UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

MAYCON PEREIRA DE SOUZA

Um estudo do Método de Gradientes Conjugados não Lineares

Campinas

2017

Maycon Pereira de Souza

Um estudo do Método de Gradientes Conjugados não Lineares

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática Aplicada e Computacional.

Orientadora: Sandra Augusta Santos

Este exemplar corresponde à versão final da Dissertação defendida pelo aluno Maycon Pereira de Souza e orientada pela Profa. Dra. Sandra Augusta Santos.

Campinas

2017

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

[ ] Tese	[ ] Artigo Científico
[x] Dissertação	[ ] Capítulo de Livro
[ ] Monografia – Especialização	[ ] Livro
[ ] TCC - Graduação	[ ] Trabalho Apresentado em Evento
[ ] Produto Técnico e Educacional - Tipo: _____	

Nome Completo do Autor: **Maycon Pereira de Souza**
Matrícula: **1992623**
Título do Trabalho**: Um estudo do Método de Gradientes Conjugados não Lineares**

Autorização - Marque uma das opções

1. (x) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ( ) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ___/___/_____ (Embargo);
3. ( ) Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

( ) O documento está sujeito a registro de patente.
( ) O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
( ) Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

gov.br Documento assinado digitalmente
MAYCON PEREIRA DE SOUZA
Data: 13/05/2026 18:52:04-0300
Verifique em https://validar.iti.gov.br

Local: **Uruaçu.** Data: **13/05/2026.**

_____
Assinatura do Autor e/ou Detentor dos Direitos Autorais

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

So89e

Souza, Maycon Pereira de, 1985-
Um estudo do método de gradientes conjugados não lineares / Maycon Pereira de Souza. – Campinas, SP : [s.n.], 2017.

Orientador: Sandra Augusta Santos.
Dissertação (mestrado profissional) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Otimização irrestrita. 2. Métodos do gradiente conjugado. 3. Métodos iterativos (Matemática). I. Santos, Sandra Augusta,1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: A study on the nonlinear conjugate gradient method
Palavras-chave em inglês:
Unrestricted optimization
Conjugate gradient methods
Iterative methods (Mathematics)
Área de concentração: Matemática Aplicada e Computacional
Titulação: Mestre em Matemática Aplicada e Computacional
Banca examinadora:
Sandra Augusta Santos [Orientador]
Douglas Soares Gonçalves
Maria Aparecida Diniz Ehrhardt
Data de defesa: 27-04-2017
Programa de Pós-Graduação: Matemática Aplicada e Computacional

Dissertação de Mestrado Profissional defendida em 27 de abril de 2017 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). SANDRA AUGUSTA SANTOS

Prof(a). Dr(a). MARIA APARECIDA DINIZ EHRHARDT

Prof(a). Dr(a). DOUGLAS SOARES GONÇALVES

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

Aos meus pais, à minha esposa Cleidiane e aos meus filhos Davi e Estêvão. E sobretudo a Deus, que me deu a vida e a oportunidade de estudar e de sempre querer ir além.

"Somos na realidade o que somos diante de Deus."

São Francisco de Assis.

Agradecimentos

À minha esposa Cleidiane, a quem amo muito, por toda compreensão e força dadas durante esse estudo.

Aos meus filhos, Davi e Estêvão, que são as jóias da minha vida.

A meus pais e irmãos, que em tudo colaboraram para a conclusão desta conquista.

À minha orientadora Sandra Augusta Santos, por gentilmente me aceitar como aluno e por todo apoio e suporte dados ao longo dessa dissertação.

À banca pelas valiosas sugestões e trabalho dedicados à avaliação do presente estudo.

Ao programa institucional de bolsas de qualificação de servidores do Instituto Federal de Goiás.

E sobretudo a Deus, por sua misericórdia infinita que está sempre sobre todos nós.

Resumo

Neste trabalho estudamos alguns métodos iterativos para resolver o problema de minimização irrestrita de funções suaves, entre os quais o Método da Máxima Descida, o Método de Gradientes Conjugados Lineares e o Método de Gradientes Conjugados não Lineares. Neste último analisamos geometricamente, em problemas bidimensionais, a variabilidade dos resultados em decorrência do ponto inicial, da precisão requerida e da escolha dos parâmetros associados à busca linear. Também observamos que os escalares $\beta_{(i)}$, responsáveis pela conjugação das direções geradas, coincidem quando a função é quadrática estritamente convexa. Já no caso não linear geral, de classe C^1, analisamos quatro fórmulas para os escalares que produzem resultados distintos, ilustrados com o problema tridiagonal de Broyden, em instâncias com dimensão variável entre 2 e 45.

Palavras-chave: otimização irrestrita. gradientes conjugados lineares. gradientes conjugados não lineares.

Abstract

This work describes a study on iterative methods for solving the unconstrained minimization of smooth functions. More specifically, we address the steepest descent method, the linear conjugate gradient method and the nonlinear conjugate gradient method. A geometrical study of bidimensional problems is provided, illustrating distinct results as a consequence of the initial point, the demanded precision and the parameters that define the line search. Additionally, we have prepared a study involving four different expressions for the scalar in charge for maintaining the conjugacy of the directions. Although the results are equivalent in the linear case, there are significant differences in the nonlinear case, illustrated with several instances of the Broyden tridiagonal problem, with dimension from 2 up to 45.

Keywords: unconstrained optimization. linear conjugate gradients. nonlinear conjugate gradients.

Lista de ilustrações

Lista de tabelas

Lista de símbolos

$\mathbb{R}^{m\times n}$	Espaço vetorial das matrizes de ordem $m \times n$.
$\mathbf{I}_n$	Matriz identidade de ordem n.
$A^{n\times n}$	Matriz quadrada de ordem n.
$A^{m\times n}$	Matriz retangular de ordem $m \times n$.
x^T	Transposto do vetor x.
$x^T y$	Produto interno entre os vetores x e y.
$\|x\|$	Norma Euclidiana do vetor x.

Sumário

Introdução

Neste trabalho estudamos o Método dos Gradientes Conjugados para otimização irrestrita, desde sua fundamentação teórica, definição do algoritmo propriamente dito para minimização de uma função quadrática estritamente convexa e extensão do método para funções não lineares gerais, de classe C^1.

Assim, o nosso objetivo é resolver o problema:

$$\begin{aligned} &\text{minimizar} \quad f(x) \\ &\text{sujeito a } x \in \mathbb{R}^n. \end{aligned} \tag{1}$$

É conhecido que há vários outros métodos iterativos para a resolução do problema acima. Como exemplo, temos o Método da Máxima Descida, estudado no Capítulo 1 e o Método de Newton, que não será abordado. Vale dizer, no entanto, que o Método de Newton, ao fazer uso das derivadas segundas e necessitar da resolução de um sistema linear a cada iteração, acaba tornando-se caro do ponto de vista computacional, embora, por utilizar informações de segunda ordem, possa convergir com maior rapidez. Por outro lado, o Método da Máxima Descida, como veremos, é menos custoso, pois só utiliza derivadas primeiras, mas não é eficiente. Já o Método dos Gradientes Conjugados, objeto deste trabalho, é um método que não é dispendioso pois faz uso de produtos matriz-vetor ao invés de fatorações e com ele temos certeza que encontraremos a solução em até n passos, se a função quadrática for estritamente convexa.

Estudamos as propriedades de direções conjugadas, a partir das quais se faz uma busca linear. Recaímos, assim, a cada iteração do método, em um problema de minimização unidirecional que pode ser abordado, por exemplo, pela busca inexata envolvendo a Condição de Armijo. Cabe ressaltar que as propriedades de convergência do método de Gradientes Conjugados dependem da eficácia da busca unidirecional. De fato, a convergência em n passos do caso quadrático emprega busca linear exata.

Veremos também que há outras variantes para a escolha de um escalar essencial utilizado para definir a direção do método, tais como as propostas por (POLAK; RIBIERE, 1969) e (POLYAK, 1969); (FLETCHER; REEVES, 1964); (DAI; YUAN, 1999) e (HESTENES; STIEFEL, 1952). Para o caso em que a função é quadrática estritamente convexa, todas as variantes coincidem, mas para funções não lineares gerais, as diferentes escolhas produzem resultados distintos.

O Método dos Gradientes Conjugados começou a ser estudado na década de 50, sendo que foi proposto como um método direto para resolver sistemas lineares, não como um método iterativo. Muitos autores citam o artigo de (HESTENES; STIEFEL,

1952) como referência base ao que foi desenvolvido posteriormente.

Segundo (SHEWCHUK, 1994) o Método de Gradientes Conjugados foi abandonado durante a década de 60, e apenas nos anos 70 foi retomado pela comunidade científica. O método é utilizado principalmente para a resolução de problemas de grande porte, nos quais não é razoável sequer fazer as n iterações que podem ser demandadas no caso quadrático estritamente convexo.

No Capítulo 1 falaremos sobre o Método da Máxima Descida bem como apresentaremos a construção do seu algoritmo, a noção de direções de descida e as buscas exata e inexata. No Capítulo 2 discorreremos sobre o Método dos Gradientes Conjugados Lineares e apresentaremos alguns testes computacionais. Já no Capítulo 3 falaremos sobre o Método de Gradientes Conjugados não Lineares e faremos uma discussão sobre os exemplos apresentados. O Capítulo 4 contém nossas considerações finais, seguida das referências utilizadas e de um Apêndice, no qual incluímos explicitamente os algoritmos descritos e utilizados neste trabalho.

0.1 Notações e conceitos preliminares

Apresentaremos a seguir algumas definições que serão usadas ao longo dos próximos capítulos.

Seja um ponto $x^* \in \mathbb{R}^n$. Dizemos que x^* é ponto crítico de uma função $f \in C^1$ se $\nabla f(x^*) = 0$. Dizemos que x^* é minimizador global de (1), se $f(x^*) \leqslant f(x),\ \forall x \in \mathbb{R}^n$. Dizemos que x^* é minimizador local de (1), se existe uma vizinhança $\mathbb{U}$ de x^* tal que $f(x^*) \leqslant f(x),\ \forall x \in \mathbb{U}$. Dizemos que x^* é ponto de sela para o problema (1) se x^* é ponto crítico, mas não é nem minimizador local nem maximizador local (definição análoga, trocando a desigualdade por $\geqslant$).

Dizemos que uma matriz $A \in \mathbb{R}^{n\times n}$ é definida positiva se, para todo vetor $x \in \mathbb{R}^n$ não nulo, a seguinte desigualdade for satisfeita

$$x^T Ax > 0. \tag{2}$$

Por exemplo, no plano, geometricamente falando, seja uma função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x) = x^T Ax$. Se A é uma matriz definida positiva então seu gráfico será um parabolóide com concavidade voltada para cima, e o seu ponto de mínimo será o seu vértice.

Definimos o produto interno entre os vetores x e $y \in \mathbb{R}^n$, e escrevemos $x^T y$, como sendo $x^T y = \sum_{i=1}^{n}(x_i)(y_i)$; é imediato vermos que $x^T y = y^T x$.

Sejam as matrizes A e $B \in \mathbb{R}^{n\times n}$, podemos verificar as seguintes igualdades $(AB)^T = B^T A^T$, veja por exemplo (KÜHLKAMP, 2005), página 35. E também, se

ambas forem invertíveis, $(AB)^{-1} = B^{-1}A^{-1}$. De fato, basta notar que $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A(\mathbf{I}_n)A^{-1} = \mathbf{I}_n$ e $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}(\mathbf{I}_n)B = \mathbf{I}_n$, e como a inversa de uma matriz é única, logo está verificada a igualdade, veja por exemplo (BOLDRINI et al., 1980), página 75.

Ao longo deste texto utilizaremos a notação a seguir para nos referirmos ao gradiente de uma função $f(x)$ de classe C^1,

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial}{\partial x_1} f(x) \\ \dfrac{\partial}{\partial x_2} f(x) \\ \vdots \\ \dfrac{\partial}{\partial x_n} f(x) \end{bmatrix}.$$

Uma vez estabelecidos estes conceitos preliminares, podemos agora estudar o Método da Máxima Descida para quadráticas estritamente convexas.

1 Máxima Descida

1.1 Preâmbulo

Faremos neste capítulo um estudo detalhado sobre o Método da Máxima Descida, que é um método iterativo[1] para minimizar irrestritamente funções suaves. Quando aplicado a uma função quadrática estritamente convexa, reduz-se a um método para resolver sistemas lineares da forma:

$$Ax = b, \tag{1.1}$$

onde $A \in \mathbb{R}^{n\times n}$ é uma matriz simétrica e definida positiva, $x \in \mathbb{R}^n$ é o vetor solução e $b \in \mathbb{R}^n$ é o vetor de termos independentes.

O sistema (1.1) surge naturalmente ao se tentar minimizar a função quadrática estritamente convexa dada a seguir, conforme (NOCEDAL; WRIGHT, 2006). Chamamos de função quadrática estritamente convexa a toda função $f : \mathbb{R}^n \to \mathbb{R}$ escrita na forma:

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \tag{1.2}$$

com $A \in \mathbb{R}^{n\times n}$, simétrica e definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

De fato, seja

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \text{ e } c \in \mathbb{R}.$$

Substituindo em (1.2) temos:

$$f(x) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c,$$

e fazendo algumas multiplicações e agrupando convenientemente os termos chegamos em:

$$\begin{aligned} f(x) = (1/2)[a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + a_{21}x_1x_2 + a_{22}x_2^2 + \cdots + \\ a_{2n}x_2x_n + \cdots + a_{n1}x_1x_n + a_{n2}x_2x_n + \cdots + a_{nn}x_n^2] - b_1x_1 - b_2x_2 - \cdots - b_nx_n + c \end{aligned} \tag{1.3}$$

[1] Método iterativo é um procedimento que gera uma sequência de soluções aproximadas que vão melhorando conforme iterações são executadas.

e derivando (1.3) segue:

$$\nabla f(x) = \begin{bmatrix} a_{11}x_1 + \frac{a_{12}x_2}{2} + \cdots + \frac{a_{1n}x_n}{2} + \frac{a_{21}x_2}{2} + \cdots + \frac{a_{n1}x_n}{2} + \cdots & -b_1 \\ \frac{a_{12}x_1}{2} + \frac{a_{21}x_1}{2} + a_{22}x_2 + \cdots + \frac{a_{2n}x_n}{2} + \cdots + \frac{a_{n2}x_n}{2} + \cdots & -b_2 \\ \vdots & \\ \frac{a_{1n}x_1}{2} + \cdots + \frac{a_{2n}x_2}{2} + \cdots + \frac{a_{n1}x_1}{2} + \cdots + a_{nn}x_n + \cdots & -b_n \end{bmatrix}$$

que pode ser escrita como:

$$\nabla f(x) = \frac{1}{2}\begin{bmatrix} a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n \end{bmatrix} + \frac{1}{2}\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

ou

$$\nabla f(x) = \frac{1}{2}\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \frac{1}{2}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

consequentemente

$$\nabla f(x) = \frac{1}{2}A^T x + \frac{1}{2}Ax - b. \tag{1.4}$$

Portanto de (1.2) concluímos (1.4), e como A é uma matriz simétrica, isto é, $A = A^T$, temos:

$$\nabla f(x) = Ax - b, \tag{1.5}$$

e igualando a derivada de (1.5) a zero temos o sistema (1.1).

No resultado a seguir mostramos que o fato da matriz A ser simétrica e definida positiva implica que resolver o sistema (1.1) é equivalente a minimizarmos a função dada em (1.2).

Teorema 1.1. *Se $p \in \mathbb{R}^n$ é um ponto qualquer e f é a função dada em (1.2), com A simétrica e definida positiva, então $f(x) < f(p)$, onde x é solução de $Ax = b$.*

Demonstração. Seja p um ponto arbitrário pertencendo ao domínio de f e x a solução de (1.1), vamos mostrar que $f(x) < f(p)$, isto é, x é o minimizador global de f.
Seja A uma matriz simétrica e definida positiva, e seja e o erro que se comete ao se

aproximar de x, isto é, $e = p - x \Rightarrow p = x + e$. Aplicando em (1.2) temos as seguintes igualdades

$$\begin{aligned} f(x+e) &= \frac{1}{2}(x+e)^T A(x+e) - b^T(x+e) + c \\ &= \frac{1}{2}(x^T + e^T)(Ax + Ae) - b^T x - b^T e + c \\ &= \frac{1}{2}x^T Ax + \frac{1}{2}x^T Ae + \frac{1}{2}e^T Ax + \frac{1}{2}e^T Ae - b^T x - b^T e + c \\ &= \frac{1}{2}x^T Ax + -b^T x + c + \frac{1}{2}x^T Ae + \frac{1}{2}e^T Ax + \frac{1}{2}e^T Ae - b^T e, \end{aligned}$$

e como $Ax = b$, segue que

$$f(x+e) = \frac{1}{2}x^T Ax - b^T x + c + \frac{1}{2}x^T Ae + \frac{1}{2}e^T b + \frac{1}{2}e^T Ae - b^T e$$

e de $(x^T Ae)^T = e^T A^T x = e^T Ax = e^T b$, temos

$$f(x+e) = \frac{1}{2}x^T Ax - b^T x + c + e^T b + \frac{1}{2}e^T Ae - b^T e.$$

De $e^T b = b^T e$, temos:

$$f(x+e) = \frac{1}{2}x^T Ax - b^T x + c + \frac{1}{2}e^T Ae,$$

isto é:

$$f(x+e) = f(x) + \frac{1}{2}e^T Ae,$$

como $e^T Ae > 0$, $\forall e \neq 0$, pois A é definida positiva, logo x minimiza f, isto é, $f(p) > f(x)$. □

Assim podemos tanto resolver o sistema (1.1) ou minimizar a função dada em (1.2) que chegaremos a mesma solução.

Cabe ressaltar que o valor de c em (1.2) não altera seu ponto ótimo, que é a solução exata, mas apenas o valor ótimo. Assim sendo o que de fato é relevante são as matrizes A e b.

1.2 Busca Exata - Descrição do Método da Máxima Descida

Sabemos que dentre todas as direções ao longo das quais a função f cresce, a direção do gradiente é a de crescimento máximo, veja por exemplo (LIMA, 1976), página 140. Logo, se queremos minimizar a função dada em (1.2) é bastante natural tomar a direção oposta ao gradiente. O Método da Máxima Descida é um método iterativo, em que

uma sequência de pontos $\{x_{(i)}\}$ é gerada. Para produzir tal sequência, anda-se na direção oposta à do gradiente, com um controle de passo, conforme veremos mais adiante.

Temos de (1.5) que

$$-\nabla f(x) = b - Ax,$$

e chamamos de resíduo no i-ésimo iterando ao vetor

$$r_{(i)} = b - Ax_{(i)} = -\nabla f(x_{(i)}),$$

ou seja, o resíduo será a direção de Máxima Descida. Chamamos de erro na iteração i à diferença,

$$e_{(i)} = x_{(i)} - x^*,$$

onde x^* denota a solução do problema, e indica quão longe estamos da solução. Logo, se o erro for nulo, significa que estamos na solução do problema. Podemos ver também de

$$e_{(i)} = x_{(i)} - x^* \Rightarrow -Ae_{(i)} = -Ax_{(i)} + Ax^* \Rightarrow r_{(i)} = -Ae_{(i)}.$$

Podemos, portanto, pensar no resíduo como o erro que se comete quando se transforma o erro dado no domínio, por A, no mesmo espaço que contém b, veja por exemplo, (SHEWCHUK, 1994).

À medida que a sequência $\{x_{(i)}\}$ é gerada, fazemos um controle para verificar se a norma do resíduo $r_{(i)}$ é suficientemente pequena, de acordo com uma tolerância pré estabelecida. Esta tolerância estabelece um limiar de aceitação para um ponto que se aproxime da solução do problema.

Para compreendermos bem este método, o apresentaremos com um exemplo.

Exemplo 1.1. Tomaremos $n = 2$ e

$$A = \begin{bmatrix} 4 & 3 \\ 3 & 12 \end{bmatrix}, \; b = \begin{bmatrix} -3 \\ 1 \end{bmatrix} \text{ e } c = 5.$$

Escrito na forma de (1.2), a função quadrática fica

$$f(x) = 2x_1^2 + 3x_1x_2 + 6x_2^2 + 3x_1 - x_2 + 5. \tag{1.6}$$

Derivando (1.6) e igualando a zero, chegamos ao seguinte sistema:

$$\begin{cases} 4x_1 + 3x_2 = -3 \\ 3x_1 + 12x_2 = 1, \end{cases} \tag{1.7}$$

cuja solução é $(x_1, x_2)^T = (-1, 1/3)^T$. Tome agora como ponto de partida $x_{(0)} = (8, 0)^T$ e tolerância de 10^{-5}.

Assim, o nosso objetivo, é encontrar uma sequência de valores $x_{(1)}, x_{(2)}, ..., x_{(i)}$ que tende à solução do Exemplo 1.1, de tal modo que valha a seguinte inequação modular $\|r_{(i)}\| \leqslant 10^{-5}$. Como a sequência de vetores e obtida iterativamente, temos

$$x_{(i+1)} = x_{(i)} + \alpha_{(i)} r_{(i)}, \tag{1.8}$$

onde o escalar $\alpha_{(i)} \in \mathbb{R}^+$ é o tamanho do passo e $r_{(i)}$ é a direção que devemos tomar na i-ésima iteração. Como o Método da Máxima Descida sempre toma a direção oposta ao gradiente no ponto, basta apenas sabermos qual deve ser o tamanho do passo $\alpha_{(i)}$. Vamos interromper brevemente a apresentação do Exemplo 1.1, para deduzirmos uma fórmula explícita para o tamanho do passo.

Para encontrarmos o tamanho do passo, que na verdade será o melhor passo possível, por isso chamamos de busca exata, vamos minimizar a função a valores reais $\varphi(\alpha) = f(x_{(i)} + \alpha r_{(i)})$. Sabemos que uma condição necessária para minimizarmos uma função escalar é igualarmos a sua derivada a zero, e aplicando a regra da cadeia temos (GUIDORIZZI, 2000):

$$\frac{d}{d\alpha}\varphi(\alpha) = \nabla f(x_{(i)} + \alpha r_{(i)})^T \frac{d}{d\alpha}(x_{(i)} + \alpha r_{(i)}) = \nabla f(x_{(i)} + \alpha r_{(i)})^T r_{(i)} = 0,$$

de onde seguem as igualdades

$$\begin{aligned}
(b - A(x_{(i)} + \alpha r_{(i)}))^T r_{(i)} &= 0 \\
(b - Ax_{(i)})^T r_{(i)} - \alpha (Ar_{(i)})^T r_{(i)} &= 0 \\
-\alpha (Ar_{(i)})^T r_{(i)} &= -(b - Ax_{(i)})^T r_{(i)} \\
\alpha r_{(i)}^T (Ar_{(i)}) &= r_{(i)}^T r_{(i)} \\
\alpha &= \frac{r_{(i)}^T r_{(i)}}{r_{(i)}^T A r_{(i)}}.
\end{aligned}$$

Portanto, o Método da Máxima Descida com busca exata é dado por:

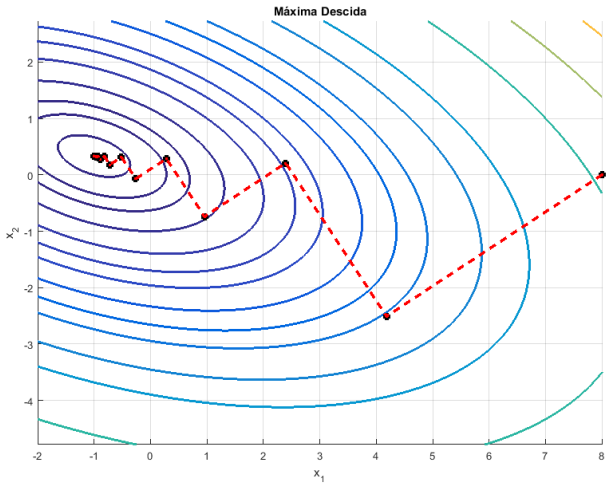
$$r_{(i)} = b - Ax_{(i)}, \tag{1.9}$$

$$\alpha_{(i)} = \frac{r_{(i)}^T r_{(i)}}{r_{(i)}^T A r_{(i)}}, \tag{1.10}$$

$$x_{(i+1)} = x_{(i)} + \alpha_{(i)} r_{(i)}.$$

Na dedução acima vimos que $r_{(i+1)}^T r_{(i)} = 0$, ou seja, $\nabla f(x_{(i+1)}) \perp \nabla f(x_{(i)})$. Assim duas direções consecutivas no Método da Máxima Descida com busca linear exata são perpendiculares, o que produz o seu caracteristico "zigue-zague", veja Figura 1.

Figura 1 – Ilustração da trajetória dos iterados do Método de Máxima Descida para o Exemplo 1.1.



Agora podemos voltar ao nosso Exemplo 1.1. Calculando a primeira iteração temos:

$$r_{(0)} = \underbrace{\begin{bmatrix} -3 \\ 1 \end{bmatrix}}_{b} - \underbrace{\begin{bmatrix} 4 & 3 \\ 3 & 12 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 8 \\ 0 \end{bmatrix}}_{x_{(0)}} = \begin{bmatrix} -35 \\ -23 \end{bmatrix},$$

$$\alpha_{(0)} = \frac{r_{(0)}^T r_{(0)}}{r_{(0)}^T A r_{(0)}} \simeq 0.1090$$

e

$$x_{(1)} = x_{(0)} + \alpha_{(0)} r_{(0)} \simeq \begin{bmatrix} 4.1817 \\ -2.5091 \end{bmatrix}.$$

Como $\|r_{(0)}\| \simeq 41.8807 \geqslant 10^{-5}$, logo o processo deve continuar.

