Revista Eletrônica Paulista de Matemática

ISSN 2316-9664
Volume 15, jul. 2019
Iniciação Científica

Taxas de convergência para métodos de otimização

Convergence rates for optimization methods

Quezia Cristiane de Oliveira Maia Carvalho
Universidade Federal de Uberlândia
queziamaia.qm@gmail.com

Evaneide Alves Carneiro
Universidade Federal de Uberlândia
eva.ac@ufu.br

Resumo

O propósito do presente trabalho é apresentar algumas noções básicas acerca de métodos de otimização em geral. São mencionadas definições importantes dos diferentes tipos de métodos e os principais problemas relacionados a utilização e análise teórica deles. Em especial, tratamos sobre os conceitos que classificam os métodos, algumas noções de convergência, taxas de convergência e regras de parada. No mais, demonstramos uma relação entre duas taxas de convergência, que é um dos principais indicadores de eficiência de um método iterativo, apresentamos um exemplo de uma das taxas de convergências citadas e ainda analisamos a taxa de convergência do método de Newton para equações com o intuito de exemplificar o exposto. Os métodos de Newton, muito eficientes na área da matemática computacional, são um meio utilizado para resolver sistemas de equações não-lineares. Estes admitem também uma interpretação voltada para a otimização, a fim de desenvolver técnicas de globalização.
Palavras-chave: Otimização. Taxas de convergência. Regras de parada.

Abstract

The purpose of this paper is to present some basic notions about optimization methods in general. Some important definitions of the different types of methods and the main problems related to their use and theoretical analysis are mentioned. In particular, we deal with the concepts that classify the methods, some notions of convergence, convergence rates and stopping rules. Moreover, we show a relation between two convergence rates, which is one of the main efficiency indicators of an iterative method. We present an example of one of the convergence rates and analyze the convergence rate of the Newton's method for equations to exemplify the above. Newton's method are used to solve systems of nonlinear equations and admit an optimization interpretation.
Keywords: Optimization. Convergence rates. Stopping rules.

Artigo recebido em set. 2018 e aceito em maio 2019.

1 Introdução

O presente trabalho é um recorte da iniciação científica "Introdução a Otimização: Métodos unidimensionais e Métodos de Newton".

A otimização matemática, que consiste basicamente em procurar soluções ideais (maximizantes ou minimizantes) para um certo problema, é uma área que tem crescido bastante nos últimos anos, apesar de ser relativamente recente. Tal crescimento da área conduziu a um aumento da especialização e da diversificação. Assim sendo, podem-se observar especializações em otimização não linear, otimização discreta, otimização combinatória, e mais recentemente em otimização estocástica. O crescimento deve-se em parte ao fato que os problemas de otimização surgem em várias aplicações práticas de ciências, como engenharia e economia. Entretanto, muitas vezes tais problemas são complexos, de difícil solução e envolvem reduções de custos significativas, melhorias do tempo de processos, ou uma melhor alocação de recursos em atividades.

Muitos processos podem necessitar de uma alocação otimizada de recursos, podendo incluir capital, equipamentos, tarefas, e devem ser corretamente alocados nas quantidades, nos tempos e na sequência para a obtenção do melhor resultado possível. Para resolver o problema em questão, implementamos então um método iterativo que busca uma solução para o problema de acordo com seus dados e alguns critérios que auxiliam na resolução mais rápida e precisa. Um exemplo de um desses critérios é a taxa de convergência, um dos principais indicadores de eficiência de um método, pois mostra a proximidade dos pontos da função objetivo do problema à solução esperada.

