

# Implementação de um Método de Lagrangianos Aumentados com Informação de Primeira Ordem

E. G. Birgin<sup>1</sup>, D. S. Marcondes<sup>2</sup>  
 USP, São Paulo, SP

**Resumo.** Métodos de Lagrangianos Aumentados são amplamente estudados e usados para solucionar problemas de Otimização não linear com restrições de igualdade e desigualdades. **Algencan** é um algoritmo de Lagrangiano Aumentado com salvaguardas bem estabelecido com vários resultados de complexidade, convergências e de desempenho computacional. Os subproblemas do Algencan são resolvidos utilizando um método de restrições ativas chamado de Gencan. Neste trabalho mostramos resultados numéricos para duas modificações do Gencan utilizando somente informação de primeira ordem e comparamos com a versão corrente de Gencan.

**Palavras-chave.** Otimização não linear. Métodos de Lagrangianos Aumentados. Experimentos Numéricos. MINRES. Métodos de Gradientes Conjugados.

## 1 Introdução

Estamos interessados em problemas de programação não linear definidos da seguinte forma:

$$\begin{aligned} & \text{Minimizar} && f(x) \\ & \text{sujeita a} && h(x) = 0 \\ & && g(x) \leq 0 \\ & && x \in \Omega, \end{aligned} \tag{1}$$

onde  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  são contínuas e  $\Omega \subset \mathbb{R}^n$  é um conjunto fechado. Daqui em diante,  $\|\cdot\|$  representa a norma Euclidiana,  $\|\cdot\|_\infty$  representa a norma infinito e  $v_+$  significa  $\max\{0, v\}$ . O conjunto  $\mathbb{R}_+$  será o conjunto dos números reais não negativos.

Otimização não linear é uma área da matemática amplamente estudada, tanto na parte teórica quanto na parte prática [22]. Existem vários métodos para a resolução de problemas modelados como (1), por exemplo, o método de Lagrangiano Aumentado baseados em Powell-Hestenes-Rockafellar (PHR) [16, 26, 27] que utiliza a função de Lagrangiano Aumentado dada por:

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[ h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[ \left( g_i(x) + \frac{\mu_i}{\rho} \right)_+ \right]^2 \right\} \tag{2}$$

para todo  $x \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}_+^p$ .

Chamamos de métodos Lagrangianos Aumentados baseados em PHR, para resolver (1), métodos iterativos que a cada iteração minimiza aproximadamente a função objetivo somada com um termo que penaliza quando alguma restrição é violada, ou seja, a cada iteração externa, dado valor de  $\rho$ , é minimizada a função  $L_\rho$ .

<sup>1</sup>egbirgin@ime.usp.br

<sup>2</sup>diaulas@ime.usp.br

Dentre os algoritmos de Lagrangianos Aumentados baseado em PHR, destacamos o algoritmo Algencan descrito no Algoritmo 1.1. Algencan, introduzido em [2, 3], não utiliza variáveis de folga nas desigualdades, utiliza salvaguardas e admite restrições gerais no conjunto  $\Omega$ . No caso em que  $x \in \Omega$  são restrições de caixas, os subproblemas de minimização com restrições de caixas do Algencan são resolvidos utilizando um algoritmo de restrições ativas, o Gencan [7]. A adoção de um método de restrições ativas para resolver os subproblemas do Algencan maximiza o potencial de se beneficiar de novos desenvolvimentos para a minimização com caixas.

**Algoritmo 1.1.** Sejam  $\rho_1 \geq 0$ ,  $\tau \in (0, 1)$ ,  $\gamma > 1$ ,  $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^m$ ,  $\bar{\mu}^1 \in [0, \mu_{\max}]^p$ ,  $\{\varepsilon_k\}_{k=1}^{\infty}$  dados. Faça  $k \leftarrow 1$ .