Tabela 1 – Algumas iterações do Exemplo 1.1 (considerando seis casas decimais).

Iteração $_{(i)}$	**Vetor** $x_{(i)}$	**Resíduo**
1	$[4.181739 \quad -2.509142]^T$	22.214163
2	$[2.396491 \quad 0.207537]^T$	15.805304
3	$[0.955526 \quad -0.739383]^T$	8.383358
⋮	⋮	⋮
10	$[-0.931105 \quad 0.330781]^T$	0.320593
11	$[-0.960334 \quad 0.311574]^T$	0.170047
⋮	⋮	⋮
29	$[-0.999993 \quad 0.333329]^T$	2.640360×10^{-5}
30	$[-0.999995 \quad 0.333333]^T$	1.878607×10^{-5}
31	$[-0.999997 \quad 0.333332]^T$	9.964402×10^{-6}

Vemos da Tabela 1 que o processo termina na 31^a iteração, pois $\|r_{(31)}\| \leqslant 10^{-5}$

É bom enfatizar que o Método da Máxima Descida não é eficiente quando a matriz A do sistema (1.1) é mal condicionada.

Definição 1.1. *Sejam λ_{min} e λ_{max}, respectivamente, o menor e o maior autovalor da matriz A. Definimos o número de condição espectral κ da matriz A como o quociente*

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}}. \tag{1.11}$$

O fato de uma matriz ser mal condicionada está intimamente relacionado com o número de condição espectral κ assumir valores grandes, veja por exemplo (WATKINS, 2002). Para ilustrarmos essa afirmação e o mau desempenho do método de máxima descida neste caso, vejamos outro exemplo.

Exemplo 1.2. Tomaremos agora um exemplo com $n = 3$. Sejam

$$A = \begin{bmatrix} 11 & 30 & 0 \\ 30 & 600 & 0 \\ 0 & 0 & 1 \end{bmatrix} \; e \; b = \begin{bmatrix} 47 \\ -390 \\ 3 \end{bmatrix}.$$

A solução exata do sistema $Ax = b$ é $x^* = (x_1, x_2, x_3)^T = (7, -1, 3)^T$. Porém, aplicando o Método da Máxima Descida neste exemplo, e considerando o ponto inicial $x_{(0)} = (0, 0, 0)^T$ e Tol $= 10^{-5}$, temos que são necessárias 537 iterações, veja a Tabela 2 onde foram exibidas apenas as iterações relevantes para o exemplo.

Tabela 2 – Algumas iterações do Exemplo 1.2 (considerando seis casas decimais).

Iteração $_{(i)}$	**Vetor** $x_{(i)}$	**Resíduo**
100	$[7.000130 \quad -0.997440 \quad 2.703749]^T$	1.569946
⋮	⋮	⋮
150	$[7.000039 \quad -0.999217 \quad 2.909489]^T$	0.479650
⋮	⋮	⋮
250	$[7.000003 \quad -0.999927 \quad 2.991551]^T$	0.044771
⋮	⋮	⋮
350	$[7.000000 \quad -0.999993 \quad 2.999211]^T$	0.004179
⋮	⋮	⋮
536	$[7.000000 \quad -0.999999 \quad 2.999990]^T$	5.074997×10^{-5}
537	$[6.999999 \quad -1.000000 \quad 2.999990]^T$	9.734970×10^{-6}

Observe que esta matriz é mal condicionada, pois seus autovalores são $\lambda_1 = 1$, $\lambda_2 \simeq 9.47$ e $\lambda_3 \simeq 601.52$, logo $\kappa = \dfrac{601.52}{1} = 601.52$.

Com este exemplo podemos perceber que a convergência na aplicação do Método da Máxima Descida fica extremamente comprometida. O mau condicionamento faz com

que o progresso da sequência de iterandos fique bastante lento, assim como a diminuição do resíduo, observe isto na Tabela 2. Vale ainda mencionar que, no caso geral em $\mathbb{R}^n$, à medida que a razão (1.11) cresce, as hipersuperfícies de nível da quadrática estritamente convexa cuja Hessiana é a matriz A ficam mais e mais alongadas, de maneira que o fenômeno do "zigue-zague" se acentua. Por outro lado, se $\lambda_{\min} = \lambda_{\max}$, isto é, se $\kappa = 1$, então o Método da Máxima Descida com busca exata converge em uma iteração, independentemente da dimensão do problema. Retomaremos essa questão no final da Seção 2.2.

1.3 Busca Inexata - Condição de Armijo

Com base no livro de (MARTÍNEZ; SANTOS, 1995), falaremos agora sobre um método para resolver o problema geral de minimização sem restrições

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeito a } x \in \mathbb{R}^n, \end{aligned}$$

em que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função suave.

Para definirmos esse método tomaremos direções ao longo das quais a função $f(x)$ decresce, isto é, queremos escolher pontos tais que $f(x_{(i+1)}) < f(x_{(i)})$, desde que $\nabla f(x_{(i)}) \neq 0$. Assim, falaremos sobre direção de descida. Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida para f, a partir de $x \in \mathbb{R}^n$, se existir $\epsilon > 0$ tal que, para todo $t \in (0, \epsilon]$ tivermos

$$f(x + td) < f(x). \tag{1.12}$$

O próximo teorema nos mostra que direções que formam um ângulo maior que 90^0 com o gradiente são direções de descida.

Teorema 1.2. *Se $\nabla f(x)^T d < 0$ então d é direção de descida para f, a partir de x.*

Demonstração. Sabemos que $\nabla f(x)^T d = \dfrac{\partial f}{\partial d}(x) = \lim_{t \to 0} \dfrac{f(x + td) - f(x)}{t}$, e como por hipótese $\nabla f(x)^T d < 0$, então para todo $t > 0$ suficientemente pequeno, temos que $\dfrac{f(x + td) - f(x)}{t} < 0$, isto é, $f(x + td) < f(x)$. □

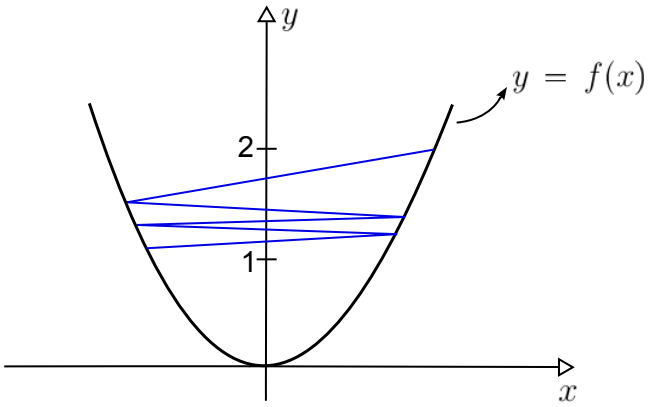
No Apêndice A.1 apresentamos um algoritmo geral utilizando direções de descida. Porém, a direção ser apenas de descida não é suficiente para que tal algoritmo, aplicado ao nosso problema de minimização de f, encontre o ponto mínimo. De fato, o algoritmo dado no Apêndice A.1 é impotente até para nos conduzir a pontos estacionários, isto é, aqueles nos quais a derivada se anula.

Ilustramos tal afirmação no seguinte exemplo.

Exemplo 1.3. Considere a função escalar de uma variável real $f(x) = x^2$, cujo mínimo é $x^* = 0$. Temos que a imagem da sequência $x_{(i)} = (-1)^i(1 + 2^{-i})$ converge para 1, que não é o seu mínimo. Veja também que a sequência $f(x_{(i)})$ é descrescente pois,

$$f(x_{(i+1)}) = \left[(-1)^{(i+1)}\left(1 + \frac{1}{2^{i+1}}\right)\right]^2 = (-1)^{2i+2}\left(1 + \frac{2}{2^{i+1}} + \frac{1}{2^{2i+2}}\right) = 1 + \frac{2}{2^{i+1}} + \frac{1}{2^{2(i+1)}}$$
$$< 1 + \frac{2}{2^i} + \frac{1}{2^{2i}} = (-1)^{2i}\left(1 + \frac{2}{2^i} + \frac{1}{2^{2i}}\right) = \left[(-1)^i\left(1 + \frac{1}{2^i}\right)\right]^2 = f(x_{(i)}).$$

Figura 2 – Gráfico da função $f(x) = x^2$ com alguns elementos da sequência $\{(x_{(i)}, f(x_{(i)}))\}$ gerada no Exemplo 1.3.



Assim sendo devemos procurar uma alternativa para impedir que o problema detectado no Exemplo 1.3 ocorra. Podemos conseguir um decréscimo suficiente ao longo das direções de descida, por meio da denominada condição de Armijo, conforme (MARTÍNEZ; SANTOS, 1995).

Seja então um ponto $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ uma direção de descida a partir de x e $c_1 \in (0, 1)$. A condição de Armijo consiste em encontrar um escalar $t > 0$ tal que seja válida a seguinte desigualdade

$$f(x + td) \leqslant f(x) + c_1 t \nabla f(x)^T d. \tag{1.13}$$

Para funções não lineares gerais, não temos uma fórmula fechada para o tamanho do passo, como a obtida em (1.10). Mesmo que f seja uma quadrática estritamente convexa, mas mal condicionada, a condição de Armijo pode ser uma alternativa interessante, pois, esta procura uma boa redução da função objetivo, sem tentar de fato minimizá-la. Na verdade, a redução é proporcional ao tamanho do passo, fato este que caracteriza a busca inexata. Logo, no método da busca inexata queremos apenas que a função objetivo tenha uma redução suficiente e esta redução se dá através do ajuste do tamanho do passo. Explicando melhor, na busca exata, como no Método da Máxima Descida para quadráticas

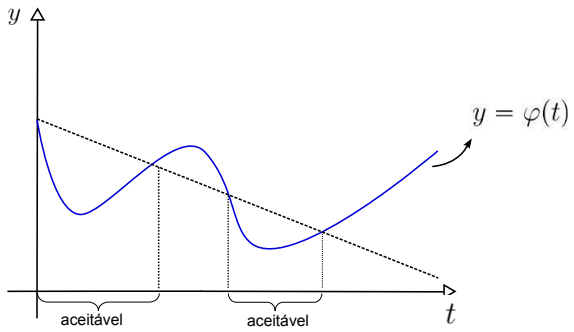
estritamente convexas, nós conseguíamos o passo ideal, dado em (1.10). Já na busca inexata não nos preocupamos em conseguir um passo ideal, mas apenas em encontrar um valor para o passo que já seja suficiente para satisfazer a condição de Armijo. Na implementação da busca inexata com a condição de Armijo começamos com $t = 1$, e se necessário, diminuimos sistematicamente o valor de t até que (1.13) seja verificado, veja por exemplo (RIBEIRO; KARAS, 2014).

Ao longo do texto usaremos o símbolo α para indicarmos que estamos fazendo uma busca exata, e o símbolo t para indicarmos que estamos fazendo uma busca inexata.

Tanto o Método da Máxima Descida com busca exata para funções quadráticas estritamente convexas quanto o Método da Máxima Descida com busca inexata para funções suaves não lineares, via condição de Armijo, são globalmente convergentes, veja por exemplo (MARTÍNEZ; SANTOS, 1995). Isto significa que não importa qual seja o ponto inicial que tomemos na primeira iteração, o método sempre convergirá para um ponto estacionário do problema, isto é, um ponto no qual o gradiente se anula.

No gráfico da Figura 3, podemos ver os intervalos admissíveis para a condição de Armijo. Entretanto, como no Exemplo 1.3, o decréscimo produzido em f pode ser tão lento que apenas a condição de descida não é suficiente para encontrarmos um minimizador para f, isto porque podemos tomar, prematuramente, passos extremamente pequenos.

Figura 3 – Intervalos admissíveis para a Condição de Armijo.



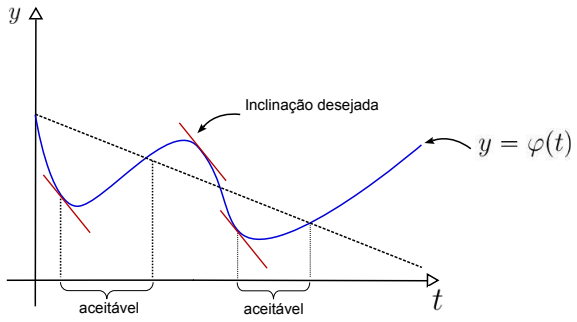
Logo devemos procurar uma alternativa para evitar que isso ocorra, uma possibilidade é a condição de curvatura ou condição de Wolfe. A condição de Wolfe pede que, dada uma direção de descida d, para algum $c_2 \in (c_1, 1)$, em que $c_1 \in (0, 1)$ é a constante de Armijo, seja válida a desigualdade:

$$\nabla f(x + td)^T d \geqslant c_2 \nabla f(x)^T d. \tag{1.14}$$

Se tomarmos o valor absoluto em (1.14) teremos a chamada condição forte de Wolfe, que é:

$$|\nabla f(x + td)^T d| \leqslant c_2 |\nabla f(x)^T d|. \tag{1.15}$$

Figura 4 – Intervalos admissíveis para a Condição de Wolfe, conjuntamente à Condição de Armijo.



Em (NOCEDAL; WRIGHT, 2006) são sugeridos para as constantes c_1 da condição de Armijo e c_2 da condição de Wolfe, os valores $c_1 = 10^{-4}$ e $c_2 = 0.1$.

O algoritmo para efetuar a busca inexata com a condição de Armijo conjuntamente com Wolfe forte está no Apêndice A.2. Veja que deste modo, estamos usando o processo de *backtracking*, mas que há outras estratégias disponíveis, por exemplo, envolvendo interpolação. Para mais detalhes, sugerimos as referências (DENNIS JR; SCHNABEL, 1996) e (NOCEDAL; WRIGHT, 2006).

Agora podemos dar início ao estudo do método de Gradientes Conjugados Lineares, assunto este que será tratado no próximo capítulo.

2 Gradientes Conjugados Lineares

cap2

2.1 Método de Gradientes Conjugados Lineares

O primeiro ingrediente essencial para estabelecermos o Método de Gradientes Conjugados Lineares é a definição de vetores A-conjugados, dada a seguir.

Definição 2.1. *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz e simétrica definida positiva. Dizemos que os vetores não nulos $d_{(0)}, d_{(1)}, \ldots, d_{(k)} \in \mathbb{R}^n$ são A-conjugados se*

$$d_{(i)}^T A d_{(j)} = 0, \tag{2.1}$$

para todo $i, j = 0, 1, \ldots, k$, com $i \neq j$. (RIBEIRO; KARAS, 2014).

Por exemplo, os vetores

$$d_{(0)} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \text{ e } d_{(1)} = \begin{bmatrix} -3 \\ 3/2 \end{bmatrix}$$

são A-conjugados se considerarmos a matriz

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}.$$

De fato, pois

$$d_{(0)}^T A d_{(1)} = \begin{bmatrix} 2 & 5 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -3 \\ 3/2 \end{bmatrix} = 0.$$

Observe que se $A = \mathbf{I}_n$, então os vetores A-conjugados são ortogonais no sentido usual.

O próximo teorema nos diz que vetores A-conjugados são linearmente independentes, o que é extremamente relevante, para evitar que tomemos direções repetidas, como ocorre no Método da Máxima Descida para minimizar funções quadráticas estritamente convexas.

Teorema 2.1. *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz definida positiva. Um conjunto qualquer de vetores não nulos A-conjugados é linearmente independente.*

Demonstração. Sejam $d_{(0)}, d_{(1)}, \ldots, d_{(k)} \in \mathbb{R}^n$ vetores não-nulos A-conjugados, e considere as constantes $a_0, a_1, \ldots, a_k \in \mathbb{R}$ na combinação linear:

$$a_0 d_{(0)} + a_1 d_{(1)} + \ldots + a_k d_{(k)} = 0.$$

Dado $i \in \{0, 1, \ldots, k\}$, multiplicando os dois membros da igualdade acima por $(d_{(i)}^T A)$, e usando o fato que os vetores são A-conjugados, temos:

$$(d_{(i)}^T A)(a_0 d_{(0)} + a_1 d_{(1)} + \ldots + a_k d_{(k)}) = (d_{(i)}^T A)0$$

$$a_i (d_{(i)}^T A d_{(i)}) = 0,$$

e como A é definida positiva e $d_{(i)} \neq 0$, segue que $a_i = 0$, portanto os vetores são linearmente independentes. □

Dado um ponto inicial $x_{(0)} \in \mathbb{R}^n$ e um conjunto de direções conjugadas $\{d_{(0)}, d_{(1)}, \ldots, d_{(n-1)}\}$, utilizaremos a relação de recorrência

$$x_{(i+1)} = x_{(i)} + \alpha_{(i)} d_{(i)}. \tag{2.2}$$

Vamos estabelecer um processo iterativo no qual todas as direções de busca sejam A-conjugadas. O nosso primeiro objetivo é determinar $\alpha_{(i)} \in \mathbb{R}^+$ em (2.2), e para isso, vamos supor que $e_{(i+1)}$ seja ortogonal a $d_{(i)}$, com o intuito de esgotar a direção $d_{(i)}$ e evitar o "zigue-zague" da máxima descida. Assim temos:

$$\begin{aligned}
d_{(i)}^T e_{(i+1)} &= 0 \\
d_{(i)}^T (e_{(i)} + \alpha_{(i)} d_{(i)}) &= 0 \\
d_{(i)}^T e_{(i)} + \alpha_{(i)} d_{(i)}^T d_{(i)} &= 0 \\
\alpha_{(i)} d_{(i)}^T d_{(i)} &= -d_{(i)}^T e_{(i)} \\
\alpha_{(i)} &= \frac{-d_{(i)}^T e_{(i)}}{d_{(i)}^T d_{(i)}}.
\end{aligned}$$

Mas se soubéssemos quanto vale $e_{(i)}$, saberíamos a resposta do problema de minimização, já que o termo do erro envolve o vetor solução. Para contornarmos essa dificuldade, ao invés de supor ortogonalidade, vamos pedir que $e_{(i+1)}$ e $d_{(i)}$ sejam A-conjugados, isto é,

$$\begin{aligned}
d_{(i)}^T A e_{(i+1)} &= 0 \\
d_{(i)}^T A (e_{(i)} + \alpha_{(i)} d_{(i)}) &= 0 \\
d_{(i)}^T A e_{(i)} + \alpha_{(i)} d_{(i)}^T A d_{(i)} &= 0 \\
\alpha_{(i)} d_{(i)}^T A d_{(i)} &= -d_{(i)}^T A e_{(i)} \\
\alpha_{(i)} &= \frac{-d_{(i)}^T A e_{(i)}}{d_{(i)}^T A d_{(i)}}.
\end{aligned}$$

Veja também, que $d_{(i)}$ ser A-conjugado a $e_{(i+1)}$ é o mesmo que encontrar o mínimo da função a valores reais $\varphi(\alpha) = f(x_{(i)} + \alpha d_{(i)})$. De fato, igualando a derivada a zero, segue: $\frac{d}{d\alpha} f(x_{(i)} + \alpha d_{(i)}) = \nabla f(x_{(i)} + \alpha d_{(i)})^T \frac{d}{d\alpha}(x_{(i)} + \alpha d_{(i)}) = -r_{(i+1)}^T d_{(i)} = A^T e_{(i+1)}^T d_{(i)} = d_{(i)}^T A e_{(i+1)} = 0$.

E como $r_{(i)} = -Ae_{(i)}$ da expressão anteriormente obtida para o passo temos:

$$\alpha_{(i)} = \frac{d_{(i)}^T r_{(i)}}{d_{(i)}^T A d_{(i)}}. \tag{2.3}$$

Observe que se em (2.3) tomássemos a direção $d_{(i)}$ como a do resíduo, teríamos (1.10), que é justamente o passo dado no Método da Máxima Descida.

Veja ainda que

$$\begin{aligned} r_{(i+1)} &= -Ae_{(i+1)} \\ &= -A(e_{(i)} + \alpha_{(i)} d_{(i)}) \\ &= -Ae_{(i)} - \alpha_{(i)} A d_{(i)} \\ &= r_{(i)} - \alpha_{(i)} A d_{(i)}. \end{aligned} \tag{2.4}$$

A equação (2.4) nos diz que podemos expressar o novo resíduo em termos do resíduo anterior e do produto $Ad_{(i)}$ já calculado para obter $\alpha_{(i)}$, economizando um produto matriz-vetor. Esta economia é relevante em problemas de grande porte.

Veja também que $d_{(i)}^T r_{(i)} = -\nabla f(x_{(i)})^T d_{(i)}$, pois $-d_{(i)}^T \nabla f(x_{(i)}) = d_{(i)}^T r_{(i)}$, já que $r_{(i)} = -\nabla f(x_{(i)})$. Assim, uma outra maneira de olharmos para o tamanho do passo dado em (2.3) é escrevermos como

$$\alpha_{(i)} = -\frac{\nabla f(x_{(i)})^T d_{(i)}}{d_{(i)}^T A d_{(i)}}. \tag{2.5}$$

2.2 Descrição do Método

O Método de Gradientes Conjugados sugere que, dada uma aproximação inicial $x_{(0)}$ para a solução de (1.1) tomemos um conjunto de direções conjugadas $\{d_{(0)}, d_{(1)}, \ldots d_{(n-1)}\}$, onde a primeira direção é a de Máxima Descida e todas as outras são A-conjugadas entre si. Vamos mostrar que, em no máximo n iterações, teremos encontrado a solução de (1.1). Explicando melhor, inicialmente, dado um ponto inicial $x_{(0)}$ e seu resíduo $r_{(0)}$, escolhemos a primeira direção de busca $d_{(0)} = r_{(0)}$, ao longo da qual será feita uma busca linear exata para alcançarmos um novo ponto $x_{(1)}$. Agora, de posse das informações de $d_{(0)}, r_{(0)}$ e $r_{(1)}$, calculamos uma nova direção de busca $d_{(1)}$, conjugada a $d_{(0)}$, ao longo da qual buscaremos um novo ponto $x_{(2)}$, e assim, sucessivamente, até que seja atinjida a solução de (1.1).

Iremos mostrar que o algoritmo com essas características minimiza a função quadrática estritamente convexa dada em (1.2) em no máximo n passos.

Teorema 2.2. *Considere a função quadrática dada em (1.2) e seu minimizador x^*, dado pela solução do sistema (1.1). Dados $x_{(0)} \in \mathbb{R}^n$ e um conjunto de direções A-conjugadas $\{d_{(0)}, d_{(1)}, \ldots, d_{(n-1)}\}$, a sequência definida em (2.2)-(2.3), satisfaz $x_{(n)} = x^*$.*

Demonstração. Pelo Teorema 2.1 temos que os n vetores do conjunto $\{d_{(0)}, d_{(1)}, \ldots, d_{(n-1)}\}$ são linearmente independentes, logo eles formam uma base de $\mathbb{R}^n$. Portanto, podemos escrever qualquer vetor em $\mathbb{R}^n$ como combinação linear destes vetores. Assim, existem escalares $c_{(k)} \in \mathbb{R}$, com $k \in \{0, 1, \ldots, n-1\}$ de modo que seja válida a igualdade

$$x^* - x_{(0)} = \sum_{k=0}^{n-1} c_{(k)} d_{(k)}. \tag{2.6}$$

Para $i \in \{0, 1, \ldots, n-1\}$ arbitrário, multiplique a igualdade (2.6) por $d_{(i)}^T A$, temos:

$$d_{(i)}^T A(x^* - x_{(0)}) = (d_{(i)}^T A) \sum_{k=0}^{n-1} c_{(k)} d_{(k)},$$

e como as direções são A-conjugadas, chegamos em

$$d_{(i)}^T A(x^* - x_{(0)}) = c_{(i)} d_{(i)}^T A d_{(i)},$$

o que nos conduz a

$$c_{(i)} = \frac{d_{(i)}^T A(x^* - x_{(0)})}{d_{(i)}^T A d_{(i)}} = \frac{d_{(i)}^T A x^* - d_{(i)}^T A x_{(0)}}{d_{(i)}^T A d_{(i)}}. \tag{2.7}$$

Por outro lado, pela definição de $x_{(i)}$ em (2.2)-(2.3), temos:

$$x_{(i)} = x_{(0)} + \alpha_{(0)} d_{(0)} + \alpha_{(1)} d_{(1)} + \cdots + \alpha_{(i-1)} d_{(i-1)},$$

que multiplicando por $d_{(i)}^T A$, temos $d_{(i)}^T A x_{(i)} = d_{(i)}^T A x_{(0)}$, isto pois as direções são A-conjugadas. E substituindo em (2.7) e usando (1.5), obtemos as seguintes igualdades:

$$\begin{aligned} c_{(i)} &= \frac{d_{(i)}^T A x^* - d_{(i)}^T A x_{(i)}}{d_{(i)}^T A d_{(i)}} = \frac{d_{(i)}^T (A x^* - A x_{(i)})}{d_{(i)}^T A d_{(i)}} = \frac{d_{(i)}^T (b - A x_{(i)})}{d_{(i)}^T A d_{(i)}} \\ &= -\frac{d_{(i)}^T \nabla f(x_{(i)})}{d_{(i)}^T A d_{(i)}} = -\frac{\nabla f(x_{(i)})^T d_{(i)}}{d_{(i)}^T A d_{(i)}} = \alpha_{(i)}. \end{aligned}$$

Portanto, de (2.6) segue que

$$x^* = x_{(0)} + \sum_{k=0}^{n-1} \alpha_{(k)} d_{(k)} = x_{(n)}.$$

□

Apresentaremos agora dois teoremas que nos mostrarão que a minimização consecutiva ao longo de direções A-conjugadas resulta na minimização da função quadrática estritamente convexa dada em (1.2) não apenas em uma direção, mas sim em uma variedade afim de dimensão n, dada por $x_{(0)} + [d_{(0)}, d_{(1)}, \cdots, d_{(n-1)}]$, em que usamos a notação $[v_1, \ldots, v_n]$ para representar o subespaço gerado pelos vetores $v_1, \ldots, v_n$.

