Um método inexato de decomposição para minimizar a soma de duas funções convexas

Majela Pentón

UFBA

Novembro 2019

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

- f **função objetivo**.

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{array}{l} \min f(x) \\ \text{s.a. } x \in \mathbb{R}^n \end{array}$$

- f **função objetivo**.
- O problema (P) é chamado de problema de **minimização irrestrita**.

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{array}{l} \min f(x) \\ \text{s.a. } x \in \mathbb{R}^n \end{array}$$

- f **função objetivo**.
- O problema (P) é chamado de problema de **minimização irrestrita**.

- x^* **mínimo global**: $f(x^*) \leq f(x)$ para todo $x \in \mathbb{R}^n$;

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{array}{l} \min f(x) \\ \text{s.a. } x \in \mathbb{R}^n \end{array}$$

- f **função objetivo**.
- O problema (P) é chamado de problema de **minimização irrestrita**.

- x^* **mínimo global**: $f(x^*) \leq f(x)$ para todo $x \in \mathbb{R}^n$;

  $f^* = f(x^*)$ **valor ótimo** de f.

Problema geral

Seja $f : \mathbb{R}^n \to \mathbb{R}$

$$(P) \qquad \begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

- f **função objetivo**.
- O problema (P) é chamado de problema de **minimização irrestrita**.

- x^* **mínimo global**: $f(x^*) \leq f(x)$ para todo $x \in \mathbb{R}^n$;

  $f^* = f(x^*)$ **valor ótimo** de f.
- x^* **mínimo local**: se existe V vizinhança de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in V$.

Teorema 1 (Condição necessária de primeira ordem)

Se x^* é um mínimo local de f em $\mathbb{R}^n$ e f é uma função diferenciável, então

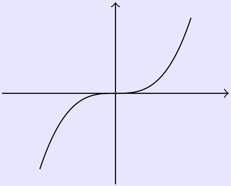
$$\nabla f(x^*) = 0.$$

Teorema 1 (Condição necessária de primeira ordem)

Se x^* é um mínimo local de f em $\mathbb{R}^n$ e f é uma função diferenciável, então

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

Não é suficiente



$f(x) = x^3$

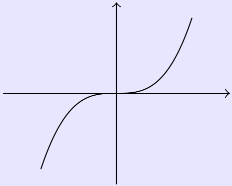
$f'(0) = 0$

Teorema 1 (Condição necessária de primeira ordem)

Se x^* é um mínimo local de f em $\mathbb{R}^n$ e f é uma função diferenciável, então

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

Não é suficiente



$f(x) = x^3$

$f'(0) = 0$

Definição 1

Todo x tal que $\nabla f(x) = 0$ é chamado de **ponto estacionário** *do problema.*

Teorema 2 (Condição necessária de segunda ordem)

Se f é uma função duas vezes diferenciável e x^* é um mínimo local do problema, então x^* é um ponto estacionário ($\nabla f(x^*) = 0$) e a matriz Hessiana de f no ponto x^* é semidefinida positiva ($Hf(x^*) \succeq 0$).

Teorema 2 (Condição necessária de segunda ordem)

Se f é uma função duas vezes diferenciável e x^* é um mínimo local do problema, então x^* é um ponto estacionário ($\nabla f(x^*) = 0$) e a matriz Hessiana de f no ponto x^* é semidefinida positiva ($Hf(x^*) \succeq 0$).

Teorema 3 (Condição suficiente de segunda ordem)

Se f é uma função duas vezes diferenciável e x^* é um ponto estacionário ($\nabla f(x^*) = 0$) tal que a Hessiana de f em x^* é definida positiva ($Hf(x^*) \succ 0$), então x^* é um mínimo local de f.

Método do gradiente

Minimização irrestrita diferenciável:

$$\begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

- f é diferenciável no $\mathbb{R}^n$.

Método do gradiente

Minimização irrestrita diferenciável:

$$\begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

- f é diferenciável no $\mathbb{R}^n$.

Método do gradiente: escolhe um ponto inicial $x_0 \in \mathbb{R}^n$ e itera:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad k = 0, 1, 2, \ldots$$

Método do gradiente

Minimização irrestrita diferenciável:

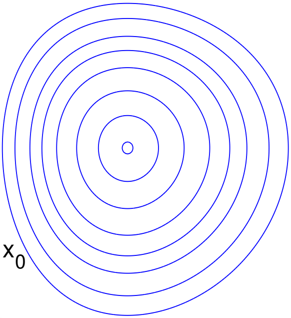
$$\min f(x)$$
$$\text{s.a. } x \in \mathbb{R}^n$$

- f é diferenciável no $\mathbb{R}^n$.