**Passo 1.** Compute  $x^k \in \Omega$  uma solução aproximada de **Minimizar**  $L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k)$  **sujeito a**  $x \in \Omega$  satisfazendo:

$$\|P_{\Omega}(x^k - \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) - x^k\| \leq \varepsilon_k.$$

**Passo 2.** Defina  $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k)$  e  $\mu^{k+1} = (\bar{\mu}^k + \rho_k g(x^k))_+$ .

**Passo 3.** Defina, para todo  $i = 1, \dots, p$ ,  $V_i^k = \min\{-g_i(x^k), \bar{\mu}_i^k / \rho_k\}$ .

**Passo 4.** Escolha  $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ ,  $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$ .

**Passo 5.** Faça  $k \leftarrow k + 1$  e vá para o Passo 1.

O algoritmo Gencan divide a região viável em faces disjuntas e, em cada iteração, existe um critério para decidir se a minimização deve ser efetuada na face atual ou se essa face deve ser abandonada. A saída de uma face é feita dando um passo na direção do gradiente espectral projetado e a minimização dentro de uma face, na versão corrente, é feita com um método de Newton truncado que usa gradientes conjugados lineares, ou seja, um método do tipo Newton que utiliza gradientes conjugados lineares para resolver o sistema Newtoniano linear. Provas de convergências dos algoritmos Algencan e Gencan podem ser encontradas em [3, 4, 9].

Um dos problemas práticos mais relevantes onde o Gencan foi utilizado é a construção de um ponto inicial para simulações de dinâmica molecular, através do empacotamento de moléculas em um espaço limitado. Esse problema pode ser modelado como um grande problema de viabilidade, que foi abordado reformulando-o como um grande problema de minimização com restrições de caixas. Foi assim que o software Packmol foi concebido [1].

Neste trabalho, estamos interessados em comparar formas de computar  $x^k \in \Omega$  no Passo 1, do Algoritmo 1, quando temos somente informação de primeira ordem. As comparações serão feitas usando os métodos de gradientes conjugados de Hager-Zhang [14, 15] e Newton Truncado calculando a direção aproximada de Newton usando gradientes conjugados lineares (usado na versão corrente do Gencan) e usando o método MINRES [23].

## 2 Algencan

A primeira versão do que mais tarde foi chamado de Algencan foi desenvolvida em 2002 juntamente com a introdução de um método de restrições ativas para minimização com restrições chamado de Gencan [7]. Em geral, a presença do Algencan, do Gencan ou do SPG nas últimas décadas possibilitou abordar diversas aplicações e/ou problemas interessantes [1, 9]. Por essa razão, repensar, revisar e aprimorar o Algencan e o Gencan têm sido, e continua sendo, uma tarefa constante.

A versão 1 do Algencan corresponde às publicações de 2007/2008, abordando o trabalho sobre a teoria da convergência dos métodos Lagrangianos Aumentados utilizando a qualificação de restrições CPLD [2, 3]. A versão 2 incorporou a aceleração (tentativa de resolução de um sistema não

linear KKT, com o método de Newton, entre iterações do método do Lagrangiano Aumentado) e pré-condicionadores que utilizam explicitamente que os subproblemas correspondem à minimização da função Lagrangiana Aumentada [6, 10]. A versão 3 corresponde a uma reimplementação completa e coincidiu com a publicação do livro, em 2014 [9]. A versão 4, de 2020, [5] trata dos resultados de complexidade do Algencan e da comparação com o Ipopt [28] considerando que a informação de segunda ordem estava disponível. Nesta versão, algumas modificações foram feitas no Algencan, utilizando ideias aprendidas pelos autores no desenvolvimento de métodos com regularização cúbica [8].

A versão corrente do Gencan utiliza o método de Newton Truncado utilizando gradientes conjugados lineares para o cálculo da direção aproximada de Newton [7] e é possível usar informação de primeira ordem ou de primeira e segunda ordem. Desejamos comparar alternativas diferentes, a essa versão corrente, que solucione o subproblema de Lagrangianos Aumentados e que utilize somente informação de primeira ordem.

