Otimização não linear

Isabel Espírito Santo

Departamento de Produção e Sistemas

Escola de Engenharia

Universidade do Minho

iapinho@dps.uminho.pt

Métodos numéricos de resolução

problema sem restrições

$$\min_{x \in \mathbb{R}^n} f(x) \quad (n > 1)$$

- Métodos de procura direta;
- Métodos do gradiente.

Métodos de procura direta:

- só usam informação da função objetivo f;
- são apropriados para **problemas não diferenciáveis** (embora possam ser usados em problemas diferenciáveis);
- método de Nelder-Mead (destina-se a problemas de otimização multidimensionais).

Métodos do gradiente

Métodos do gradiente:

- usam informação da função e das derivadas (gradiente ou/e Hessiana);
- só podem ser usados na resolução de **problemas diferenciáveis**;
- convergem mais rapidamente do que os métodos de procura direta;
- geram uma sucessão de aproximações $\{x^{(k)}\}$ à solução:

$$\boxed{x^{(k+1)} = x^{(k)} + \alpha^k \, d^{(k)}}$$

em que $d^{(k)}$ (vector) é a **direção de procura** (ou passo) e α^k (escalar) é o **comprimento do passo**. A equação iterativa para o cálculo da direção de procura é diferente para cada método.

Método de Newton

Derivando em ordem a d e igualando a zero (define a condição de 1^{a}ordem para o mínimo da quadrática, $\nabla q\,(d) = 0$), obtém-se

$$\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})\, d = 0$$
$$\Updownarrow$$
$$\nabla^2 f(x^{(k)})\, d = -\nabla f(x^{(k)}) \qquad (2)$$

A solução do **sistema linear** (2), d, é a direção de procura:

- se a dimensão do problema (n) for pequena ou média, usa-se o método directo e estável - EGPP;
- se n for grande, usa-se um método iterativo - gradientes conjugados.

Método de Newton

A nova aproximação $x^{(k)} + d$ não é necessariamente o minimizante de $f(x)$ e o processo deve ser repetido.

As equações iterativas do **Método de Newton** (na forma básica) são

$$\begin{cases} \nabla^2 f(x^{(k)})\, d_N^{(k)} = -\nabla f(x^{(k)}), & \textbf{(sistema Newton)} \\ x^{(k+1)} = x^{(k)} + d_N^{(k)} & \text{para } k = 1, 2, \ldots \end{cases}$$

em que $d_N^{(k)}$ é a direção Newton.

Propriedades do método de Newton

- o método de Newton tem convergência
  - **local** (a convergência para a solução só é garantida se a aproximação inicial $x^{(1)}$ estiver na vizinhança da solução);
  - **quadrática**

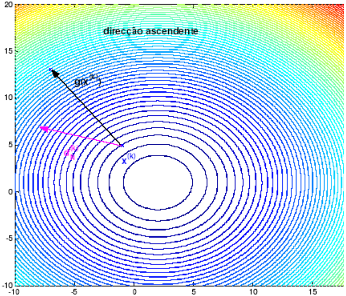
$$\|x^{(k+1)} - x^*\| \leq \gamma \|x^{(k)} - x^*\|^2, \, \gamma > 0;$$

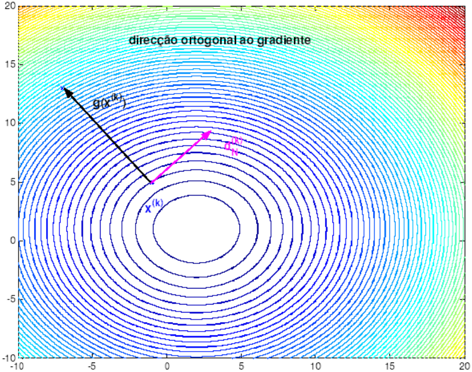
- o método de Newton possui a propriedade da **terminação quadrática**, i.e., se $f(x)$ $(x \in \mathbb{R}^n)$ for uma função quadrática e convexa o método de Newton necessita no máximo de n iterações para encontrar a solução.

