Mathematica - Revista eletrônica de divulgação matemática.
ISSN: 3086-6782. **V. 1** (2025) - pp. 23 - 38

Métodos do Gradiente e de Newton em Otimização Quadrática Convexa

Alex Augusto Nunes Machado
Simone Aparecida Miloca - Unioeste

(*Recebido em 29/11/2021. Aceito em 03/12/2021. Publicado em 30/11/2025*)

Resumo: Este trabalho tem por objetivos apresentar uma breve contextualização e formalização de conceitos envolvendo problemas pertencentes a uma área da Matemática denominada Otimização. O interesse está em uma classe particular nomeada Programação Quadrática Irrestrita o qual aparece em diversas situações práticas tanto na fase de modelagem como em processos de busca de soluções em outros métodos de otimização. Para tanto, descrevemos resultados teóricos envolvendo condições de otimalidade para problemas irrestritos bem como apresentamos dois algoritmos de busca direcionais, os métodos do Gradiente e de Newton, para então justificar matematicamente os motivos pelos quais o método de Newton resolve um problema dessa classe em apenas uma iteração, bem como demonstrar a fórmula utilizada para determinar o tamanho do passo no Método do Gradiente.

Palavras-chave: Programação Quadrática; Método de Newton; Método do Gradiente.

1 Introdução

Conforme apontam Ribeiro e Karas (2013), direta ou indiretamente a noção de otimização faz parte do dia-a-dia do ser humano moderno, dado que as pessoas responsáveis pela tomada de decisão (em qualquer campo do conhecimento) deparam-se com problemas nos quais necessitam obter resultados não somente precisos mas também eficientes.

Nesse sentido, conforme pontuam Hillier e Lieberman (2013, p. 1), desde o desenvolvimento da sociedade industrializada, e sobretudo com a eclosão da segunda guerra mundial na qual os militares britânicos e norte-americanos visando alocar os recursos da melhor maneira possível "(...) convocaram grande número de cientistas para aplicar uma abordagem científica para lidar com este e outros problemas táticos e estratégicos.", a demanda por ferramentas que auxiliem nesse processo decisório por parte de organizações, públicas ou privadas, nos mais diversos campos, sejam no âmbito científico ou empresarial, cresceu substancialmente. É neste contexto que a Otimização Matemática se aplica.

Isto posto, o que vem a ser esse campo de estudo denominado *Otimização Matemática*? Luenberger e Ye (2016) em sua introdução escrevem que,

> "O conceito de otimização agora está bem enraizado como um princípio subjacente à análise de muitos problemas complexos de decisão ou alocação. [...] aborda-se um problema de decisão complexo, envolvendo a seleção de valores para uma série de variáveis inter-relacionadas, focando a atenção em um único objetivo projetado para quantificar o desempenho e medir a qualidade da decisão. Este objetivo é maximizado (ou minimizado, dependendo da formulação) sujeito às restrições que podem limitar a seleção dos valores das variáveis de decisão." (LUENBERGER; YE, 2016, p. 1, tradução do autor)

Já, Ribeiro e Karas (2013) apontam que a Otimização consiste em encontrar pontos de mínimo ou máximo de uma função $f\colon A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ que está sujeita a determinadas restrições (outras funções). Tal função f é denominada *função objetivo* e emerge, quando pessoas responsáveis pela tomada de decisão deparam-se com problemas nos quais necessitam obter resultados precisos e, além disso,

> "(...) um único aspecto adequado de um problema pode ser isolado e caracterizado por um objetivo, seja lucro ou perda em um ambiente de negócios, velocidade ou distância em um problema físico, retorno esperado no ambiente de investimentos de risco (...)" (LUENBERGER; YE, 2016, p. 1, tradução do autor).

Com isso, por meio de um processo de modelagem é possível traduzir a situação que demanda ser resolvida para o "mundo matemático". Essa tradução, de modo geral, é representada por

$$\begin{array}{ll} \text{Otimizar} & f(x) \\ \text{sujeito a} & x \in \Omega \subset \mathbb{R}^n \end{array}. \tag{1}$$

no qual Ω representa um conjunto de restrições que podem ser de igualdade ou desigualdade. Portanto, (1) formaliza em termos matemáticos o que vem a ser um problema de Otimização Matemática.

Assim, percebe-se então que o modelo consiste em uma espécie de representação do "mundo real" no "mundo matemático". Esse processo de tradução como indicam Goldbarg e Luna (2005, p. 9) envolve "a percepção do elaborador do modelo (ou equipe de elaboração), uma faculdade cognitiva de alto nível. As *fórmulas* ou *equações* do modelo não existem prontas e acabadas na natureza, elas têm que ser identificadas ou criadas.". Todavia, embora tal processo vise a extração da essência do problema, conforme salientam Hillier e Lieberman (2013), vale destacar que "a correspondência entre experiência e modelo formal está longe de ser perfeita: a tradução está sujeita a erros, simplificações e falhas de comunicação." (MARTÍNEZ; SANTOS, 1995, p. 1).

Além disso, deve-se observar que o conjunto $\Omega \subset \mathbb{R}^n$ (e a função objetivo) pode assumir diversos formatos, uma vez que sua obtenção está diretamente ligada ao processo de formulação do modelo matemático. Nesse sentido conforme apontam Ribeiro e Karas (2013) e Goldbarg e Luna (2005) é possível agrupar os problemas de otimização em algumas categorias, cada qual possuindo técnicas específicas para sua resolução.