A relevância da pesquisa na área, vem do fato de existirem muitos problemas em aberto, noções não totalmente compreendidas, conceitos estabelecidos em cima de simples conjeturas, atraindo o interesse de muitos pesquisadores motivados pela consciência de estarem participando da construção de uma das mais promissoras áreas de conhecimento da atualidade. Como parte de uma pesquisa de iniciação científica, este trabalho é uma introdução aos conceitos e técnicas da área, e a principal referência utilizada foi o livro de Izmailov (2007), em relação ao qual a contribuição é a demonstracão da relação entre duas taxas de convergência e os exemplos apresentados, que acreditamos contribuir para a compreensão do tema exposto.

2 Classificação dos métodos de otimização

Consideremos o problema dado em formato geral

$$\min f(x) \quad \text{sujeito a} \quad x \in D = \left\{ x \in \Omega \;\middle|\; \begin{array}{l} h(x) = 0, \\ g_i(x) \leq 0. \end{array} \right\}, \tag{1}$$

onde $\Omega \subset \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^l$ e $g : \mathbb{R}^n \to \mathbb{R}^m$, com $g(x) = (g_1(x), g_2(x), \ldots g_m(x))$. A solução desse problema depende das condições sobre o conjunto D e da regularidade das funções envolvidas.

Dizemos que (1) é um *problema de minimização convexo* quando $D \subset \mathbb{R}^n$ é um conjunto convexo e $f : D \to \mathbb{R}$ é uma função convexa no conjunto D. A importância da convexidade fica evidenciada no resultado a seguir, cuja demonstração pode ser encontrada em Izmailov (2007).

Teorema .1 Teorema de minimização convexa
Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \to \mathbb{R}$ uma função convexa em D.

Então todo minimizador local em (1) *é global. Além disso, o conjunto de minimizadores é convexo.*
Se f é estritamente convexa, não pode haver mais de um minimizador.

Ressaltamos que quando a função é diferenciável, a convexidade admite várias caracterizações que são muito úteis para determinar se uma função é convexa ou não. Por exemplo, se $\Omega \subset \mathbb{R}^n$ é um conjunto convexo e aberto e $f : \Omega \to \mathbb{R}$ é diferenciável em Ω, então f é convexa em Ω se e somente se, para todo $x \in \Omega$ e todo $y \in \Omega$,

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle.$$

Este resultado também pode ser encontrado em Izmailov (2007)

Todo método para resolução numérica do problema (1) utiliza o cálculo de valores da função objetivo e das restrições do problema proposto, e em alguns casos, o método pode envolver também uma ou mais derivadas destas funções. Tendo em mente essa informação, passamos às classificações dos métodos:

Um método é chamado *passivo* quando os pontos selecionados para calcular os valores ditos acima são escolhidos segundo uma estratégia que não leva em consideração as informações dadas ao longo do processo iterativo. Em contrapartida, um método é chamado *sequencial* quando cada ponto é gerado a partir da informação obtida nos pontos anteriores. É fato que os métodos sequenciais, geralmente, são mais eficientes, por isso a grande maioria dos métodos computacionais são desse tipo.

Um método sequencial gera uma sequência de pontos no $\mathbb{R}^n$, do tipo $x^1, x^2, \ldots, x^k, \ldots$, chamados de *iterandos do método*. A sequência $\{x^k\}$, também chamada *trajetória*, não necessariamente contém todos os pontos em que os valores das funções e derivadas são calculados, ela pode ter apenas alguns pontos selecionados especialmente. Ao se obter uma nova aproximação $\{x^{k+1}\}$ a partir de $\{x^k\}$ dizemos que houve uma *iteração* do método.

Frequentemente, métodos são descritos no formato de um esquema iterativo, como

$$x^{k+1} = \Psi_k(x^k), \quad k = 0, 1, \ldots, \tag{2}$$

onde Ψ_k é operador com valores em $\mathbb{R}^n$, definido em um conjunto $U_k \subset \mathbb{R}^n$. Ressaltamos que é natural supor que na iteração $k+1$ este operador seja conhecido. Assim, se o esquema iterativo é fixado, a escolha do ponto inicial $x^0 \in U^0$ define uma única trajetória do método se, e só se,

$$\Psi_{k-1}(\Psi_{k-2}(\ldots \Psi_0(x^0) \ldots)) \in U_k \quad \forall k = 1, 2, \ldots.$$

A *ordem* de um método é definida pela ordem máxima das derivadas da função objetivo e das restrições usadas no problema. Por exemplo, se um método utiliza apenas as derivadas primeiras então este é um método de primeira ordem.