Teorema 2.3. *Dados $x_{(0)} \in \mathbb{R}^n$ e um conjunto de direções A-conjugadas $\{d_{(0)}, d_{(1)}, \ldots, d_{(n-1)}\}$, considere a sequência definida em (2.2)-(2.3). Então, para todo $j < i$ temos*

$$\nabla f(x_{(i)})^T d_{(j)} = 0. \tag{2.8}$$

Demonstração. Seja $\varphi : \mathbb{R} \to \mathbb{R}$ dada por $\varphi(\alpha) = f(x_{(i)} + \alpha d_{(i)})$ e seja $x_{(i+1)}$ dado por (2.2)-(2.3). Como uma condição necessária para minimizarmos uma função escalar suave é que sua derivada seja zero, aplicando a regra da cadeia temos:

$$\nabla f(x_{(i+1)})^T d_{(i)} = \nabla f(x_{(i)} + \alpha_{(i)} d_{(i)})^T d_{(i)} = \varphi'(\alpha_{(i)}) = 0. \tag{2.9}$$

Provaremos o resultado por indução em i. Para $i = 0$, temos que $\nabla f(x_{(1)})^T d_{(0)} = 0$ por (2.9). Temos por hipótese de indução que o teorema vale para $i - 1$, assim queremos mostrar que vale para i. Para $j = i - 1$, temos que $\nabla f(x_{(i)})^T d_{(j)} = 0$, pela relação (2.9). Provaremos agora a afirmação para $j < i - 1$, e usando o fato que $\nabla f(x_{(i)}) = A(x_{(i-1)} + \alpha_{(i-1)} d_{(i-1)}) - b = Ax_{(i-1)} + \alpha_{(i-1)} Ad_{(i-1)} - b = Ax_{(i-1)} - b + \alpha_{(i-1)} Ad_{(i-1)} = \nabla f(x_{(i-1)}) + \alpha_{(i-1)} Ad_{(i-1)} \Rightarrow \nabla f(x_{(i)}) = \nabla f(x_{(i-1)}) + \alpha_{(i-1)} Ad_{(i-1)}$ e substituindo esta expressão em (2.8) temos:

$$\begin{aligned} \nabla f(x_{(i)})^T d_{(j)} &= (\nabla f(x_{(i-1)}) + \alpha_{(i-1)} Ad_{(i-1)})^T d_{(j)} \\ &= \nabla f(x_{(i-1)})^T d_{(j)} + \alpha_{(i-1)} d_{(i-1)}^T Ad_{(j)} \\ &= \nabla f(x_{(i-1)})^T d_{(j)} = 0, \end{aligned}$$

pela hipótese de indução e pelo fato das direções serem A-conjugadas. □

O Teorema 2.3 nos diz que o gradiente em uma iteração (i) qualquer é perpendicular a todas as direções que foram geradas anteriormente.

Teorema 2.4. *Dados $x_{(0)} \in \mathbb{R}^n$ e um conjunto de direções A-conjugadas $\{d_{(0)}, d_{(1)}, \ldots, d_{(k-1)}\}$, considere a sequência definida em (2.2)-(2.3). Então o ponto $x_{(k)}$ minimiza a função quadrática estritamente convexa sobre a variedade afim $C = x_{(0)} + [d_{(0)}, d_{(1)}, \cdots, d_{(k-1)}]$.*

A demonstração deste teorema pode ser encontrada, por exemplo, em (RIBEIRO; KARAS, 2014), página 112.

O Teorema 2.4 nos fala que, embora o método minimize unidirecionalmente a cada iteração, como consequência da conjugação, conseguimos minimizar uma função quadrática estritamente convexa na variedade afim determinada pelo ponto inicial e pelas direções A-conjugadas consideradas até a k-ésima. Dessa forma, após n iterações, obterá o minimizador irrestrito desejado em todo o espaço $\mathbb{R}^n$.

Já sabemos como calcular o tamanho do passo no Método de Gradientes Conjugados Lineares, veremos agora como podemos gerar direções que sejam A-conjugadas.

Para isto, dados $x_{(0)} \in \mathbb{R}^n$, definamos de maneira recursiva

$$d_{(0)} = -\nabla f(x_{(0)}) = r_{(0)}, \tag{2.10}$$

e para $i = 0, 1, \cdots, n-2$, tome

$$d_{(i+1)} = r_{(i+1)} + \beta_{(i)} d_{(i)}, \tag{2.11}$$

onde $x_{(i+1)}$ é dado por (2.2)-(2.3) e $\beta_{(i)}$ é calculado de tal modo que a direção $d_{(i)}$ seja A-conjugada a $d_{(i+1)}$, isto e, $d_{(i)}^T A d_{(i+1)} = 0$. Assim, temos

$$\begin{aligned}
d_{(i)}^T A(r_{(i+1)} + \beta_{(i)} d_{(i)}) &= 0 \\
d_{(i)}^T A r_{(i+1)} + \beta_{(i)} d_{(i)}^T A d_{(i)} &= 0 \\
\beta_{(i)} d_{(i)}^T A d_{(i)} &= -d_{(i)}^T A r_{(i+1)} \\
\beta_{(i)} &= \frac{-d_{(i)}^T A r_{(i+1)}}{d_{(i)}^T A d_{(i)}} \\
\beta_{(i)} &= \frac{-r_{(i+1)}^T A d_{(i)}}{d_{(i)}^T A d_{(i)}}
\end{aligned} \tag{2.12}$$

De posse de todas essas informações, podemos agora escrever o algoritmo do Método de Gradientes Conjugados Lineares.

$d_{(0)} = r_{(0)}$, faça $i = 0$.
Enquanto $\|r_{(i)}\| >$ Tol e i $<$ Maxiter, faça
$$\alpha_{(i)} = \frac{r_{(i)}^T d_{(i)}}{d_{(i)}^T A d_{(i)}},$$
$$r_{(i+1)} = r_{(i)} - \alpha_{(i)} A d_{(i)},$$
$$x_{(i+1)} = x_{(i)} + \alpha_{(i)} d_{(i)},$$
$$\beta_{(i)} = \frac{-r_{(i+1)}^T A d_{(i)}}{d_{(i)}^T A d_{(i)}},$$
$$d_{(i+1)} = r_{(i+1)} + \beta_{(i)} d_{(i)},$$
$$i = i + 1.$$

Note que o algoritmo do Método de Gradientes Conjugados Lineares pode convergir para a solução do problema (1.1) sem ter que de fato utilizar n iterações, conforme estabelece o seguinte resultado.

Teorema 2.5. *Se a matriz A possui apenas r autovalores distintos, então o Método de Gradientes Conjugados Lineares encontrará a solução do problema (1.1) em no máximo r iterações.*

Demonstração. Ver Teorema 5.4 de (NOCEDAL; WRIGHT, 2006), página 115. □

Observe ainda que as direções geradas pelo Método dos Gradientes Conjugados Lineares são de fato direções de descida, pois

$$\begin{aligned}\nabla f(x_{(i)})^T d_{(i)} &= \nabla f(x_{(i)})^T (r_{(i)} + \beta_{(i-1)} d_{(i-1)}) \\ &= \nabla f(x_{(i)})^T (-\nabla f(x_{(i)}) + \beta_{(i-1)} d_{(i-1)}) \\ &= -\nabla f(x_{(i)})^T \nabla f(x_{(i)}) + \beta_{(i-1)} \nabla f(x_{(i)})^T d_{(i-1)} \\ &= -\nabla f(x_{(i)})^T \nabla f(x_{(i)}) \\ &= -\|\nabla f(x_{(i)})\|^2 < 0,\end{aligned}$$

pois $\nabla f(x_{(i)})^T d_{(i-1)} = 0$ de acordo com o Teorema 2.3. Além disso, o custo fundamental das etapas do método é uma única multiplicação matriz-vetor $Ad_{(i)}$, que deve ser armazenada em um vetor e utilizada para calcular $\alpha_{(i)}$, $r_{(i+1)}$ e $\beta_{(i)}$.

Como destaca (NOCEDAL; WRIGHT, 2006), o Método dos Gradientes Conjugados é um método de direções conjugadas no qual não é necessário fazermos uso de todos os vetores gerados anteriormente a $d_{(i)}$. Ou seja, para calcularmos $d_{(i)}$, só precisamos saber $d_{(i-1)}$, não precisamos dos vetores $d_{(0)}, d_{(1)}, \cdots, d_{(i-2)}$, e isto faz com que o método não precise guardar informações desnecessárias.

É bom salientarmos que o cálculo feito em (2.12) para encontrar o escalar $\beta_{(i)}$, foi dado em sua forma original. Apresentaremos outras quatro versões para o mesmo, que serão estabelecidas no Teorema 2.6, e mostraremos também que para uma função quadrática estritamente convexa elas coincidem.

Teorema 2.6. *Sejam $x_{(i)} \in \mathbb{R}^n$ e $d_{(i)} \in \mathbb{R}^n$ gerados pelo Algoritmo de Gradientes Conjugados Lineares. Com a notação $y_{(i)} = -r_{(i+1)} + r_{(i)}$, as seguintes expressões são equivalentes:*

I) $\beta_{(i)} = \dfrac{-r_{(i+1)}^T A d_{(i)}}{d_{(i)}^T A d_{(i)}}$;

II) $\beta_{(i)}{}^{1} = \dfrac{r_{(i+1)}^T r_{(i+1)}}{r_{(i)}^T r_{(i)}}$;

III) $\beta_{(i)}{}^{2} = \dfrac{-r_{(i+1)}^T y_{(i)}}{r_{(i)}^T r_{(i)}}$;

IV) $\beta_{(i)}{}^{3} = \dfrac{r_{(i+1)}^T r_{(i+1)}}{d_{(i)}^T y_{(i)}}$;

[1] Método devido a Fletcher e Reeves (FR).
[2] Método devido a Polak, Ribière e Polyak (PRP).
[3] Método devido a Dai e Yuan (DY).

V) $\beta_{(i)}$[4] $= \dfrac{-r_{(i+1)}^T y_{(i)}}{d_{(i)}^T y_{(i)}}$.

Demonstração. ***I) ⇒ II)*** Como $r_{(i+1)} = b - Ax_{(i+1)} = b - A(x_{(i)} + \alpha_{(i)} d_{(i)}) = b - Ax_{(i)} - \alpha_{(i)} Ad_{(i)} = r_{(i)} - \alpha_{(i)} Ad_{(i)} \Rightarrow r_{(i+1)} - r_{(i)} = -\alpha_{(i)} Ad_{(i)}$, logo

$$Ad_{(i)} = \frac{-r_{(i+1)} + r_{(i)}}{\alpha_{(i)}}. \tag{2.13}$$

E substituindo (2.13) na expressão de $\beta_{(i)}$ dada em ***I)*** temos

$$\beta_{(i)} = \frac{-r_{(i+1)}^T(-r_{(i+1)} + r_{(i)})}{d_{(i)}^T(-r_{(i+1)} + r_{(i)})}, \tag{2.14}$$

que pode ser escrito como

$$\beta_{(i)} = \frac{-r_{(i+1)}^T(-r_{(i+1)} + r_{(i)})}{-d_{(i)}^T r_{(i+1)} + d_{(i)}^T r_{(i)}}, \tag{2.15}$$

ou então

$$\beta_{(i)} = \frac{r_{(i+1)}^T r_{(i+1)} - r_{(i+1)}^T r_{(i)}}{-d_{(i)}^T r_{(i+1)} + d_{(i)}^T r_{(i)}} \tag{2.16}$$

$$= \frac{r_{(i+1)}^T r_{(i+1)}}{r_{(i)}^T r_{(i)}}, \tag{2.17}$$

já que $-d_{(i)}^T r_{(i+1)} = -r_{(i+1)}^T d_{(i)} = \nabla f(x_{(i+1)})^T d_{(i)} = 0$ pelo Teorema 2.3; também vale a relação $r_{(i+1)}^T r_{(i)} = 0$ e ainda $d_{(i)}^T r_{(i)} = r_{(i)}^T d_{(i)} = r_{(i)}^T r_{(i)}$.

II) ⇒ III) Usando (2.15) segue que

$$\beta_{(i)} = \frac{-r_{(i+1)}^T y_{(i)}}{r_{(i)}^T r_{(i)}}, \tag{2.18}$$

pois $-d_{(i)}^T r_{(i+1)} = 0$, $d_{(i)}^T r_{(i)} = r_{(i)}^T r_{(i)}$ e $y_{(i)} = -r_{(i+1)} + r_{(i)}$.

III) ⇒ IV) Temos por (2.14) que

$$\beta_{(i)} = \frac{r_{(i+1)}^T r_{(i+1)} - r_{(i+1)}^T r_{(i)}}{d_{(i)}^T(-r_{(i+1)} + r_{(i)})},$$

$$= \frac{r_{(i+1)}^T r_{(i+1)}}{d_{(i)}^T y_{(i)}},$$

pois $r_{(i+1)}^T r_{(i)} = 0$.

IV) ⇒ V) Dada em (2.14), desde que $y_{(i)} = -r_{(i+1)} + r_{(i)}$.

V) ⇒ I) Temos que a relação (2.13) pode ser escrita como

$$\alpha_{(i)} Ad_{(i)} = -r_{(i+1)} + r_{(i)} = y_{(i)}, \tag{2.19}$$

[4] Método devido a Hestenes e Stiefel (HS).

logo substituindo (2.19) na expressão de $\beta_{(i)}$ dada em ***V)*** temos:

$$\beta_{(i)} = \frac{-r_{(i+1)}^T A d_{(i)}}{d_{(i)}^T A d_{(i)}},$$

que é exatamente o que queremos. □

Os valores dos escalares $\beta_{(i)}$ do item ***II)*** até ***V)*** do Teorema 2.6 estão sintetizados na Tabela 3.

Tabela 3 – Fórmulas dos Betas.

		Numerador	
$\beta_{(i)}$		$r_{(i+1)}^T r_{(i+1)}$	$-r_{(i+1)}^T y_{(i)}$
Denominador	$r_{(i)}^T r_{(i)}$	FR	PRP
	$d_{(i)}^T y_{(i)}$	DY	HS

Para completar as informações e contextualizar a convergência do método, associando-a ao número de condição espectral κ da matriz positiva definida A, cf. Definição 1.1, apresentamos as seguintes relações. Definindo a norma da energia $\|e\|_A = \sqrt{e^T A e}$, é possível mostrar (ver, p.ex. (SHEWCHUK, 1994)) que o Método da Máxima Descida satisfaz

$$\|e_{(i)}\|_A \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^i \|e_{(o)}\|_A, \quad \forall i = 1, 2, \ldots$$

enquanto o Método de Gradientes Conjugados Lineares satisfaz

$$\|e_{(i)}\|_A \leqslant 2\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^i \|e_{(o)}\|_A, \quad \forall i = 1, 2, \ldots$$

Na próxima seção foram feitos alguns testes computacionais do Método dos Gradientes Conjugados Lineares utilizando-se o escalar $\beta_{(i)}$ formulado por Fletcher e Reeves.

2.3 Testes Computacionais

Nesta seção descrevemos alguns testes computacionais utilizando o Método de Gradientes Conjugados Lineares. Para isto, primeiro implementamos no MATLAB versão (R2015a) a geração de matrizes $A \in \mathbb{R}^{n \times n}$ simétricas e definidas positivas. Escrevemos a matriz A da forma $A = QDQ^T$, isto é, como um produto de três matrizes, onde a matriz

Q é calculada através da fatoração QR de uma matriz quadrada de dimensão n com componentes geradas aleatoriamente, para produzirmos a matriz ortogonal Q, onde Q^T é a sua transposta e D é uma matriz diagonal formada pelos autovalores de A, veja, por exemplo, (WATKINS, 2002). Para construirmos a matriz D, fizemos uma correspondência entre um valor x qualquer do intervalo $(0, 1)$, dado de forma aleatória pelo comando rand no MATLAB, e um ponto y do intervalo $(\lambda_{min}, \lambda_{max})$, onde, respectivamente, λ_{min} e λ_{max} representam o menor e o maior autovalores da matriz A, da seguinte maneira:

$$\frac{x-0}{1-0} = \frac{y-\lambda_{min}}{\lambda_{max}-\lambda_{min}} \Rightarrow y = \lambda_{min} + x(\lambda_{max} - \lambda_{min}).$$

Desta maneira geramos a matriz A, por meio da sua diagonalização. Note que temos uma base de $\mathbb{R}^n$ formada por autovetores ortogonais, e como todos os autovalores são positivos, pois pertencem ao intervalo $(\lambda_{min}, \lambda_{max})$ e $\lambda_{min} > 0$, logo a matriz A é definida positiva. Veja por exemplo (ANTON; BUSBY, 2006), para o resultado que uma matriz simétrica é definida positiva se, e somente se, todos os seus autovalores são positivos. O algoritmo completo usado para construir uma matriz simétrica definida positiva está no Apêndice A.3.

Agora podemos apresentar nossas conclusões para alguns experimentos, nos quais mantivemos alguns valores fixos e variamos outros, como detalhado a seguir.

i) Fixando λ_{min}, λ_{max} e Tol (tolerância), e variando apenas n, que é a ordem da matriz A.
Conclusão: Ao deixarmos fixos $\lambda_{min} = 0.5$, $\lambda_{max} = 50$, e Tol$= 10^{-7}$, à medida que aumentamos n, no conjunto $n \in \{2, 3, 6, 9, 13, 23, 33, 43, 53, 63, 73, 83, 93, 103, 113, 123, 133, 233\}$, o número de iterações pouco cresce, pois seria como se estivéssemos apenas com um único *cluster*, onde entendemos por *cluster* ao agrupamento de autovalores.

ii) Fixando λ_{min}, λ_{max} e n, e variando Tol.
Conclusão: Ao deixarmos fixos $\lambda_{min} = 0.5$, $\lambda_{max} = 50$ e $n = 10$, à medida que diminuimos Tol, no seguinte conjunto $\{2, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-13}, 10^{-23}, 10^{-33}, 10^{-43}\}$, o problema tende a não ter solução pelo Método de Gradientes Conjugados Lineares, isto é, o número máximo de iterações tende a ser atingido, pois estamos querendo que o resíduo seja muito pequeno.

iii) Fixando n e Tol e deixando λ_{max} muito próximo de λ_{min} (fixo), isto é, fizemos λ_{max} tender a λ_{min}.
Conclusão: Ao deixarmos fixos $n = 10$ e Tol $= 10^{-4}$ e se fizermos λ_{max}, no conjunto $\{0.6, 0.55, 0.501, 0.5000000001\}$ próximo de $\lambda_{min} = 0.5$, o número de iterações tende a ser 1, isto é, temos convergência imediata.

iv) Fixando n e Tol e deixando λ_{min} (fixo) muito distante de λ_{max}, ou seja, fizemos com que o valor absoluto entre λ_{min} e λ_{max} fosse cada vez maior.
Conclusão: Ao deixarmos fixos $n = 10$ e Tol $= 10^{-4}$ e se fizermos $\lambda_{min} = 0.5$ muito distante

de λ_{max}, no conjunto $\lambda_{max} \in \{5 \times 10^8, 5 \times 10^9, 5 \times 10^{10}, 5 \times 10^{11}\}$, temos que o número de iterações tende a crescer até atingir o número máximo permitido.

v) Fixando n, Tol e λ_{max} (em um número não tão grande) e deixando λ_{min} muito próximo de zero.
Conclusão: Aqui fixamos $n = 10$, Tol $= 10^{-8}$ e $\lambda_{max} = 50$, e tomamos λ_{min} no conjunto $\{10^{-4}, 10^{-6}, 10^{-8}, 10^{-9}, 10^{-10}, 10^{-11}, 10^{-12}, 10^{-16}, 10^{-20}, 10^{-30}\}$. Há convergência pela norma do resíduo, porém não à solução esperada[5]. O problema começou a convergir para outro valor justamente quando λ_{min} ficou menor do que Tol, enquanto λ_{min} era maior ou igual a Tol havia convergência para a solução desejada. Observe que as matrizes geradas neste teste **v)** são mal condicionadas.

Com os testes computacionais acima pudemos constatar, a menos que a matriz fosse mal condicionada, que o método encontrou a solução do problema em no máximo n iterações.

No Método de Gradientes Conjugados Lineares, o tamanho do passo dado em (2.5) foi calculado de forma explícita, minimizando exatamente a quadrática estritamente convexa ao longo da direção de busca. Porém, não podemos usar exatamente a mesma abordagem para minimizarmos funções não quadráticas. De acordo com (SHEWCHUK, 1994), Gradientes Conjugados podem ser utilizados não apenas para minimizar uma função quadrática estritamente convexa, mas qualquer função suave para a qual conseguimos calcular seu gradiente. No Método de Gradientes Conjugados fazemos uso apenas da derivada primeira, fato este que permite que o método, com as devidas adaptações, seja amplamente utilizado para se minimizar uma função suave. No próximo capítulo veremos alguns exemplos da utilização do Método de Gradientes Conjugados para funções não lineares mais gerais.

[5] Construímos o problema para que a solução fosse o vetor $x = (1, 1, \cdots, 1)^T$, pois queríamos resolver $Ax = b$, e como b é dado, escrevemos $b = A * ones(n, 1) = \sum_{j=1}^{n} A(:, j)$.

3 Gradientes Conjugados não Lineares

Já vimos que podemos utilizar o Método de Gradientes Conjugados para minimizarmos uma função quadrática estritamente convexa, agora veremos que, também, sob um ponto de vista análogo, poderemos utilizar os Gradientes Conjugados para uma função suave qualquer.

Basicamente, há duas diferenças essenciais entre o Método de Gradientes Conjugados não Lineares e o Método de Gradientes Conjugados Lineares. Primeiro, é inviável tentar achar o tamanho do passo na forma explícita, assim sendo faz-se uso, por exemplo, da busca inexata dada pela condição de Armijo, vista em (1.13) combinada com a condiçao de Wolfe, conforme (1.15). Segundo, o Método de Gradientes Conjugados não Lineares não termina necessariamente em n passos. Dessa forma, é usual considerar uma reinicialização das direções de busca a cada quantidade preestabelecida de passos (RIBEIRO; KARAS, 2014).

De acordo com o exposto anteriormente, um algoritmo para o Método de Gradientes Conjugados não Lineares tem a seguinte sequência de passos:

1) Dados um ponto inicial $x_{(0)}$, $t_{min} > 0$ e Tol, faça $d_{(0)} = -\nabla f(x_{(0)})$ e $i = 0$.

Enquanto $\|\nabla f(x_{(i)}\| >$ Tol, faça

2) Encontre $t > t_{\min}$ que satisfaça a condição de decréscimo suficiente (Armijo) e a condição de curvatura (Wolfe forte) em $x_{(i)} + td_{(i)}$.

3) Se $t > t_{\min}$, defina $x_{(i+1)} = x_{(i)} + td_{(i)}$. Caso contrário, $x_{(i+1)} = x_{(i)}$.

4) Se $((i+1)\text{mod}(n) \neq 0)$ e $(t > t_{\min})$, calcule $\beta_{(i)}$ por alguma expressão dentre as estabelecidas no Teorema 2.6. Caso contrário, faça $\beta_{(i)} = 0$.

5) $d_{(i+1)} = -\nabla f(x_{(i+1)}) + \beta_{(i)} d_{(i)}$.

6) $i = i + 1$.

Observe que fazermos $\beta_{(i)} = 0$ em 4) dado acima, significa tomarmos a direção no ponto corrente como o oposto ao gradiente, o que corresponde a retomarmos a direção do resíduo. Este procedimento de reinicialização é muito utilizado em problemas não lineares gerais para se evitar erros de arredondamento e apagar informações velhas que podem não ser mais úteis. Além disso, como no caso não linear não temos garantia que as direções geradas sejam de descida, é a forma de assegurar a boa definição do algoritmo. A sequência gerada está bem definida, pois muda de ponto sempre que a busca linear é bem sucedida, e se mantém no ponto corrente caso haja falha na busca. Neste último caso,

prossegue com a direção de máxima descida, a qual também é utilizada a cada n passos, como reinicialização.

Note que há quatro variantes para o valor de $\beta_{(i)}$ dadas pelo Teorema 2.5, e dependendo do valor escolhido para $\beta_{(i)}$ pode-se encontrar o mínimo da função mais rapidamente ou não. De fato, a busca por outras fórmulas para esse escalar tem sido tema de pesquisa atual, veja por exemplo o artigo de (DAI; KOU, 2013). Também se tivéssemos adotado outra maneira para o cálculo do tamanho do passo, ao invés das escolhas em 2) dado acima, teríamos outras versões para o método em questão.

Apresentaremos a seguir alguns exemplos do Método de Gradientes Conjugados não Lineares. Os exemplos de 3.1 até 3.4 foram feitos utilizando-se $\beta_{(i)}$ formulado por Fletcher e Reeves.