Algumas das categorias possíveis são: programação convexa, otimização linearmente restrita, programação separável, programação não convexa, programação geométrica, pro-

gramação fracionária e programação quadrática, sendo que esta última será o tema deste trabalho.

Mais especificamente, serão apresentados alguns conceitos e resultados básicos do campo de Otimização Matemática, bem como comentado brevemente a respeito dos algoritmos direcionais, em particular os Método do Gradiente e de Newton. Na sequência, as seções 4, 5 e 6 serão responsáveis por formalizar o problema de Programação Quadrática Irrestrita e provar matematicamente os motivos pelos quais o método de Newton necessita de apenas uma iteração para resolver problemas dessa classe além de demonstrar a fórmula utilizada para estabelecer o tamanho do passo no Método do Gradiente.

2 Condições de otimalidade

Visando estabelecer os critérios para verificar se algum determinado ponto $x \in \Omega$ é o que fornece o melhor valor para uma função f dada, isto é, formalizar matematicamente o que foi discutido na Introdução, esta seção apresenta algumas definições e resultados necessários para o avanço à Programação Quadrática Irrestrita.

Definição 1. Sejam $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ e $x^* \in \Omega \subset \mathbb{R}^n$. Definimos x^* como sendo um ***minimizador local de f em*** $\boldsymbol{\Omega}$ se, e só se, existe $\delta > 0$, tal que

$$f(x^*) \leq f(x),$$

para todo $x \in B(x^*, \delta) \cap \Omega$, com $B(x^*, \delta)$ sendo uma bola de centro x^* e raio δ.

Definição 2. Sejam $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ e $x^* \in \Omega \subset \mathbb{R}^n$. Definimos x^* como sendo um ***minimizador global de f em*** $\boldsymbol{\Omega}$ se, e só se,

$$f(x^*) \leq f(x),$$

para todo $x \in \Omega$.

Isto posto, uma pergunta natural é: "toda função $f\colon A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ possui um valor x^* que a otimiza?" A Proposição 3, conhecida como "*Teorema de Weierstrass*", apresenta a resposta à esse questionamento, e sua demonstração pode ser encontrada em Ribeiro e Karas (2013, p. 36).

Proposição 3. *Se $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ é uma função contínua e Ω é um conjunto compacto não vazio, então existe minimizador global de f em Ω.*

Ou seja, sob certas circunstâncias, é possível garantir a existência de um minimizador global. Em outras palavras, isto significa que se dado um problema, seja ele de *qualquer natureza*, puder ser modelado matematicamente de tal modo que seu domínio seja um subconjunto compacto (limitado e fechado) de $\mathbb{R}^n$, então pode-se garantir a existência de um valor que minimiza f, independente da natureza do problema que o gerou.

Ademais, vale destacar que embora o Teorema de Weierstrass faça referência a minimização de uma função, tal proposição também pode ser aplicada no contexto de maximização de f, dado que maximizar f é o mesmo que minimizar $-f$. De fato, suponha que x^* seja maximizador global de $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, assim, por definição temos que para todo $x \in \Omega$,

$$f(x^*) \geq f(x)$$

logo, para todo $x \in \Omega$,

$$-f(x^*) \leq -f(x).$$

Portanto, segue por definição que x^* é minimizador global de $-f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. A Figura 1 ilustra essas ideias.

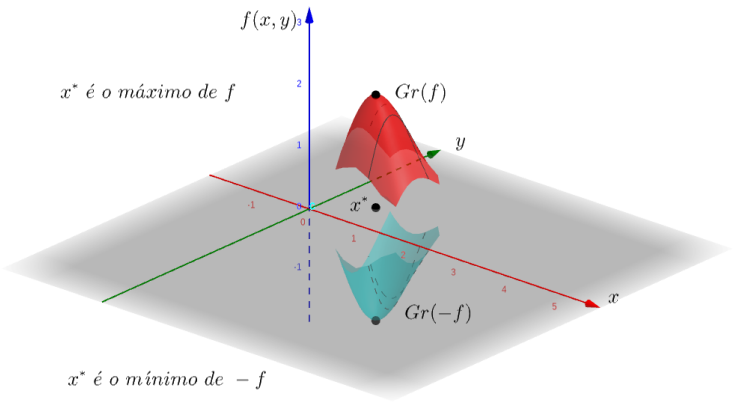


Figura 1: Minimizar f é equivalente a maximizar $-f$

Diante disso, a proposição 4 estabelece os critérios que permitem afirmar se algum x pertencente ao domínio de f pode ser um *candidato* a minimizador (ou maximizador) local.

Proposição 4. *Seja $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ diferenciável em $x^* \in \mathbb{R}^n$. Se x^* é minimizador local de f, então*

$$\nabla f(x^*) = 0.$$

Prova. Suponha $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ diferenciável em $x^* \in \Omega$ e x^* minimizador local de f em Ω. Além disso, suponha $d \in \mathbb{R}^n \backslash \{0\}$.