Noções de convergência

Um método tem *convergência finita* quando qualquer ponto inicial $x^0 \in U_0$ forma uma trajetória tal que um de seus elementos x^k é uma solução do problema para algum k finito. Já nos métodos com *convergência assintótica*, usados em problemas com estrutura mais geral, como não há certeza que x^k seja uma solução exata $\bar{x}$ do problema, prova-se que ele é uma boa aproximação da solução para um k suficientemente grande.

Como essa aproximação pode depender de elementos variados do problema, a convergência recebe diferentes nomes de acordo com cada aproximação, as quais listaremos a seguir:

1. A *convergência em relação às variáveis do problema* ocorre quando

$$\{x^k\} \to \bar{x} \quad (k \to \infty),$$

a partir do ponto inicial x^0.

2. A *convergência ao conjunto de soluções*, quando

$$dist(x^k, S) \to 0 \quad (k \to \infty),$$

onde S é o conjunto de soluções do problema. Além disso, quando S é um conjunto compacto, esta propriedade se resume à existência de pontos de acumulação de $\{x^k\}$, considerando que todo ponto de acumulação desta sequência é uma solução do problema. No mais, quando S é ilimitado, é possível que $\{x^k\}$ munido da propriedade acima não possua pontos de acumulação.

3. A *convergência em relação à função objetivo* acontece quando

$$dist(x^k, D) \to 0,\ f(x^k) \to \bar{v} \quad (k \to \infty),$$

onde $\bar{v}$ é o valor ótimo do problema. Vale mencionar que essa propriedade não garante que a trajetória $\{x^k\}$ do método convirja.

4. A *convergência em relação ao gradiente* ocorre quando

$$\{f'(x^k)\} \to 0 \quad (k \to \infty).$$

Essa convergência é mais proveitosa em problemas irrestritos, pois um resultado para métodos de otimização irrestrita garante que todo ponto de acumulação de $\{x^k\}$ é estacionário.

5. Por fim, temos a *convergência global* quando $U_0 = \mathbb{R}^n$ e tem-se convergência a partir de qualquer ponto $x^0 \in \mathbb{R}^n$. Porém, quando a convergência só pode ser garantida a partir de pontos estacionários ou pontos suficientemente próximos à solução, falamos de *convergência local* do método.

Destacamos que é sempre bom que métodos com convergência local sejam combinados com outro método que forneça um ponto inicial x^0, o que chamamos de *globalização da convergência*, a fim de que haja uma aplicação bem sucedida do método local.

De acordo com Izmailov (2007), de modo geral, uma maneira de solucionar um problema relativamente difícil é reduzi-lo a uma sequência de problemas mais simples, seja ela finita ou infinita. O impasse aqui é o fato de que nem sempre é possível encontrar métodos eficientes para a resolução de cada problema simples. Um exemplo dessa estratégia é o Método Clássico de Newton que reduz a resolução de um sistema de equações não-lineares à solução de uma sequência de equações lineares. Vê-se então que o desenvolvimento de métodos para problemas mais complexos exige estudo prévio de métodos que solucionam problemas mais básicos.