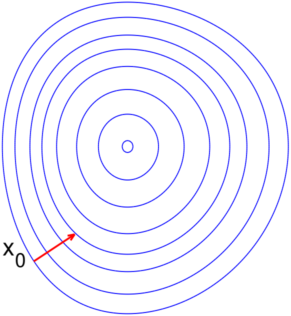
Método do gradiente: escolhe um ponto inicial $x_0 \in \mathbb{R}^n$ e itera:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad k = 0, 1, 2, \ldots$$

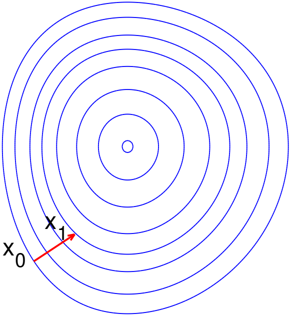
- $\alpha_k > 0$ é o **comprimento de passo**.



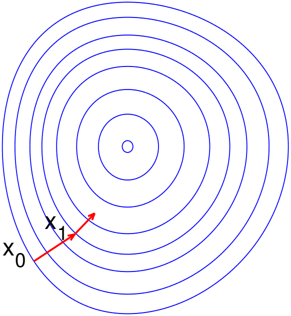
x_0



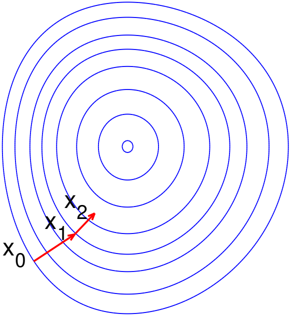
$-\nabla f(x_0)$



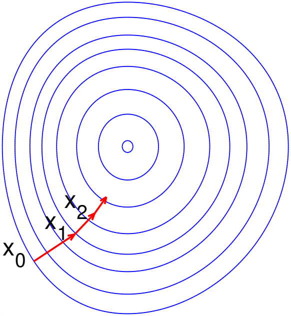
$$x_1 = x_0 - \alpha_0 \nabla f(x_0)$$



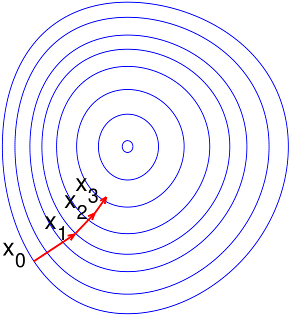
$-\nabla f(x_1)$



$$x_2 = x_1 - \alpha_1 \nabla f(x_1)$$



$$-\nabla f(x_2)$$



$$x_3 = x_2 - \alpha_2 \nabla f(x_2)$$

O método do gradiente é um método de descida: **o passo** α_k é escolhido tal que $f(x_{k+1}) \leq f(x_k)$

O método do gradiente é um método de descida: **o passo** α_k é escolhido tal que $f(x_{k+1}) \leq f(x_k)$

- $\alpha_k = \alpha$ constante para todo k.

O método do gradiente é um método de descida: **o passo** α_k é escolhido tal que $f(x_{k+1}) \leq f(x_k)$

- $\alpha_k = \alpha$ constante para todo k.
- busca linear de Armijo: com $\sigma \in (0, 1)$

$$f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k) - \sigma \alpha_k \left\| \nabla f(x_k) \right\|^2 .$$

O método do gradiente é um método de descida: **o passo** α_k é escolhido tal que $f(x_{k+1}) \leq f(x_k)$

- $\alpha_k = \alpha$ constante para todo k.
- busca linear de Armijo: com $\sigma \in (0,1)$

$$f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k) - \sigma \alpha_k \left\| \nabla f(x_k) \right\|^2 .$$

- busca linear exata: $\alpha_k = \arg\min_{\alpha} f(x_k - \alpha \nabla f(x_k))$.

O método do gradiente é um método de descida: **o passo** α_k é escolhido tal que $f(x_{k+1}) \leq f(x_k)$

- $\alpha_k = \alpha$ constante para todo k.
- busca linear de Armijo: com $\sigma \in (0, 1)$

$$f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k) - \sigma \alpha_k \|\nabla f(x_k)\|^2 .$$

- busca linear exata: $\alpha_k = \arg\min_{\alpha} f(x_k - \alpha \nabla f(x_k))$.

Convergência a um ponto estacionário

Sob hipóteses adequadas, temos que todo **ponto de acumulação** da sequência (x_k), gerada pelo método do gradiente, é um **ponto estacionário do problema**.

Em geral, em otimização (não linear), o nosso objetivo é bastante moderado: encontrar apenas um mínimo local do problema. Mesmo assim, esse objetivo pode não ser alcançado com o método de gradiente.