Com isso, uma vez que x^* é minimizador local, segue por definição, que existe $\delta > 0$ tal que para todo $x \in B(x^*, \delta) \cap \Omega$,

$$f(x^*) \leq f(x),$$

e portanto, em particular, para $x = x^* + td$ com $t \in (0, \delta)$, isto é, para todo $t \in (0, \delta)$,

$$f(x^*) \leq f(x^* + td).$$

Assim,

$$0 \leq f(x^* + td) - f(x^*).$$

Além disso, fazendo uso da fórmula de Taylor de ordem 1 , juntamente com a mudança de variável $td = x - x^*$, temos que

$$f(x^* + td) = f(x^*) + t\nabla f(x^*)^T d + r(t),$$

para $\lim\limits_{t \longrightarrow 0} \frac{r(t)}{t} = 0$.

Logo,

$$\begin{aligned} 0 \leq f(x^* + td) - f(x^*) &\Longrightarrow 0 \leq f(x^*) + t\nabla f(x^*)^T d + r(t) - f(x^*) \\ &\Longrightarrow 0 \leq t\nabla f(x^*)^T d + r(t) \Longrightarrow \frac{0}{t} \leq \frac{t\nabla f(x^*)^T d}{t} + \frac{r(t)}{t} \\ &\Longrightarrow 0 \leq \nabla f(x^*)^T d + \frac{r(t)}{t} \Longrightarrow \lim_{t \longrightarrow 0} 0 \leq \lim_{t \longrightarrow 0} \left(\nabla f(x^*)^T d + \frac{r(t)}{t} \right) \\ &\Longrightarrow 0 \leq \nabla f(x^*)^T d + \lim_{t \longrightarrow 0} \frac{r(t)}{t}, \end{aligned}$$

e como $\lim\limits_{t \longrightarrow 0} \frac{r(t)}{t} = 0$, temos que

$$\nabla f(x^*)^T d \geq 0. \tag{D.6}$$

Diante disso, suponha por absurdo que $\nabla f(x^*) \neq 0$, logo tomando $d = -\nabla f(x^*)$ em (D.6), temos que

$$\nabla f(x^*)^T(-\nabla f(x^*)) \geq 0 \Longrightarrow -||\nabla f(x^*)||^2 \geq 0 \Longrightarrow ||\nabla f(x^*)||^2 \leq 0,$$

o que é absurdo, portanto, não se pode supor $\nabla f(x^*) \neq 0$, logo $\nabla f(x^*) = 0$.

□

Posta em outras palavras, essa proposição informa que dado *qualquer* x no domínio de f se $\nabla f(x)$ não for nulo, então x *não* pode ser sequer um candidato a minimizador local. Sendo assim, elementos de $\mathbb{R}^n$ que satisfazem essa proposição (também conhecida como *condição necessária de 1ª ordem*) recebem um nome especial, o qual será formalizado com a definição seguinte.

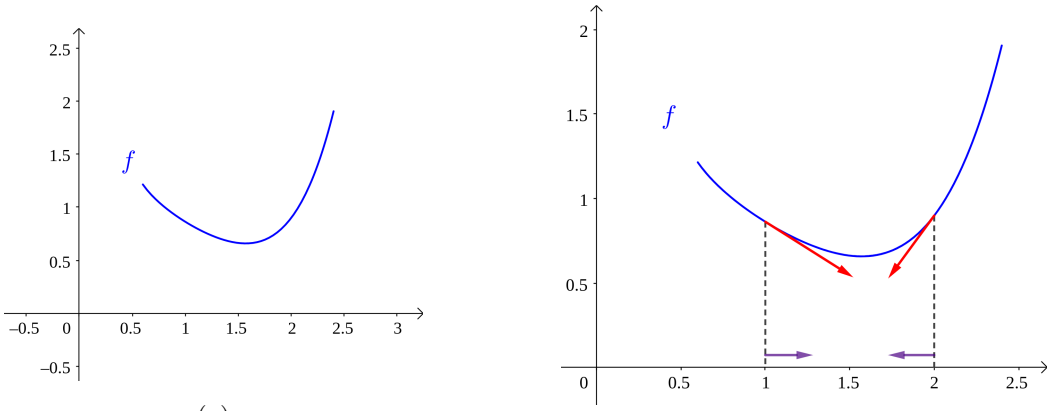
Definição 5. Sejam $f\colon \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ diferenciável em Ω e $x^* \in \Omega$. Definimos x^* como sendo ***ponto crítico*** ou (*estacionário*) de f se, e só se, $\nabla f(x^*) = 0$.

Isto posto, convém enfatizar que as funções (objetivo e/ou restrições) envolvidas podem não ser tão elementares, ou mesmo que sejam, podem demandar muitas variáveis, e assim, a solução direta pode não ser uma boa opção.

Diante disso, entram em cena os *métodos de busca* que funcionam como um meio de, a partir de um palpite inicial $x_0 \in \mathbb{R}^n$ para a solução do problema, encontrar outros pontos, mais precisamente uma sequência de pontos de $\mathbb{R}^n$ que, idealmente, conforme destacam Luenberger e Ye (2016) e Ribeiro e Karas (2013), geram uma sequência de pontos de $\mathbb{R}^n$ que se aproximam da solução ótima. Diante disso, a próxima seção será destinada a apresentar alguns desses métodos de busca.

3 Algoritmos com buscas direcionais

Imaginemos a seguinte situação, encontrar o mínimo de uma dada função f que possui o seu gráfico ilustrado na Figura 2a.