Uma alternativa é usar o método MINRES, no lugar dos gradientes conjugados lineares, para a resolução do sistema Newtoniano linear. Existem vários artigos na literatura recente sobre o MINRES [17–21], que esclarecem algumas propriedades teóricas, questões de complexidade computacional e apresentam comparações com o método de gradientes conjugados lineares (no contexto da utilização de um ou outro para resolver sistemas Newtonianos lineares de um método de Newton para minimização sem restrições). Neste texto, chamaremos o algoritmo de Newton truncado utilizando MINRES para resolver o sistema Newtoniano de Gencan-MR.

Outra alternativa, é trocar o método de Newton truncado pelo método de gradientes conjugados não linear. Existem muitas opções de métodos de gradientes conjugados não linear como, por exemplo, de Fletcher-Reeves [12], de Polak-Ribière-Polyak [24, 25], de Dai-Yuan [11], etc. Optamos pelo método de gradientes conjugados de Hager-Zhang (chamaremos de CGHZ) [14], que tem se destacado em relação aos outros métodos de gradientes conjugados [15]. A implementação de CGHZ dos autores é sofisticada, incluindo uma busca linear complexa que procura tamanhos de passos que satisfaçam a condição de Wolfe ou a condição de Wolfe aproximada.

### 3 Experimentos Numéricos

Nesta seção estamos interessados em comparar a versão corrente do Gencan, Gencan-MR e o algoritmo CGHZ para quando não temos acesso à informação de segunda ordem. Na versão corrente do Gencan e no Gencan-MR foi usado o método de quocientes incrementais para os cálculos dos produtos Hessiana-vetor necessários.

Foram implementados o Gencan-MR em Fortran 90 e comparamos com a implementação dos autores para verificar a se nossa implementação estava correta. Estendemos o MINRES para a minimização com restrições usando uma estratégia de restrições ativas. Implementamos, também, em Fortran 90 o método CGHZ e validamos comparando com a implementação dos autores em linguagem C. Todos os experimentos foram realizados utilizando um computador com processador 5.3 GHz Intel Core i9-13900K e 128 GB de memória RAM, executados em Ubuntu (versão 23.04). Os códigos foram compilados usando compilador GFortran (versão 12.3.0) com a opção de otimizador -O3 ativa.

Primeiramente comparamos a qualidade do iterado final da versão corrente do Gencan e do Gencan-MR, considerando todos os 378 problemas sem restrições e com restrições de caixas da coleção CUTEst [13] com as suas dimensões predefinidas. Na Tabela 1 são apresentados os resultados da contagem do número de problemas para os quais o valor final é semelhante ao melhor valor encontrado, com tolerância  $f_{tol}$  relativa se o melhor valor for superior a 1 e tolerância  $f_{tol}$  absoluta se o melhor valor for inferior a 1 para cada método, ou seja, contamos o número de vezes que:

$$f_i \leq f^{\min} + f_{\text{tol}} \max\{1, |f^{\min}|\}, \text{ onde } f^{\min} = \min\{f_1, f_2\}. \quad (3)$$

Tabela 1: Contagem do número de problemas para os quais o valor final é semelhante ao melhor valor encontrado pelos algoritmos: versão corrente do Gencan e Gencan-MR.

	$f_{\text{tol}}$								
	0.1	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$10^{-7}$	$10^{-8}$	0
Gencan (Versão corrente)	346	339	331	329	327	322	322	321	193
Gencan-MR	362	352	348	343	340	336	326	324	229

Adicionalmente, contamos em quantos problemas os algoritmos pararam com sucesso, ou seja, por identificar que o problema é ilimitado ( $f(x^k) \leq -10^{12}$ ) ou porque encontrou um ponto com gradiente projetado “pequeno” ( $\|P_{\Omega}(x^k - \nabla f(x^k)) - x^k\|_{\infty} \leq 10^{-8}$ ). Esses números são 277 para a versão corrente do Gencan e 303 para o Gencan-MR.