Em geral, em otimização (não linear), o nosso objetivo é bastante moderado: encontrar apenas um mínimo local do problema. Mesmo assim, esse objetivo pode não ser alcançado com o método de gradiente.

Exemplo 1

$$f(y, z) = \frac{1}{2}y^2 + \frac{1}{4}z^4 - \frac{1}{2}z^2$$

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0, 0)$, $(0, 1)$, $(0, -1)$.

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0, 0)$, $(0, 1)$, $(0, -1)$.
- Mínimos locais: $(0, 1)$ e $(0, -1)$.

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0,0)$, $(0,1)$, $(0,-1)$.
- Mínimos locais: $(0,1)$ e $(0,-1)$.
- O ponto $(0,0)$ não é mínimo local nem máximo local

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0,0)$, $(0,1)$, $(0,-1)$.
- Mínimos locais: $(0,1)$ e $(0,-1)$.
- O ponto $(0,0)$ não é mínimo local nem máximo local

Ponto inicial: $x_0 = (1,0)$

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0, 0)$, $(0, 1)$, $(0, -1)$.
- Mínimos locais: $(0, 1)$ e $(0, -1)$.
- O ponto $(0, 0)$ não é mínimo local nem máximo local

Ponto inicial: $x_0 = (1, 0)$

$$\Rightarrow \quad \nabla f(x_0) = (1, 0) \quad \Rightarrow x_1 = (1, 0) - \alpha_k(1, 0) \quad \Rightarrow (x_1)_2 = 0.$$

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0, 0)$, $(0, 1)$, $(0, -1)$.
- Mínimos locais: $(0, 1)$ e $(0, -1)$.
- O ponto $(0, 0)$ não é mínimo local nem máximo local

Ponto inicial: $x_0 = (1, 0)$

$$\Rightarrow \quad \nabla f(x_0) = (1, 0) \quad \Rightarrow x_1 = (1, 0) - \alpha_k(1, 0) \quad \Rightarrow (x_1)_2 = 0.$$

Então $(x_k)_2 = 0$ para todo $k = 1, 2, \ldots$

$$x = (y, z), \quad \nabla f(x) = \left(y, z^3 - z\right), \quad Hf(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3z^2 - 1 \end{pmatrix}$$

- Pontos estacionários: $(0,0)$, $(0,1)$, $(0,-1)$.
- Mínimos locais: $(0,1)$ e $(0,-1)$.
- O ponto $(0,0)$ não é mínimo local nem máximo local

Ponto inicial: $x_0 = (1,0)$

$$\Rightarrow \quad \nabla f(x_0) = (1,0) \quad \Rightarrow x_1 = (1,0) - \alpha_k(1,0) \quad \Rightarrow (x_1)_2 = 0.$$

Então $(x_k)_2 = 0$ para todo $k = 1, 2, \ldots$
O método converge para o ponto $(0,0)$ que **não é mínimo local**.

Só podemos garantir convergência a um **ponto estacionário** do problema.

Funções convexas diferenciáveis

- A condição necessária de otimalidade de primeira ordem **é fraca**.

Funções convexas diferenciáveis

- A condição necessária de otimalidade de primeira ordem **é fraca**.
- Podemos tentar restringir a análise a um conjunto de funções diferenciáveis com propriedades *'boas'*.

Funções convexas diferenciáveis

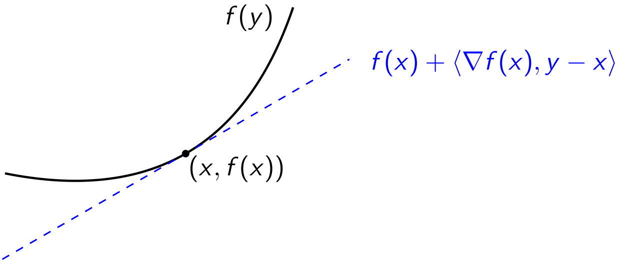
- A condição necessária de otimalidade de primeira ordem **é fraca**.
- Podemos tentar restringir a análise a um conjunto de funções diferenciáveis com propriedades *'boas'*.

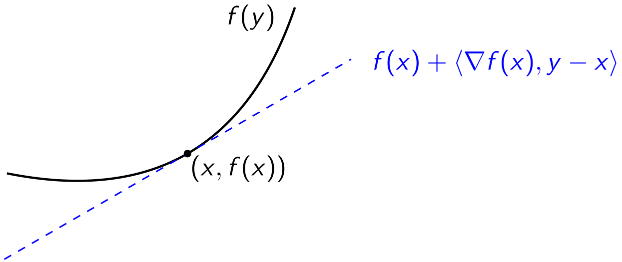
Definição 2 (Função convexa)

Uma função $f : \mathbb{R}^n \to \mathbb{R}$ *diferenciável é dita convexa no* $\mathbb{R}^n$ *se*

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

para todo $x, y \in \mathbb{R}^n$


$f(y)$
$f(x) + \langle \nabla f(x), y - x \rangle$
$(x, f(x))$



Equivalente a definição geral de funções convexas

Definição 3

$f : \mathbb{R}^n \to \mathbb{R}$ é convexa se

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

para todo $\theta \in [0,1]$ e $x, y \in \mathbb{R}^n$.