(a) (b)

Figura 2: Minimização por bissecção

Intuitivamente podemos inferir que se escolhermos $x^* = 1$ então esse valor fará $\frac{d}{dx}f(1) < 0$ pois, visualmente, a função ainda pode assumir valores menores que $f(1)$ caso tomemos valores à direita de 1 para x^*. Assim, $f(1+\varepsilon) < f(1)$, fazendo com que $f(1+\varepsilon) - f(1) < 0$. Logo, $x^* = 1$ não poderá ser ponto crítico de f.

Porém, se tomarmos agora $x^* = 2$ sua derivada será positiva, pois novamente a intuição visual informa que f ainda pode assumir valores menores que $f(2)$ na condição de escolhermos algum valor à esquerda de 2. Deste modo, agora $f(2+\varepsilon) > f(2)$, e assim $f(2+\varepsilon) - f(2) > 0$, fazendo que $\frac{d}{dx}f(2) > 0$.

Diante disso, é razoável supor que embora ainda não saibamos qual é o valor que minimiza f, o intervalo $(1, 2)$ certamente conterá um valor que minimiza f, ao menos localmente. A Figura 2b ilustra essas ideias.

Ou seja, $\frac{d}{dx}f(x_0) < 0$ implica que o valor que minimiza f encontra-se à direita de x_0, em contrapartida, se $\frac{d}{dx}f(x_0) > 0$ então x^* deverá estar à esquerda de x_0.[1] Isto posto, uma estratégia para explorar essa característica consiste em

1° Tomar um intervalo $[a, b]$ de modo que $\frac{d}{dx}f(a) \cdot \frac{d}{dx}f(b) < 0$;

2° Calcular x_0 como sendo $\dfrac{b+a}{2}$;

3° Tomar o novo intervalo (a, b) de tal modo que ou $a = x_0$ se $\frac{d}{dx}f(x_0) < 0$ ou $b = x_0$ se $\frac{d}{dx}f(x_0) > 0$;

[1] Porém, deve-se observar que isso é sempre verdade para o caso de a função a ser minimizada ser convexa. Para mais informações, consultar Hillier e Lieberman (2013, p. 953).

4° Calcular x_1 como sendo $\frac{b+a}{2}$, caso x_1 seja seja a solução ótima ou seja suficientemente bom pare, se não execute uma nova iteração.

Portanto, percebamos que tal estratégia consiste em ir reduzido o intervalo de busca pela metade até encontrar a solução desejada, ou ao menos obter uma solução com a precisão tão boa quanto se deseja. Essa abordagem é conhecida como método da bissecção e um exemplo de implementação é ilustrado pelo Algoritmo 1.

Tomar um intervalo $[a, b]$;
Escolher o palpite inicial como sendo $x_0 = \frac{b+a}{2}$;
$k \leftarrow 0$;
enquanto $\nabla f(x_k) \neq 0$ **faça**
  se $\frac{d}{dx} f(x_k) < 0$ **então**
    $a \leftarrow x_k$;
  senão
    $b \leftarrow x_k$;
  fim
  $x_k = \frac{b+a}{2}$;
fim

Algoritmo 1: Exemplo genérico de implementação do método da bissecção

Notemos que o método da bissecção de fato faz uso da ideia de, a partir de um valor inicial, obter outros de modo que $f(x_i) \leq f(x_{i-1})$. Todavia, como estender essas ideias ao $\mathbb{R}^n$? Visando responder a esse questionamento temos a seguinte definição.

Definição 6. Sejam $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, $\overline{x} \in \mathbb{R}^n$ e $d \in \mathbb{R}^n \backslash \{0\}$ uma direção. Definimos d como sendo uma ***direção de descida para f a partir de*** $\overline{\boldsymbol{x}}$ se, e só se, existe $\delta > 0$ tal que $f(\overline{x}+td) < f(\overline{x})$ para todo $t \in (0, \delta)$.

Isto posto, embora a Definição 6 estabeleça como identificar *se dada direção d* é ou não de descida, a proposição seguinte vem para ajudar nesse processo de verificação.

Proposição 7. *Seja* $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$. *Se* $\nabla f(\overline{x})^T d < 0$, *então d é uma direção de descida para f a partir de* $\overline{x}$.

Prova. Suponha $\nabla f(\overline{x})^T d < 0$, assim da definição de derivada direcional temos que

$$\frac{\partial}{\partial d} f(\overline{x}) < 0,$$

e como existe a derivada direcional na direção d, então f restrita à direção d é contínua neste ponto $\overline{x}$, e portanto, para $t \neq 0$,

$$\lim_{t \to 0} \frac{f(\overline{x} + td) - f(\overline{x})}{t} < 0.$$

Logo, da definição de limite, existe $\delta > 0$ tal que

$$\frac{f(\overline{x} + td) - f(\overline{x})}{t} < 0,$$

para todo $t \in (-\delta, \delta)$. Assim, $f(\overline{x} + td) < f(\overline{x})$ para o caso particular em que $t \in (0, \delta)$.

□

Pondo em outras palavras a Proposição 7 está informando que caso a derivada na direção de d seja negativa, então existirão nessa mesma direção vetores de $\mathbb{R}^n$ de modo que o valor de f aplicado em tais vetores sejam menores que $f(\overline{x})$.