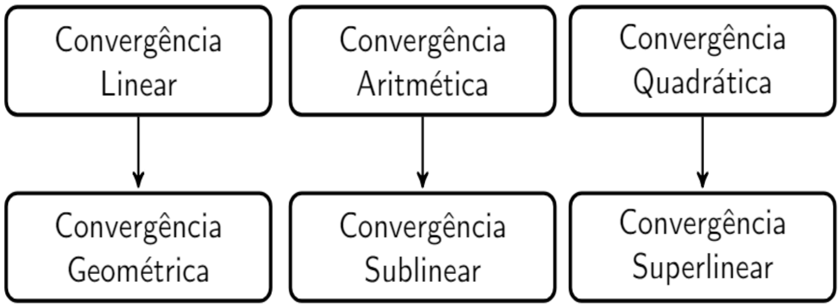
3 Taxas de convergência

A *taxa de convergência* é um dos principais indicadores de eficiência de um método iterativo, por isso é uma das propriedades mais importantes para a análise e avaliação de métodos com convergência assintótica.

Suponhamos que um dado método seja convergente, local ou globalmente, a uma solução $\bar{x}$. Resultados sobre taxa de convergência fornecem uma garantia de velocidade na redução da distância $\|x^{k+1} - \bar{x}\|$ em relação a $\|x^k - \bar{x}\|$ quando k é suficientemente grande. Neste caso, supomos também que $x^0 \in \mathbb{R}^n$ é suficientemente próximo a $\bar{x}$, que $x^k \neq \bar{x}$ para todo k e que todas as constantes a seguir não dependem de x^0.
Sabendo isso, denominamos as seguintes taxas de convergência:

1. *Convergência linear:*
   Se existe $q \in (0,1)$ tal que
   $$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = q.$$
2. *Convergência sublinear:*
   Quando
   $$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = 1.$$
   o que a torna mais lenta que a linear.
3. *Convergência superlinear:*
   Quando $q = 0$ na relação da convergência linear.
4. *Convergência aritmética:*
   Ocorre quando
   $$\limsup_{k\to\infty} k\|x^k - \bar{x}\| \leq C,$$
   onde $C > 0$ é uma constante fixa. Mostraremos adiante que este é um exemplo de propriedade que garante apenas taxa sublinear.
5. *Convergência geométrica:*
   Quando
   $$\limsup_{k\to\infty} \frac{\|x^k - \bar{x}\|}{q^k} \leq C,$$
   onde $q \in (0,1)$.
6. *Convergência quadrática:*
   Verifica-se quando
   $$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|^2} \leq C,$$
   Essa taxa é um caso particular da taxa superlinear.

Cabe mencionar que a convergência quadrática da sequência $\{x^k\} \subset \mathbb{R}^n$ a um ponto $\bar{x}$, a princípio não implica na convergência quadrática e nem superlinear das sequências $\{x_i^k\}$ a $\bar{x}_i, i = 1, \ldots, n$. Outra observação a se fazer é que a taxa linear implica convergência geométrica, mas a recíproca não acontece.



Provaremos então que (6) $\Rightarrow$ (3):

$$\begin{aligned}
&\limsup_{k\to\infty}\frac{\|x^{k+1}-\bar{x}\|}{\|x^k-\bar{x}\|}\\
&=\limsup_{k\to\infty}\frac{\|x^{k+1}-\bar{x}\|}{\|x^k-\bar{x}\|^2}\cdot\|x^k-\bar{x}\|\\
&\leq\limsup_{k\to\infty}\frac{\|x^{k+1}-\bar{x}\|}{\|x^k-\bar{x}\|^2}\cdot\limsup_{k\to\infty}\|x^k-\bar{x}\|\\
&\leq C\cdot 0=0
\end{aligned}$$