3.1 Alguns Exemplos

No Apêndice A.4 apresentamos o algoritmo utilizado para resolver os quatro exemplos desta seção.

Vale destacar que em todos os testes tomamos a constante da condição de Wolfe forte como sendo $c_2 = 0.99$, diferente da sugerida por (NOCEDAL; WRIGHT, 2006), com o intuito de aumentarmos o comprimento do intervalo de passos aceitáveis, pois $c_2 = 0.1$ restringe excessivamente tal intervalo. Ilustramos o efeito da escolha dessa constante na Figura 5, com três possibilidades para c_2.

No entanto, a escolha $c_2 = 0.1$, assegura que as direções geradas são de fato direções de descida, isto se utilizarmos $\beta_{(i)}$ de Fletcher e Reeves, conforme estabelece o Lema 5.6 de (NOCEDAL; WRIGHT, 2006), página 125, enunciado a seguir.

Lema 3.1. *Suponha que o algoritmo de Gradientes Conjugados não Lineares, com $\beta_{(i)}$ de Fletcher-Reeves, é implementado com um tamanho de passo t que satisfaz a condição de Wolfe forte (1.15) com $0 < c_2 < 1/2$. Então o método gerará direções de descida $d_{(i)}$ tais que*

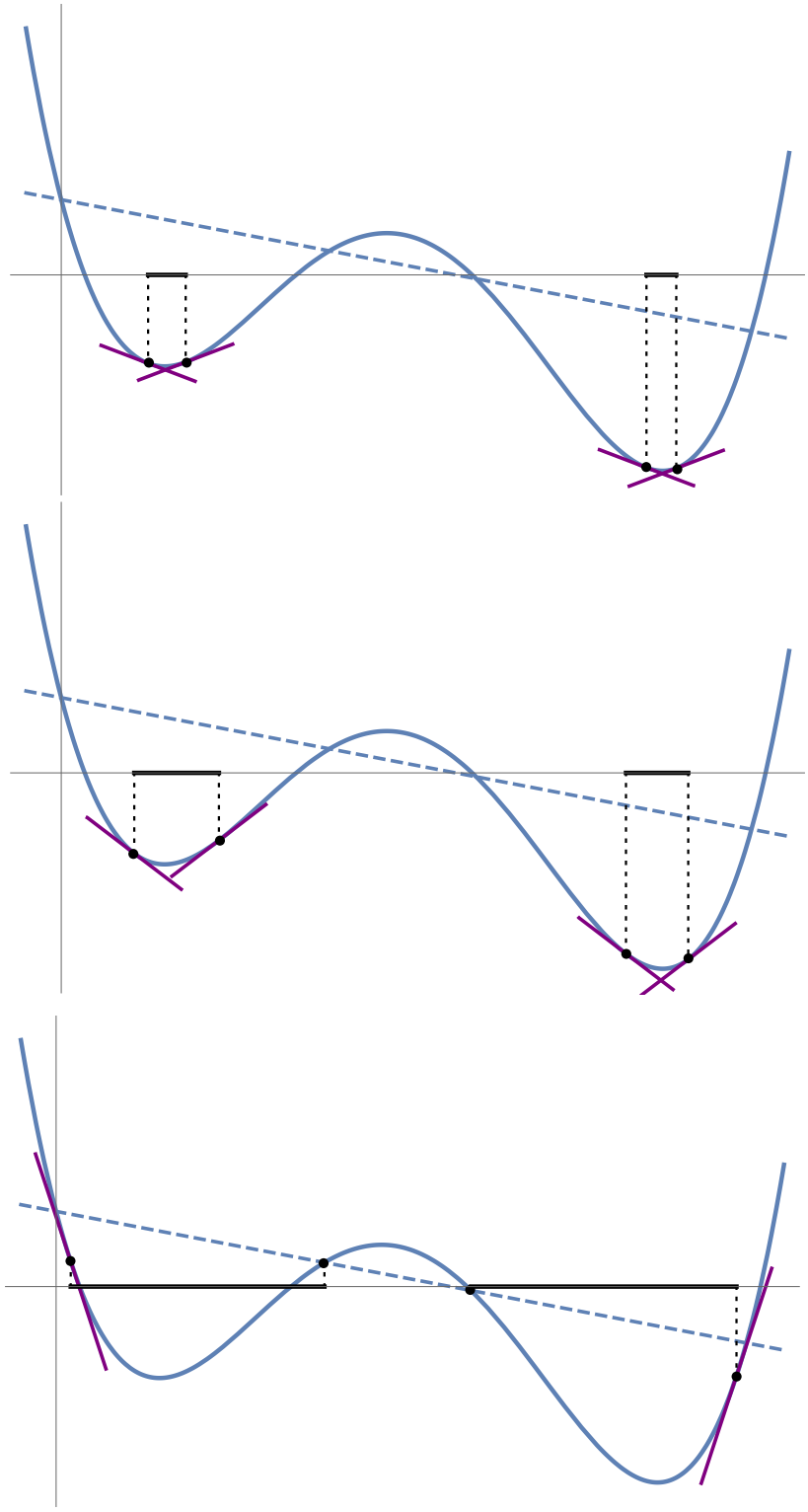
$$\frac{-1}{1-c_2} \leqslant \frac{\nabla f(x_{(i)})^T d_{(i)}}{\|\nabla f(x_{(i)})\|^2} \leqslant \frac{2c_2 - 1}{1-c_2}, \quad \forall i = 0, 1, \ldots$$

Exemplo 3.1. Tomaremos como exemplo a função definida em (RIBEIRO; KARAS, 2014), página 160, dada por:

$$f(x_1, x_2) = \frac{1}{2}((x_1^2 - x_2)^2 + (1 - x_1)^2).$$

Essa função possui o seu mínimo em $(x_1, x_2)^T = (1, 1)^T$, com valor 0, isto é, $f(1, 1) = 0$.

Figura 5 – Variação do intervalo admissível (em destaque) para as condições de Armijo (com $c_1 = 0.01$, reta tracejada) e de Wolfe forte (com as inclinações delimitadoras indicadas), com diferentes valores para a constante c_2. No primeiro gráfico usamos $c_2 = 0.1$, no segundo $c_2 = 0.2$, e no terceiro, $c_2 = 0.8$. A função é $\varphi(t) = 50 - (6250t)/21 + (3525t^2)/14 - (975t^3)/14 + (125t^4)/21$.



Faremos agora uma análise tomando alguns pontos iniciais e variaremos apenas a tolerância no Método de Gradientes Conjugados para funções não quadráticas utilizando busca de Armijo mais condição de Wolfe forte. Aqui utilizamos as seguintes constantes: $c_1 = 0.01$ (constante de Armijo), $c_2 = 0.99$ (constante de Wolfe forte), veja a relação (1.15), $\gamma = 0.5$ (redução no tamanho de passo), $t_{min} = 10^{-2}$ (tamanho do passo mínimo, se este valor é atingido também fazemos $\beta_{(i)} = 0$) e Maxiter $= 10^3$ (número máximo de iterações permitidas).

Selecionamos abaixo alguns pontos iniciais e em seguida apresentamos tabelas descrevendo a tolerância (Tol), o número de iterações (Niter) e o número de avaliações de funções (Naval), bem como algumas figuras relacionadas a esses pontos.

1) $x_{(0)} = (2, 2)^T$;

2) $x_{(0)} = (-1, 0)^T$;

3) $x_{(0)} = (0, -2)^T$;

4) $x_{(0)} = (1, 0.8)^T$;

5) $x_{(0)} = (0.999, 0.999)^T$;

6) $x_{(0)} = (-10, -10)^T$.

Tabela 4 – Resultados para o ponto $x_{(0)} = (2, 2)^T$.

Tol	Niter	Naval
10^{-3}	30	74
10^{-5}	82	201
10^{-7}	118	284
10^{-9}	169	404
10^{-11}	215	514
10^{-13}	259	617
10^{-15}	280	663

Figura 6 – Sequência dos iterados a partir de $x_{(0)} = (2, 2)^T$.

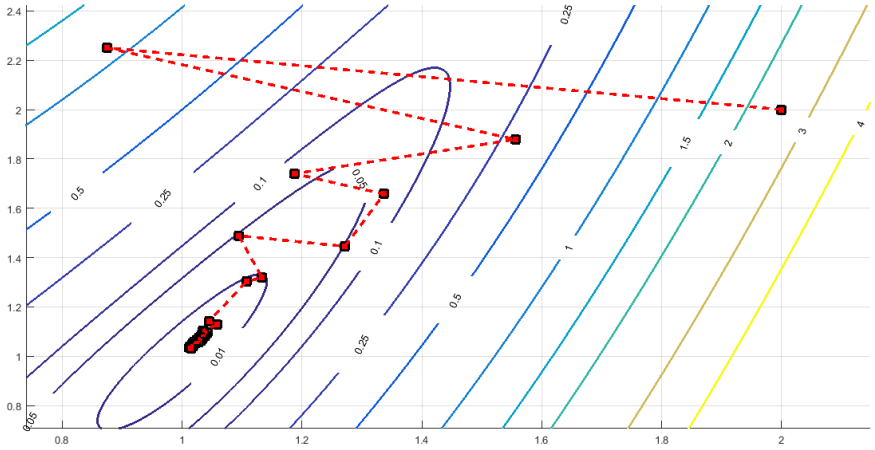


Figura 7 – Evolução da função objetivo começando em $x_{(0)} = (2, 2)^T$.

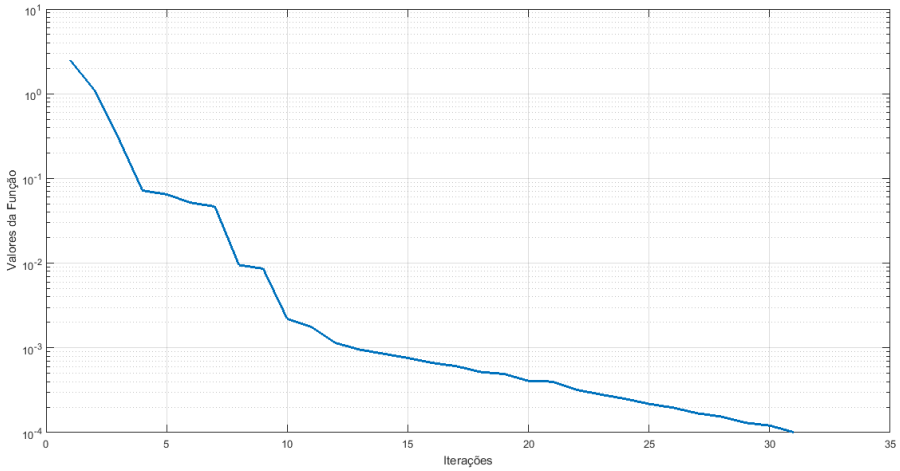


Tabela 5 – Resultados para o ponto $x_{(0)} = (-1, 0)^T$.

Tol	Niter	Naval
10^{-3}	7	14
10^{-5}	46	113
10^{-7}	95	231
10^{-9}	154	377
10^{-11}	205	496
10^{-13}	244	586
10^{-15}	272	653

Figura 8 – Sequência dos iterados a partir de $x_{(0)} = (-1, 0)^T$.

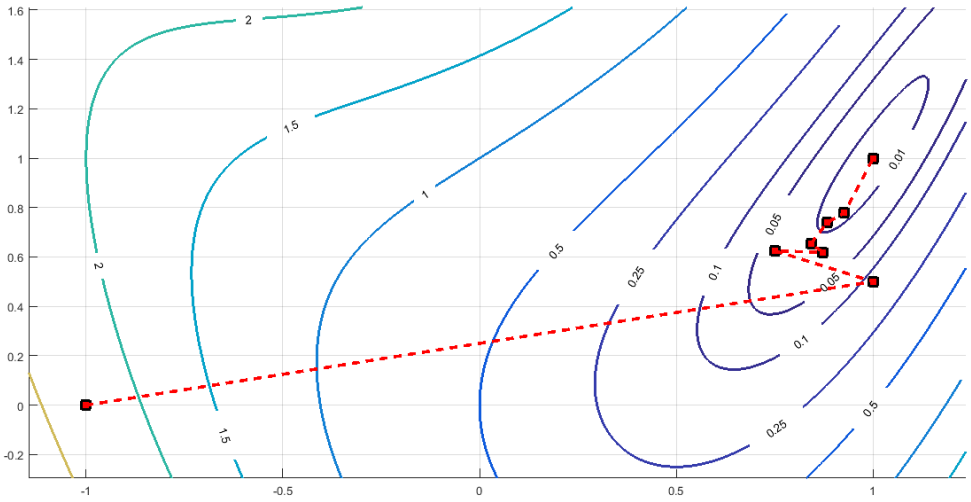


Figura 9 – Evolução da função objetivo começando em $x_{(0)} = (-1, 0)^T$.

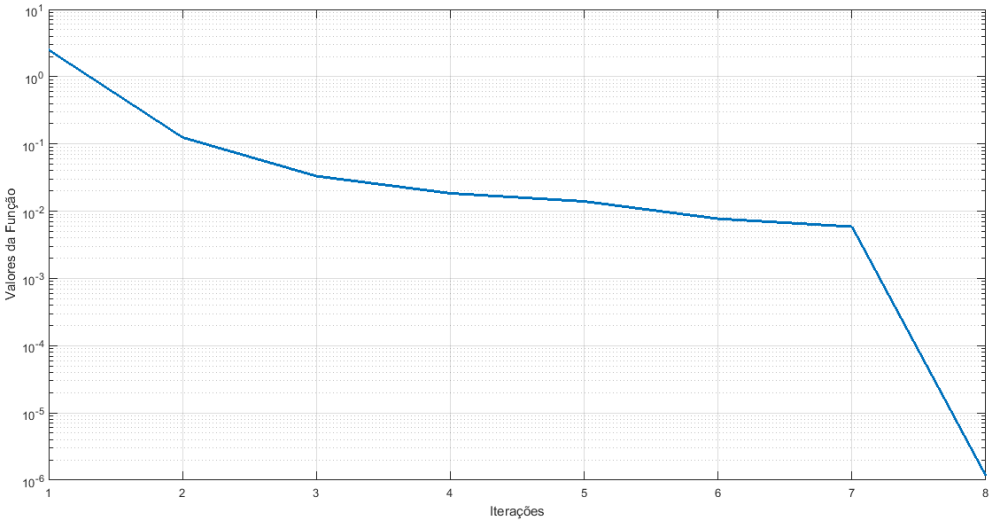


Tabela 6 – Tabela de iterações para o ponto $x_{(0)} = (0, -2)^T$.

Tol	Niter	Naval
10^{-3}	36	87
10^{-5}	83	198
10^{-7}	135	324
10^{-9}	180	430
10^{-11}	199	475
10^{-13}	265	636
10^{-15}	307	737

Figura 10 – Sequência dos iterados a partir de $x_{(0)} = (0, -2)^T$.

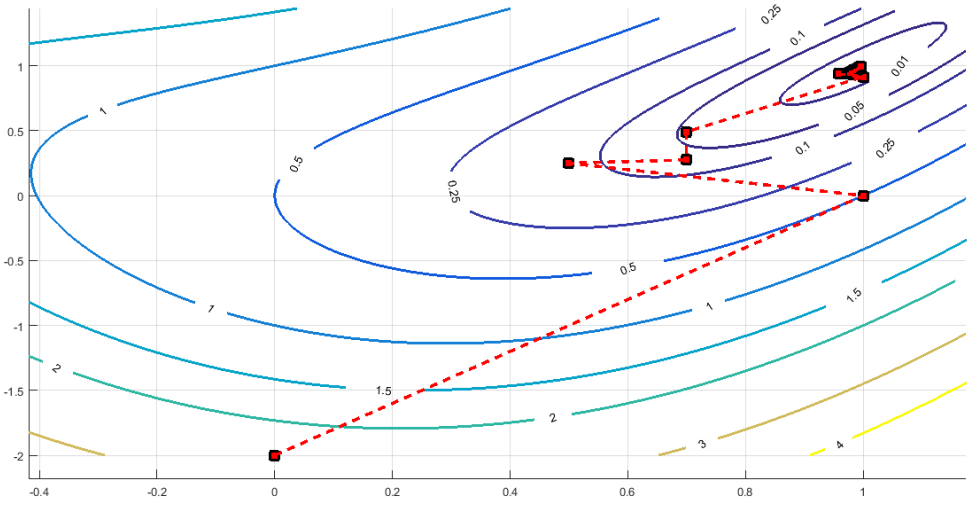


Figura 11 – Evolução da função objetivo começando em $x_{(0)} = (0, -2)^T$.

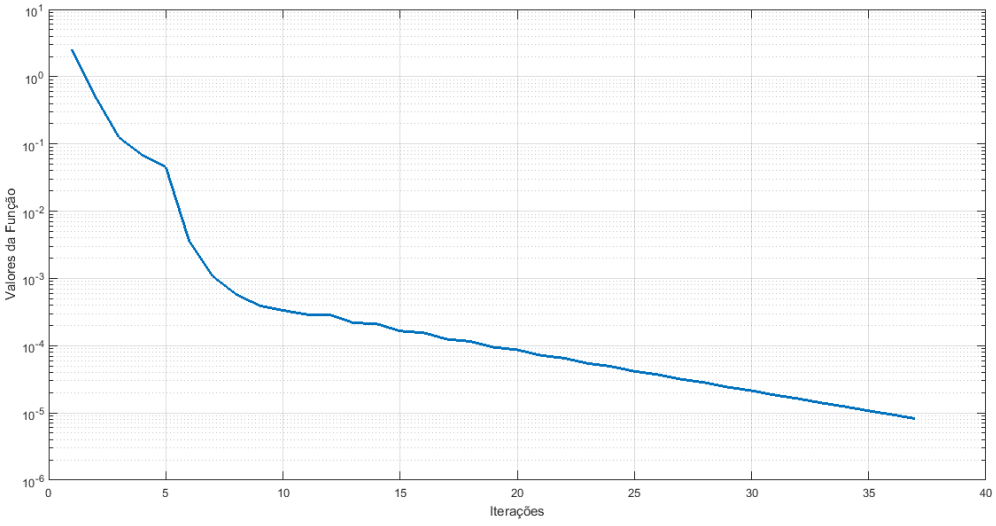


Tabela 7 – Resultados para o ponto $x_{(0)} = (1, 0.8)^T$.

Tol	Niter	Naval
10^{-3}	57	142
10^{-5}	115	283
10^{-7}	117	280
10^{-9}	187	454
10^{-11}	237	575
10^{-13}	289	698
10^{-15}	331	793

Figura 12 – Sequência dos iterados a partir de $x_{(0)} = (1, 0.8)^T$.

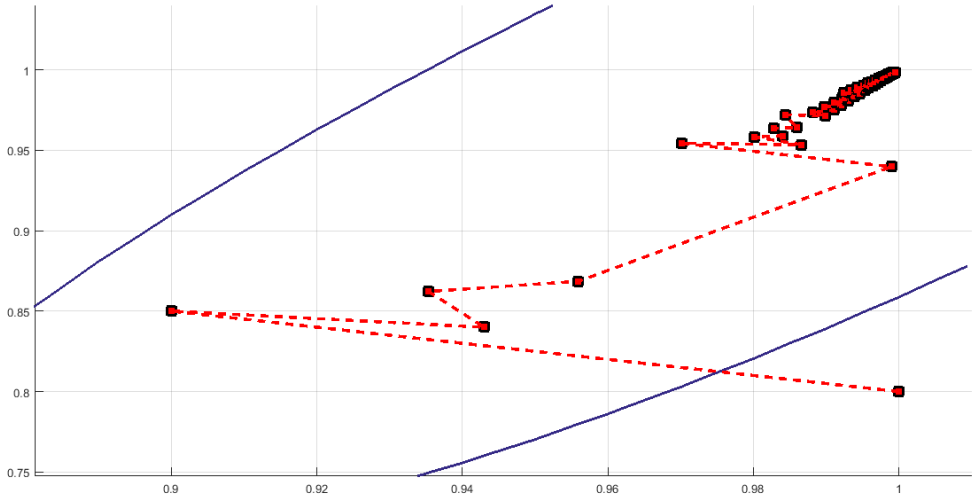


Figura 13 – Evolução da função objetivo começando em $x_{(0)} = (1, 0.8)^T$.

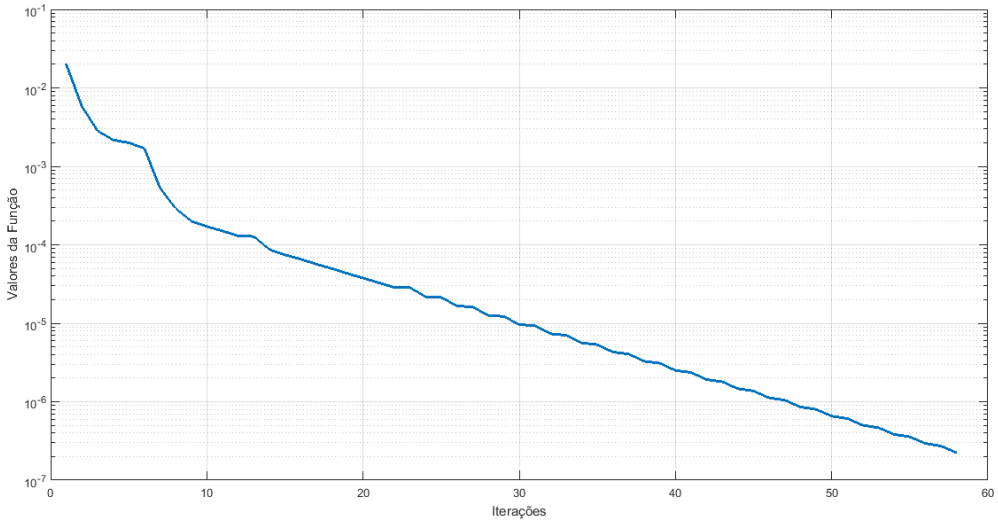


Tabela 8 – Resultados para o ponto $x_{(0)} = (0.999, 0.999)^T$.

Tol	Niter	Naval
10^{-3}	70	178
10^{-5}	138	348
10^{-7}	206	517
10^{-9}	256	634
10^{-11}	300	737
10^{-13}	345	843
10^{-15}	1000	1000

Figura 14 – Sequência dos iterados a partir de $x_{(0)} = (0.999, 0.999)^T$.

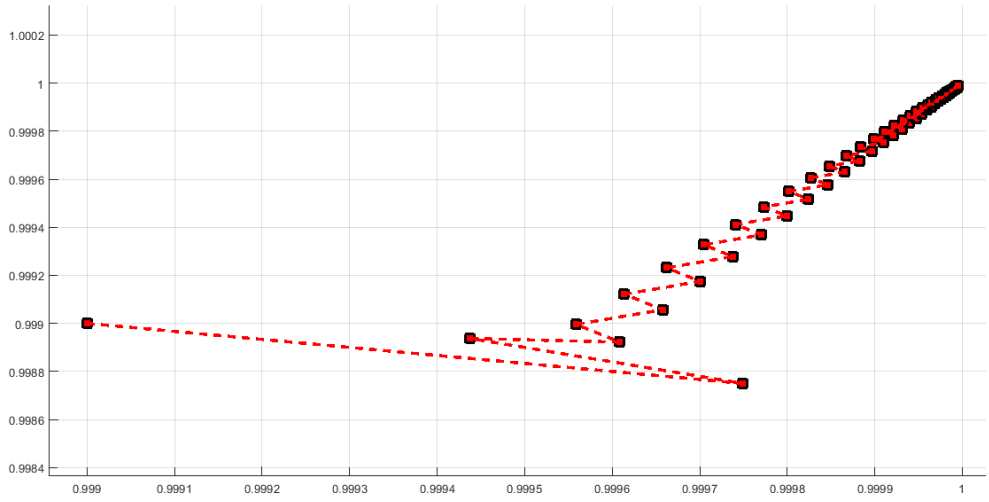


Figura 15 – Evolução da função objetivo começando em $x_{(0)} = (0.999, 0.999)^T$.

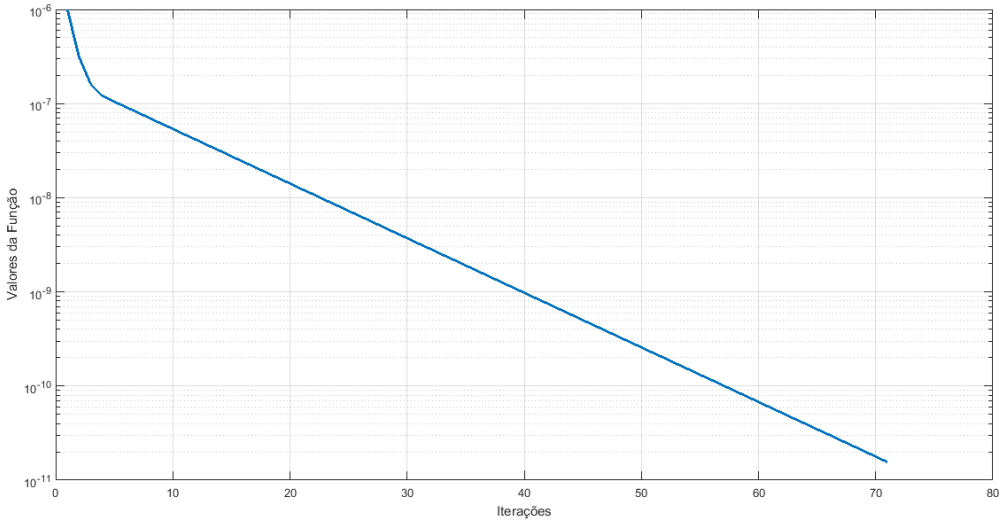


Tabela 9 – Resultados para o ponto $x_{(0)} = (-10, -10)^T$.

