Convergência global de algoritmos de descida: um pouco além de Armijo, ângulo e proporcionalidade

Douglas S. Gonçalves

Department of Applied Mathematics
University of Campinas (UNICAMP)

Campinas, Outubro, 2012

Partially supported by CNPq, FAPESP

O problema

$f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.a} & x \in \Omega = \mathbb{R}^n \end{array}$$

Solução x^*

$$f(x^*) \leq f(x), \quad \forall x \in \Omega.$$

Algoritmos de Descida

$$x_{k+1} = x_k + t_k\, d_k$$

Dado um ponto inicial x^0, espera-se que

$$f(x_{k+1}) < f(x_k), \qquad k = 0, 1, \ldots,$$

O problema

$f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável

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

Solução x^*

$$f(x^*) \leq f(x), \quad \forall x \in \Omega.$$

Algoritmos de Descida

$$x_{k+1} = x_k + t_k\, d_k$$

Dado um ponto inicial x^0, espera-se que

$$f(x_{k+1}) < f(x_k), \quad k = 0, 1, \ldots,$$

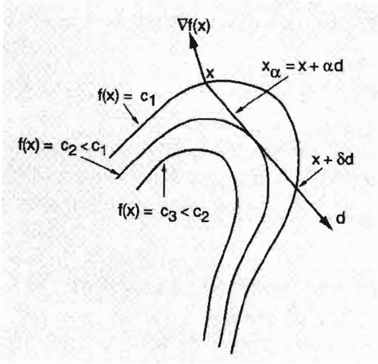
Direções de descida

$$\nabla f(x)^T d < 0$$

Expansão Taylor (Ordem 1)

$$f(x + t\,d) = f(x) + t\nabla f(x)^T d + o(t)$$

$\exists t > 0$ tal que $f(x + td) < f(x)$



Bertsekas, 1999

Métodos de direções de descida

$$x_{k+1} = x_k + t_k\, d_k,$$

onde $\forall k,\quad \nabla f(x_k)^T d_k < 0$ (se $\nabla f(x_k) \neq 0$)

$$f(x_k + t_k\, d_k) < f(x_k), \qquad k = 0, 1, \ldots.$$

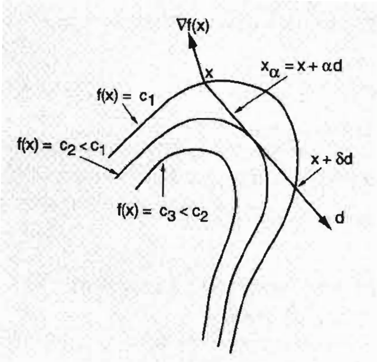
Direções de descida

$$\nabla f(x)^T d < 0$$

Expansão Taylor (Ordem 1)

$$f(x + t\, d) = f(x) + t \nabla f(x)^T d + o(t)$$

$\exists t > 0$ tal que $f(x + td) < f(x)$



Bertsekas, 1999

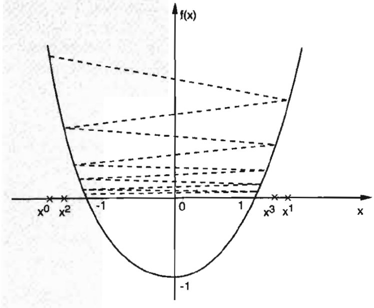
Métodos de direções de descida

$$x_{k+1} = x_k + t_k\, d_k,$$

onde $\forall k, \quad \nabla f(x_k)^T d_k < 0$ (se $\nabla f(x_k) \neq 0$)

$$f(x_k + t_k\, d_k) < f(x_k), \qquad k = 0, 1, \ldots .$$

Decréscimo simples não é suficiente ...



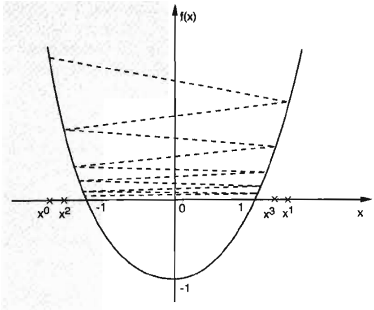
Bertsekas, 1999

Precisamos de um decréscimo substancial!

Condição de Armijo

$$f(x_k + t_k\, d_k) \leq f(x_k) + \sigma\, t\, \nabla f(x_k)^T d_k, \qquad \sigma \in (0, 1)$$

Decréscimo simples não é suficiente ...