Limitações do método de Newton

- a direção $d_N^{(k)}$ (solução do sistema Newton) pode **não** ser **descendente** para f em $x^{(k)}$, ou seja,

$$\nabla f(x^{(k)})^T \, d_N^{(k)} > 0 \quad \text{ou} \quad \nabla f(x^{(k)})^T \, d_N^{(k)} = 0$$

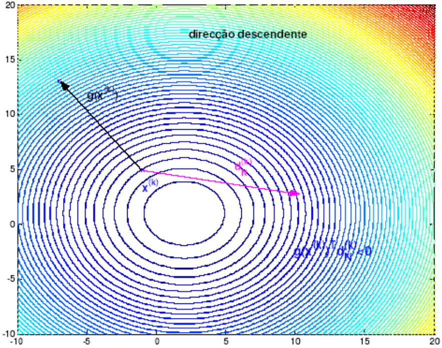




Limitações do método de Newton

- A direção $d_N^{(k)}$, ainda que seja **descendente** ($\nabla f(x^{(k)})^T d_N^{(k)} < 0$), pode ser muito grande e se $x^{(k+1)} = x^{(k)} + d_N^{(k)}$ não se verifica

$$f(x^{(k)} + d_N^{(k)}) < f(x^{(k)})$$



Limitações do método de Newton

- Além disso, a matriz $\nabla^2 f(x^{(k)})$ (matriz dos coeficientes do sistema Newton) pode ser singular, o que significa que o sistema Newton não tem solução ou tem uma infinidade de soluções

$$\Updownarrow$$

$$\nexists d_N^{(k)} \text{ (única)}$$

Desvantagens do método de Newton

- Cálculo das segundas derivadas:
  - se a expressão de f é complicada, estas tornam-se difíceis de calcular;
  - exigem um grande esforço de cálculo quando n é grande.
- A convergência é local.

Para ultrapassar a convergência local

deve implementar-se uma técnica de globalização para garantir que o método converge para a solução, a partir de qualquer aproximação inicial.

Técnicas de globalização

Os métodos do gradiente (quando convergem) convergem para um ponto estacionário ($\nabla f(x) = 0$).

Porquê implementar uma técnica de globalização?

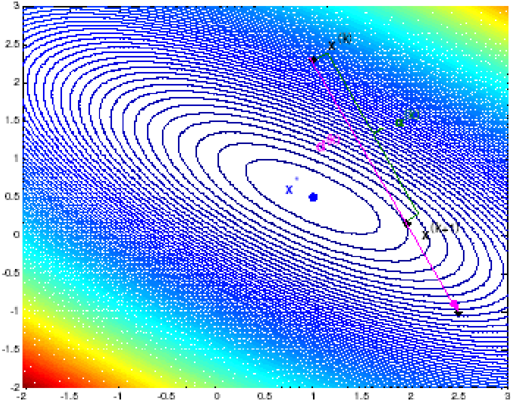
- para garantir que o método converge, qualquer que seja a aproximação inicial $x^{(1)}$ (i.e., $x^{(1)}$ pode estar fora da região de convergência do método);
- para garantir que o método converge para um ponto estacionário que é **minimizante**.

1. Procura unidimensional (line search) $\begin{cases} \text{exata} \\ \text{aproximada} \end{cases}$
2. Região de confiança (trust region)
3. Filtro

Procura unidimensional exata

Dados $x^{(k)}$ e $d^{(k)}$, calcular $\alpha^{(k)}$:

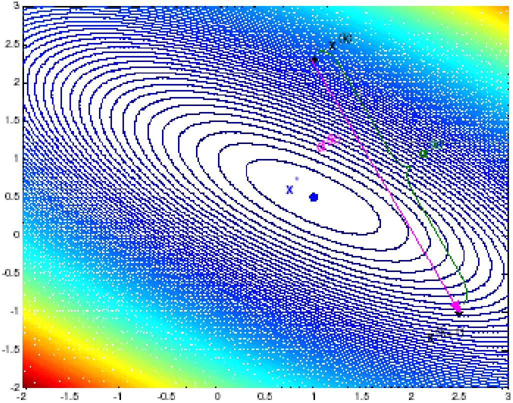
$$\begin{cases} \text{comprimento do passo ótimo: } \alpha^{(k)} = \arg\min\limits_{\alpha} f\left(x^{(k)} + \alpha d^{(k)}\right) \\ \\ \text{nova aproxinação à solução: } x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)} \end{cases}$$



Procura unidimensional aproximada

Dados $x^{(k)}$ e $d^{(k)}$, calcular $\alpha^{(k)}$ (comprimento do passo) que origina uma redução significativa no valor de f na nova aproximação (isto é, satisfaz a **condição de Armijo**):

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + \mu \alpha^{(k)} \nabla f(x^{(k)})^T d^{(k)}$$



Critério de Armijo

O objetivo é exigir uma redução significativa (usando a condição de Armijo) no valor da função objetivo, i.e.,

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + \mu\, \alpha^{(k)}\, \nabla f(x^{(k)})^T\, d^{(k)}$$

com $0 < \mu < \frac{1}{2}$.