Tol	Niter	Naval
10^{-3}	7	36
10^{-5}	21	63
10^{-7}	38	100
10^{-9}	78	193
10^{-11}	127	312
10^{-13}	173	420
10^{-15}	221	535

Figura 16 – Sequência dos iterados a partir de $x_{(0)} = (-10, -10)^T$.

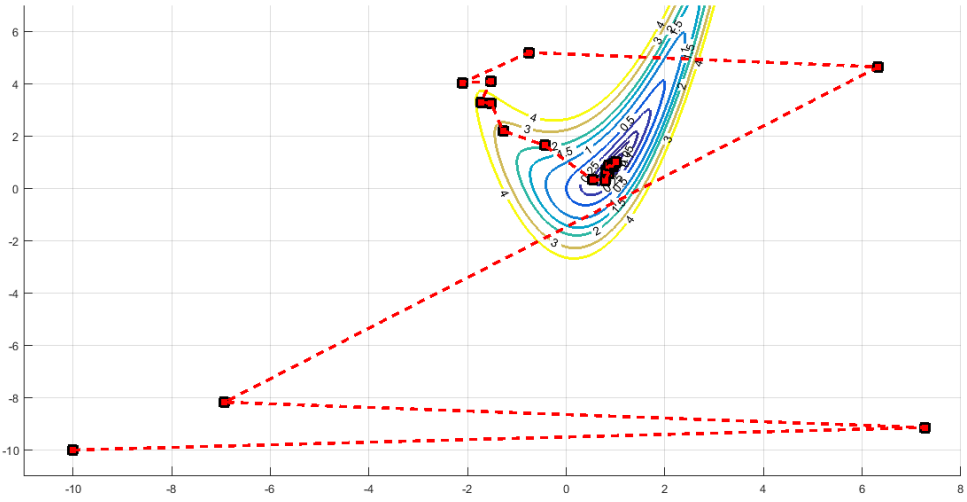
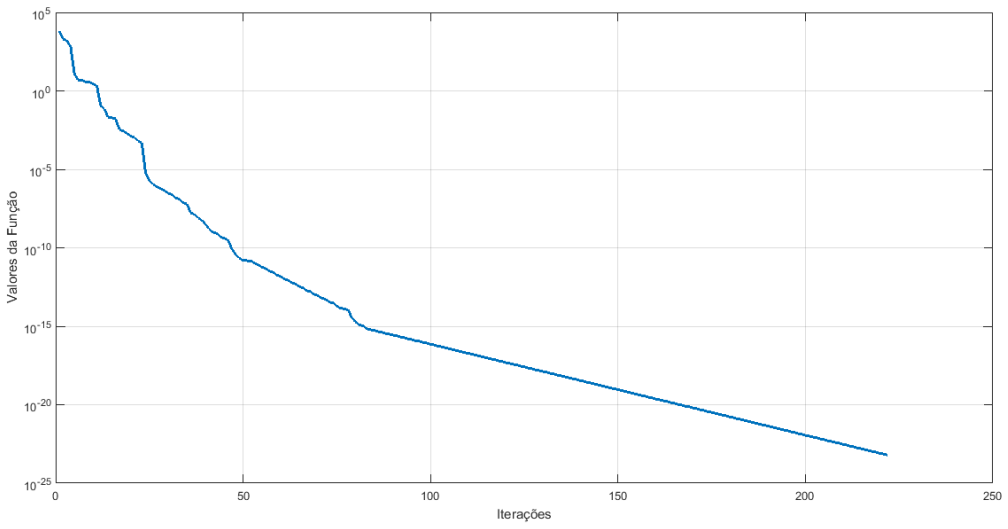


Figura 17 – Evolução da função objetivo começando em $x_{(0)} = (-10, -10).^T$



3.1.1 Algumas observações para o Exemplo 3.1: estar mais próximo da solução não significa convergir mais rápido

Veja que os pontos 1), 2) e 3) estão todos na mesma curva de nível, ou seja, $f(x_{(0)}) = 2.5$, porém, a partir desses pontos o método não converge com o mesmo número de iterações.

Observe que para o ponto 5) foi atingida a capacidade de precisão da máquina ao fazermos Tol $= 10^{-15}$.

Podemos notar, pelos resultados obtidos para os pontos, 1), 2), 3), 4) e 5), não necessariamente quando o ponto está mais próximo da solução, o método convergirá mais rapidamente, isto é, com menos iterações. Para este exemplo, com um ponto inicial bastante próximo da solução, o método gastou mais iterações.

Note que o ponto 5), mesmo estando mais próximo da solução do que o ponto 1), quando o método começou com o ponto 5), gastou mais iterações para chegar na solução com a mesma precisão requerida. Isto se deu, possivelmente, pelo fato de que para o ponto 5) foi necessário fazer o recomeço do método mais vezes o que deixou seu comportamente muito semelhante à Máxima Descida em uma situação de mau condicionamento.

Veja que o ponto 6) foi o que apresentou melhor desempenho, do ponto de vista do número de iterações efetuadas, em relação aos demais pontos. Para este ponto reduzimos o tamanho do passo mínimo para $t_{min} = 10^{-3}$. Essa mudança foi feita para acomodar o fato da norma do vetor gradiente neste ponto inicial valer 2.214×10^3, fazendo com que o primeiro passo da busca precisasse ser inferior a 0.01 para haver progresso efetivo.

Exemplo 3.2. Seja a função $f : \mathbb{R}^2 \to \mathbb{R}$ dada por:

$$f(x_1, x_2) = (x_1 - 3)^2 + (x_2 + 3)^2(x_2 - x_1)^2, \tag{3.1}$$

através do cálculo, sabemos que esta função possui dois pontos mínimos, a saber $(x_1, x_2)^T = (3, 3)^T$ e $(x_1, x_2)^T = (3, -3)^T$, com valor $f(x_1, x_2) = 0$, e também possui um ponto de sela em $(x_1, x_2)^T \simeq (-0.08, -1.54)^T$.

Aqui utilizamos $c_1 = 0.01,\ c_2 = 0.99,\ \gamma = 0.5,\ t_{min} = 10^{-4}$ e Maxiter $= 10^3$. Assim como no Exemplo 3.1, faremos uma análise em alguns pontos iniciais.

1) $x_{(0)} = (-3, 1)^T$;

2) $x_{(0)} = (4, 3)^T$;

3) $x_{(0)} = (-3, 3)^T$;

4) $x_{(0)} = (-3, 2)^T$;

5) $x_{(0)} = (0, -1.5)^T$.

Tabela 10 – Resultados para o ponto $x_{(0)} = (-3, 1)^T$.

Tol	Niter	Naval
10^{-3}	2	10
10^{-5}	2	10
10^{-7}	2	10
10^{-9}	2	10
10^{-11}	2	10
10^{-13}	2	10
10^{-15}	2	10

Figura 18 – Sequência dos iterados a partir de $x_{(0)} = (-3, 1)^T$.

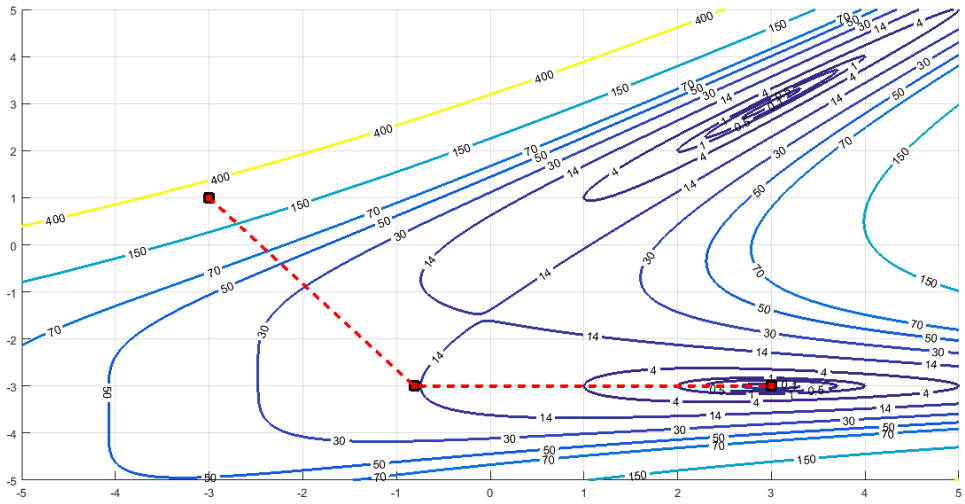


Figura 19 – Evolução da função objetivo começando em $x_{(0)} = (-3, 1)^T$.

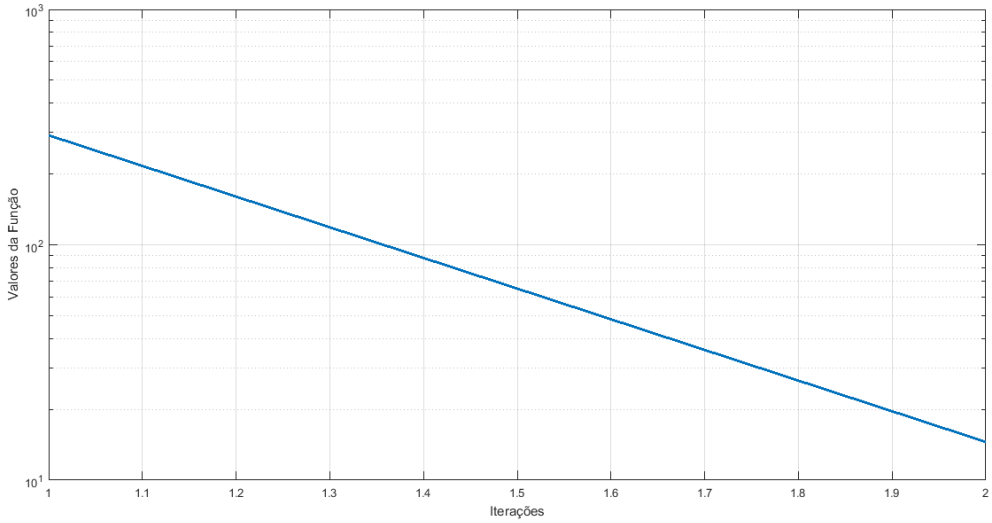


Tabela 11 – Resultados para o ponto $x_{(0)} = (4, 3)^T$.

Tol	Niter	Naval
10^{-3}	24	164
10^{-5}	45	297
10^{-7}	125	828
10^{-9}	207	1376
10^{-11}	336	2233
10^{-13}	398	2642
10^{-15}	504	3343

Figura 20 – Sequência dos iterados a partir de $x_{(0)} = (4, 3)^T$.

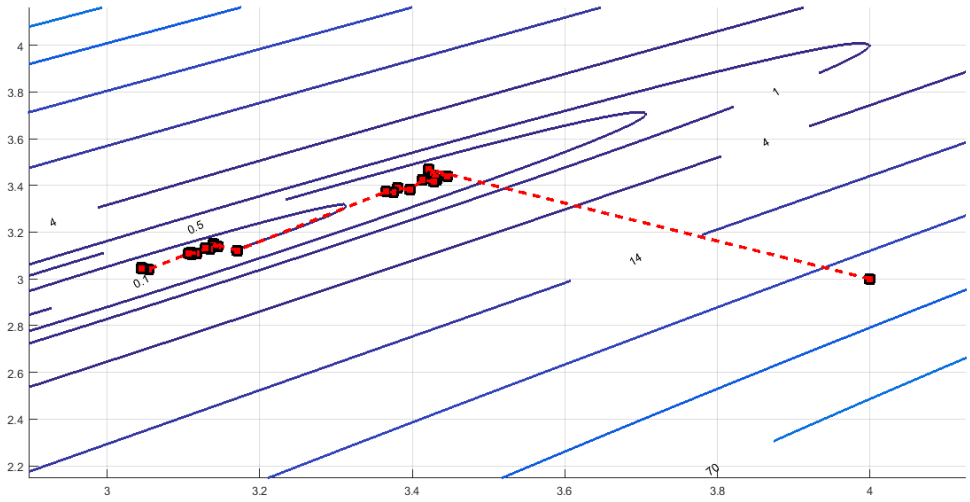


Figura 21 – Evolução da função objetivo começando em $x_{(0)} = (4, 3)^T$.

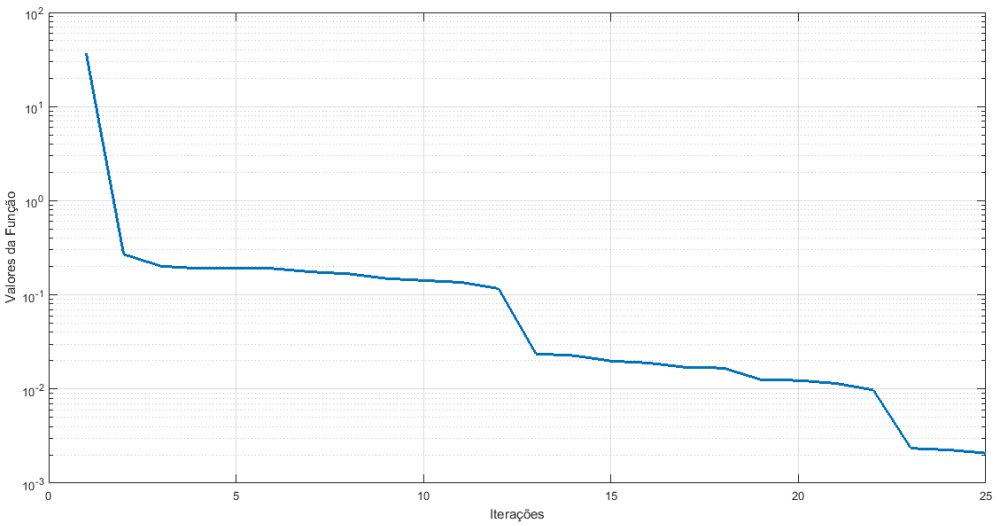


Tabela 12 – Resultados para o ponto $x_{(0)} = (-3, 3)^T$.

Tol	Niter	Naval
10^{-3}	24	138
10^{-5}	51	305
10^{-7}	72	440
10^{-9}	151	964
10^{-11}	243	1579
10^{-13}	299	1948
10^{-15}	432	2835

Figura 22 – Sequência dos iterados a partir de $x_{(0)} = (-3, 3)^T$.

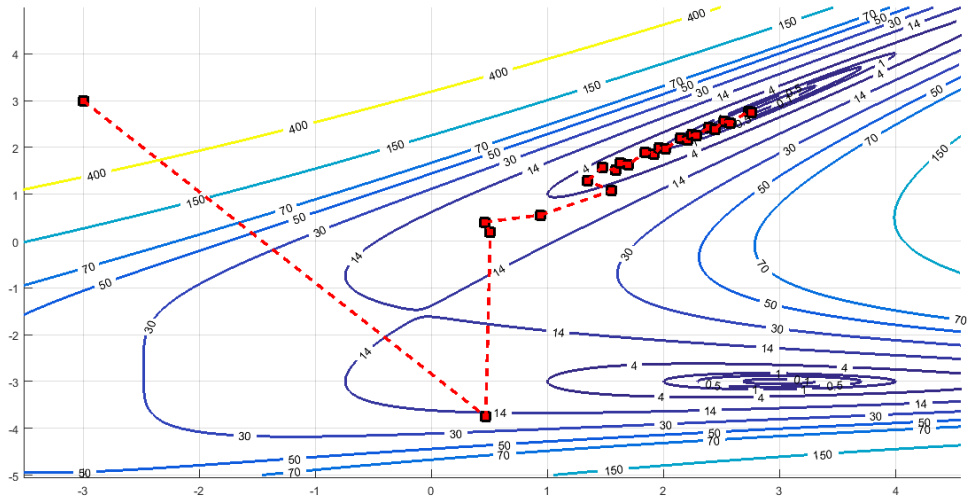


Figura 23 – Evolução da função objetivo começando em $x_{(0)} = (-3, 3)^T$.

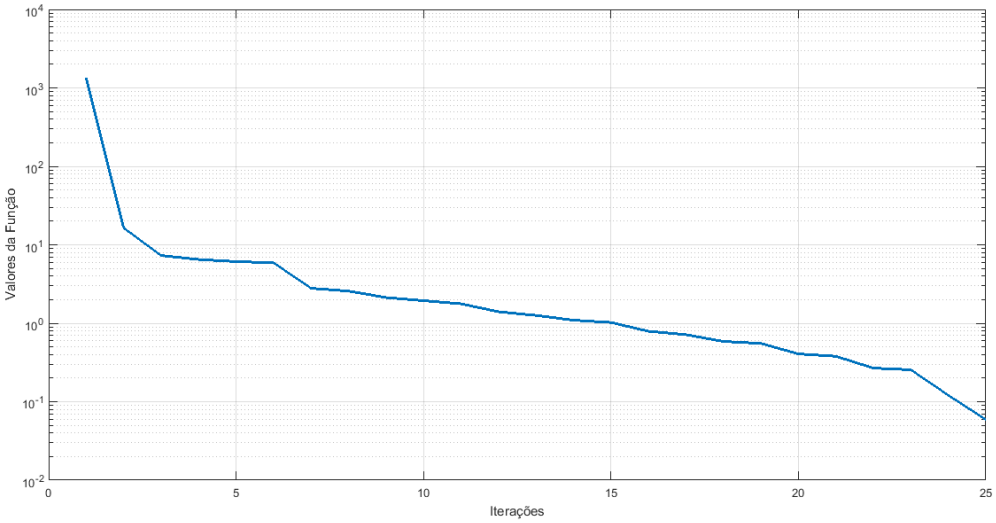


Tabela 13 – Resultados para o ponto $x_{(0)} = (-3, 2)^T$.

Tol	Niter	Naval
10^{-3}	14	92
10^{-5}	52	334
10^{-7}	88	569
10^{-9}	182	1194
10^{-11}	252	1657
10^{-13}	352	2322
10^{-15}	445	2941

Figura 24 – Sequência dos iterados a partir de $x_{(0)} = (-3, 2)^T$.

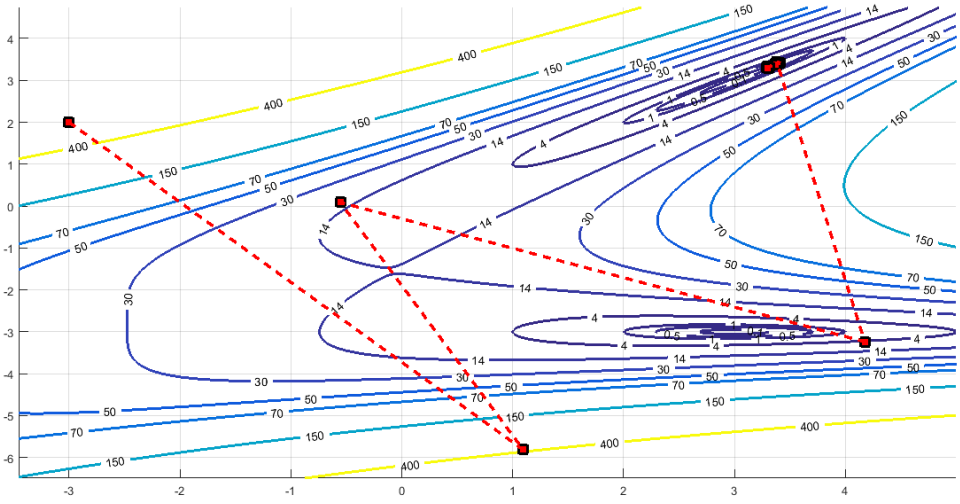


Figura 25 – Evolução da função objetivo começando em $x_{(0)} = (-3, 2)^T$.

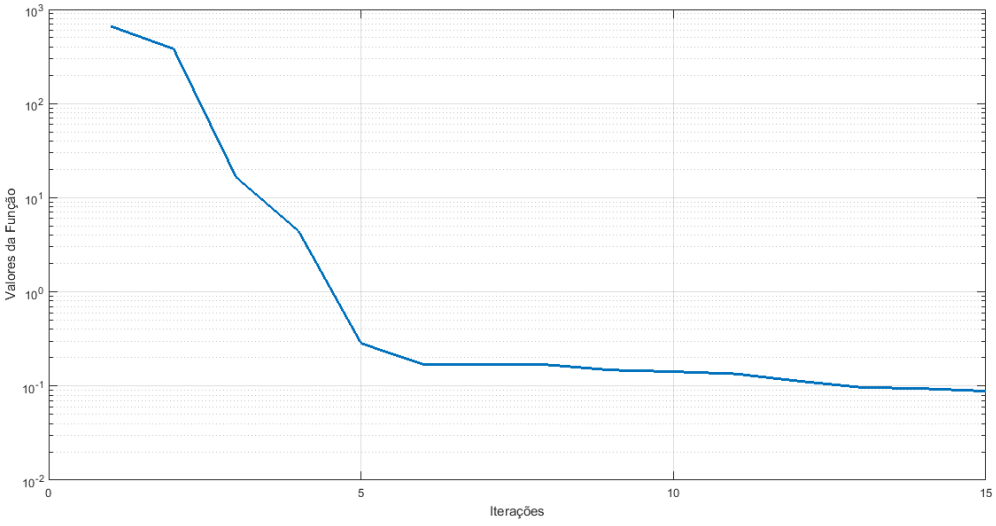


Tabela 14 – Resultados para o ponto $x_{(0)} = (0, -1.5)^T$.

Tol	Niter	Naval
10^{-3}	148	624
10^{-5}	199	840
10^{-7}	285	1208
10^{-9}	420	1785
10^{-11}	407	1725
10^{-13}	388	1637
10^{-15}	1000	1000

Figura 26 – Sequência dos iterados a partir de $x_{(0)} = (0, -1.5)^T$.

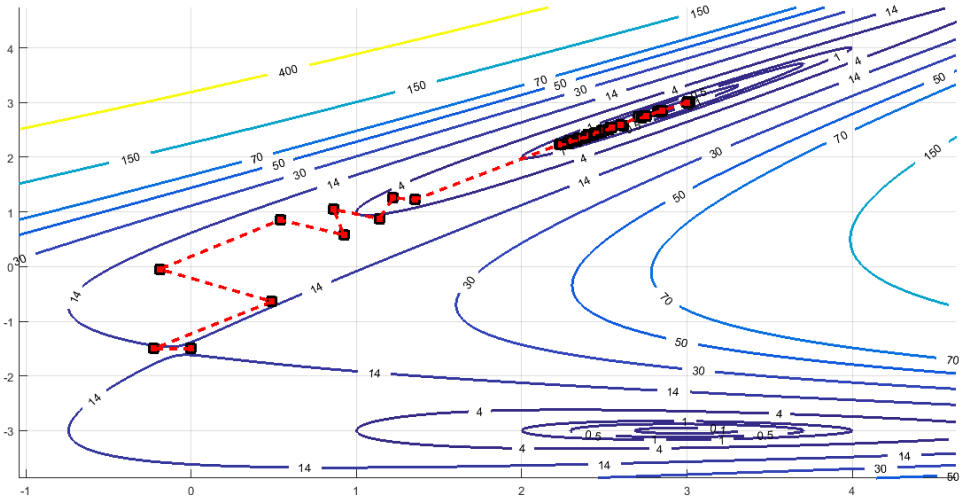
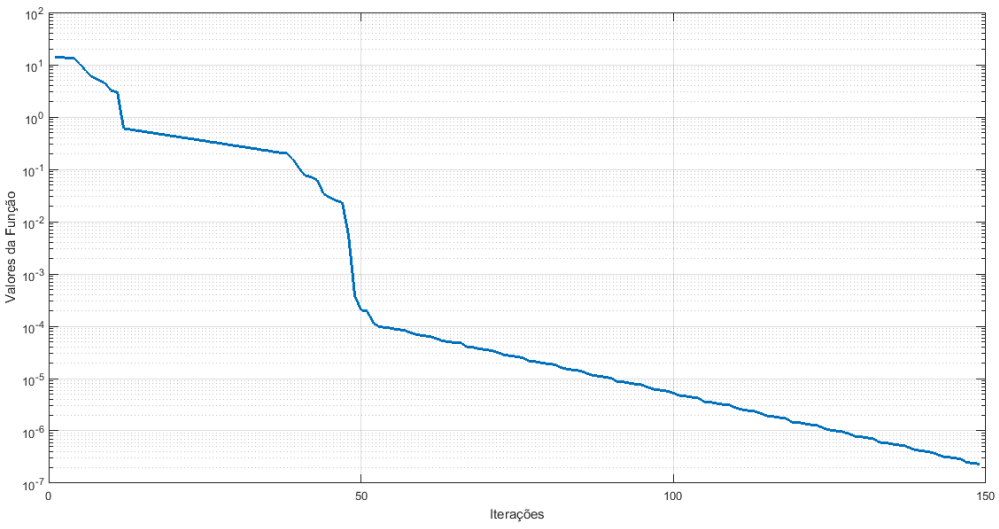


Figura 27 – Evolução da função objetivo começando em $x_{(0)} = (0, -1.5)^T$.