Na Figura 1 temos o perfil de desempenho do tempo de execução, em segundos, comparando versão corrente do Gencan e Gencan-MR em 260 problemas sem restrições ou com restrições de caixas, em que ambos os algoritmos pararam com sucesso encontrando uma solução com valor equivalente de  $f$  considerando  $f_{\text{tol}} = 0.1$ .

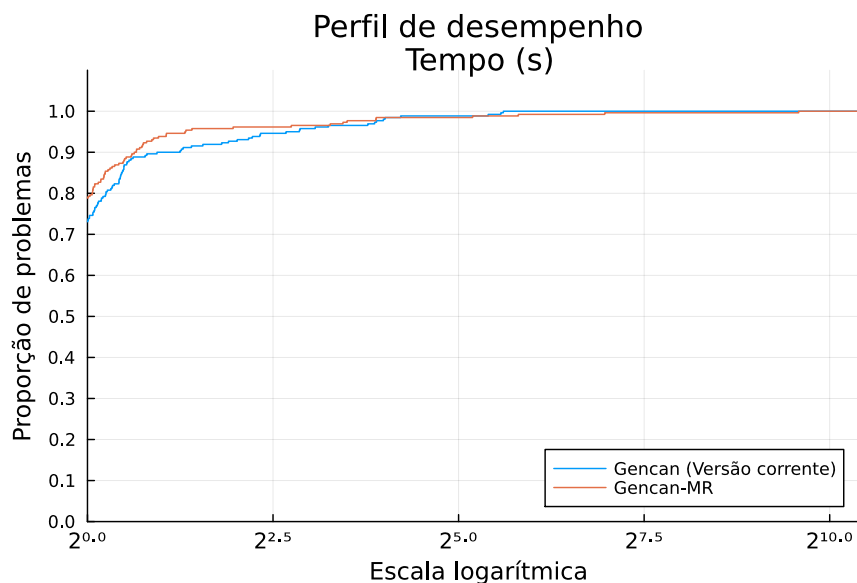


Figura 1: Perfil de desempenho comparando a versão corrente do Gencan e o Gencan-MR em 260 problemas, sem restrições ou com restrições de caixas, onde ambos os algoritmos pararam com sucesso e encontraram uma solução com valor equivalente de  $f$  considerando  $f_{\text{tol}} = 0.1$ . Fonte: dos autores.

Comparamos, também, a qualidade do iterado final da versão corrente do Gencan, do Gencan-MR e do CGHZ, considerando todos os 234 problemas sem restrições da coleção CUTEst com as suas dimensões predefinidas e apresentamos os resultados na Tabela 2. Nestes experimentos, a

versão corrente do Gencan parou com sucesso em 165 problemas, o Gencan-MR em 179 e CGHZ em 198.

Tabela 2: Contagem do número de problemas para os quais o valor final é semelhante ao melhor valor encontrado pelos algoritmos: versão corrente do Gencan, Gencan-MR e CGHZ.

	$f_{tol}$								
	0.1	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$10^{-7}$	$10^{-8}$	0
Gencan (Versão corrente)	205	199	192	190	190	186	186	186	78
Gencan-MR	215	206	204	198	196	193	185	184	113
CGHZ	217	215	213	210	207	205	204	203	88

Na Figura 2 temos o perfil de desempenho do tempo de execução, em segundos, comparando versão corrente do Gencan, Gencan-MR e CGHZ em 147 problemas sem restrições, onde os três algoritmos pararam com sucesso encontrando uma solução com valor equivalente de  $f$  considerando  $f_{tol} = 0.1$ .