Condição suficiente de otimalidade

Se f é uma função convexa, diferenciável, e $x^* \in \mathbb{R}^n$ é um ponto estacionário

Condição suficiente de otimalidade

Se f é uma função convexa, diferenciável, e $x^* \in \mathbb{R}^n$ é um ponto estacionário

$$\Rightarrow f(x) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle = f(x^*) \quad \forall x \in \mathbb{R}^n$$

Condição suficiente de otimalidade

Se f é uma função convexa, diferenciável, e $x^* \in \mathbb{R}^n$ é um ponto estacionário

$$\Rightarrow f(x) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle = f(x^*) \quad \forall x \in \mathbb{R}^n$$

então x^* **é mínimo global**.

Condição suficiente de otimalidade

Se f é uma função convexa, diferenciável, e $x^* \in \mathbb{R}^n$ é um ponto estacionário

$$\Rightarrow f(x) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle = f(x^*) \quad \forall x \in \mathbb{R}^n$$

então x^* **é mínimo global**.

Para funções convexas diferenciáveis a condição necessária de primeira ordem também é suficiente

Método do gradiente para funções convexas

Problema de minimização convexa

Seja $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável e convexa em $\mathbb{R}^n$

$$\begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

Método do gradiente para funções convexas

Problema de minimização convexa

Seja $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável e convexa em $\mathbb{R}^n$

$$\begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

Convergência

- Se (x_k) é gerada pelo método do gradiente, então a sequência toda converge a uma solução global.

Método do gradiente para funções convexas

Se ∇f é L-**Lipschitz contínuo**:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \mathbb{R}^n$$

Método do gradiente para funções convexas

Se ∇f é L-**Lipschitz contínuo**:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \mathbb{R}^n$$

Método do gradiente: $\alpha = \frac{1}{L}$

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \quad k = 0, 1, 2, \ldots$$

Método do gradiente para funções convexas

Se ∇f é L-**Lipschitz contínuo**:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \mathbb{R}^n$$

Método do gradiente: $\alpha = \frac{1}{L}$

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \quad k = 0, 1, 2, \ldots$$

- $f(x_k) - f^* \leq C/k$

Método do gradiente para funções convexas

Se ∇f é L-**Lipschitz contínuo**:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \mathbb{R}^n$$

Método do gradiente: $\alpha = \frac{1}{L}$

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \quad k = 0, 1, 2, \ldots$$

- $f(x_k) - f^* \leq C/k$
- Dizemos que o método do gradiente tem **taxa de convergência** $\mathcal{O}(1/k)$. Ou seja, o $\#$ de iterações para obter uma ϵ-**solução** (no sentido de $f(x_k) - f^* \leq \epsilon$) é $\mathcal{O}(1/\epsilon)$.

Método do gradiente para funções convexas

Se ∇f é L-**Lipschitz contínuo**:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \mathbb{R}^n$$

Método do gradiente: $\alpha = \frac{1}{L}$

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \quad k = 0, 1, 2, \ldots$$

- $f(x_k) - f^* \leq C/k$
- Dizemos que o método do gradiente tem **taxa de convergência** $\mathcal{O}(1/k)$. Ou seja, o # de iterações para obter uma ϵ-**solução** (no sentido de $f(x_k) - f^* \leq \epsilon$) é $\mathcal{O}(1/\epsilon)$.

Prós: Ideia simples, fácil implementação, usualmente as iterações são baratas.
Contras: Não pode ser aplicado a funções não diferenciáveis.

Funções convexas não diferenciáveis

Subgradiente de uma função convexa

Se $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa, dizemos que $v \in \mathbb{R}^n$ é um **subgradiente** de f no ponto $x \in \mathbb{R}^n$ se

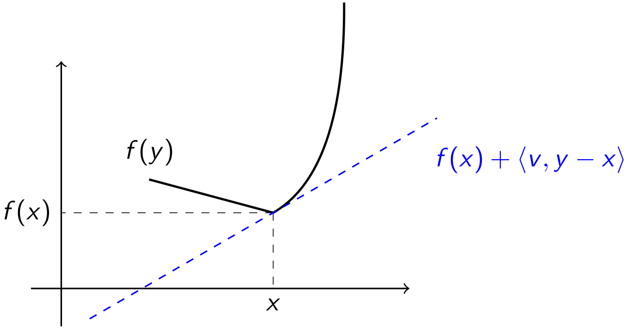
$$f(y) \geq f(x) + \langle v, y - x \rangle, \qquad \forall y \in \mathbb{R}^n.$$