□

Exemplo .1 *Consideremos a sequência* $\{x^k\}\subset\mathbb{R}^2$ *assim gerada: Para todo k, tomamos* $x_2^{k+1}=(x_2^k)^2$. *Para k ímpar, tomamos* $x_1^{k+1}=x_1^k$ *e para k par, tomamos* $x_1^{k+1}=(x_2^{k+1})^2$ *e seja* $x_2^1\in(0,1)$. *Mostraremos que a sequência* $\{x^k\}$ *converge para* $\bar{x}=(0,0)$ *com taxa quadrática. Para isso, lembremos que a convergência quadrática é dada por*

$$\limsup_{k\to\infty}\frac{\|x^{k+1}-\bar{x}\|}{\|x^k-\bar{x}\|^2}\leq C.$$

Sabemos que para k ímpar, ocorre

$$\begin{cases}\|x^{k+1}-\bar{x}\|=\|(x_1^k,(x_2^k)^2)-(0,0)\|=\sqrt{(x_1^k)^2+(x_2^k)^4}\\ \|x^k-\bar{x}\|=\|(x_1^k,x_2^k)-(0,0)\|=\sqrt{(x_1^k)^2+(x_2^k)^2}\end{cases}.$$

E considerando que

$$(x_1^k)=(x_2^k)^2=x_1^{k+1}\ e\ (x_1^k)^2=(x_2^k)^4\leq(x_2^k)^2,$$

temos

$$\limsup_{k\to\infty}\frac{\sqrt{(x_1^k)^2+(x_2^k)^4}}{(x_1^k)^2+(x_2^k)^2}=\limsup_{k\to\infty}\frac{\sqrt{2(x_2^k)^4}}{(x_2^k)^4+(x_2^k)^2}=\limsup_{k\to\infty}\frac{\sqrt{2}(x_2^k)^2}{(x_2^k)^2+(1+(x_2^k))^2}\leq\sqrt{2}.$$

Agora, para k par, temos

$$\begin{cases}\|x^{k+1}-\bar{x}\|=\|((x_2^{k+1})^2,(x_2^{k+1})-(0,0))\|=\sqrt{(x_2{}^k+1)^4+(x_2^{k+1})^2}\\ \|x^k-\bar{x}\|=\sqrt{(x_1^k)^2+(x_2^k)^2}\end{cases}.$$

Usando o fato de que

$$x_1^k = x_2^k = x_1^{k-1} = (x_2^{k-1})^2,\ k \geq 4 \ \Rightarrow (x_1^k)^2 = (x_2^k)^2 = x_2^{k+1}$$

$$\limsup_{k\to\infty} \frac{\sqrt{(x_2^{k+1})^2 \cdot (x_2^{k+1})^2 + 1}}{(x_1^k)^2 + (x_2^k)^2} = \limsup_{k\to\infty} \frac{x_2^{k+1}\sqrt{(x_2^{k+1})^2 + 1}}{(x_1^k)^2 + (x_2^k)^2}$$

$$= \limsup_{k\to\infty} \frac{\sqrt{(x_2^{k+1})^2 + 1}}{(x_1^k)^2 + (x_2^k)^2} \leq \frac{\sqrt{2}}{\frac{(x_1^k)^2}{x_2^{k+1}}} + \frac{(x_2^k)^2}{x_2^{k+1}} \leq \frac{\sqrt{2}}{1+1} \leq \frac{\sqrt{2}}{2}.$$

As estimativas de taxa de convergência geram informação útil para comparação qualitativa de diferentes métodos, entretanto, é importante ressaltar que esta taxa não é a única variável a se considerar, pois bem se sabe que é preciso levar em conta o custo computacional de uma iteração. Outra propriedade bem relevante de um método é a *estabilidade*, definida pela capacidade de lidar com problemas cujos dados são fornecidos de maneira inexata ou com perturbações. Além dessa, ainda é importante, em alguns casos, que a sequência gerada seja viável, ou em termos matemáticos $x^k \in D \ \forall \ k$, pois em várias aplicações práticas, a função f ou as restrições g, h estão bem definidas apenas em pontos viáveis, isto é, a função objetivo ou as restrições só estão bem definidas nos pontos $(x^1, x^2, \ldots, x^n)$ interiores a D (região viável) que é o conjunto de todos os pontos que satisfazem as restrições de um problema de programação linear. Sendo assim, para a solução do problema, é melhor que se escolha um método que produza pontos viáveis em cada iteração.