3.1.2 Algumas observações para o Exemplo 3.2: a trajetória descrita pelos iterandos não muda com a variação da precisão

No ponto 1) foram nescessárias apenas 2 iterações para se chegar na solução exata, e o método converge para a solução em $(x_1, x_2)^T = (3, -3)^T$. Veja que na Figura 19, o decréscimo na função objetivo se deu com taxa linear.

Em relação ao ponto 2), mesmo estando mais próximo de qualquer um dos dois pontos de mínimo, o método precisou de mais iterações para convergir, no caso, para o ponto $(x_1, x_2)^T = (3, 3)^T$.

Já em relação ao ponto 3) parece que vai convergir para o ponto $(x_1, x_2)^T = (3, -3)^T$ e depois converge para o ponto $(x_1, x_2)^T = (3, 3)^T$.

Quanto ao ponto 4), o comportamento da sequência gerada é realmente muito peculiar, pois esta chega muito próximo do ponto $(x_1, x_2)^T = (3, -3)^T$ e também, depois, acaba por convergir para o ponto $(x_1, x_2)^T = (3, 3)^T$.

Em relação ao ponto 5), (este ponto está bem próximo do ponto sela, que é um ponto estacionário) para que tivéssemos convergência foi necessário utilizar $\gamma = 0.3$, e este também foi um dos pontos que utilizou mais iterações para cada valor da tolerância (observe a Tabela 14). Veja ainda que usando Tol $= 10^{-9}$ tivemos 420 iterações e usando Tol $= 10^{-11}$ tivemos 407 iterações; essa diminuição no número de iterações, mesmo exigindo mais precisão se deve, possivelmente, a erros de arredondamento.

Com base no comportamento obtido com este ponto pode-se observar que mesmo se um ponto estiver próximo de alguma solução, não significa que o método irá convergir para esta solução. E também mesmo se variarmos a tolerância, a trajetória descrita pelos iterandos não muda, isto é, continua no mesmo caminho.

Exemplo 3.3. Tomaremos como exemplo a função de Broyden tridiagonal, veja (MORÉ; GARBOW; HILLSTROM, 1981), que é dada por:

$$f(x) = F(x)^T F(x), \tag{3.2}$$

com $F : \mathbb{R}^n \to \mathbb{R}^n$ e $F(x) = [f_1(x), f_2(x), \cdots, f_n(x)]^T$, onde,

$$\begin{aligned} f_1(x) &= (3 - 2x_1)x_1 - 2x_2 + 1, \\ f_i(x) &= (3 - 2x_i)x_i - x_{i-1} - 2x_2 + 1, \quad \text{para } i \in \{2, \ldots, n-1\}, \\ f_n(x) &= (3 - 2x_n)x_n - x_{n-1} + 1. \end{aligned}$$

E seu gradiente é:

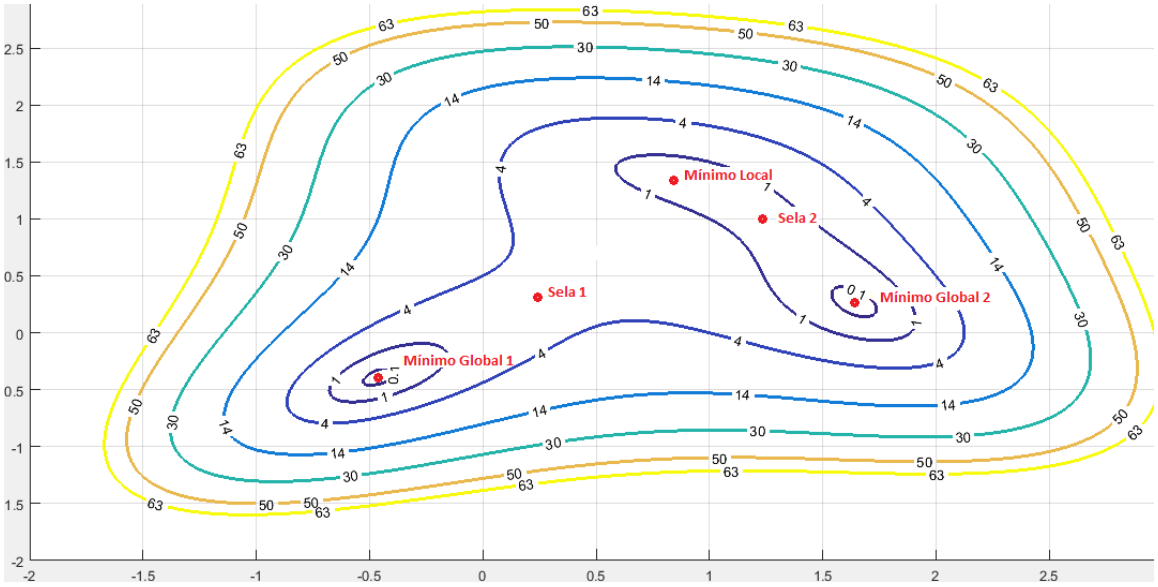
$$\nabla f(x) = 2\begin{bmatrix} (3-4x_1)f_1(x) - f_2(x) \\ -2f_1(x) + (3-4x_2)f_2(x) - f_3(x) \\ \vdots \\ -2f_{n-2}(x) + (3-4x_{n-1})f_{n-1}(x) - f_n(x) \\ -2f_{n-1}(x) + (3-4x_n)f_n(x) \end{bmatrix}.$$

Neste exemplo, consideraremos o caso $n = 2$, isto é:

$$f(x_1, x_2) = ((3-2x_1)x_1 - 2x_2 + 1)^2 + ((3-2x_2)x_2 - x_1 + 1)^2, \tag{3.3}$$

sabemos que esta função possui dois pontos de mínimo globais, que são $(x_1, x_2)^T \simeq (-0.45, -0.38)^T$ (Mínimo Global 1) e $(x_1, x_2)^T \simeq (1.64, 0.26)^T$ (Mínimo Global 2), com valor $f(x_1, x_2) = 0$, um ponto de mínimo local $(x_1, x_2)^T \simeq (0.97, 1.30)^T$ e dois pontos de sela $(x_1, x_2)^T \simeq (0.37, 0.41)^T$ (Sela 1) e $(x_1, x_2)^T = (1.25, 1)^T$ (Sela 2), veja Figura 28.

Figura 28 – Pontos onde a derivada primeira se anula.



Para este exemplo utilizamos $c_1 = 0.01$, $c_2 = 0.99$, $\gamma = 0.5$, $t_{min} = 10^{-4}$ e Maxiter $= 10^3$, e os pontos iniciais escolhidos foram:

1) $x_{(0)} = (2, 2)^T$;

2) $x_{(0)} = (3, 3)^T$;

3) $x_{(0)} = (0.5, 0.5)^T$;

4) $x_{(0)} = (-0.8, -0.7)^T$.

Tabela 15 – Resultados para o ponto $x_{(0)} = (2, 2)^T$.

Tol	Niter	Naval
10^{-3}	12	70
10^{-5}	17	103
10^{-7}	23	142
10^{-9}	29	181
10^{-11}	35	219
10^{-13}	43	270
10^{-15}	49	308

Figura 29 – Sequência dos iterados a partir de $x_{(0)} = (2, 2)^T$.

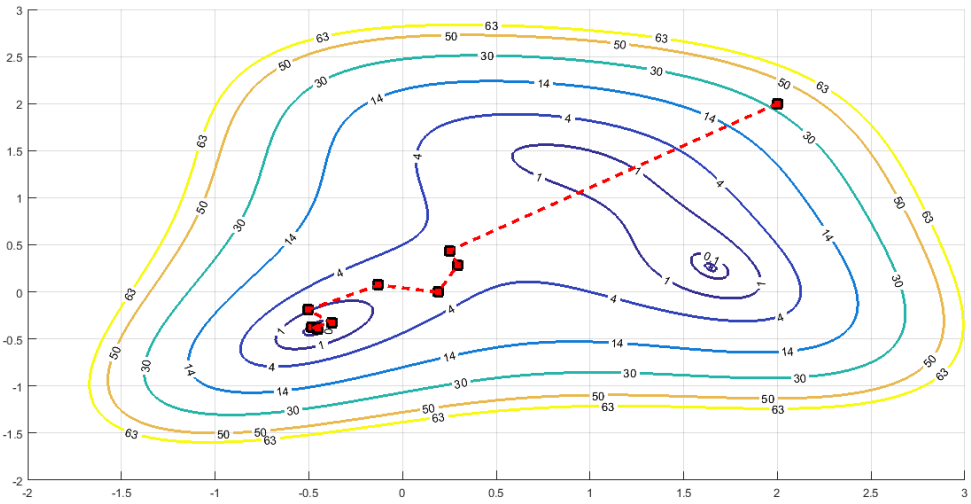


Figura 30 – Evolução da função objetivo começando em $x_{(0)} = (2, 2)^T$.

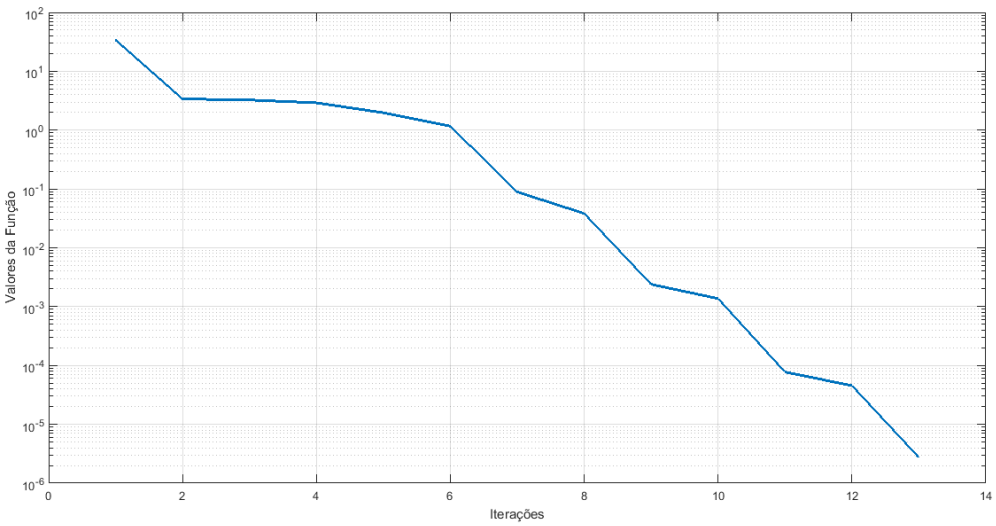


Tabela 16 – Resultados para o ponto $x_{(0)} = (3,3)^T$.

Tol	Niter	Naval
10^{-3}	4	24
10^{-5}	12	59
10^{-7}	23	106
10^{-9}	35	158
10^{-11}	46	207
10^{-13}	1000	1000
10^{-15}	1000	1000

Figura 31 – Sequência dos iterados a partir de $x_{(0)} = (3,3)^T$.

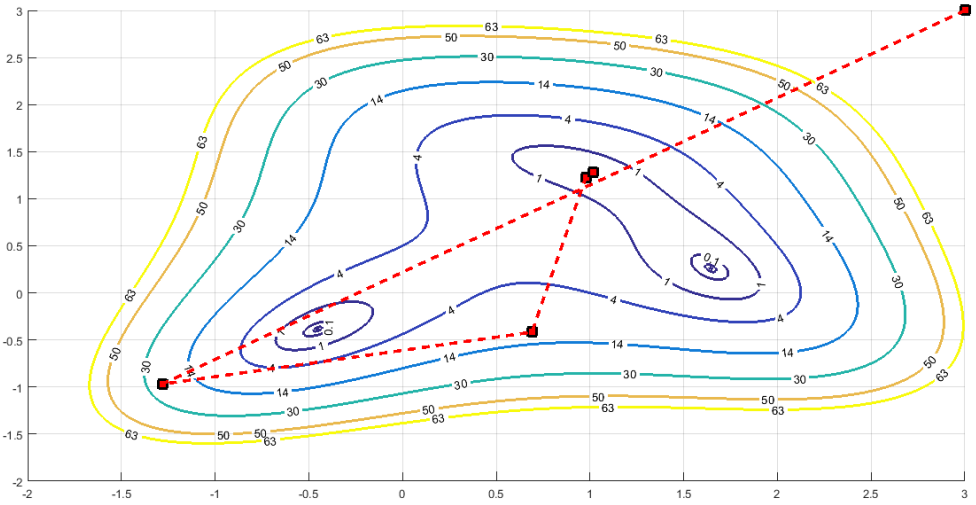


Figura 32 – Evolução da função objetivo começando em $x_{(0)} = (3,3)^T$.

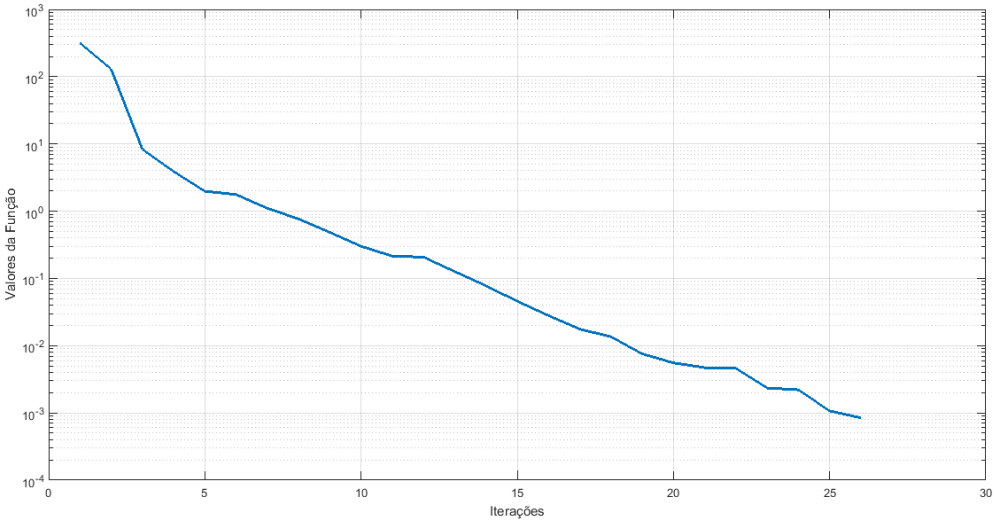


Figura 33 – Com $\gamma = 0.8$ a sequência converge para o Mínimo Global 1.

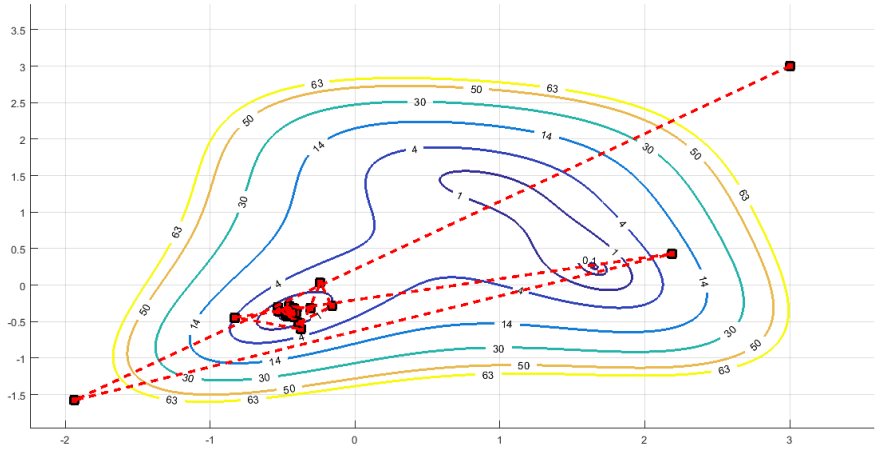


Tabela 17 – Resultados para o ponto $x_{(0)} = (0.5, 0.5)^T$.

Tol	Niter	Naval
10^{-3}	19	100
10^{-5}	31	166
10^{-7}	42	226
10^{-9}	1000	1000
10^{-11}	1000	1000
10^{-13}	1000	1000
10^{-15}	1000	1000

Figura 34 – Sequência dos iterados a partir de $x_{(0)} = (0.5, 0.5)^T$.

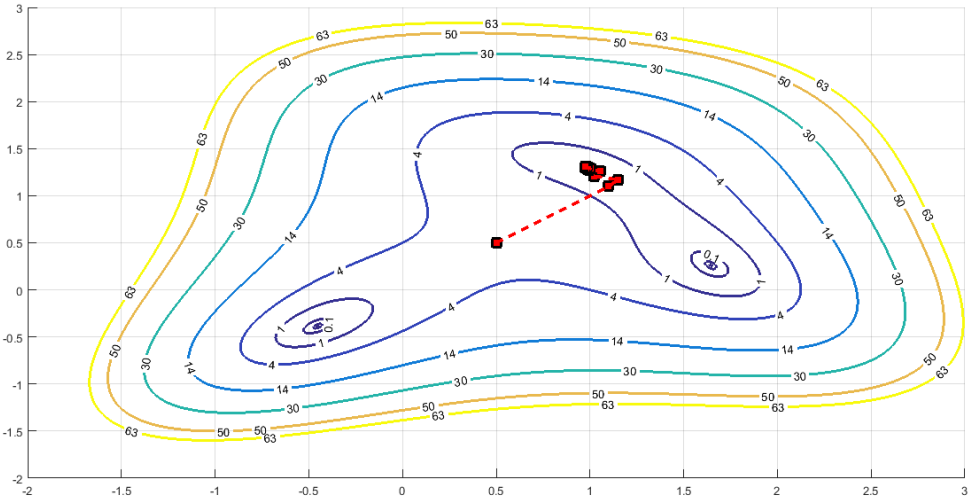


Figura 35 – Evolução da função objetivo começando em $x_{(0)} = (0.5, 0.5)^T$.

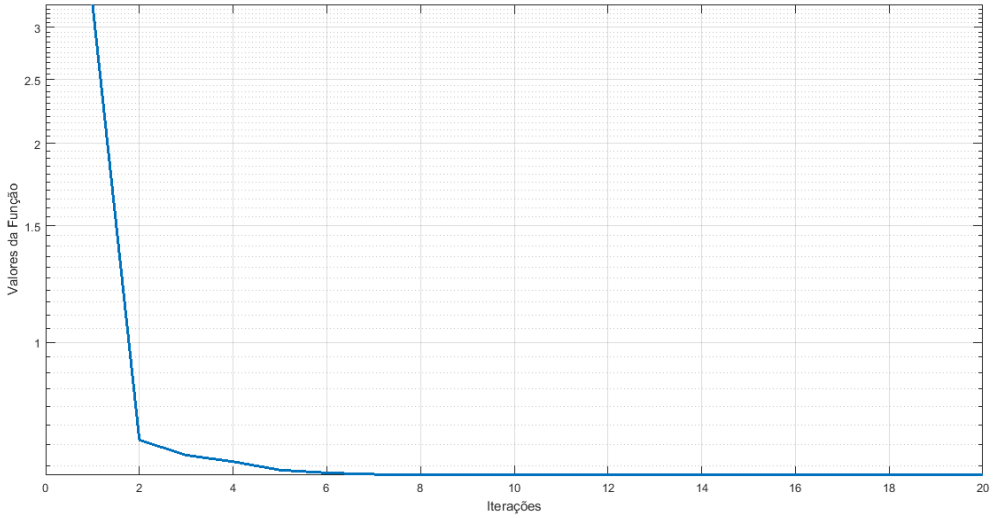


Tabela 18 – Resultados para o ponto $x_{(0)} = (-0.8, -0.7)^T$.

Tol	Niter	Naval
10^{-3}	12	57
10^{-5}	17	82
10^{-7}	24	118
10^{-9}	29	145
10^{-11}	34	171
10^{-13}	41	207
10^{-15}	46	234

Figura 36 – Sequência dos iterados a partir de $x_{(0)} = (-0.8, -0.7)^T$.

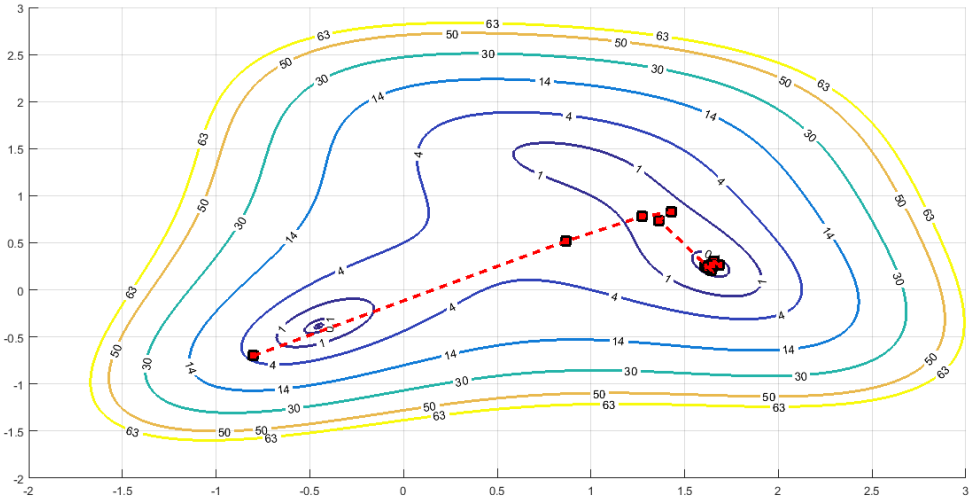


Figura 37 – Evolução da função objetivo começando em $x_{(0)} = (-0.8, -0.7)^T$.

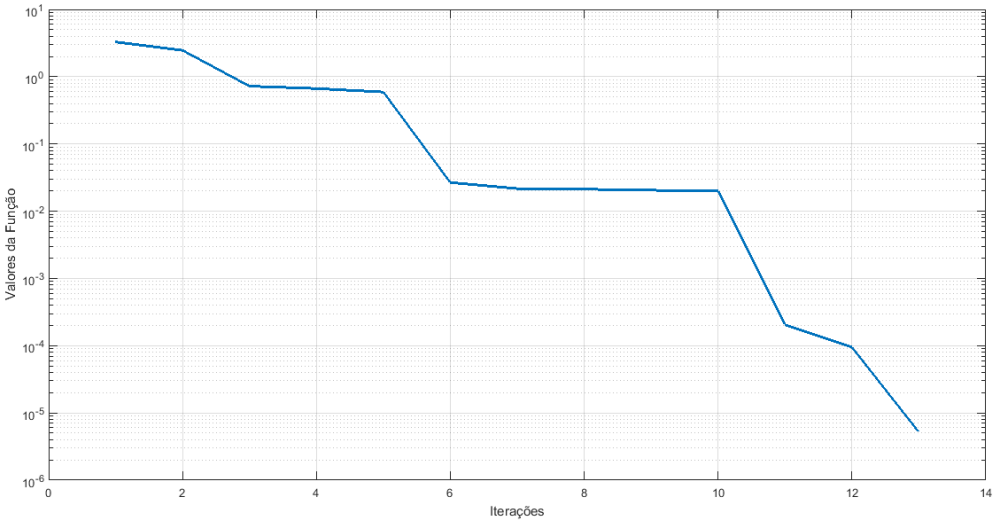
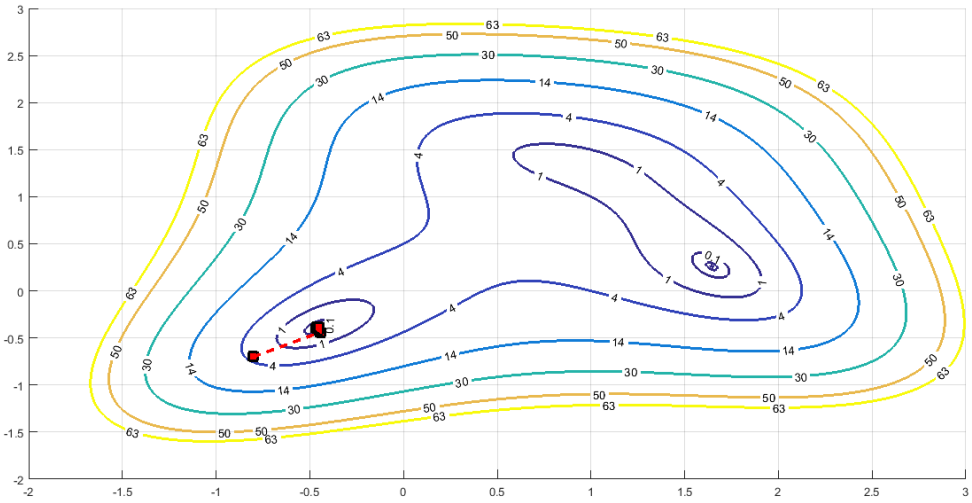


Figura 38 – Com $\gamma = 0.3$ a sequência converge para o Mínimo Global 1.



3.1.3 Algumas observações para o Exemplo 3.3: a influência do valor de γ

Com o ponto 1) os iterandos fazem uma trajetória peculiar, parece que contornam a Sela 1 e convergem para o Mínimo Global 1.

Com o ponto 2), ao usarmos $\gamma = 0.5$ a sequência converge para o Mínimo Local, veja Figura 31, e ao utilizarmos $\gamma = 0.8$ a sequência converge para o Mínimo Global 1, Figura 33, na qual podemos observar que na segunda iteração o ponto chega bem próximo do Mínimo Global 2. Observe que não houve qualquer progresso ao utilizarmos Tol $= 10^{-13}$ ou Tol $= 10^{-15}$.