Funções convexas não diferenciáveis

Subgradiente de uma função convexa

Se $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa, dizemos que $v \in \mathbb{R}^n$ é um **subgradiente** de f no ponto $x \in \mathbb{R}^n$ se

$$f(y) \geq f(x) + \langle v, y - x \rangle, \qquad \forall y \in \mathbb{R}^n.$$



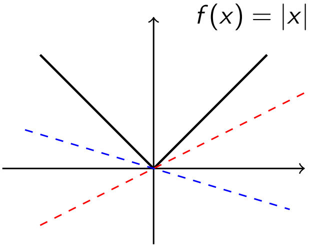
- O subgradiente sempre existe
- Se f é diferenciável em x, então $v = \nabla f(x)$ é o único subgradiente de f em x.

- O subgradiente sempre existe
- Se f é diferenciável em x, então $v = \nabla f(x)$ é o único subgradiente de f em x.

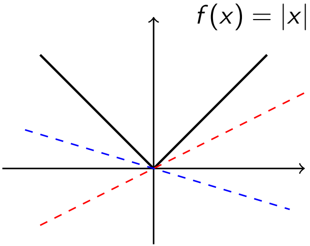
Subdiferencial

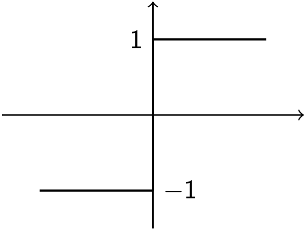
O conjunto de todos os subgradientes de f em x se chama o **subdiferencial** de f em x, e é denotado por $\partial f(x)$.

Exemplo:



Exemplo:





$$\partial f(x) = \begin{cases} 1 & \text{se } \; x > 0 \\ [-1, 1] & \text{se } \; x = 0 \\ -1 & \text{se } \; x < 0 \end{cases}$$

Teorema 4 (Condição de otimalidade)

Se f é uma função convexa, então x^* é um mínimo de f se e somente se

$$0 \in \partial f(x^*).$$

Teorema 4 (Condição de otimalidade)

Se f é uma função convexa, então x^* é um mínimo de f se e somente se

$$0 \in \partial f(x^*).$$

Segue da definição que se $v = 0$ é subgradiente:

$$f(y) \geq f(x^*) + \langle 0, y - x^* \rangle = f(x^*)$$

pegando $v = 0 \in \partial f(x^*)$.

Teorema 4 (Condição de otimalidade)

Se f é uma função convexa, então x^* é um mínimo de f se e somente se

$$0 \in \partial f(x^*).$$

Segue da definição que se $v = 0$ é subgradiente:

$$f(y) \geq f(x^*) + \langle 0, y - x^* \rangle = f(x^*)$$

pegando $v = 0 \in \partial f(x^*)$.

Observe a analogia com o caso diferenciável, onde

$$\partial f(x) = \{\nabla f(x)\}$$

Método do subgradiente

Minimização irrestrita convexa

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

Método do subgradiente

Minimização irrestrita convexa

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

- $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa não diferenciável.

Método do subgradiente

Minimização irrestrita convexa

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

- $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa não diferenciável.

Método do subgradiente: similar ao método do gradiente, mas substituindo gradientes por subgradientes.

Método do subgradiente

Minimização irrestrita convexa

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

- $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa não diferenciável.

Método do subgradiente: similar ao método do gradiente, mas substituindo gradientes por subgradientes.
Escolha um ponto inicial $x_0 \in \mathbb{R}^n$ e itera

$$x_{k+1} = x_k - \alpha_k v_k, \quad k = 0, 1, 2, \ldots,$$

Método do subgradiente

Minimização irrestrita convexa

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

- $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa não diferenciável.

Método do subgradiente: similar ao método do gradiente, mas substituindo gradientes por subgradientes.
Escolha um ponto inicial $x_0 \in \mathbb{R}^n$ e itera

$$x_{k+1} = x_k - \alpha_k v_k, \quad k = 0, 1, 2, \ldots,$$

- $v_k \in \partial f(x_k)$;
- $\alpha_k > 0$.

Observação:

- O método do subgradiente não é necessariamente um método de decida, logo guardamos em cada iteração o melhor iterado

$$f(x_k^{best}) = \min_{i=1,\dots,k} f(x_k).$$

Observação:

- O método do subgradiente não é necessariamente um método de decida, logo guardamos em cada iteração o melhor iterado

$$f(x_k^{best}) = \min_{i=1,\ldots,k} f(x_k).$$

Escolhas do tamanho do passo

- Constante: $\alpha_k = \alpha$ para todo $k \in \mathbb{N}$.