Sendo assim, de posse dessas duas ferramentas o Algoritmo 2 estabelece uma estratégia genérica para minimizar funções de $\mathbb{R}^n$.

Escolher x_0;
$k \leftarrow 0$;
enquanto $\nabla f(x_k) \neq 0$ **faça**
  Obter d_k tal que $\nabla f(x_k)^T d_k < 0$;
  Obter $t_k > 0$ tal que $f(x_k + t_k d_k) < f(x_k)$;
  $x_{k+1} \leftarrow x_k + t_k d_k$;
  $k \leftarrow k + 1$;
fim

Algoritmo 2: Algoritmo de descida

Colocando em outros termos, o que o Algoritmo 2 faz em sua essência é, partindo de um ponto inicial, encontrar uma direção na qual a imagem de f diminui, isto é, obter uma direção de descida, para então encontrar o novo ponto inicial e recomeçar o processo levando em consideração essa atualização.

Na prática, a condição $\nabla f(x_k) = 0$ é substituída por algum critério de parada[2], seja pelo fato de que encontrar x_k que atenda a especificação pode requerer um alto custo computacional, ou simplesmente pela solução desejada não necessitar ser *a melhor*, mas sim uma *suficientemente boa*.

Além disso, conforme colocam Ribeiro e Karas (2013), as escolhas de d_k e t_k não podem ser arbitrárias pois, se assim o fossem, poderiam gerar uma sequência de pontos que, embora possuam pontos de acumulação estes podem não ser estacionários, isto é, podem não ser bons candidatos a minimizadores.

Diante disso, uma pergunta natural que emerge é: "*Como escolher d_k e t_k de modo conveniente ao propósito de minimizar f mais rapidamente?*", a qual pode ser parafraseada em: "*Como escolher as direções e as amplitudes de cada passo em direção ao mínimo de modo eficiente?*".

[2]Por exemplo, substituir $\nabla f(x_k) = 0$ por $\nabla f(x_k) < \varepsilon$ em que ε é escolhido de modo que x_k é suficientemente próximo da melhor solução possível.

Pois bem, no que se refere a escolha da direção em problemas de minimização, é natural a escolha de $-\nabla f(x_k)$ pois, da teoria do cálculo em $\mathbb{R}^n$, sabe-se que o vetor gradiente aponta no sentido de maior crescimento de f. Sendo assim, visando ilustrar essas ideias, consideremos a seguinte situação:

$$\text{minimizar} \quad f(x) = -2x_1x_2 - 2x_2 + x_1^2 + 2x_2^2.$$

Para tanto, deve-se primeiramente estabelecer um critério de parada, nesse caso $||\nabla f(x_k)|| < \varepsilon = 0,1$ e também um palpite inicial para a solução, nesse caso $x_0 = (0,0)$. Assim, dado que

$$\frac{d}{dx_1}f = -2x_2 + 2x_1 \text{ e } \frac{d}{dx_2}f = -2x_1 + 4x_2 - 2,$$

temos que

$$||\nabla f(x_0)|| = ||\nabla f(0,0)|| = ||(-0+0, -0+0-2)|| = ||(0,-2)|| = \sqrt{0^2 + (-2)^2} = 2 > \varepsilon.$$

Portanto, iniciamos a ***primeira iteração*** com

$$-\nabla f(x_0) = -\nabla f(0,0) = -(0,-2) = (0,2),$$

e assim, da definição de direção de descida (Definição 6), temos que

$$\begin{aligned} f(x_0 + t_0(-\nabla f(x_0))) &< f(x_0) \\ &\iff f((0,0) + t_0(0,2)) < f(0,0) \\ &\iff f(0,2t_0) = -2\cdot 0\cdot(2t_0) - 2\cdot(2t_0) + 0^2 + 2(2t_0)^2 \\ &\qquad\qquad < -2\cdot 0\cdot 0 - 2\cdot 0 + 0^2 + 2\cdot 0^2 \\ &\iff -4t_0 + 8t_0^2 < 0. \end{aligned}$$

Ou seja, qualquer t_0 tomado de modo que $-4t_0 + 8t_0^2 < 0$ já seria suficiente para "descermos" no sentido do minimizador. Todavia, é razoável optar pelo t_0 de tal modo que essa descida seja a mais rápida possível, sendo assim recaímos em outro problema de minimização, desta vez um de apenas uma variável.

Neste caso, fazendo uso da Proposição 4 para o caso particular de funções de uma variável, temos que

$$\frac{d}{dt_0}(-4t_0 + 8t_0^2) = 0 \iff -4 + 16t_0 = 0 \iff t_0 = \frac{1}{4}.$$

Portanto, o tamanho do passo no sentido de $-\nabla f(x_0)$ será $t_0 = \frac{1}{4}$, isto é,

$$x_1 = (0,0) + \frac{1}{4}(0,2) = \left(0, \frac{1}{2}\right).$$

Logo, encerra-se a primeira iteração e verifica-se novamente o critério de parada.

$$\begin{aligned} ||\nabla f(x_1)|| &= \left|\left|\nabla f\left(0,\frac{1}{2}\right)\right|\right| = \left|\left|\left(-2\cdot\frac{1}{2} + 2\cdot 0, -2\cdot 0 + 4\cdot\frac{1}{2} - 2\right)\right|\right| = \left|\left|\left(-1, -\frac{3}{2}\right)\right|\right| \\ &= \sqrt{(-1)^2 + \left(-\frac{3}{2}\right)^2} = \sqrt{\frac{4+9}{4}} = \frac{\sqrt{13}}{2} \simeq 1,803 > \varepsilon. \end{aligned}$$

Com isso, devido ao fato de que x_1 não satisfaz ao critério de parada, repete-se o processo aplicado na iteração 1 de modo a obter x_k que atenda ao critério de parada estabelecido.