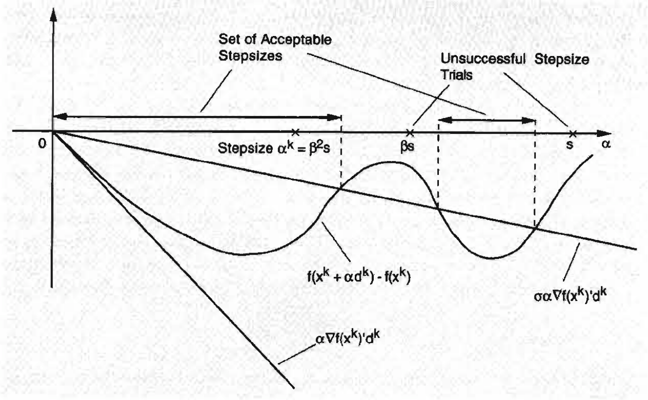
Bertsekas, 1999

Precisamos de um decréscimo substancial!

Condição de Armijo

$$f(x_k + t_k\, d_k) \leq f(x_k) + \sigma\, t\, \nabla f(x_k)^T d_k, \qquad \sigma \in (0,1)$$

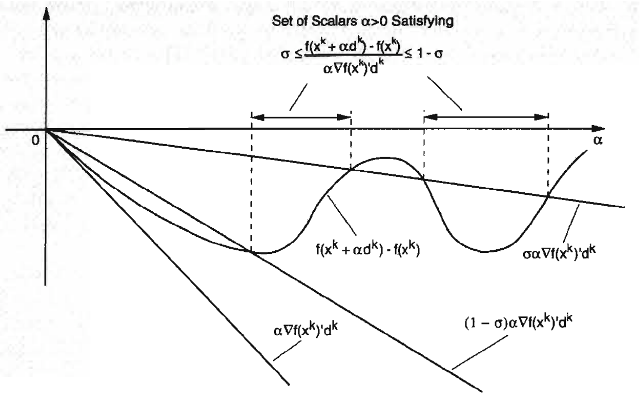
Armijo



Bertsekas, 1999

$$\frac{f(x_k + t_k\, d_k) - f(x_k)}{t_k\, \nabla f(x_k)^T d_k} \geq \sigma$$

Goldstein

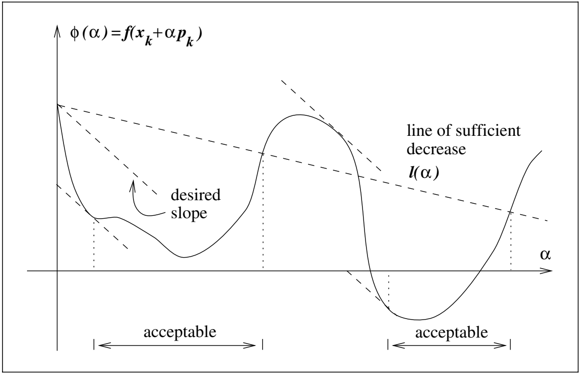


Bertsekas, 1999

$$\sigma \leq \frac{f(x_k + t_k\, d_k) - f(x_k)}{t_k\, \nabla f(x_k)^T d_k} \leq 1 - \sigma, \qquad \sigma \in (0, \frac{1}{2})$$

Wolfe

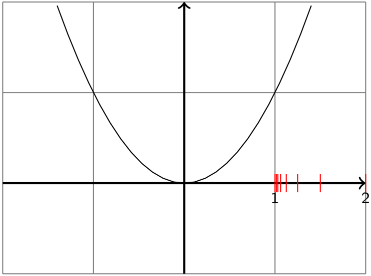
1. $f(x_k + t_k\, d_k) \leq f(x_k) + c_1\, t_k\, \nabla f(x_k)^T d_k,$ $\qquad c_1 \in (0,1)$
2. $\nabla f(x_k + t_k\, d_k)^T d_k \geq c_2 \nabla f(x_k)^T d_k,$ $\qquad c_2 \in (c_1, 1)$



Nocedal & Wright, 1999

Proporcionalidade

Apenas Armijo não basta!



$$x_k = 1 + \frac{1}{2^k}, \quad t_k = 1$$

$$d_k = x_{k+1} - x_k, \quad \sigma = \frac{1}{2}$$

$$\left(1 + \frac{1}{2^{k+1}}\right)^2 < \left(1 + \frac{1}{2^k}\right)^2 - \frac{1}{2} 2 \left(1 + \frac{1}{2^k}\right) \frac{1}{2^{k+1}}$$

- Às vezes passos pequenos são inevitáveis
- Devemos ao menos tentar passos grandes (sobretudo longe da solução)

$$\|d_k\| \geq \beta \|\nabla f(x_k)\|, \quad \beta > 0$$

Ângulo

$$f(x_{k+1}) = f(x_k) + t_k \, \nabla f(x_k)^T d_k + o(t_k)$$

- Se $\nabla f(x_k)^T d_k$ tem magnitude substancial, o progresso pode ser substancial
- Por outro lado, problemas à vista quando

$$\frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \, \|d_k\|} = \cos\theta_k \to 0$$