3.1 Regras de parada

Mesmo que um método tenha convergência assintótica, nenhum método computacional é infinito, por isso apresentamos a seguir o conceito de *regra de parada*. As regras mais precisas são as que se relacionam com as taxas de convergência ou outras propriedades relacionadas a ela, pois quando a taxa de convergência é conhecida, podemos estimar o número de iterações a se fazer até que a solução alcance uma precisão dada. O problema em questão é que na maioria dos problemas, a taxa de convergência é local, o que impossibilita estabelecer o número de iterações necessárias para se encontrar um ponto a partir do qual a taxa vale. E ainda, as taxas de convergência apresentam certas constantes q e C com valores desconhecidos que limitam as aplicações práticas dessas taxas para obtenção de regras de parada.

Nas resoluções computacionais, usamos regras de parada diferentes das ditas acima, porque são mais fáceis de implementar, apesar de apresentarem menor precisão. Tais regras levam em consideração o valor obtido em x^k e talvez em alguns pontos que o antecedem, além do comportamento das sequências x^k e $\{f(x^k)\}$ e as condições de otimalidade.

São tipos de regras de parada:

1. Ao fixar um número pequeno $\varepsilon > 0$, o método se encerra quando ao fazer a iteração $(k+1)$, tem-se
$$\|x^{k+1} - \bar{x}\| \leq \varepsilon.$$
Isso nos garante que o passo do método está curto e que não há mais como melhorar a qualidade de aproximação a uma solução.

2. Apresentamos uma taxa que se desenvolve de maneira parecida com a anterior
$$\|f(x^{k+1})-f(\bar{x})\| \leq \varepsilon.$$

3. Quando f é diferenciável em um problema irrestrito, temos a taxa
$$\|f'(x^{k+1})\| \leq \varepsilon.$$

Deixamos claro que as taxas de convergência apresentadas acima não garantem em nenhum sentido uma aproximação do iterando x^{k+1} a uma solução do problema. Mesmo assim, elas são muito úteis em problemas computacionais, porque não há métodos mais eficientes que possam ser usados.

Na prática, mesmo que nenhuma regra de parada seja satisfeita em um problema de grande porte, o método é parado por exaustão do tempo disponível, o que torna o número máximo de iterações ou o número máximo de avaliações da função objetivo uma regra de parada implícita. Pode ser que nesse caso a aproximação gerada não seja tão próxima da real solução, porém às vezes isso é o melhor que pode se obter com as ferramentas que se tem em mãos.

Como vimos que há alguns impasses quanto às regras de parada, o ideal é que elas sejam usadas em combinação, segundo uma hierarquia estabelecida entre elas. Essa hierarquia e vários parâmetros relacionados a ela é uma questão que abrange mais a área da computação que da ciência exata.

4 Análise da taxa de convergência do método de Newton

Essa seção diferencia o presente trabalho do publicado na caderno de trabalhos completos e resumos do ERMAC 2018, citado nas referências.

O método de Newton, que descrevemos abaixo, é uma ferramenta para a solução de sistemas de equações não-lineares, o que pode ser aplicado a problemas de otimização, na determinação dos possíveis pontos extremos de uma função derivável.

Especificamente, O método de Newton clássico é introduzido para o problema de achar um $x \in \mathbb{R}^n$ tal que
$$\Phi(x) = 0, \tag{3}$$
onde $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ é diferenciável. Seja $x^k \in \mathbb{R}^n$ uma aproximação a uma solução qualquer $\bar{x}$ da equação (3). Podemos então fazer uma aproximação de (3) pela sua linearização
$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \tag{4}$$
A relação (4) é denominada *equação de iteração do método de Newton*.

Algoritmo .1 O método de Newton
Escolher $x^0 \in \mathbb{R}^n$ *e tomar* $k = 0$.