Em resumo, esse processo fornece uma ideia de como fazer uso dessa estratégia de minimização envolvendo o vetor gradiente. Devido à sua importância essa estratégia recebe o nome de ***Método do Gradiente***, também conhecida como ***Método de Cauchy***. O Algoritmo 3 resume essa abordagem.

Escolher o critério de parada $\varepsilon > 0$;
Escolher x_0;
$k \leftarrow 0$;
enquanto $||\nabla f(x_k)|| > \varepsilon$ **faça**
  Definir $d_k = -\nabla f(x_k)$;
  Obter $t_k > 0$ tal que $f(x_k + t_k d_k) < f(x_k)$;
  $x_{k+1} \leftarrow x_k + t_k d_k$;
  $k \leftarrow k + 1$;
fim

Algoritmo 3: Método do Gradiente

Conforme destacam Ribeiro e Karas (2013, p. 91), este método converge globalmente, e sua demonstração pode ser encontrada na página 91 deste mesmo livro.

Isto posto, outra característica importante que o exemplo anterior ilustra é o fato de que a determinação do tamanho de passo k em sua respectiva iteração não é algo tão elementar, pois isso requer a resolução de outro problema de minimização, em particular minimizar f na direção d_k. E, conforme aponta Friedlander (1994, p.33) esse processo recebe o nome de ***busca linear exata***.

Diante disso, dado que obter t_k tamanhos de passo exige a resolução de k problemas de minimização (cada qual associada a sua respectiva direção d_k), quando deseja-se utilizar a abordagem da busca linear exata para a obtenção do tamanho do passo na iteração k, a implementação do Método do Gradiente requer o uso de um segundo algoritmo de minimização. Porém, diferentemente do problema inicial, este "subproblema" consiste na minimização de funções de apenas uma variável, e assim podemos fazer uso de algoritmos já bem estabelecidos como o método da bissecção e o método da seção áurea.

Todavia, nem sempre a aplicação da busca linear exata para cada iteração é conveniente, seja por tal técnica demandar um alto custo computacional, ou mesmo pelo fato de que a minimização do subproblema associado a iteração k simplesmente não é possível. Sendo assim, entra em cena a ***busca inexata*** que, ao invés de obter o *melhor* tamanho de passo na direção d_k, visa encontrar um t_k *suficientemente* bom, isto é, não almeja minimizar o subproblema associado a iteração k, mas sim "[...] busca uma boa redução da função ao longo da direção, sem tentar minimizá-la." Ribeiro e Karas (2013, p.77). Algumas das estratégias utilizadas na busca inexata são: Regra de Armijo, Regra de Goldstein, Regra de Wolfe e Regra de Wolfe

forte[3].

Todavia, tendo em vista o que o foco deste trabalho, além de ser de caráter introdutório, abarca somente problemas de programação quadrática irrestrita, a estratégia escolhida para a determinação do tamanho de cada passo será a busca exata, pois como será discutido na seção seguinte, encontrar o tamanho do passo t_k nessa classe de problemas é sempre possível.

Voltemos ao problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{sujeito a} & x \in \mathbb{R}^n \end{array}. \tag{2}$$

Como discutido na Seção 2, para resolvermos (2) devemos encontrar algum x^* de tal modo que $\nabla f(x^*) = 0$. Diante disso, visando alcançar esse objetivo, devemos então escolher não só uma direção de descida (d_k) mas também o tamanho do passo que será dado nessa direção (t_k). Uma das estratégias para isso consiste no Método do Gradiente, que apoia-se na ideia de "seguir" a direção de $-\nabla f(x_k)$ em cada iteração.

Todavia, uma outra abordagem consiste em tomar x_{k+1} como sendo o ponto que minimiza a aproximação de Taylor de segunda ordem[4] na iteração k. Essa estratégia é conhecida como ***Método de Newton***[5] e é implementada pelo Algoritmo 4.

Escolher o critério de parada $\varepsilon > 0$;
Escolher x_0;
$k \leftarrow 0$;
enquanto $||\nabla f(x_k)|| > \varepsilon$ **faça**
  Definir $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$;
  $x_{k+1} \leftarrow x_k + d_k$;
  $k \leftarrow k + 1$;
fim

Algoritmo 4: Método de Newton

Salientamos aqui que o cálculo da direção d_k do Algoritmo 4 é feito geralmente resolvendo um sistema de equações lineares que tem um custo computacional menor que o necessário para a inversão de matrizes (RIBEIRO; KARAS, 2013).

4 Programação quadrática irrestrita

De acordo com Hillier e Lieberman (2013), na classe de problemas de programação quadrática as restrições são lineares, porém sua função objetivo possui lei de formação contendo

[3]Para mais detalhes, ver Shi e Shen (2005).