Exemplo

$$f(x) = \tfrac{1}{2}\|x\|_2^2, \qquad \nabla f(x) = x \qquad\qquad d_k = U_k(-x_k)$$

$$U_k = \begin{bmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{bmatrix}, \quad \theta_k = \tfrac{\pi}{2}\left(1 - \tfrac{1}{k^2+1}\right), \qquad\qquad \beta = 1.$$

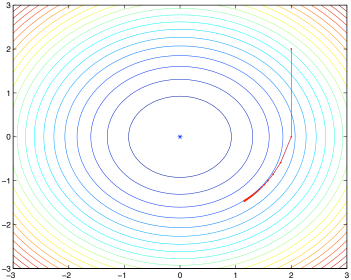
Para $\sigma \in (0, \frac{1}{2})$, podemos mostrar que

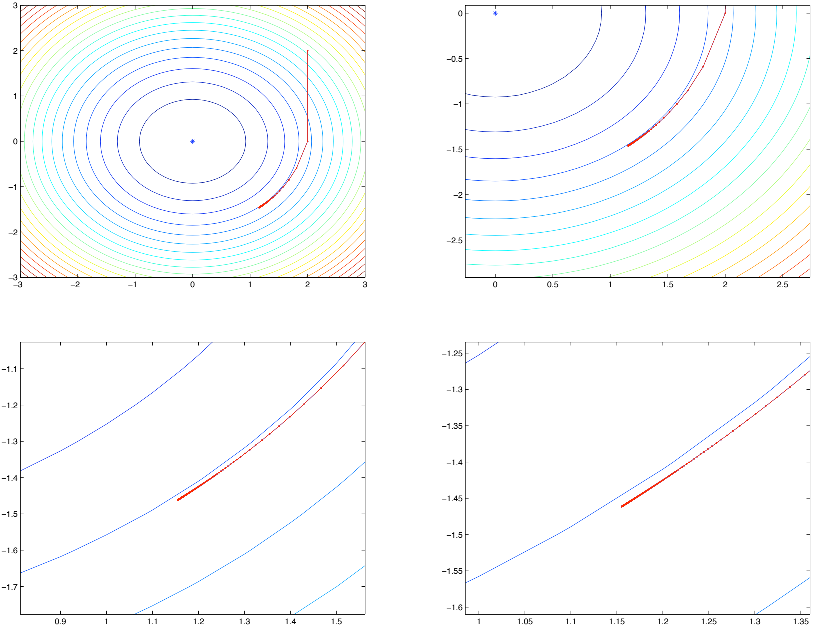
$$\frac{1}{2}\left\|x_k - t\,U_k x_k\right\|^2 \leq \frac{1}{2}\left\|x_k\right\|^2 - \sigma\, t\, x_k^T U_k x_k$$

é satisfeita para

$$0 < t < \frac{x_k^T U_k x_k}{\left\|x_k\right\|^2} = \cos\theta_k$$

Usando $t_k = \frac{x_k^T U_k x_k}{1+\|x_k\|^2}$, $x_0 = (2,2)^T$





3
2
1
0
−1
−2
−3
−3 −2 −1 0 1 2 3
0
−0.5
−1
−1.5
−2
−2.5
0 0.5 1 1.5 2 2.5
−1.1
−1.2
−1.3
−1.4
−1.5
−1.6
−1.7
0.9 1 1.1 1.2 1.3 1.4 1.5
−1.25
−1.3
−1.35
−1.4
−1.45
−1.5
−1.55
−1.6
1 1.05 1.1 1.15 1.2 1.25 1.3 1.35

Zoutendijk

Theorem (Nocedal & Wright, 1999, Zoutendijk, 1976)

Considere uma iteração do tipo $x_{k+1} = x_k + t_k d_k$*, onde* d_k *é uma direção de descida e* t_k *satisfaz as condições de Wolfe. Suponha que* f *é limitada inferiormente e continuamente diferenciável em um aberto* $\mathcal{A}$ *contendo o conjunto de nível* $\left\{x \parallel f(x) \leq f(x^0)\right\}$*, onde* x^0 *é o ponto inicial. Assuma também que o gradiente é Lipschitz em* $\mathcal{A}$*.*

Então

$$\sum_{k \geq 0} \cos^2 \theta_k \left\|\nabla f(x_k)\right\|^2 < \infty.$$

Consequência

- $\cos^2 \theta_k \left\|\nabla f(x_k)\right\|^2 \to 0$
- Se $\cos \theta_k \geq \delta > 0, \quad \forall k$ então $\|\nabla f(x_k)\| \to 0.$

Zoutendijk

Theorem (Nocedal & Wright, 1999, Zoutendijk, 1976)

Considere uma iteração do tipo $x_{k+1} = x_k + t_k d_k$, onde d_k é uma direção de descida e t_k satisfaz as condições de Wolfe. Suponha que f é limitada inferiormente e continuamente diferenciável em um aberto $\mathcal{A}$ contendo o conjunto de nível $\{x \parallel f(x) \leq f(x^0)\}$, onde x^0 é o ponto inicial. Assuma também que o gradiente é Lipschitz em $\mathcal{A}$.