Com o ponto 3) o comportamento da sequência gerada é extremamente peculiar, pois só houve convergência quando utilizamos $\gamma = 0.6$, $\gamma = 0.8$ ou $\gamma = 0.9$, e essa convergência foi para o ponto de Mínimo Local. Ao utilizarmos $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ ou $\gamma = 0.7$, não houve progresso algum na função objetivo. A Tabela 17 foi construída utilizando $\gamma = 0.6$. Notamos ainda que não houve qualquer progresso ao utilizarmos Tol $= 10^{-9}$ ou valores menores.

Quanto ao ponto 4) a sequência de iterandos converge para o Mínimo Global 2, isto se utilizarmos $\gamma = 0.5$, veja Figura 36. Já se utilizarmos $\gamma = 0.3$, a sequência converge para o Minimo Global 1, observe a Figura 38.

Exemplo 3.4. Neste exemplo analisamos a função definida por

$$f(x_1, x_2) = x_1 e^{(-x_1^2 - x_2^2 + 5)}.$$

Esta função possui seu ponto de mínimo em $(x_1, x_2)^T = \left(\frac{-\sqrt{2}}{2}, 0\right)^T$, com valor $f(x_1, x_2) \simeq -63.65$.

Também neste exemplo utilizamos $c_1 = 0.01$, $c_2 = 0.99$, $\gamma = 0.5$, $t_{min} = 10^{-4}$ e Maxiter $= 10^3$, e os pontos iniciais escolhidos foram:

1) $x_{(0)} = (-1.5, 0.8)^T$;

2) $x_{(0)} = (-1, -1)^T$;

3) $x_{(0)} = (2, 2)^T$;

4) $x_{(0)} = (-1, 1.3)^T$.

Tabela 19 – Resultados para o ponto $x_{(0)} = (-1.5, 0.8)^T$.

Tol	Niter	Naval
10^{-3}	10	80
10^{-5}	12	97
10^{-7}	15	122
10^{-9}	16	132
10^{-11}	1000	1000
10^{-13}	1000	1000
10^{-15}	1000	1000

Figura 39 – Sequência dos iterados a partir de $x_{(0)} = (-1.5, 0.8)^T$.

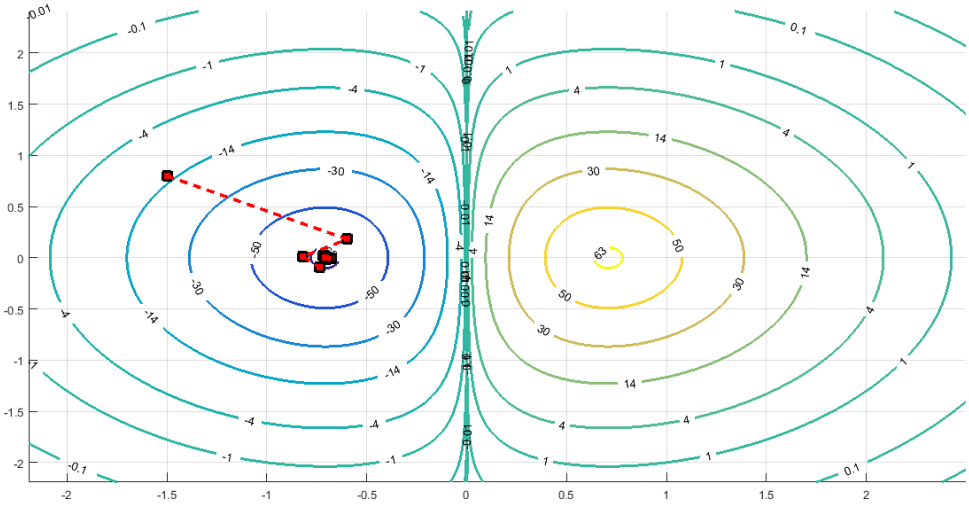


Figura 40 – Evolução da função objetivo começando em $x_{(0)} = (-1.5, 0.8)^T$.

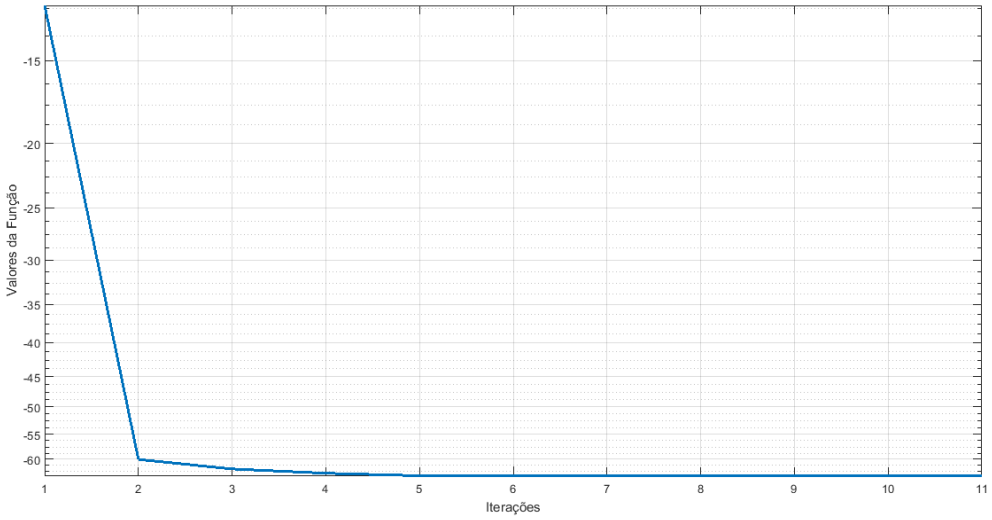


Tabela 20 – Resultados para o ponto $x_{(0)} = (-1, -1)^T$.

Tol	Niter	Naval
10^{-3}	17	133
10^{-5}	23	182
10^{-7}	29	230
10^{-9}	1000	1000
10^{-11}	1000	1000
10^{-13}	1000	1000
10^{-15}	1000	1000

Figura 41 – Sequência dos iterados a partir de $x_{(0)} = (-1, -1)^T$.

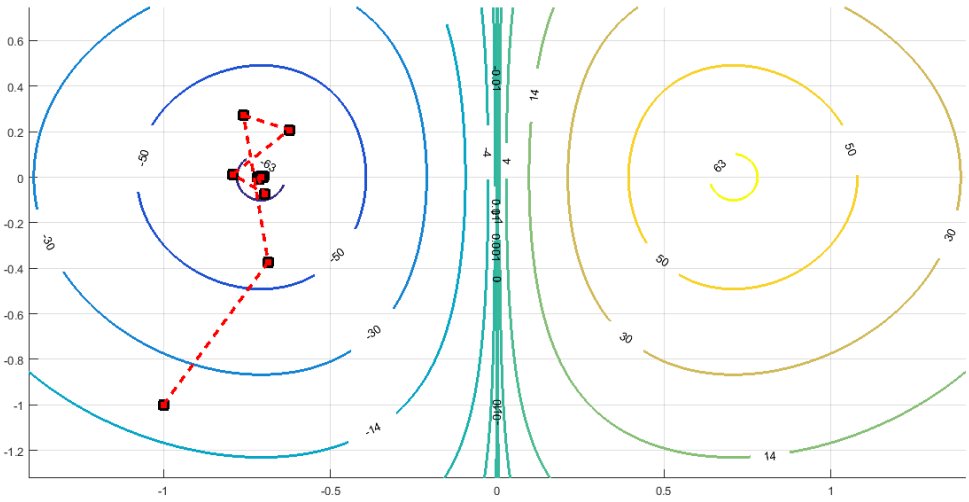


Figura 42 – Evolução da função objetivo começando em $x_{(0)} = (-1, -1)^T$.

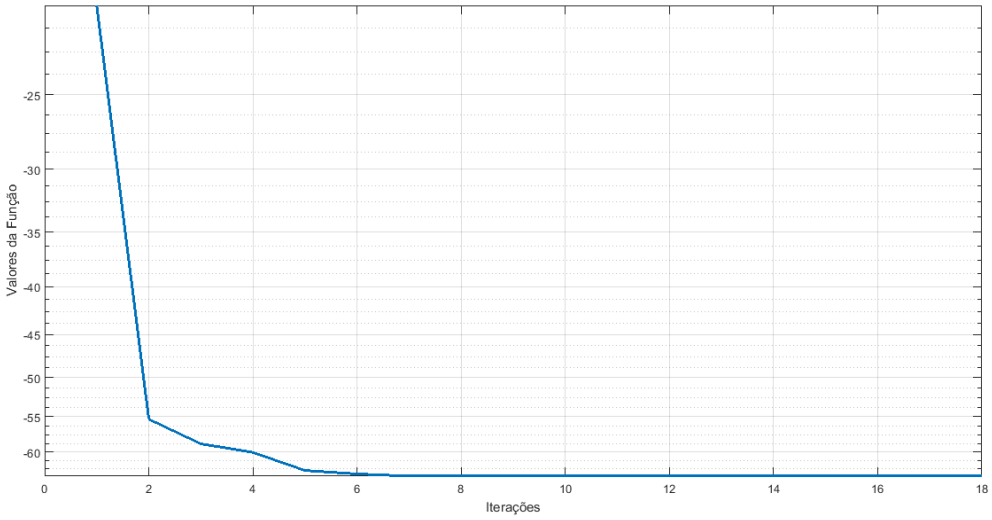


Tabela 21 – Resultados para o ponto $x_{(0)} = (2, 2)^T$.

Tol	Niter	Naval
10^{-3}	1000	1000
10^{-5}	1000	1000
10^{-7}	1000	1000
10^{-9}	1000	1000
10^{-11}	1000	1000
10^{-13}	1000	1000
10^{-15}	1000	1000

Figura 43 – Sequência dos iterados a partir de $x_{(0)} = (2, 2)^T$.

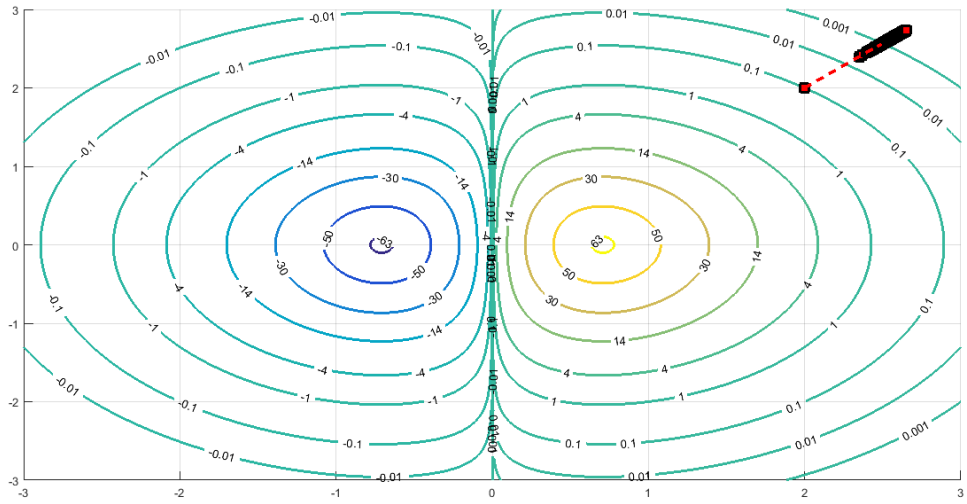


Figura 44 – Evolução da função objetivo começando em $x_{(0)} = (2, 2)^T$.

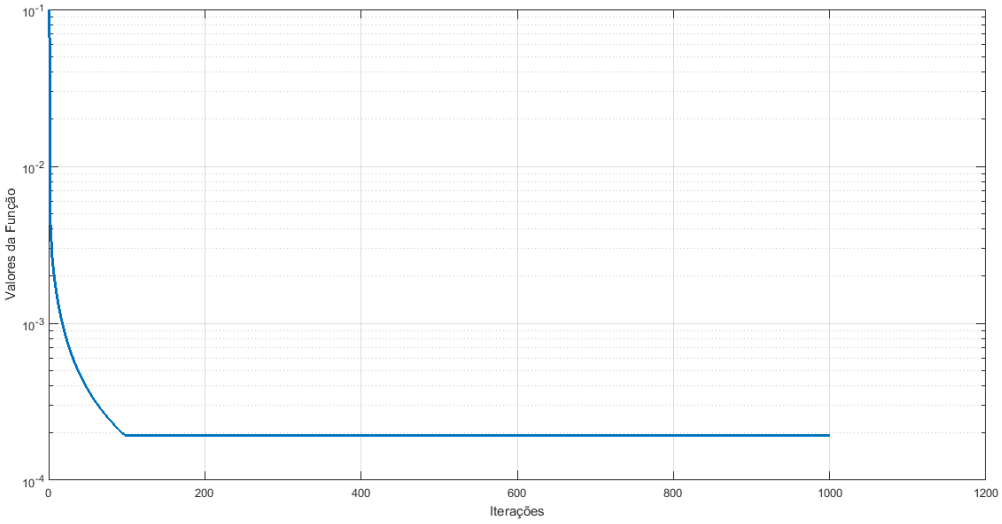


Tabela 22 – Resultados para o ponto $x_{(0)} = (-1, 1.3)^T$.

Tol	Niter	Naval
10^{-3}	106	4982
10^{-5}	152	7187
10^{-7}	195	9275
10^{-9}	223	10622
10^{-11}	239	11416
10^{-13}	257	12311
10^{-15}	1000	1000

Figura 45 – Sequência dos iterados a partir de $x_{(0)} = (-1, 1.3)^T$.

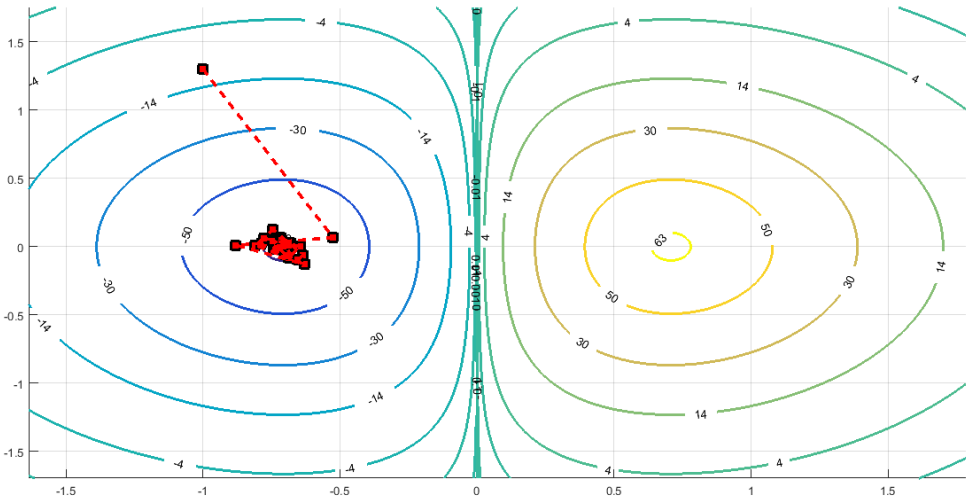
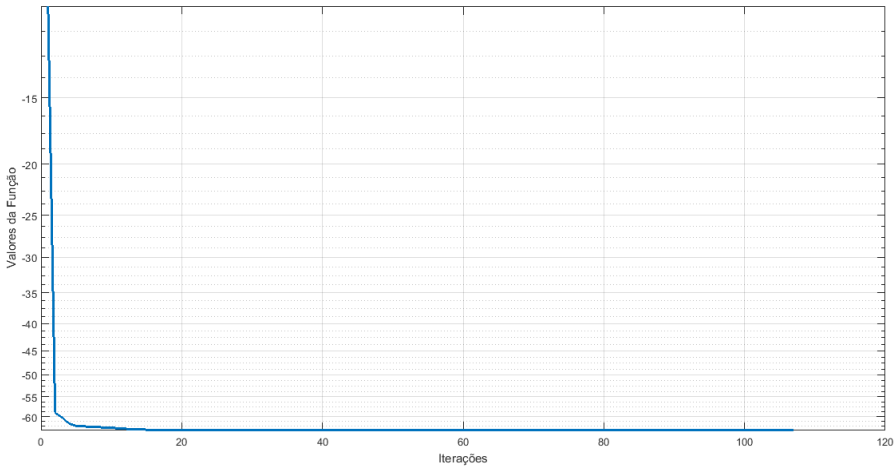


Figura 46 – Evolução da função objetivo começando em $x_{(0)} = (-1, 1.3)^T$.



3.1.4 Algumas observações para o Exemplo 3.4: começou do lado errado

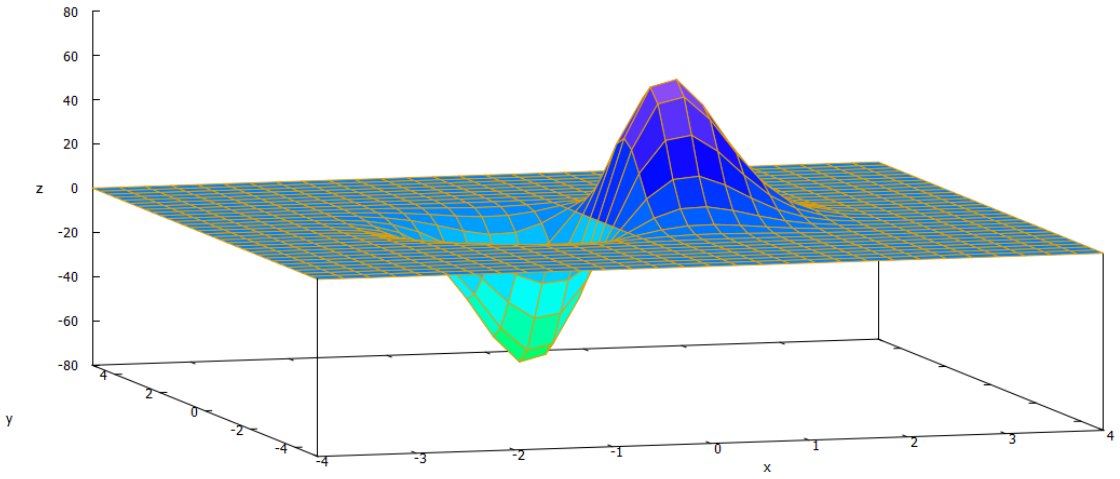
No ponto 1) os iterandos pouco avançam na função objetivo, veja que a função não conseguiu calcular para Tol $= 10^{-11}$ ou valores menores.

No ponto 2) também os iterandos pouco avançam na função objetivo, não houve progresso na função objetivo para Tol $= 10^{-9}$ ou valores menores.

No ponto 3) nenhum dos valores já utilizados $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ fizeram com que a sequência de iterandos convergisse para o ponto $(x_1, x_2)^T = \left(\frac{-\sqrt{2}}{2}, 0\right)^T$, isto se deu pelo fato que começamos com a sequência pelo lado oposto da montanha, veja Figura 47. Na verdade, menos de 100 iterações já foram suficientes para que a norma do gradiente ficasse extremamente pequena.

No ponto 4) só conseguimos que a sequência de iterandos convergisse para o valor correto utilizando $\gamma = 0.9$.

Figura 47 – Gráfico da função do Exemplo 3.4



Os exemplos 3.5 e 3.6 foram feitos com o algoritmo dado no Apêndice A.4 mudando apenas o escalar $\beta_{(i)}$ de acordo com a Tabela 3.

Exemplo 3.5. Agora tomaremos a função de Broyden tridiagonal, dada em (3.3), porém utilizaremos os quatro distintos valores de $\beta_{(i)}$ dados pelo Teorema 2.6.

Em cada tabela a seguir informamos o ponto inicial, o número de iterações para os quatro valores do escalar $\beta_{(i)}$, bem como γ, t_{min} e Tol. Deixamos fixos $c_1 = 0.01$, $c_2 = 0.99$ e Maxiter $= 10^3$.

Tabela 23 – Iterações do método com as escolhas: $x_{(0)} = (-1, -1)^T$, $\gamma = 0.5$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	5	9	5	9
10^{-5}	12	17	18	23
10^{-7}	24	22	27	33
10^{-9}	37	34	37	44
10^{-11}	1000	47	48	1000
10^{-13}	1000	1000	1000	1000
10^{-15}	1000	1000	1000	1000

Tabela 24 – Iterações do método com as escolhas: $x_{(0)} = (2,2)^T$, $\gamma = 0.5$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	12	13	14	11
10^{-5}	17	18	22	22
10^{-7}	23	24	28	30
10^{-9}	29	32	36	38
10^{-11}	35	39	43	45
10^{-13}	43	46	50	53
10^{-15}	49	51	56	59

Tabela 25 – Iterações do método com as escolhas: $x_{(0)} = (2,1)^T$, $\gamma = 0.3$ e $t_{min} = 10^{-3}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	9	13	7	12
10^{-5}	15	17	12	19
10^{-7}	24	23	17	28
10^{-9}	32	31	27	37
10^{-11}	39	41	34	45
10^{-13}	48	50	39	54
10^{-15}	52	62	43	62

Tabela 26 – Iterações do método com as escolhas: $x_{(0)} = (-3,4)^T$, $\gamma = 0.8$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	14	14	20	22
10^{-5}	45	42	46	73
10^{-7}	75	71	69	116
10^{-9}	103	100	89	143
10^{-11}	133	128	110	172
10^{-13}	162	157	131	196
10^{-15}	186	186	151	215

Tabela 27 – Iterações do método com as escolhas: $x_{(0)} = (1,-3)^T$, $\gamma = 0.6$ e $t_{min} = 10^{-3}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	8	16	8	11
10^{-5}	14	22	14	20
10^{-7}	19	28	20	27
10^{-9}	25	37	28	32
10^{-11}	30	43	35	40
10^{-13}	38	51	41	47
10^{-15}	44	56	49	54

3.1.5 Algumas observações para o Exemplo 3.5: a escolha para o $\beta_{(i)}$ faz diferença (problemas bidimensionais)

Na Tabela 23, todos os valores de $\beta_{(i)}$ fizeram com que a sequência de iterandos convergisse para o Mínimo Local, veja Figura 28. Observe que para cada valor da tolerância, o escalar $\beta_{(i)}$ dado por Hestenes e Stiefel foi o menos eficiente.

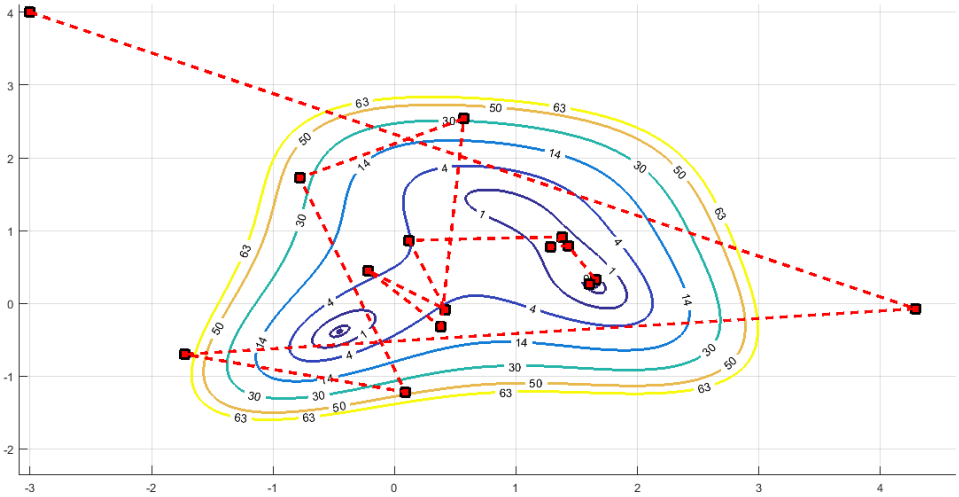
Na Tabela 24, considerando todos os $\beta_{(i)}$, houve convergência para o Mínimo Global 1, veja Figura 28, se não considerarmos a Tolerância de 10^{-3}, podemos colocar os $\beta_{(i)}$ em ordem crescente de eficiência[1] da seguinte maneira: HS, DY, PRP, FR.

Já na Tabela 25, o Mínimo Global 1 foi atingido considerando todos os $\beta_{(i)}$, e o $\beta_{(i)}$ dado por Day e Yuan foi o mais eficiente.

Com relação à Tabela 26, utilizando-se os $\beta_{(i)}$ dados por Fletcher e Reeves; Polak, Ribière e Polyak; Hestenes e Stiefel, tivemos que a sequência de iterandos convergiu para o Mínimo Global 2, veja a Figura 48. Já o $\beta_{(i)}$ dado por Dai e Yuan fez com que a sequência convergisse para o Mínimo Global 1, observe a Figura 49. Repare que o $\beta_{(i)}$ dado por Hestenes e Stiefel teve o pior desempenho.

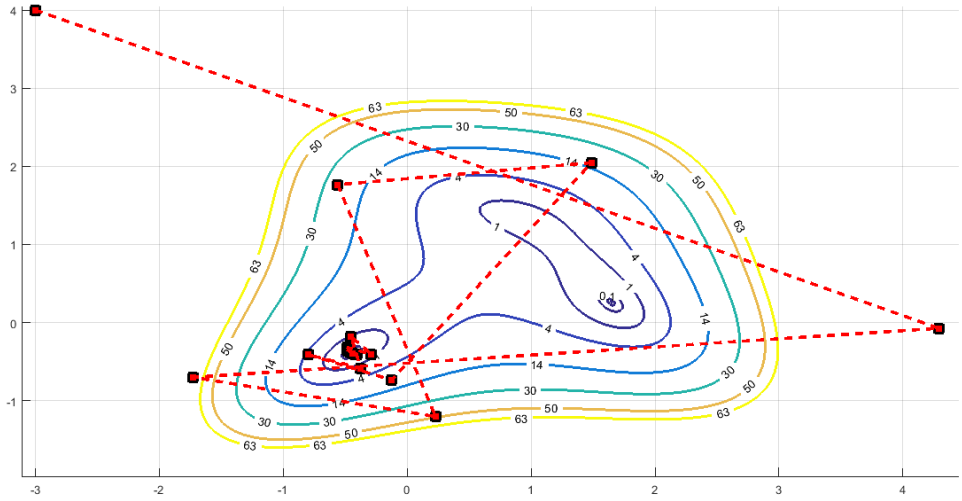
Na Tabela 27, houve convergência para o Mínimo Global 1 para todos os $\beta_{(i)}$. Repare que o $\beta_{(i)}$ dado por Fletcher e Reeves teve o melhor desempenho, já o dado por Polak, Ribière e Polyak teve o pior.