Observação:

- O método do subgradiente não é necessariamente um método de decida, logo guardamos em cada iteração o melhor iterado

$$f(x_k^{best}) = \min_{i=1,\dots,k} f(x_k).$$

Escolhas do tamanho do passo

- Constante: $\alpha_k = \alpha$ para todo $k \in \mathbb{N}$.
- Quadrado somável, mas não somável:

$$\alpha_k > 0, \qquad \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Observação:

- O método do subgradiente não é necessariamente um método de decida, logo guardamos em cada iteração o melhor iterado

$$f(x_k^{best}) = \min_{i=1,\ldots,k} f(x_k).$$

Escolhas do tamanho do passo

- Constante: $\alpha_k = \alpha$ para todo $k \in \mathbb{N}$.
- Quadrado somável, mas não somável:

$$\alpha_k > 0, \qquad \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Diferença com o método do gradiente: as escolhas do passo são prefixadas, não calculadas em cada iteração.

Resultados de convergência

- Passo constante:
  Se $f^\star = -\infty$, então $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se $f^\star > -\infty$, então $\liminf_{k\to\infty} f(x_k) \leq f^\star + \alpha c^2/2$.

Resultados de convergência

- Passo constante:
  Se $f^\star = -\infty$, então $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se $f^\star > -\infty$, então $\liminf_{k\to\infty} f(x_k) \leq f^\star + \alpha c^2/2$.
- Quadrado somável, mas não somável: $\liminf_{k\to\infty} f(x_k) = f^\star$.

Resultados de convergência

- Passo constante:
  Se $f^\star = -\infty$, então $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se $f^\star > -\infty$, então $\liminf_{k\to\infty} f(x_k) \leq f^\star + \alpha c^2/2$.
- Quadrado somável, mas não somável: $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se existe solução do problema então $x_k \to x^\star$, onde $x^\star$ é uma solução.

Resultados de convergência

- Passo constante:
  Se $f^\star = -\infty$, então $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se $f^\star > -\infty$, então $\liminf_{k\to\infty} f(x_k) \leq f^\star + \alpha c^2/2$.
- Quadrado somável, mas não somável: $\liminf_{k\to\infty} f(x_k) = f^\star$.
  Se existe solução do problema então $x_k \to x^\star$, onde $x^\star$ é uma solução.

Taxa de convergência

O método do subgradiente tem taxa de convergência $\mathcal{O}(1/\sqrt{k})$. Ou seja, para obter $f(x_k^{best}) - f^* \leq \epsilon$ é necessário $\mathcal{O}(1/\epsilon^2)$ iterações.

Prós: Ampla aplicação: se podemos computar subgradientes, então podemos minimizar (quase) qualquer função convexa. Fácil implementação.
Contras: A taxa de convergência $\mathcal{O}(1/\sqrt{k})$ é muito lenta.

Prós: Ampla aplicação: se podemos computar subgradientes, então podemos minimizar (quase) qualquer função convexa. Fácil implementação.
Contras: A taxa de convergência $\mathcal{O}(1/\sqrt{k})$ é muito lenta.

Método do gradiente: para obter $f(x_k) - f^* \leq 1/100 \Rightarrow$ 100 iterações.
Método do subgradiente: para obter $f(x_k^{best}) - f^* \leq 1/100 \Rightarrow$ 10000 iterações.

Prós: Ampla aplicação: se podemos computar subgradientes, então podemos minimizar (quase) qualquer função convexa. Fácil implementação.
Contras: A taxa de convergência $\mathcal{O}(1/\sqrt{k})$ é muito lenta.

Método do gradiente: para obter $f(x_k) - f^* \leq 1/100 \Rightarrow$ 100 iterações.
Método do subgradiente: para obter $f(x_k^{best}) - f^* \leq 1/100 \Rightarrow$ 10000 iterações.

Pode-se melhorar?

Prós: Ampla aplicação: se podemos computar subgradientes, então podemos minimizar (quase) qualquer função convexa. Fácil implementação.
Contras: A taxa de convergência $\mathcal{O}(1/\sqrt{k})$ é muito lenta.

Método do gradiente: para obter $f(x_k) - f^* \leq 1/100 \Rightarrow 100$ iterações.
Método do subgradiente: para obter $f(x_k^{best}) - f^* \leq 1/100 \Rightarrow$ 10000 iterações.

Pode-se melhorar? Em geral não!
Mas considerando funções da forma

$$f(x) = h(x) + g(x)$$

onde h é convexa e diferenciável e g é convexa.
Para muitos problemas a taxa de convergência $\mathcal{O}(1/k)$ do método do gradiente pode ser recuperada com um algoritmo simples.