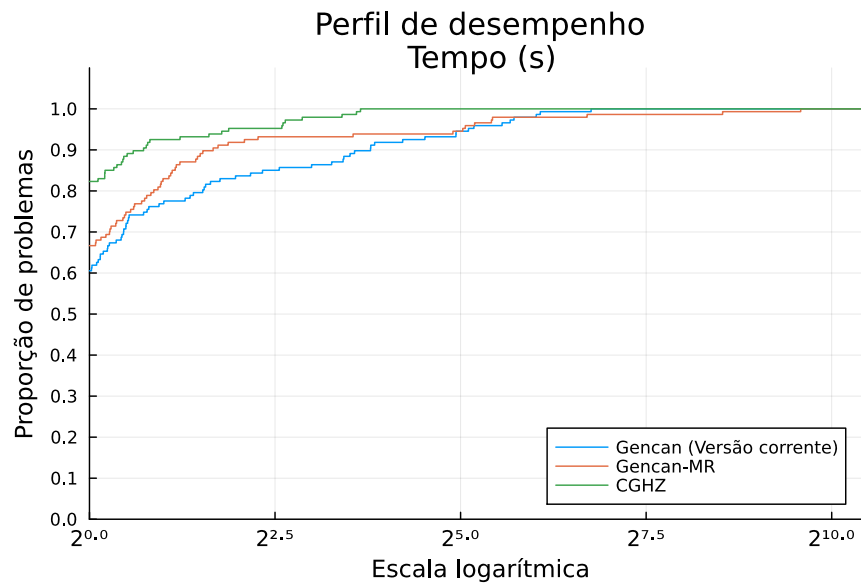


Figura 2: Perfil de desempenho comparando a versão corrente do Gencan, Gencan-MR e CGHZ em 147 problemas irrestritos onde os três algoritmos pararam com sucesso e encontraram uma solução com valor equivalente de  $f$  considerando  $f_{tol} = 0.1$ . Fonte: dos autores.

## 4 Considerações Finais

Neste trabalho relembramos um pouco a história do desenvolvimento de Algencan e testamos abordagens diferentes para resolver o subproblema associado a esse algoritmo.

Foram implementados os algoritmos de gradientes conjugados não linear de Hager-Zhang e Newton truncado utilizando MINRES para a solução do sistema Newtoniano linear e, também,

foram realizados experimentos numéricos que compararam esses algoritmos com a versão corrente do Gencan. Esses experimentos foram realizados utilizando problemas da coleção CUTEst comparando valores de funções objetivos, critérios de parada e tempo de execução desses algoritmos.

Com os resultados apresentados concluímos que Gencan-MR “é um pouco melhor” se comparado à versão corrente do Gencan e que CGHZ é mais robusto e eficiente do que versão corrente do Gencan e Gencan-MR nos problemas testados.

## Referências

- [1] R. Andrade, E. G. Birgin, J. M. Martínez e L. Martínez. “PACKMOL: A package for building initial configurations for molecular dynamics simulations”. Em: **Journal of Computational Chemistry** 30 (2009), pp. 2157–2164.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. “Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification”. Em: **Mathematical Programming** 111 (2008), pp. 5–32.
- [3] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. “On Augmented Lagrangian Methods with General Lower-Level Constraints”. Em: **SIAM Journal on Optimization** 18 (2008), pp. 1286–1309.
- [4] R. Andreani, G. Haeser, M. L. Schuverdt e P. J. S. Silva. “Two New Weak Constraint Qualifications and Applications”. Em: **SIAM Journal on Optimization** 22 (2012), pp. 1109–1135.
- [5] E. G. Birgin e J. M. Martínez. “Complexity and performance of an Augmented Lagrangian algorithm”. Em: **Optimization Methods and Software** 35 (2020), pp. 885–920.
- [6] E. G. Birgin e J. M. Martínez. “Improving ultimate convergence of an augmented Lagrangian method”. Em: **Optimization Methods and Software** 23 (2008), pp. 177–195.
- [7] E. G. Birgin e J. M. Martínez. “Large-Scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients”. Em: **Computational Optimization and Applications** 23 (2002), pp. 101–125.
- [8] E. G. Birgin e J. M. Martínez. “On Regularization and Active-set Methods with Complexity for Constrained Optimization”. Em: **SIAM Journal on Optimization** 28 (2018), pp. 1367–1395.
- [9] E. G. Birgin e J. M. Martínez. **Practical Augmented Lagrangian Methods for Constrained Optimization**. Fundamentals of Algorithms. Philadelphia: Society for Industrial and Applied Mathematics, 2014.
- [10] E. G. Birgin e J. M. Martínez. “Structured minimal-memory inexact quasi-Newton method and secant preconditioners for augmented Lagrangian optimization”. Em: **Computational Optimization and Applications** 39 (2008), pp. 1–16.
- [11] Y. H. Dai e Y. Yuan. “A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property”. Em: **SIAM Journal on Optimization** 10 (1999), pp. 177–182.
- [12] R. Fletcher e C. M. Reeves. “Function minimization by conjugate gradients”. Em: **The Computer Journal** 7 (1964), pp. 149–154.
- [13] N. I. M. Gould, D. Orban e Ph. L. Toint. “CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization”. Em: **Computational Optimization and Applications** 60 (2014), pp. 545–557.

