k) + \tilde{\mathbf{c}}^k + \mu_k \boldsymbol{\varphi}^k] + \mathbf{t}_0^k\} \quad (14)$$

$$\tilde{\mathbf{d}}_{\mathbf{x}}^k = -\boldsymbol{\theta}_k^{-1} [\nabla f(\mathbf{x}^k) + \tilde{\mathbf{c}}^k + \mu_k \boldsymbol{\varphi}^k + \mathbf{Jg}(\mathbf{x}^k)^t \boldsymbol{\lambda}_0^k + \mathbf{Jg}(\mathbf{x}^k)^t \tilde{\mathbf{d}}_{\lambda_0}^k] \quad (15)$$

$$\tilde{\mathbf{d}}_{z_1}^k = \mathbf{Jh}(\mathbf{x}^k) \tilde{\mathbf{d}}_{\mathbf{x}}^k + \mathbf{t}_1^k; \tilde{\mathbf{d}}_{z_2}^k = -\mathbf{Jh}(\mathbf{x}^k) \tilde{\mathbf{d}}_{\mathbf{x}}^k + \mathbf{t}_2^k; \tilde{\mathbf{d}}_{z_3}^k = \tilde{\mathbf{d}}_{\mathbf{x}}^k + \mathbf{t}_3^k; \tilde{\mathbf{d}}_{z_4}^k = -\tilde{\mathbf{d}}_{\mathbf{x}}^k + \mathbf{t}_4^k \quad (16)$$

$$\tilde{\mathbf{d}}_{\lambda_v}^k = -\bar{\mathbf{Z}}_{v_k}^{-1} \boldsymbol{\Lambda}_{v_k} \tilde{\mathbf{d}}_{z_v^k} + \mu_k \bar{\mathbf{Z}}_{v_k}^{-1} \boldsymbol{\delta}_v^k - \bar{\mathbf{Z}}_{v_k}^{-1} \mathbf{D}_{z_v^k} \mathbf{d}_{\lambda_v}^k - \boldsymbol{\lambda}_v^k, v = 1, \dots, 4 \quad (17)$$

em que

$$\tilde{\mathbf{c}}^k = \mathbf{c}^k + \mathbf{Jh}(\mathbf{x}^k)^t (\bar{\mathbf{Z}}_{1k}^{-1} \mathbf{D}_{z_1^k} \mathbf{d}_{\lambda_1}^k - \bar{\mathbf{Z}}_{2k}^{-1} \mathbf{D}_{z_2^k} \mathbf{d}_{\lambda_2}^k) + \bar{\mathbf{Z}}_{3k}^{-1} \mathbf{D}_{z_3^k} \mathbf{d}_{\lambda_3}^k - \bar{\mathbf{Z}}_{4k}^{-1} \mathbf{D}_{z_4^k} \mathbf{d}_{\lambda_4}^k \quad (18)$$

Um ponto a se destacar é que, na abordagem aqui apresentada, os procedimentos previsor e corretor são realizados na mesma iteração. Além disso, o parâmetro de barreira também é utilizado no passo previsor.

2.2 Cálculo da nova solução e do comprimento do passo

Com as direções corrigidas, a partir da solução $\boldsymbol{\omega}^k$ numa iteração k , calcula-se um novo ponto $\boldsymbol{\omega}^{k+1} = \boldsymbol{\omega}^k + \alpha \tilde{\mathbf{d}}_{\boldsymbol{\omega}}^k$. O escalar α é um fator tipicamente igual a 0,995, cuja função é evitar que o novo ponto

esteja exatamente na fronteira da região relaxada, o que causaria erros devido à inexistência do logaritmo. Uma justificativa mais rica pode ser encontrada em [7]. O vetor de direções corrigidas é expresso por $\tilde{\mathbf{d}}_{\omega}^k = (\alpha_P \tilde{\mathbf{d}}_x^k, \alpha_P \tilde{\mathbf{d}}_z^k, \alpha_D \tilde{\mathbf{d}}_{\lambda}^k)^t$, onde os escalares α_P e α_D são calculados segundo Granville [2]:

$$\alpha_P = \min \left\{ \begin{array}{l} \min_{(z_1)_i > 0 \text{ e } (\tilde{\mathbf{d}}_{z_1})_i < 0} -\frac{(z_1)_i}{(\tilde{\mathbf{d}}_{z_1})_i}, \quad \min_{(z_2)_i > 0 \text{ e } (\tilde{\mathbf{d}}_{z_2})_i < 0} -\frac{(z_2)_i}{(\tilde{\mathbf{d}}_{z_2})_i}, \\ \min_{(z_3)_j > 0 \text{ e } (\tilde{\mathbf{d}}_{z_3})_j < 0} -\frac{(z_3)_j}{(\tilde{\mathbf{d}}_{z_3})_j}, \quad \min_{(z_4)_j > 0 \text{ e } (\tilde{\mathbf{d}}_{z_4})_j < 0} -\frac{(z_4)_j}{(\tilde{\mathbf{d}}_{z_4})_j} \end{array} \right\} \quad (19)$$

$$\alpha_D = \min \left\{ \begin{array}{l} \min_{(\lambda_1)_i > 0 \text{ e } (\tilde{\mathbf{d}}_{\lambda_1})_i < 0} -\frac{(\lambda_1)_i}{(\tilde{\mathbf{d}}_{\lambda_1})_i}, \quad \min_{(\lambda_2)_i > 0 \text{ e } (\tilde{\mathbf{d}}_{\lambda_2})_i < 0} -\frac{(\lambda_2)_i}{(\tilde{\mathbf{d}}_{\lambda_2})_i}, \\ \min_{(\lambda_3)_j > 0 \text{ e } (\tilde{\mathbf{d}}_{\lambda_3})_j < 0} -\frac{(\lambda_3)_j}{(\tilde{\mathbf{d}}_{\lambda_3})_j}, \quad \min_{(\lambda_4)_j > 0 \text{ e } (\tilde{\mathbf{d}}_{\lambda_4})_j < 0} -\frac{(\lambda_4)_j}{(\tilde{\mathbf{d}}_{\lambda_4})_j} \end{array} \right\} \quad (20)$$

com $i = 1, \dots, r$ e $j = 1, \dots, n$.

2.3 Parâmetro de barreira, estimadores dos multiplicadores de Lagrange e critério de parada

O parâmetro de barreira foi atualizado pela seguinte heurística:

$$\mu_{k+1} = \beta \mu_k, \beta \in [0, 1] \quad (21)$$

Os estimadores dos multiplicadores de Lagrange foram atualizados pela estratégia de baixo custo computacional desenvolvida por Pinheiro [3]:

$$\delta_i^{k+1} = \lambda_i^k \quad (22)$$

Como critério de parada, verifica-se a viabilidade primal-dual e as condições de folgas complementares sobre a nova solução, ω^{k+1} , isto é,

$$\|\nabla L(\omega^{k+1})\|_{\infty} < \varepsilon \quad (23)$$

onde $\varepsilon > 0$ é a precisão adotada. Quando (23) ocorre, o método deve parar, pois ω^{k+1} é a solução ótima com precisão desejada. Senão é realizada uma nova iteração dos procedimentos previsor e corretor.

2.4 A estratégia de convergência global

Neste trabalho, adotou-se uma estratégia de convergência global, proposta por Pinheiro [3], a qual é uma variante do método de Levenberg-Marquardt. Em problemas de otimização não-convexos, a matriz dual normal, θ_k pode não ser definida positiva. Isto pode acarretar a instabilidade do método, bem como a convergência para máximos locais.

Pode-se verificar se a matriz θ_k é definida positiva utilizando a Decomposição de Cholesky. Se a decomposição não é possível, a diagonal da matriz é perturbada, de modo a torná-la definida positiva, da seguinte maneira:

$$\hat{\theta}_k = \theta_k + \gamma_k I_n \quad (24)$$

onde I_n é a matriz identidade de ordem n (igual à de θ_k) e $\gamma_k \in \mathbb{R}_+$. O número γ_k é denominado parâmetro de amortecimento. Este valor é incrementado enquanto a decomposição não puder ser finalizada, pela seguinte heurística:

$$\gamma_{p+1} = \gamma_p \left(\frac{\mu_{k-1} + \sqrt{(\sqrt{5} - 1)^2 \mu_k^2 + \mu_{k-1}^2}}{2\mu_{k-1}} \right) \quad (25)$$

Quando a decomposição for finalizada, define-se $\gamma_k = \gamma_{p+1}$. Ao terminar uma iteração, o valor do parâmetro de amortecimento é atualizado pela seguinte heurística, a qual é uma modificação daquela proposta por Bazaraa et al. [1]:

- i. Se $L(\omega^k) - L(\omega^{k+1}) < 0,25$, então $\gamma_{k+1} = \frac{1}{3}\gamma_k$;
- ii. Se $L(\omega^k) - L(\omega^{k+1}) > 0,75$, então γ_k é atualizado pela heurística (25);
- iii. Se $0,25 \leq L(\omega^k) - L(\omega^{k+1}) \leq 0,75$, então $\gamma_{k+1} = \gamma_k$.