Minimização da soma de duas funções convexas

Minimizando a soma de duas funções

$$\min_{x \in \mathbb{R}^n} h(x) + g(x)$$

Minimização da soma de duas funções convexas

Minimizando a soma de duas funções

$$\min_{x \in \mathbb{R}^n} h(x) + g(x)$$

- $h, g : \mathbb{R}^n \to \mathbb{R}$ são funções convexas e h é diferenciável.

Minimização da soma de duas funções convexas

Minimizando a soma de duas funções

$$\min_{x \in \mathbb{R}^n} h(x) + g(x)$$

- $h, g : \mathbb{R}^n \to \mathbb{R}$ são funções convexas e h é diferenciável.

Operador proximal

Dado $\alpha > 0$, $\mathbf{prox}_{\alpha g} : \mathbb{R}^n \to \mathbb{R}^n$

$$\mathbf{prox}_{\alpha g}(z) := \arg\min_{x \in \mathbb{R}^n} \alpha g(x) + \frac{1}{2} \|x - z\|^2 .$$

Método do gradiente proximal: escolha $x_0 \in \mathbb{R}^n$ e itera

$$x_{k+1} = \mathbf{prox}_{\alpha_k g}(x_k - \alpha_k \nabla h(x_k)), \quad k = 0, 1, \ldots$$

- $\alpha_k > 0$.

Método do gradiente proximal: escolha $x_0 \in \mathbb{R}^n$ e itera

$$x_{k+1} = \mathbf{prox}_{\alpha_k g}(x_k - \alpha_k \nabla h(x_k)), \quad k = 0, 1, \ldots$$

- $\alpha_k > 0$.

Para $f(x) = h(x) + g(x)$ supondo que

- ∇h é Lipschitz contínuo com $L > 0$
- $\mathbf{prox}_{\alpha_k g}$ pode ser avaliado

e usando passo constante $\alpha_k = 1/L$, o método do gradiente proximal tem taxa de convergência $\mathcal{O}(1/k)$.

Método do gradiente proximal: escolha $x_0 \in \mathbb{R}^n$ e itera

$$x_{k+1} = \textbf{prox}_{\alpha_k g}(x_k - \alpha_k \nabla h(x_k)), \quad k = 0, 1, \ldots$$

- $\alpha_k > 0$.

Para $f(x) = h(x) + g(x)$ supondo que

- ∇h é Lipschitz contínuo com $L > 0$
- $\textbf{prox}_{\alpha_k g}$ pode ser avaliado

e usando passo constante $\alpha_k = 1/L$, o método do gradiente proximal tem taxa de convergência $\mathcal{O}(1/k)$.

É recuperada a taxa de convergência do método do gradiente, mas tem o custo da avaliação do **prox**.

Para muitas funções g, importantes na prática, o $\mathbf{prox}_g$ tem forma fechada, mas

Para muitas funções g, importantes na prática, o $\mathbf{prox}_g$ tem forma fechada, mas e se não podemos avaliar **prox**?

Para muitas funções g, importantes na prática, o $\mathbf{prox}_g$ tem forma fechada, mas e se não podemos avaliar **prox**?

Avaliar $\mathbf{prox}_{\alpha g}(z) = \arg\min_{x \in \mathbb{R}^n} \alpha g(x) + 1/2 \|x - z\|^2$ é equivalente a resolver o **sistema proximal**: encontrar x, $v \in \mathbb{R}^n$ tais que

$$\begin{cases} v \in \partial g(x), \\ \alpha v + x - z = 0. \end{cases} \tag{1}$$

Para muitas funções g, importantes na prática, o $\mathbf{prox}_g$ tem forma fechada, mas e se não podemos avaliar **prox**?

Avaliar $\mathbf{prox}_{\alpha g}(z) = \arg\min_{x\in\mathbb{R}^n} \alpha g(x) + 1/2\,\|x-z\|^2$ é equivalente a resolver o **sistema proximal**: encontrar x, $v \in \mathbb{R}^n$ tais que

$$\begin{cases} v \in \partial g(x), \\ \alpha v + x - z = 0. \end{cases} \tag{1}$$

Soluções aproximadas:

- Dado $r \geq 0$, um ponto (x, v) é uma r-**solução aproximada** de (1) no ponto (α, z) se $v \in \partial g(x)$ e

$$\|\alpha v + x - z\| \leq r.$$

- Dado $\sigma \in [0, 1)$ a tripla (x, v) é uma σ-**solução aproximada** de (1) no ponto (α, z) se $v \in \partial g(x)$

$$\|\alpha v + x - z\| \leq \sigma\,\|x - z\|.$$

Método do gradiente proximal inexato: similar ao método do gradiente proximal, mas avaliando o $\mathbf{prox}_g$ de forma inexata.