Nota: se a direção $d^{(k)}$ usada for **descendente para** f, ou seja, $\nabla f(x^{(k)})^T\, d^{(k)} < 0$, existe um valor de $\alpha^{(k)} \in (0, 1]$ que verifica esta condição.

Algoritmo do critério de Armijo para calcular $\alpha^{(k)}$

Dados $x^{(k)}$, $d^{(k)}$, $\nabla f\left(x^{(k)}\right)$, $f\left(x^{(k)}\right)$ e μ

1. $\alpha \leftarrow 1$
2. $\bar{x} \leftarrow x^{(k)} + \alpha d^{(k)}$
3. <u>se</u> $\left(f(\bar{x}) \leq f\left(x^{(k)}\right) + \mu\, \alpha\, \nabla f\left(x^{(k)}\right)^T d^{(k)}\right)$ <u>então</u>

   fazer $\alpha^{(k)} = \alpha$

   <u>senão</u>

   $\alpha \leftarrow \alpha/2$ e voltar a 2.

Algoritmo genérico de um **método do gradiente**

ler aproximação inicial $x^{(1)} \in \mathbb{R}^n$

$k \leftarrow 0$

repetir

- $k \leftarrow k + 1$
- calcular $d^{(k)}$ (direção de procura ou passo)
  - (por exemplo, a direção do método de segurança de Newton ou a direção do método quasi-Newton)
- calcular $\alpha^{(k)}$ (comprimento do passo)
  - (usando, por exemplo, o critério de Armijo)
- definir $x^{(k+1)} \leftarrow x^{(k)} + \alpha^{(k)} d^{(k)}$

até (CP = verdadeiro)

Solução: $\begin{cases} x^* \approx x^{(k+1)} \\ f^* \approx f\left(x^{(k+1)}\right) \end{cases}$

Critério de paragem (CP)

Parar o processo iterativo se

$$\underbrace{\|\nabla f(x^{(k+1)})\|_2}_{\text{medida de estacionaridade}} \leq \varepsilon$$

ε constante positiva próxima de zero.

Método de segurança de Newton

É possível ultrapassar as limitações do método de Newton, implementando, em cada iteração, as seguintes **sugestões** - garantem que a direção calculada é **descendente** para a função $f \quad \Rightarrow \quad$ algoritmo de segurança de Newton.

Em qualquer iteração k:

- Quando $\nabla^2 f(x^{(k)})$ é singular
  $\Rightarrow$ usar $d_{SN}^{(k)} = -\nabla f\left(x^{(k)}\right)$
  (em que $-\nabla f\left(x^{(k)}\right)$ é a direção de descida máxima e é descendente para f)

Método de segurança de Newton

- Quando $d_N^{(k)}$ é ortogonal ao gradiente
  $\Leftrightarrow \left| \nabla f(x^{(k)})^T d_N^{(k)} \right| \leq \eta$
  (com $\eta > 0 \, (\approx 0)$)
  $\Rightarrow$ usar $d_{SN}^{(k)} = -\nabla f \left(x^{(k)} \right)$
- Quando $d_N^{(k)}$ é ascendente
  $\Leftrightarrow \nabla f(x^{(k)})^T d_N^{(k)} > \eta$
  $\Rightarrow$ usar $d_{SN}^{(k)} = -d_N^{(k)}$
- Quando $d_N^{(k)}$ é descendente
  $\Rightarrow$ usar $d_{SN}^{(k)} = d_N^{(k)}$

NOTA: $d_{SN}^{(k)}$ é descendente, para todo o k.