[4]Vale destacar que também há a estratégia de utilizar a aproximação de Taylor de primeira ordem. Nesse caso utilizam-se propriedades da Jacobiana para chegar ao mesmo resultado. Ao leitor interessado em mais detalhes recomenda-se a leitura de Ribeiro e Karas (2013, p. 96)

[5]Mais especificamente Método de Newton Puro, pois $t_k = 1$ para todo k. Todavia, para uma discussão teórica mais detalhada desse método sugere-se a leitura de Ribeiro e Karas (2013) e Martins e Pereira (2010).

termos quadráticos ou produto entre duas variáveis. Em particular, consideremos o caso de programação quadrática irrestrita, caracterizada matricialmente por,

$$\begin{array}{cc} \text{Minimizar} & f(x) = \frac{1}{2}x^T Ax + b^T x + c \\ \text{sujeito a} & x \in \mathbb{R}^n, \end{array} \tag{3}$$

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

Notemos que

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Ax + b^T x + c \\ &= \frac{1}{2}\left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) + \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c \\ &= \frac{1}{2}\left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \right) \\ &\quad + \begin{bmatrix} b_1x_1 & b_2x_2 & \dots & b_nx_n \end{bmatrix} + c \\ &= \frac{1}{2}\left(\begin{bmatrix} a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + a_{21}x_1x_2 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \right. \\ &\quad \left. + \cdots + a_{n1}x_1x_n + a_{n2}x_2x_n + \cdots + a_{nn}x_n^2 \end{bmatrix} \right) \\ &\quad + \begin{bmatrix} b_1x_1 & b_2x_2 & \dots & b_nx_n \end{bmatrix} + c \\ &\equiv \frac{1}{2}(a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n, a_{21}x_1x_2 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ &\quad + \cdots + a_{n1}x_1x_n + a_{n2}x_2x_n + \cdots + a_{nn}x_n^2) \\ &\quad + b_1x_1 + b_2x_2 + \cdots + b_nx_n + c. \end{aligned}$$

Deste modo, segue da Proposição 4 que se x^* minimiza f, então

$$\nabla f(x^*) = 0 \Longrightarrow \begin{bmatrix} \frac{\partial}{\partial x_1}f(x^*) \\ \frac{\partial}{\partial x_2}f(x^*) \\ \vdots \\ \frac{\partial}{\partial x_n}f(x^*) \end{bmatrix} = 0.$$

Com isso, derivando f analiticamente em relação a $x_1, x_2, \dots, x_n$, obtemos que

$$\begin{aligned} \frac{\partial}{\partial x_1}f &= \frac{1}{2}(2a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + a_{21}x_2 + 0 + \cdots + 0 \\ &\qquad + \cdots + a_{n1}x_n + 0 + \cdots + 0) + b_1 \\ &= \frac{1}{2}(2a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + a_{21}x_2 + \cdots + a_{n1}x_n) + b_1 \\ &= \frac{1}{2}(2a_{11}x_1 + (a_{12} + a_{21})x_2 + \cdots + (a_{1n} + a_{n1})x_n) + b_1, \end{aligned}$$

e como A é simétrica, segue que

$$\begin{aligned}\frac{\partial}{\partial x_1}f &= \frac{1}{2}(2a_{11}x_1 + 2a_{21}x_2 + \cdots + 2a_{n1}x_n) + b_1\\ &= \frac{1}{2}\cdot 2(a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n) + b_1\\ &= a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n + b_1.\end{aligned}$$

Analogamente, mostra-se que $\frac{\partial}{\partial x_2}f = a_{21}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n + b_2, \ldots, \frac{\partial}{\partial x_n}f = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + b_n$. Ou seja,

$$\begin{aligned}\begin{bmatrix}\frac{\partial}{\partial x_1}f(x_1)\\ \frac{\partial}{\partial x_2}f(x_2)\\ \vdots\\ \frac{\partial}{\partial x_n}f(x_n)\end{bmatrix} &= \begin{bmatrix}a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n + b_1\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n + b_2\\ \vdots\\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + b_n\end{bmatrix} = \begin{bmatrix}a_{11} & a_{21} & \cdots & a_{n1}\\ a_{21} & a_{22} & \cdots & a_{n2}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\cdot\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} + \begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix}\\ &= \begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\cdot\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} + \begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix} \qquad \text{(A ser simétrica)}\\ &= Ax + b.\end{aligned}$$

E portanto, em particular para x^*, temos que

$$\nabla f(x^*) = 0 \iff Ax^* + b = 0.$$

Isto é, uma forma de solucionar problemas de programação quadrática irrestrita consiste na resolução de um sistema linear, dado que nessa classe de problemas, obter x^* é teoricamente possível desde que A seja inversível.

Como poderá ser constado na Seção 5, o método de Newton explora essa característica para resolver problemas quadráticos (aqui sem restrições) em apenas uma iteração.