Figura 48 – Sequência dos iterados obtidos com $\beta_{(i)}$ dado por Fletcher e Reeves para o ponto $x_{(0)} = (-3, 4)^T$ da Tabela 26.



[1] Considerando o número de iterações.

Figura 49 – Sequência dos iterados obtidos com $\beta_{(i)}$ dado por Dai e Yuan para o ponto $x_{(0)} = (-3, 4)^T$ da Tabela 26.



Exemplo 3.6. Tomaremos novamente a função de Broyden tridiagonal, dada em (3.2), porém aqui faremos para distintos valores de $n \in \{15, 20, 25, 30, 35, 40, 45\}$, logo não teremos mais o recurso visual. Entretanto, mesmo assim podemos analisar a convergência da função, onde utilizaremos os quatro distintos valores de $\beta_{(i)}$ dados pelo Teorema 2.6.

Para todos os casos a seguir tomamos como ponto inicial o vetor coluna em $\mathbb{R}^n$ com todas as componentes iguais a -1, e apresentamos tabelas com o número de iterações demandado para os quatro valores do escalar $\beta_{(i)}$, juntamente com o valor da função objetivo no ponto obtido, bem como informamos nas legendas as escolhas correspondentes para $n, \gamma,\ t_{min}$ e Tol. Os valores para γ e $t_{\min}$ foram escolhidos empiricamente, de modo a permitir bons resultados para as quatro variantes de $\beta_{(i)}$ e para todos os valores considerados para a tolerância de parada. Deixamos fixos $c_1 = 0.01,\ c_2 = 0.99$ e Maxiter $= 10^3$.

Tabela 28 – Iterações do método com as escolhas: $n = 15$, $\gamma = 0.5$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	46 $\|$ 1.176×10^{-5}	18 $\|$ 1.637×10^{-5}	28 $\|$ 3.112×10^{-5}	22 $\|$ 1.033×10^{-5}
10^{-5}	53 $\|$ 4.090×10^{-9}	27 $\|$ 2.135×10^{-9}	38 $\|$ 2.245×10^{-9}	34 $\|$ 4.588×10^{-10}
10^{-7}	66 $\|$ 5.7960×10^{-14}	44 $\|$ 1.871×10^{-13}	53 $\|$ 2.461×10^{-13}	42 $\|$ 2.435×10^{-13}
10^{-9}	80 $\|$ 1.347×10^{-17}	53 $\|$ 2.093×10^{-17}	70 $\|$ 4.011×10^{-17}	57 $\|$ 2.961×10^{-17}
10^{-11}	92 $\|$ 5.029×10^{-21}	61 $\|$ 2.788×10^{-21}	88 $\|$ 1.729×10^{-21}	75 $\|$ 2.032×10^{-21}
10^{-13}	102 $\|$ 2.698×10^{-25}	82 $\|$ 2.443×10^{-25}	93 $\|$ 4.112×10^{-25}	96 $\|$ 1.391×10^{-25}
10^{-15}	124 $\|$ 2.550×10^{-29}	96 $\|$ 2.420×10^{-29}	101 $\|$ 4.353×10^{-29}	111 $\|$ 1.375×10^{-29}

Tabela 29 – Iterações do método com as escolhas: $n = 20$, $\gamma = 0.7$ e $t_{min} = 10^{-3}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	39 $\|$ 2.672×10^{-5}	30 $\|$ 1.103×10^{-5}	29 $\|$ 1.580×10^{-5}	34 $\|$ 1.237×10^{-5}
10^{-5}	61 $\|$ 6.739×10^{-10}	50 $\|$ 1.943×10^{-9}	46 $\|$ 1.052×10^{-9}	48 $\|$ 9.853×10^{-10}
10^{-7}	76 $\|$ 1.691×10^{-13}	72 $\|$ 1.790×10^{-13}	65 $\|$ 1.004×10^{-13}	62 $\|$ 8.106×10^{-14}
10^{-9}	94 $\|$ 6.517×10^{-18}	94 $\|$ 1.763×10^{-17}	75 $\|$ 1.145×10^{-17}	76 $\|$ 4.945×10^{-18}
10^{-11}	119 $\|$ 6.578×10^{-22}	116 $\|$ 1.713×10^{-21}	94 $\|$ 1.184×10^{-21}	90 $\|$ 6.967×10^{-22}
10^{-13}	134 $\|$ 1.476×10^{-25}	139 $\|$ 1.344×10^{-25}	110 $\|$ 1.053×10^{-25}	108 $\|$ 7.244×10^{-26}
10^{-15}	149 $\|$ 1.810×10^{-29}	160 $\|$ 2.293×10^{-29}	123 $\|$ 1.089×10^{-29}	128 $\|$ 1.207×10^{-29}

Tabela 30 – Iterações do método com as escolhas: $n = 25$, $\gamma = 0.2$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	35 $\|$ 6.648×10^{-5}	23 $\|$ 1.767×10^{-5}	36 $\|$ 2.902×10^{-5}	23 $\|$ 2.596×10^{-5}
10^{-5}	46 $\|$ 4.952×10^{-9}	35 $\|$ 2.673×10^{-9}	45 $\|$ 2.875×10^{-9}	35 $\|$ 2.453×10^{-9}
10^{-7}	57 $\|$ 3.000×10^{-13}	46 $\|$ 2.532×10^{-13}	56 $\|$ 3.413×10^{-13}	46 $\|$ 1.499×10^{-13}
10^{-9}	73 $\|$ 1.865×10^{-17}	57 $\|$ 2.883×10^{-17}	67 $\|$ 3.038×10^{-17}	58 $\|$ 2.137×10^{-17}
10^{-11}	84 $\|$ 2.215×10^{-21}	71 $\|$ 1.714×10^{-21}	80 $\|$ 3.436×10^{-21}	71 $\|$ 1.864×10^{-21}
10^{-13}	96 $\|$ 1.896×10^{-25}	84 $\|$ 2.204×10^{-25}	91 $\|$ 1.994×10^{-25}	82 $\|$ 2.104×10^{-25}
10^{-15}	109 $\|$ 2.287×10^{-29}	95 $\|$ 2.855×10^{-29}	103 $\|$ 3.466×10^{-29}	93 $\|$ 2.070×10^{-29}

Tabela 31 – Iterações do método com as escolhas: $n = 30$, $\gamma = 0.1$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	14 $\|$ 5.293×10^{-5}	28 $\|$ 1.046×10^{-5}	13 $\|$ 5.311×10^{-5}	38 $\|$ 8.380×10^{-5}
10^{-5}	25 $\|$ 5.848×10^{-9}	42 $\|$ 3.885×10^{-9}	25 $\|$ 3.357×10^{-9}	65 $\|$ 5.140×10^{-9}
10^{-7}	41 $\|$ 4.519×10^{-13}	59 $\|$ 5.880×10^{-13}	38 $\|$ 1.158×10^{-12}	92 $\|$ 2.797×10^{-13}
10^{-9}	55 $\|$ 6.566×10^{-17}	75 $\|$ 8.010×10^{-17}	55 $\|$ 4.478×10^{-17}	121 $\|$ 9.417×10^{-17}
10^{-11}	71 $\|$ 6.823×10^{-21}	91 $\|$ 7.287×10^{-21}	67 $\|$ 5.051×10^{-21}	148 $\|$ 1.178×10^{-20}
10^{-13}	87 $\|$ 8.202×10^{-25}	111 $\|$ 3.310×10^{-25}	85 $\|$ 8.780×10^{-25}	174 $\|$ 2.538×10^{-25}
10^{-15}	102 $\|$ 8.665×10^{-29}	130 $\|$ 4.427×10^{-29}	101 $\|$ 7.380×10^{-29}	205 $\|$ 8.570×10^{-29}

Tabela 32 – Iterações do método com as escolhas: $n = 35$, $\gamma = 0.5$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	31 $\|$ 2.849×10^{-5}	18 $\|$ 3.048×10^{-5}	28 $\|$ 2.269×10^{-5}	22 $\|$ 1.846×10^{-5}
10^{-5}	45 $\|$ 4.754×10^{-10}	29 $\|$ 1.999×10^{-9}	42 $\|$ 3.739×10^{-9}	33 $\|$ 3.334×10^{-9}
10^{-7}	55 $\|$ 9.371×10^{-13}	42 $\|$ 6.331×10^{-13}	52 $\|$ 3.523×10^{-13}	47 $\|$ 4.747×10^{-13}
10^{-9}	67 $\|$ 7.703×10^{-17}	59 $\|$ 2.261×10^{-17}	64 $\|$ 2.823×10^{-17}	58 $\|$ 2.763×10^{-17}
10^{-11}	82 $\|$ 5.330×10^{-21}	67 $\|$ 8.910×10^{-21}	79 $\|$ 2.515×10^{-21}	76 $\|$ 2.686×10^{-21}
10^{-13}	99 $\|$ 1.764×10^{-25}	85 $\|$ 3.622×10^{-25}	88 $\|$ 2.423×10^{-25}	86 $\|$ 6.919×10^{-25}
10^{-15}	110 $\|$ 6.301×10^{-29}	99 $\|$ 3.850×10^{-29}	101 $\|$ 5.502×10^{-29}	107 $\|$ 2.673×10^{-29}

3.1.6 Algumas observações para o Exemplo 3.6: a escolha para o $\beta_{(i)}$ faz diferença (problemas com dimensões maiores)

Em todos os casos, os valores atingidos para a função objetivo são consistentes com a precisão requerida. O valor ótimo atingido na grande maioria dos casos foi da ordem

Tabela 33 – Iterações do método com as escolhas: $n = 40$, $\gamma = 0.4$ e $t_{min} = 10^{-2}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	18 $\|$ 4.203×10^{-5}	20 $\|$ 1.942×10^{-5}	28 $\|$ 2.177×10^{-5}	20 $\|$ 1.864×10^{-5}
10^{-5}	29 $\|$ 3.077×10^{-9}	34 $\|$ 1.790×10^{-9}	47 $\|$ 1.274×10^{-9}	30 $\|$ 7.149×10^{-9}
10^{-7}	42 $\|$ 2.273×10^{-13}	50 $\|$ 3.789×10^{-13}	59 $\|$ 3.996×10^{-13}	42 $\|$ 7.680×10^{-13}
10^{-9}	53 $\|$ 3.730×10^{-17}	65 $\|$ 4.454×10^{-17}	71 $\|$ 5.041×10^{-17}	51 $\|$ 5.454×10^{-17}
10^{-11}	67 $\|$ 4.937×10^{-21}	77 $\|$ 5.209×10^{-21}	89 $\|$ 3.431×10^{-21}	66 $\|$ 5.265×10^{-21}
10^{-13}	84 $\|$ 5.838×10^{-25}	91 $\|$ 5.900×10^{-25}	102 $\|$ 5.398×10^{-25}	81 $\|$ 2.000×10^{-25}
10^{-15}	103 $\|$ 2.711×10^{-29}	111 $\|$ 3.377×10^{-29}	121 $\|$ 3.830×10^{-29}	94 $\|$ 5.236×10^{-29}

Tabela 34 – Iterações do método com as escolhas: $n = 45$, $\gamma = 0.5$ e $t_{min} = 10^{-4}$.

Tol	**FR**	**PRP**	**DY**	**HS**
10^{-3}	31 $\|$ 2.705×10^{-5}	19 $\|$ 1.967×10^{-5}	25 $\|$ 2.964×10^{-5}	22 $\|$ 1.910×10^{-5}
10^{-5}	47 $\|$ 2.083×10^{-9}	27 $\|$ 8.426×10^{-10}	41 $\|$ 2.186×10^{-9}	32 $\|$ 6.143×10^{-9}
10^{-7}	60 $\|$ 2.978×10^{-13}	39 $\|$ 6.512×10^{-13}	58 $\|$ 7.691×10^{-13}	48 $\|$ 2.420×10^{-13}
10^{-9}	79 $\|$ 3.433×10^{-17}	55 $\|$ 5.585×10^{-17}	70 $\|$ 3.601×10^{-17}	63 $\|$ 9.986×10^{-17}
10^{-11}	103 $\|$ 2.810×10^{-21}	65 $\|$ 1.024×10^{-20}	77 $\|$ 3.990×10^{-21}	78 $\|$ 4.573×10^{-21}
10^{-13}	123 $\|$ 8.746×10^{-25}	78 $\|$ 3.376×10^{-25}	94 $\|$ 1.021×10^{-24}	96 $\|$ 2.628×10^{-25}
10^{-15}	137 $\|$ 7.505×10^{-29}	93 $\|$ 5.472×10^{-29}	106 $\|$ 3.914×10^{-29}	117 $\|$ 3.756×10^{-29}

de $10 \times (\text{Tol})^2$, o que nos mostra que a qualidade da solução obtida para esse problema não foi afetada pelas escolhas para $\beta_{(i)}$.

Na Tabela 28, considerando apenas o número de iterações, o mais eficiente $\beta_{(i)}$ foi dado por Polak, Ribière e Polyak, já o de Fletcher e Reeves foi o menos eficiente.

Na Tabela 29, o mais eficiente $\beta_{(i)}$ foi dado por Dai e Yuan, considerando-se Tol=10^{-3}, Tol=10^{-5}, Tol=10^{-9} e Tol=10^{-15}, já o de Hestenes e Stiefel foi mais eficiente para Tol=10^{-7}, Tol=10^{-13} e Tol=10^{-11}.

Com relação à Tabela 30, os $\beta_{(i)}$ dados por Polak, Ribière e Polyak e Hestenes e Stiefel tiveram desempenho semelhante, em termos do número de iterações efetuado.

Já para a Tabela 31, a menos de ordem de grandeza da função objetivo, podemos colocar os $\beta_{(i)}$ em ordem crescente de eficiência da seguinte maneira: HS, PRP, FR, DY.

Com relação à Tabela 32, o $\beta_{(i)}$ dado por Polak, Ribière e Polyak teve o melhor desempenho, já o de Fletcher e Reeves teve o pior desempenho, levando em consideração o número de iterações demandado.

Na Tabela 33 o mais eficiente $\beta_{(i)}$, considerando-se Tol $= 10^{-7}$ ou menor, foi dado por Hestenes e Stiefel, já o $\beta_{(i)}$ dado por Fletcher e Reeves teve desempenho melhor para Tol $= 10^{-3}$ e Tol $= 10^{-5}$.

Com relação a Tabela 34 o mais eficiente $\beta_{(i)}$ foi dado por Polak, Ribière e Polyak, já o de Fletcher e Reeves foi o menos eficiente, do ponto de vista do número de

iterações efetuado.

4 Conclusões

Neste trabalho apresentamos, no Capítulo 1, o Método da Máxima Descida, que é um método iterativo para minimizar irrestritamente funções suaves, no qual devemos tomar a direção oposta ao gradiente. Este possivelmente deve ser um dos primeiros métodos estudados para a minimização de funções, porém conforme visto, este método não é eficiente.

Já no Capítulo **??**, apresentamos o Método de Gradientes Conjugados Lineares que faz uso das direções conjugadas e que por esse fato, em no máximo n iterações encontra a solução do problema posto em (1.1). Vimos também que o método, embora trabalhe unidirecionalmente, minimiza a função quadrática estritamente convexa em variedades afins definidas pelo ponto inicial e pelas sucessivas direções conjugadas consideradas, culminando na minimização em todo o espaço $\mathbb{R}^n$.

Percebemos também que há várias variantes para a escolha do escalar $\beta_{(i)}$, que fazem com que as direções de busca sejam conjugadas e no caso em que f é uma função quadrática estritamente convexa todos esses escalares coincidem, ou seja, o método converge com o mesmo número de iterações dentro da mesma precisão, veja Teorema 2.6. Porém, para o caso em que f é uma função não linear geral, de classe C^1, tais escalares produzem resultados distintos. Com base nos testes efetuados, não chegamos a uma conclusão sobre qual escalar $\beta_{(i)}$ é mais eficiente, ou seja, a partir de um ponto inicial, considerando-se todos os parâmetros fixos, um determinado $\beta_{(i)}$ teve desempenho melhor, já para outro ponto inicial, outro $\beta_{(i)}$ foi melhor, o que pode ser comprovado pelo Exemplo 3.5. No Exemplo 3.6, em que consideramos problemas com dimensões maiores, verificamos que para valores adequados para os parâmetros que definem a busca linear, a qualidade da solução obtida não é influenciada pela escolha dos escalares $\beta_{(i)}$. Já a eficiência do método, medida em termos do número de iterações efetuadas, esta sim é afetada pela fórmula utilizada para $\beta_{(i)}$, embora os resultados não apontem por uma escolha particularmente mais eficiente. Vale dizer que fórmula alternativas para tais escalares estão presentes na literatura atual, como é o caso do artigo (DAI; KOU, 2013).

Em se tratando dos problemas não lineares, um aspecto para o qual só nos atentamos na ocasião da defesa, graças ao olhar diferenciado dos membros da banca, foi o fato de termos trabalhado sem garantia de que as direções geradas pelo Método de Gradientes Conjugados não Lineares fossem efetivamente de descida. Por um lado, a nossa escolha bastante tolerante para a constante c_2 da condição de Wolfe forte proporcionou, em geral, maior fôlego para que as buscas lineares tivessem sucesso, conforme ilustrado na Figura 5. Por outro lado, perdemos a garantia teórica de descenso, sob as condições

do Lema 3.1, pelo menos para a escolha dos escalares $\beta_{(i)}$ de Fletcher e Reeves. Para as demais escolhas de $\beta_{(i)}$ não há, a priori, nenhuma garantia de que o método gerará direções de descida, e, de fato, não chegamos a verificar tal propriedade em nossa implementação. Nesse sentido, os recomeços foram essenciais, embora muitas avaliações de função possam ter sido desperdiçadas, ao longo de direções de subida potencialmente geradas. Detectamos, portanto, um aspecto que merece um estudo mais aprofundado, o qual poderia incluir o monitoramento do comportamento das direções obtidas e estratégias distintas para se efetuar os recomeços.

O presente trabalho teve como intuito abordar alguns métodos iterativos para minimizar irrestritamente funções suaves, como o Método da Máxima Descida, o Método de Gradientes Conjudados Lineares e o Método de Gradientes Conjudados não Lineares. Em especial neste último, trabalhamos com um apelo geométrico, sobretudo em relação ao efeito dos escalares $\beta_{(i)}$, responsáveis pela manutenção das conjugações das direções e da constante γ, responsável pela redução do tamanho do passo. Tomando o nosso texto como ponto de partida, sugerimos os livros de (NOCEDAL; WRIGHT, 2006) e (MARTÍNEZ; SANTOS, 1995), para uma leitura mais detalhada e aprofundada sobre os tópicos abordados.

Referências

ANTON, H.; BUSBY, R. C. Álgebra linear contemporânea. Bookman Editora. Porto Alegre, 2006.

BOLDRINI, J. L.; COSTA, S. I.; FIGUEREDO, V.; WETZLER, H. G. Álgebra linear. Harper & Row; São Paulo, 1980.

DAI, Y.-H.; KOU, C.-X. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. *SIAM Journal on Optimization*, SIAM, v. 23, n. 1, p. 296–320, 2013.

DAI, Y.-H.; YUAN, Y. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on optimization*, SIAM, v. 10, n. 1, p. 177–182, 1999.

DENNIS JR, J.; SCHNABEL, R. *Numerical methods for unconstrained optimization and nonlinear equations*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1996.

FLETCHER, R.; REEVES, C. M. Function minimization by conjugate gradients. *The computer journal*, Br Computer Soc, v. 7, n. 2, p. 149–154, 1964.

GUIDORIZZI, H. L. Um curso de cálculo, vol. 2. Grupo Gen-LT, Rio de Janeiro, 2000.

HESTENES, M. R.; STIEFEL, E. Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Standards, v. 49, p. 409–436, 1952.

KÜHLKAMP, N. *Matrizes e sistemas de equações lineares*. Florianópolis: Editora da UFSC, 2005.

LIMA, E. L. Curso de análise, vol. 2. SBM, Rio de Janeiro, 1976.

MARTÍNEZ, J. M.; SANTOS, S. A. *Métodos computacionais de otimização, 20o. Colóquio Brasileiro de Matemática*. Rio de Janeiro: SBM, 1995.

MORÉ, J. J.; GARBOW, B. S.; HILLSTROM, K. E. Testing unconstrained optimization software. *ACM Transactions on Mathematical Software (TOMS)*, ACM, v. 7, n. 1, p. 17–41, 1981.

NOCEDAL, J.; WRIGHT, S. J. Numerical optimization 2ed. Springer. New York, 2006.

POLAK, E.; RIBIERE, G. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge*, v. 3, n. 1, p. 35–43, 1969.

POLYAK, B. T. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, Elsevier, v. 9, n. 4, p. 94–112, 1969.

RIBEIRO, A. A.; KARAS, E. W. *Otimização Contínua: aspectos teóricos e computacionais*. São Paulo: Cengage Learning, 2014.

SHEWCHUK, J. R. *An introduction to the conjugate gradient method without the agonizing pain*. Pittsburgh (PA): Carnegie-Mellon University. Department of Computer Science, 1994.

WATKINS, D. S. *Fundamentals of Matrix Computations*. 2nd. ed. New Jersey, NY, USA: John Wiley & Sons, 2002.

Apêndices

APÊNDICE A – Algoritmos

A.1 Algoritmo Geral

Algoritmo 1: Algoritmo Geral

Dados: $x_{(0)} \in \mathbb{R}^n$ (ponto inicial).
Tome $i = 0$
  Enquanto $\nabla f(x_{(i)}) \neq 0$
    Calcule $d_{(i)}$ tal que $\nabla f(x_{(i)})^T d_{(i)} < 0$
    Escolha $\alpha_{(i)} > 0$ tal que $f(x_{(i)} + \alpha_{(i)} d_{(i)}) < f(x_{(i)})$
    Faça $x_{(i+1)} = x_{(i)} + \alpha_{(i)} d_{(i)}$
    $i = i + 1$

A.2 Busca inexata de Armijo juntamente com Wolfe forte

Algoritmo 2: Busca inexata de Armijo juntamente com Wolfe forte

Dados: $x_{(i)} \in \mathbb{R}^n$ (ponto corrente), $t_{min} > 0$, $d \in \mathbb{R}^n$ (direção de descida), $\gamma \in (0, 1)$, $c_1 \in (0, 1/2)$, e $c_2 \in (c_1, 1)$.
Tome $t = 1$
  Enquanto:
  $f(x_{(i)} + td) > f(x_{(i)}) + c_1 t \nabla f(x_{(i)})^T d$ ou
  $|\nabla f(x_{(i)} + td)^T d| > c_2 |\nabla f(x_{(i)})^T d|$ e $t > t_{min}$
    $t = \gamma t$
  se $t \leqslant t_{min}$ então a busca falhou.

A.3 Construção de uma matriz simétrica definida positiva usando o Matlab

A.4 Gradientes conjugados não lineares com escalar de Fletcher Reeves

Algoritmo 3: Construção de uma matriz simétrica definida positiva usando o Matlab

Dados: n (ordem da matriz quadrada), $(lmin, lmax)$ (intervalo para o menor e o maior autovalor) e $v \in \mathbb{R}^n$ (vetor de valores aleatórios entre $(0, 1)$).

Faça: $y = lmin + v * (lmax - lmin)$

$y(1) = lmin$

$y(n) = lmax$

$y = sort(y)$ (ordena os valores entre $(lmin, lmax)$)

$M = rand(n, n)$

$[Q, R] = qr(M)$

$A = Q * diag(y) * Q'$

Algoritmo 4: Gradientes conjugados não lineares com escalar de Fletcher Reeves

Dados: $x_{(0)} \in \mathbb{R}^n$ (ponto inicial), Tol, $t_{\min}$, $c_1 = 0.01, c_2 = 0.99, \gamma \in (0, 1)$ e maxiter (numero máximo de iterações).

Calcule $r_{(0)} = -\nabla f(x_{(0)})$ e faça $i = 0$. Tome: $d_{(0)} = r_{(0)}$

Enquanto $\|r_{(i)}\| >$ Tol e $i <$ maxiter

Faça uso do algoritmo A.2 para computar o tamanho do passo t

Defina $x_{(i+1)} = x_{(i)} + td_{(i)}$

Calcule $r_{(i+1)} = -\nabla f(x_{(i+1)})$

Tome: $d_{(i+1)} = r_{(i+1)} + \beta_{(i)} d_{(i)}$,

com $\beta_{(i)} = \dfrac{r_{(i+1)}^T r_{(i+1)}}{r_{(i)}^T r_{(i)}}$ se a busca do algoritmo A.2 foi bem sucedida ou

$\beta_{(i)} = 0$ se a busca falhou

Faça $i = i + 1$