- [14] W. W. Hager e H. Zhang. “A new conjugate gradient method with guaranteed descent and an efficient line search”. Em: **SIAM Journal on Optimization** 16 (2005), pp. 170–192.
- [15] W. W. Hager e H. Zhang. “Algorithm 851: CG DESCENT, a conjugate gradient method with guaranteed descent”. Em: **ACM Transactions on Mathematical Software** 32 (2006), pp. 113–137.
- [16] M. R. Hestenes. “Multiplier and gradient methods”. Em: **Journal of Optimization Theory and Applications** 4 (1969), pp. 303–320.
- [17] A. Lim e F. Roosta. **Complexity Guarantees for Nonconvex Newton-MR Under Inexact Hessian Information**. 2023. arXiv: 2308.09912.
- [18] Y. Liu, M. W. Mahoney, F. Roosta e P. Xu. “Newton-MR: Inexact Newton Method with minimum residual sub-problem solver”. Em: **EURO Journal on Computational Optimization** 10 (2022), p. 100035.
- [19] Y. Liu e F. Roosta. **A Newton-MR algorithm with complexity guarantees for non-convex smooth unconstrained optimization**. 2023. arXiv: 2208.07095.
- [20] Y. Liu e F. Roosta. “Convergence of Newton-MR under Inexact Hessian Information”. Em: **SIAM Journal on Optimization** 31 (2021), pp. 59–90.
- [21] Y. Liu e F. Roosta. “MINRES: From Negative Curvature Detection to Monotonicity Properties”. Em: **SIAM Journal on Optimization** 32 (2022), pp. 2636–2661.
- [22] J. Nocedal e S. J. Wright. **Numerical Optimization**. 1e edition. New York, NY: Springer, 1999.
- [23] C. C. Paige e M. A. Saunders. “Solution of Sparse Indefinite Systems of Linear Equations”. Em: **SIAM Journal on Numerical Analysis** 12 (1975), pp. 617–629.
- [24] E. Polak e G. Ribiere. “Note sur la convergence de méthodes de directions conjuguées”. Em: **R.I.R.O.** 3 (1969), pp. 35–43.
- [25] B.T. Polyak. “The conjugate gradient method in extremal problems”. Em: **USSR Computational Mathematics and Mathematical Physics** 9 (1969), pp. 94–112.
- [26] M. J. D. Powell. “A method for nonlinear constraints in minimization problems”. Em: **Optimization**. Ed. por R. Fletcher. New York, NY: Academic Press, 1969, pp. 283–298.
- [27] R. T. Rockafellar. “The multiplier method of Hestenes and Powell applied to convex programming”. Em: **Journal of Optimization Theory and Applications** 12 (1973), pp. 555–562.
- [28] A. Wächter e L. T. Biegler. “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming”. Em: **Mathematical Programming** 106 (2006), pp. 25–57.