3 A estratégia para suavização de termos modulares

Dada uma função real $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, sabemos que é válida a identidade:

$$|f(\mathbf{x})| = f(\mathbf{x}) \operatorname{sgn}(f(\mathbf{x})) \quad (26)$$

onde sgn é a função sinal, definida por:

$$\operatorname{sgn}(x) = \begin{cases} -1, & \text{se } x < 0 \\ 0, & \text{se } x = 0 \\ 1, & \text{se } x > 0 \end{cases} \quad (27)$$

A função sinal, apesar de descontínua na origem, pode ser aproximada por funções de classe C^∞ com boa precisão, quando se está fora da origem. Redes neurais artificiais utilizam com frequência esse tipo de aproximação nas funções de ativação. Assim, se f for uma função de classe C^2 , um problema de minimização tiver a forma

$$\begin{array}{ll} \text{Minimizar} & |f(\mathbf{x})| \\ \text{sujeito a} & \mathbf{x} \in \Omega \end{array} \quad (28)$$

e uma função aproximante e diferenciável, \mathcal{A} , para a função sinal for conhecida, podemos substituir (28) por um problema aproximado

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \cdot \mathcal{A}(f(\mathbf{x})) \\ \text{sujeito a} & \mathbf{x} \in \Omega \end{array} \quad (29)$$

Desde que a função aproximante \mathcal{A} seja de classe C^2 , a função objetivo de (29) também o será, pois resulta da composição e produto de funções de classe C^2 . Isto significa que sobre esta função podemos aplicar os métodos clássicos de otimização, em particular o método previsor-corretor primal-dual barreira logarítmica modificada. A proximidade da solução obtida com as soluções do problema original dependerá, naturalmente, da qualidade da aproximação feita para a função sinal. Na seção a seguir, apresentamos os resultados obtidos com esta estratégia, para minimizar uma função com termos modulares. Para obter os resultados, utilizamos um aproximante da função sinal baseado na função tangente hiperbólica: $\mathcal{A}(x) = \tanh(\rho x)$, de modo que resolvemos uma sequência de problemas aproximados:

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \tanh(\rho f(\mathbf{x})) \\ \text{sujeito a} & \mathbf{x} \in \Omega \end{array} \quad (30)$$

Esta estratégia de suavização está sendo aplicada ao PDE-PV, e os resultados iniciais obtidos são promissores, sendo este problema o objeto para trabalhos futuros.

4 Resultados Numéricos

A abordagem apresentada foi aplicada ao seguinte problema-teste matemático:

$$\begin{array}{ll} \text{Minimizar} & x_1^2 + x_2 + |\operatorname{sen}(x_2) - x_1| \\ \text{sujeito a} & e^{x_1} + x_2 - 0,5 = 0 \\ & x_1^2 + x_2^2 \leq 1 \end{array} \quad (31)$$

Os resultados das iterações estão sumarizados na Tabela 1 apresentada a seguir, em que as colunas FO e FA apresentam, respectivamente, o valor da função objetivo original e aproximada no ponto. Além disso, a Figura 1 ilustra o processo de convergência.

Os parâmetros iniciais para o método foram: $\mu_0 = 5$, $\gamma_0 = 0,1$, $\beta = 0,5$ e os estimadores dos multiplicadores de Langrange iguais a 0,1. A precisão de parada foi $\epsilon = 10^{-4}$. Excetuando as variáveis x_1 e x_2 , cujos valores foram fornecidos, as demais variáveis iniciais foram calculadas assumindo que os resíduos eram nulos em (6). A região factível do problema é dada pelo arco da função exponencial em azul claro, no interior da circunferência. Observa-se que o método foi inicializado no ponto $\mathbf{x}^0 = (1, 1)^t$, o qual é infactível inclusive para a restrição de desigualdade. Ainda assim, o método foi capaz de convergir para um ponto factível, devido à utilização da função barreira modificada, que realiza uma relaxação sobre as folgas das restrições de desigualdade. Outro ponto a notar é que o método não apresentou instabilidades, mesmo quando se aproximou de pontos em que a função não é diferenciável (situados sobre a senoide em verde). O valor da função objetivo original na solução encontrada é $-0,196735$, enquanto na função aproximada é $-0,19902$, evidenciando que a função original e aproximada têm valores próximos.