1. *Calcular* $x^{k+1} \in \mathbb{R}^n$ *como uma solução da equação linear* (4).

2. *Tomar* $k := k+1$ *e retornar ao Passo* (1).

Supondo $\Phi'(x^k)$ não singular, o método de Newton é escrito na forma do esquema iterativo

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}\Phi(x^k), \quad k = 0,1,\ldots \tag{5}$$

e nesse caso a solução da equação de Newton é única.

Ressaltamos que caso $\Phi'(x)$ seja não singular, o cálculo de x^{k+1} não exige necessariamente o cálculo de sua matriz inversa, pois este é um processo consideravelmente caro do ponto de vista computacional, e além disso, existem vários métodos de resolução de sistemas de equações lineares que não envolvem a inversa da matriz associada.

O método descrito acima não necessariamente gera uma sequência convergente, como mostra o próximo exemplo.

Exemplo .2 *Seja $\Phi(x) : \mathbb{R} \to \mathbb{R}$ dada por $\Phi(x) = \dfrac{x}{\sqrt{1+x^2}}$. Então:*

$$\Phi'(x) = \frac{1}{(1+x^2)^{\frac{3}{2}}}.$$

Portanto, o método de Newton gera a sequência

$$x^{k+1} = x^k - \frac{x^k}{\sqrt{1+(x^k)^2}}((1+(x^k)^2)^{\frac{3}{2}}) = x_k - x_k(1+(x^k)^2) = -(x^k)^3.$$

A solução da equação $\Phi(x) = 0$ é $x = 0$, porém se x^0 é tal que $|x^0| > 1$, a sequência (x^k) diverge.

Portanto, como ilustrado pelo exemplo acima, para que tenhamos a garantia que o método de Newton gera uma sequência convergente, devemos indicar hipóteses sobre a função envolvida e o ponto inicial escolhido. Uma maneira de garantir a convergência local é enunciada e demonstrada a seguir.

Teorema .2 Convergência local do método de Newton
Seja $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ diferenciável numa vizinhança do ponto $\bar{x} \in \mathbb{R}^n$, com derivada contínua nesse ponto, e seja $\bar{x}$ uma solução da equação (3)*, tal que* $\det \Phi'(\bar{x}) \neq 0$*, isto é, a matriz $\Phi'(\bar{x})$ é não singular. Então para qualquer ponto inicial $x^0 \in \mathbb{R}^n$ suficientemente próximo a $\bar{x}$, o Algoritmo* (.1) *gera uma sequência $\{x^k\}$ bem definida que converge para $\bar{x}$. Nesse caso, a taxa de convergência é superlinear, e se a derivada de Φ é Lipschitz-contínua numa vizinhança de $\bar{x}$, então a taxa de convergência é quadrática.*

Demonstração .1 *Pelas propriedades de matrizes não singulares, existe uma vizinhança U do ponto $\bar{x}$ e um número $M > 0$ tais que*

$$\det \Phi'(x) \neq 0, \quad \|\Phi'(x)^{-1}\| \leq M \quad \forall x \in U. \tag{6}$$

Em particular, para $x^k \in U$ a equação (4) *tem uma única solução x^{k+1}, ou seja, a partir de um x^k, a iteração do Algoritmo* (.1) *está bem definida. Além disso, pelas equações* (5)*,* (6) *e pelo Teorema do Valor Médio, podemos tomar a vizinhança de U tal que, se $x^k \in U$, então*

$$\begin{aligned} &\|x^{k+1} - \bar{x}\| \\ &= \|x^k - \bar{x} - (\Phi'(x^k))^{-1}\Phi(x^k)\| \\ &\leq \|(\Phi'(x^k))^{-1}\| \, \|\Phi(x^k) - \Phi(\bar{x}) - \Phi'(x^k)(x^k - \bar{x})\| \\ &\leq M \sup_{t\in[0,1]} \|\Phi'(tx^k + (1-t)\bar{x}) - \Phi'(x^k)\| \, \|x^k - \bar{x}\|. \end{aligned} \tag{7}$$