Algoritmo para o cálculo da direção de segurança de Newton

Dados $x^{(k)}$ e η,

Resolver o sistema linear Newton $\nabla^2 f\left(x^{(k)}\right) d_N^{(k)} = -\nabla f\left(x^{(k)}\right)$ por EGPP

<u>se</u> (o sistema linear tem solução única - $\exists d_N^{(k)}$) <u>então</u>

    <u>se</u> $\left|\nabla f\left(x^{(k)}\right)^T d_N^{(k)}\right| \leq \eta$

    <u>então</u> $d_{SN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$

    <u>senão</u>

        <u>se</u> $\nabla f\left(x^{(k)}\right)^T d_N^{(k)} > \eta$

        <u>então</u> $d_{SN}^{(k)} \leftarrow -d_N^{(k)}$

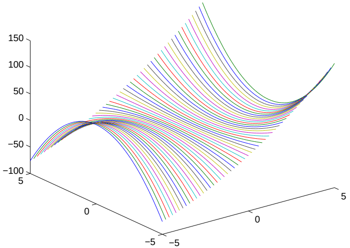
        <u>senão</u> $d_{SN}^{(k)} \leftarrow d_N^{(k)}$

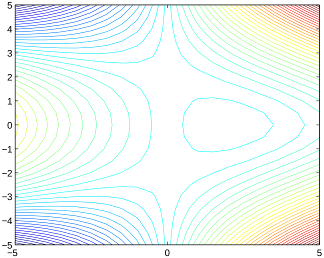
<u>senão</u>

    $d_{SN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$

Exemplo

Dada a função $f(x_1, x_2) = x_1 x_2^2 + (2 - x_1)^2$ calcule o seu mínimo usando o algoritmo de segurança de Newton. A partir de $(1, 1)$, considere $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 0.1$, $\eta = 0.0001$ e $\mu = 0.001$.





A equação para o cálculo da direção Newton - relembrar

método de Newton

$$\nabla^2 f(x^{(k)})\, d_N^{(k)} = -\nabla f(x^{(k)})$$

Para evitar o cálculo das 2^{as} derivadas para a Hessiana, $\nabla^2 f(x)$, pode usar-se uma aproximação:

$$B^{(k)} \approx \nabla^2 f(x^{(k)})$$

ou, como o sistema Newton pode ser escrito da seguinte forma - não aconselhável na prática,

$$d_N^{(k)} = -\left(\nabla^2 f(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$$

A equação para o cálculo da direção quasi-Newton

é aconselhável usar-se, em cada iteração k, uma aproximação à **inversa da Hessiana**

$$H^{(k)} \approx \left(\nabla^2 f(x^{(k)})\right)^{-1}$$

método quasi-Newton

$$d_{QN}^{(k)} = -H^{(k)}\, \nabla f(x^{(k)})$$

$d_{QN}^{(k)}$ - é a direção quasi-Newton

Método quasi-Newton

Evita-se desta forma:

- o cálculo das segundas derivadas (para formar a matriz Hessiana);
- a resolução de um sistema linear em cada iteração, substituindo-o pelo <u>produto de uma matriz por um vetor</u>.

As equações que descrevem o método quasi-Newton são:

$$\begin{cases} d_{QN}^{(k)} = -H^{(k)} \nabla f(x^{(k)}), \\ \\ x^{(k+1)} = x^{(k)} + \alpha^{(k)} d_{QN}^{(k)} \quad \text{para } k = 1, 2, ... \end{cases}$$

Características da matriz H:

- deve aproximar, o melhor possível, a inversa de $\nabla^2 f\left(x^{(k)}\right)$, ou seja, deve verificar a **condição secante**:

$$\boxed{H^{(k)} y^{(k-1)} = s^{(k-1)}}$$

com

$$y^{(k-1)} = \nabla f\left(x^{(k)}\right) - \nabla f\left(x^{(k-1)}\right)$$

(variação verificada no gradiente da iteração $k-1$ para a iteração k)

$$s^{(k-1)} = x^{(k)} - x^{(k-1)} = \alpha^{(k-1)} d_{QN}^{(k-1)}$$

(variação verificada em x)

Características da matriz H:

- deve preferencialmente ser

$$\begin{cases} \textbf{simétrica} & \text{(pois } \nabla^2 f\left(x^{(k)}\right)^{-1} \text{ também é simétrica)} \\ \textbf{definida positiva} & \text{(pois a direção } d_{QN}^{(k)} = -H^{(k)} \nabla f(x^{(k)}) \end{cases}$$

é **descendente** para f em $x^{(k)}$).