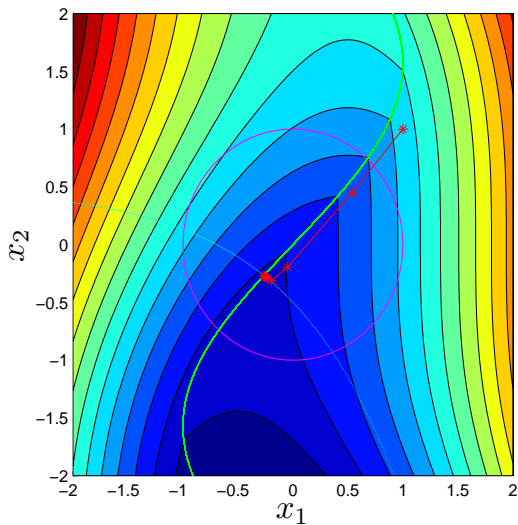


Figura 1: Curvas de nível da função objetivo e convergência do método

k	x_1^k	x_2^k	FO	FA
0	1,0	1,0	2,158529	2,024923
1	0,537915	0,448873	0,842191	0,748996
2	-0,0508989	-0,195822	-0,049557	-0,156881
3	-0,195204	-0,311958	-0,162134	-0,235092
4	-0,234609	-0,290196	-0,183623	-0,220118
5	-0,243774	-0,283628	-0,188135	-0,211164
6	-0,255546	-0,274439	-0,193674	-0,204742
7	-0,257748	-0,272788	-0,194684	-0,201951
8	-0,261963	-0,269532	-0,19659	-0,199792
9	-0,262304	-0,269277	-0,196742	-0,19902
10	-0,262287	-0,26929	-0,196735	-0,19902

Tabela 1: Iterações do método

5 Conclusão

Este trabalho apresentou um método primal-dual predictor-corrector de pontos interiores, que emprega a função barreira logarítmica modificada, para a resolução de problemas de otimização com funções objetivo com termos modulares. Para tanto, adotou-se uma estratégia que utiliza um aproximante da função sinal, para substituir o problema original por um problema aproximado. Utilizou-se, ainda, uma estratégia de convergência global, a fim de tratar a não-convexidade do problema, e obter um ponto de mínimo local. Uma implementação do método proposto, em Matlab 2011, mostrou que a estratégia foi eficiente e estável para determinar um ponto de mínimo para um problema-teste matemático, iniciando em pontos infactíveis.

Referências

- [1] M. S. Bazaraa, H. D. Sherali, C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. New Jersey: Wiley Interscience, 2006.
- [2] S. Granville. “Optimal Reactive Dispatch through Interior Point Method”, *IEEE Transactions on Power Systems*, v.9, n. 1, p. 136-146, 1994.
- [3] R. B. N. Pinheiro. *Um método predictor corretor primal-dual de pontos interiores barreira logarítmica modificada, com estratégias de convergência global e de ajuste cúbico, para problemas de programação não-linear e não-convexa*. Dissertação (mestrado) – Faculdade de Engenharia de Bauru, Universidade Estadual Paulista, Bauru, 2012.
- [4] R. Polyak. “Modified barrier functions”, *Mathematical Programming*, v. 54, n. 2, p. 177-222, 1992.
- [5] V. A. Sousa, E. C. Baptista, G. R. M. Costa. *Modified Barrier Method for Optimal Power Flow Problem*. In: IEEE Power Engineering Society General, 2004, Denver – Colorado. IEEE Catalog Number: 04CH37567C. USA: Mira Digital Publishing, 2004. v. 1. p. 1-6.
- [6] K. K. Vishwakarma, H. M. Dubey, M. Pandit, B. K. Panigrahi. “Simulated annealing approach for solving economic load dispatch problems with valve point loading effects”, *International Journal of Engineering, Science and Technology*, v. 4, n. 4, p. 60-72, 2012.
- [7] M. H. Wright. “Why a Pure Primal Newton Barrier Step May Be Infeasible”, *SIAM Journal on Optimization*, v. 5, n. 1, p. 1-12, 1995.