5 Método de Newton aplicado à programação quadrática irrestrita

Consideremos $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ tal como definido em (3). Com isso, aplicar o Algoritmo 4 faz com que a direção de descida na iteração zero seja,

$$\begin{aligned}d_0 &= -(\nabla^2 f(x_0))^{-1}\nabla f(x_0) = -(\nabla(\nabla f(x_0)))^{-1}\nabla f(x_0)\\ &= -(\nabla(Ax_0 + b))^{-1}(Ax_0 + b) = -A^{-1}(Ax_0 + b) = -x_0 - A^{-1}b.\end{aligned}$$

E com isso,

$$x_1 = x_0 + (-x_0 - A^{-1}b) = -A^{-1}b.$$

Ou seja, a abordagem de Newton aplicada a problemas de programação quadrática irrestrita minimiza f *em apenas uma iteração.* Assim, reescrevendo o Algoritmo 4 obtemos:

Escolher o critério de parada ε;
Escolher x_0;
$k \leftarrow 0$;
enquanto $||\nabla f(x_k)|| > \varepsilon$ **faça**
  Definir $d_k = -x_k - A^{-1}b$;
  $x_{k+1} \leftarrow x_k + d_k$;
  $k \leftarrow k + 1$;
fim

Algoritmo 5: Método de Newton aplicado a programação quadrática irrestrita

6 Método do Gradiente aplicado à programação quadrática irrestrita

Novamente, consideremos $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que $f(x) = \frac{1}{2}x^T Ax + b^T x + c$. Aplicando o 3 temos que a direção associada a iteração k será $d_k = -\nabla f(x_k)$, enquanto que, visando obter t_k de tal modo que a descida em direção ao minimizador seja a mais rápida possível devemos resolver o seguinte subproblema de otimização (a cada iteração):

$$\begin{array}{cc} \text{minimizar} & g(t_k) = f(x_k + t_k d_k) \\ \text{sujeito a} & t_k > 0. \end{array}$$

Diante disso, aplicando a proposição 4, temos que

$$g'(t_k) = 0.$$

Logo, obtemos que

$$\begin{aligned} \nabla f(x_k + t_k d_k)^T d_k = 0 &\iff \nabla f(x_{k+1})^T d_k = 0 \iff (Ax_{k+1} + b)^T d_k = 0 \\ &\iff (A(x_k + t_k d_k) + b)^T d_k = 0 \iff (Ax_k + t_k Ad_k + b)^T d_k = 0 \\ &\iff (Ax_k + b + t_k Ad_k)^T d_k = 0 \iff (\nabla f(x_k) + t_k Ad_k)^T d_k = 0 \\ &\iff (\nabla f(x_k)^T + t_k (Ad_k)^T) d_k = 0 \iff \nabla f(x_k)^T d_k + t_k (Ad_k)^T d_k = 0 \\ &\iff t_k (Ad_k)^T d_k = -\nabla f(x_k)^T d_k \iff t_k = -\frac{\nabla f(x_k)^T d_k}{(Ad_k)^T d_k}. \end{aligned}$$

Assim, das propriedade de transposto e do fato de que A é simétrica, segue que o tamanho do passo relativo a iteração k é dado por

$$t_k = -\frac{\nabla f(x_k)^T d_k}{d_k^T A d_k} = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T A \nabla f(x_k)}.$$

Ou seja, o Método do Gradiente aplicado a problemas de programação quadráticos é dado por:

Escolher o critério de parada ε;
Escolher x_0;
$k \leftarrow 0$;
enquanto $||\nabla f(x_k)|| > \varepsilon$ **faça**
  Definir $d_k = -\nabla f(x_k)$;
  Definir $t_k = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T A \nabla f(x_k)}$;
  $x_{k+1} \leftarrow x_k + t_k d_k$;
  $k \leftarrow k + 1$;
fim

Algoritmo 6: Método do Gradiente aplicado a programação quadrática irrestrita

7 Considerações finais

Este trabalho buscou apresentar uma breve contextualização e alguns conceitos envolvendo o campo da Otimização Matemática, além de definir uma classe particular de problemas da Otimização chamada de Programação Quadrática Irrestrita. Em especial, buscou-se mostrar que nessa classe o método de Newton resolve um problema em apenas uma iteração bem como provou-se a fórmula utilizada para determinar o tamanho do passo no Método do Gradiente.

Referências

FRIEDLANDER, A. *Elementos de programação não linear.* Campinas: Pontes, 1994.

GOLDBARG, M. C.; LUNA, H. P. L. *Otimização Combinatória e Programação Linear*: modelos e algoritmos. 2ed. ed. Rio de Janeiro: Elsevier, 2005.

HILLIER, F. S.; LIEBERMAN, G. J. *Introdução a Pesquisa Operacional.* Tradução Ariovaldo Griesi. Revisão técnica Pierre J. Ehrlich. 9. ed. Porto Alegre: AMGH, 2013.

LUENBERGER, D. G.; YE, Y. *Linear and Nonlinear Programming.* New York: Springer International Publishing Switzerland, 2016.

MARTÍNEZ, J. M.; SANTOS, S. A. Métodos computacionais de otimização. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 20., 1995, Rio de Janeiro. *Anais...* Rio de Janeiro: IMPA, 1995. Disponível em: `https://impa.br/wp-content/uploads/2017/04/20 CBM 95 03.pdf`. Acesso em: 25 jan. 2021.

MARTINS, M. H. S.; PEREIRA, R. *Cálculo III e IV.* Florianópolis: UFSC/EAD/CED/CFM, 2010.

RIBEIRO, A. A.; KARAS, E. W. *Otimização contínua*: aspectos teóricos e computacionais. São Paulo: Cengaje Learning, 2013.

SHI, Z.-J.; SHEN, J. New inexact line search method for unconstrained optimization. *Journal*

of optimization theory and applications, Springer, v. 127, n. 2, p. 425–446, 2005.