As matrizes $H^{(k)}$ são geradas através de **fórmulas de atualização** do tipo

$$H^{(k)} = H^{(k-1)} + E^{(k-1)},\ k > 1$$

e devem manter-se **simétricas** e **definidas positivas**.

Matriz H da 1^a iteração

- Para que simetria + definida positiva se conservem ao longo do processo iterativo, a matriz inicial $H^{(1)}$, para $k = 1$, deve ser também simétrica e definida positiva. Por exemplo
$$H^{(1)} = I.$$
A I não é necessariamente uma boa aproximação a $\nabla^2 f(x^{(1)})^{-1}$, mas as fórmulas de atualização de H melhoram as aproximações.
- Existem várias fórmulas de atualização para as matrizes H:
  - nem todas conservam a simetria + definida positiva;
  - as 2 fórmulas seguintes conservam simetria + definida positiva.

Fórmulas de atualização

Davidon, Fletcher e Powell - **DFP**

$$H^{(k)} = H^{(k-1)} - \frac{H^{(k-1)} y^{(k-1)} {y^{(k-1)}}^T H^{(k-1)}}{{y^{(k-1)}}^T H^{(k-1)} y^{(k-1)}} + \frac{s^{(k-1)} {s^{(k-1)}}^T}{{s^{(k-1)}}^T y^{(k-1)}}$$

Broyden, Fletcher, Goldfarb e Shanno - **BFGS**

$$H^{(k)} = \left(I - \frac{s^{(k-1)} {y^{(k-1)}}^T}{{s^{(k-1)}}^T y^{(k-1)}} \right) H^{(k-1)} \left(I - \frac{y^{(k-1)} {s^{(k-1)}}^T}{{s^{(k-1)}}^T y^{(k-1)}} \right) + \frac{s^{(k-1)} {s^{(k-1)}}^T}{{s^{(k-1)}}^T y^{(k-1)}}$$

Nota: ${y^{(k-1)}}^T s^{(k-1)} > 0$ é a condição necessária e suficiente para que as matrizes se conservem definidas positivas.

Propriedades do método quasi-Newton

- o método quasi-Newton tem convergência
  - **local** (a convergência para a solução só é garantida se a aproximação inicial $x^{(1)}$ estiver na vizinhança da solução);
  - **superlinear**
    verifica-se
    $$\|x^{(k+1)} - x^*\| \leq \gamma_k \|x^{(k)} - x^*\|$$
    com a sucessão $\{\gamma_k\} \to 0$ quando $k \to \infty$;
- o método quasi-Newton satisfaz a propriedade da **terminação quadrática** – isto é, o mínimo de uma função quadrática $q(x)\,,\; x \in \mathbb{R}^n$, obtém-se em $n,$ ou menos, do que n iterações.

Limitação do método quasi-Newton

★ Os erros de arredondamento que se cometem nos cálculos podem fazer com que $H^{(k)}$ deixe de ser definida positiva e a direção $d_{QN}^{(k)}$ deixa de ser descendente para f em $x^{(k)}$

$$\Downarrow$$

Solução: fazer $\boxed{H^{(k)} = I}$ (neste caso $H^{(k)}$ é simétrica e definida positiva)

$$\Downarrow$$

$$d_{QN}^{(k)} = -\nabla f(x^{(k)})$$

Algoritmo para o cálculo da direção quasi-Newton

Dado $x^{(k)}$
Calcular $d_{QN}^{(k)} \leftarrow -H^{(k)} \nabla f\left(x^{(k)}\right)$, sendo $H^{(k)}$ dada por:
se $k = 1$, então $H^{(k)} \leftarrow I$

senão $\begin{cases} s^{(k-1)} \leftarrow x^{(k)} - x^{(k-1)} \\ y^{(k-1)} \leftarrow \nabla f\left(x^{(k)}\right) - \nabla f\left(x^{(k-1)}\right) \\ \text{actualizar } H^{(k)} \text{ pela fórmula DFP ou BFGS} \end{cases}$

se a direção não é descendente

$\left(\nabla f\left(x^{(k)}\right)^T d_{QN}^{(k)} \geq 0\right)$

então fazer $d_{QN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$