Método do gradiente proximal inexato: similar ao método do gradiente proximal, mas avaliando o $\mathbf{prox}_g$ de forma inexata. Escolha $x_0 \in \mathbb{R}^n$ e itera para $k = 0, 1, 2, \ldots$

- $y_k = x_k - \alpha_k \nabla h(x_k)$;
- calcule $(\overline{x}_k, \overline{w}_k)$ solução aproximada de $\mathbf{prox}_g$ em (α_k, y_k);
- $x_{k+1} = y_k - \alpha_k \overline{w}_k$

Método do gradiente proximal inexato: similar ao método do gradiente proximal, mas avaliando o $\mathbf{prox}_g$ de forma inexata. Escolha $x_0 \in \mathbb{R}^n$ e itera para $k = 0, 1, 2, \ldots$

- $y_k = x_k - \alpha_k \nabla h(x_k)$;
- calcule $(\overline{x}_k, \overline{w}_k)$ solução aproximada de $\mathbf{prox}_g$ em (α_k, y_k);
- $x_{k+1} = y_k - \alpha_k \overline{w}_k$

- MGPI-A: $(\overline{x}_k, \overline{w}_k)$ é uma r_k-solução aproximada para todo $k = 1, 2, \ldots$. Neste caso a sequência (r_k) deve ser fixada com antecedência.
- MGPI-R: $(\overline{x}_k, \overline{w}_k)$ é uma σ-solução aproximada para todo $k = 1, 2, \ldots$, com $\sigma \in (0, 1)$ fixo.

O problema de remoção de borrados com variação total

O problema de remoção de borrados com variação total

O problema de remoção de borrados com variação total

Para reconstruir uma imagem $x \in \mathbb{R}^{N \times N}$ a partir de uma imagem borrada e/ou ruidosa $b \in \mathbb{R}^{N \times N}$, que descreve os dados obtidos, resolvemos o problema de minimização

$$\min_{x \in \mathbb{R}^{N \times N}} \frac{1}{2} \|Ax - b\|_F^2 + \tau \sum_{i,j=1}^{N} \|(\nabla x)_{i,j}\|_2 .$$

- $A : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ é um operador linear que gera a imagem borrada;
- $\|\cdot\|_F$ é a norma de Frobenius para matrizes;
- $\tau > 0$ parâmetro de regularização;
- $\nabla : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N}$ é o gradiente discreto

$$(\nabla_1 x)_{ij} = x_{i+1,j} - x_{i,j} \quad \text{e} \quad (\nabla_2 x)_{ij} = x_{i,j+1} - x_{i,j};$$

- $\|\cdot\|_2$ é a norma Euclideana em $\mathbb{R}^2$.

Para resolver o problema consideramos:

- $h(x) = \frac{1}{2}\|Ax - b\|_F^2$;
- $g(x) = \tau \sum_{i,j=1}^{N} \|(\nabla x)_{i,j}\|_2$;
- $\alpha_k = 1/L$ com $L = \|A^* A\|$;
- $x_0 = b$;
- $\tau = 10^{-4}$.

Method	RelDiff	FuncVal	CPU time (s)
MGPI-R			
$\sigma = 0.9$	0.00009950	0.36002654	151
$\sigma = 0.5$	0.00009952	0.35995523	181
$\sigma = 0.1$	0.00009955	0.35987295	306
MGPI-A: $\sqrt{r_k} = 1/k^q$			
$q = 1.1$	0.00009956	0.35986216	336
$q = 1.5$	0.00009957	0.35985049	1039
$q = 1.9$	0.00009957	0.35987493	1423
MGPI-A: $\sqrt{r_k} = C/k^q$			
$q = 1.1$	0.00009956	0.35991933	137
$q = 1.5$	0.00009956	0.35986485	669
$q = 1.9$	0.00009956	0.35984404	1244

Tabela: Resultados numéricos. RelDiff $= \|x_k - x_{k-1}\| \,/\, \|x_k\|$. Critério de parada RelDiff $< 10^{-4}$.

Referências:

[1] Yu. Nesterov. Introductory Lectures on Convex Optimization. Kluwer, Boston (2004).

[2] A. Izmailov, M. Solodov. Otimização, Volume 2: Métodos Computacionais. Rio de Janeiro, Brazil, Segunda Edição (2012).

[3] Millán, R.D. and Machado, M.P. Inexact proximal ϵ-subgradient methods for composite convex optimization problems. J Glob Optim (2019). https://doi.org/10.1007/s10898-019-00808-8