Notamos que

$$\sup_{t\in[0,1]} \|\Phi'(tx^k+(1-t)\bar{x})-\Phi'(x^k)\| \to 0 \quad (x^k \to \bar{x}).$$

Portanto, de (7)*, temos que para todo* $q \in (0,1)$ *existe* $\delta > 0$ *tal que* $B(\bar{x},\delta) \subset U$ *e, se* $x^k \in B(\bar{x},\delta)$*,*

$$\|x^{k+1}-\bar{x}\| \leq q\|x^k-\bar{x}\|.$$

Em particular, $x^{k+1} \in B(\bar{x},\delta)$*. Concluímos que para todo* x^0 *suficientemente próximo a* $\bar{x}$*, o Algoritmo* (.1) *gera uma sequência* $\{x^k\}$ *bem definida que converge para* $\bar{x}$*. Além disso, a taxa de convergência é superlinear, pois, quando* $x^k \to \bar{x}$*, podemos tomar* q *cada vez menor, uma vez que* $q \to 0$*.*

Quando a derivada de Φ *é Lipschitz-contínua em* U *com constante de Lipschitz* $L > 0$*, de* (7) *temos:*
se $x^k \in U$*, então*

$$\|x^{k+1}-\bar{x}\| \leq ML\|x^k-\bar{x}\|^2, \tag{8}$$

ou seja, a convergência é quadrática. □

Sob as hipóteses do Teorema (.2), a convergência local do método de Newton é bem rápida. Porém, vemos que a convergência, em geral, é apenas local. Ou seja, para um ponto inicial que não está suficientemente próximo à solução, o método pode não funcionar, mesmo que a solução satisfaça todas as hipóteses necessárias.

Uma observação a se fazer é que tanto no Teorema (.2) como em outros resultados sobre convergência quadrática de métodos Newtonianos, no lugar da hipótese de que a derivada de Φ seja Lipschitz-contínua, podemos usar uma condição um pouco mais fraca:

$$\|\Phi'(x)-\Phi'(\bar{x})\| \leq L\|x-\bar{x}\| \quad \forall x \in V,$$

onde V é uma vizinhança da solução $\bar{x}$ e $L > 0$ é uma constante.

5 Conclusões

Vimos alguns exemplos de taxas de convergência para métodos de otimização, que servem como uma pequena ilustração do assunto. Estudamos algumas relações entre diferentes tipos de convergência e analisamos a convergência do Método de Newton. Esse assunto é amplo, bastante explorado na literatura e abrange uma área de pesquisa possível para trabalhos futuros.

6 Referências bibliográficas

CARVALHO, Q. C. O. M.; CARNEIRO, E. A. Taxas de convergência para métodos de otimização. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 5., 2018, Bauru. **Caderno de trabalhos completos e resumos**... Bauru: Unesp, Faculdade de Ciências, 2018. p. 324-330. Disponível em: <https://drive.google.com/file/d/1-8SZZnioKApdmQG6BaOTltJARzfwAB9P/view>. Acesso em: 29 jan. 2019.

GENARO, R. **Métodos de otimização irrestrita e aplicações**. 2010. 50 f. Trabalho de Conclusão de Curso (Bacharelado em Matemática) – Faculdade de Ciências Integradas do Pontal, UFU, Ituiutaba, 2010.

IZMAILOV, A.; SOLODOV, M. **Otimização**. Rio de Janeiro: IMPA, 2007. v. 2.

PEREIRA, A. F. **Programação não-linear, métodos numéricos e aplicações**. 2010. 62 f. Trabalho de Conclusão de Curso (Bacharelado em Matemática) – Faculdade de Ciências Integradas do Pontal, UFU, Ituiutaba, 2010.