Então

$$\sum_{k \geq 0} \cos^2 \theta_k \left\| \nabla f(x_k) \right\|^2 < \infty.$$

Consequência

- $\cos^2 \theta_k \left\| \nabla f(x_k) \right\|^2 \to 0$
- Se $\cos \theta_k \geq \delta > 0, \quad \forall k$ então $\| \nabla f(x_k) \| \to 0$.

Gradient related

Definição (Bertsekas, 1999)

Seja $\{x_k,\ d_k\}$ uma sequência gerada por uma algoritmo de descida. A sequência de direções $\{d_k\}$ é dita gradient related se

para qualquer subsequência $\{x_k\}_{k\in\mathcal{K}}$ que converge a um ponto não-estacionário, a subsequência correspondente de direções $\{d_k\}_{k\in\mathcal{K}}$ é limitada e satisfaz

$$\lim_{k\to\infty} \sup_{k\in\mathcal{K}} \nabla f(x_k)^T d_k < 0.$$

Significado

A direção d_k não será "muito pequena" ou "muito grande" em relação a $\nabla f(x_k)$ e que o ângulo entre d_k e $\nabla f(x_k)$ não ficará tão próximo de 90^o.

Gradient related

Definição (Bertsekas, 1999)

Seja $\{x_k,\ d_k\}$ uma sequência gerada por uma algoritmo de descida. A sequência de direções $\{d_k\}$ é dita gradient related se

para qualquer subsequência $\{x_k\}_{k\in\mathcal{K}}$ que converge a um ponto não-estacionário, a subsequência correspondente de direções $\{d_k\}_{k\in\mathcal{K}}$ é limitada e satisfaz

$$\lim_{k\to\infty}\ \sup_{k\in\mathcal{K}} \nabla f(x_k)^T d_k < 0.$$

Significado

A direção d_k não será "muito pequena" ou "muito grande" em relação a $\nabla f(x_k)$ e que o ângulo entre d_k e $\nabla f(x_k)$ não ficará tão próximo de 90^o.

Proporcionalidade + Ângulo

- Proporcionalidade (condição β)

$$\|d_k\| \geq \beta \|\nabla f(x_k)\|, \qquad \beta > 0$$

- Ângulo

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \, \|d_k\|} \geq \delta, \qquad \delta \in (0,1)$$

Seja $\{x_k\}_{k\in\mathcal{K}}$, convergindo a $\bar{x}$ não estacionário.

$$\lim_{k\to\infty} \sup_{k\in\mathcal{K}} \nabla f(x_k)^T d_k \leq -\delta\,\beta\,\|\nabla f(\bar{x})\|^2 < 0.$$

E se $\{d_k\}$ é limitada então é gradient related.

Proporcionalidade + Ângulo

- Proporcionalidade (condição β)

$$\|d_k\| \geq \beta \|\nabla f(x_k)\|, \qquad \beta > 0$$

- Ângulo

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \, \|d_k\|} \geq \delta, \qquad \delta \in (0,1)$$

Seja $\{x_k\}_{k\in\mathcal{K}}$, convergindo a $\bar{x}$ não estacionário.

$$\lim_{k\to\infty} \sup_{k\in\mathcal{K}} \nabla f(x_k)^T d_k \leq -\delta\,\beta\,\|\nabla f(\bar{x})\|^2 < 0.$$

E se $\{d_k\}$ é limitada então é gradient related.

Exemplos

1. $d_k = -B_k^{-1} \nabla f(x_k)$, B_k simétrica e definida positiva

$$c_1 \|z\|^2 \le z^T B_k z \le c_2 \|z\|^2, \quad \forall z \in \mathbb{R}^n, \quad k = 0, 1, \dots,$$

para $0 < c_1 < c_2$. ("Autovalores limitados")

2. Para $p_1 \ge 0, p_2 \ge 0$

$$c_1 \|\nabla f(x_k)\|^{p_1} \le -\nabla f(x_k)^T d_k, \qquad \|d_k\| \le c_2 \|\nabla f(x_k)\|^{p_2}$$

3. Generalização de "autovalores limitados"

$$c_1 \|\nabla f(x_k)\|^{p_1} \|z\|^2 \le z^T B_k z \le c_2 \|\nabla f(x_k)\|^{p_2} \|z\|^2, \ \forall z \in \mathbb{R}^n$$

Theorem

Seja $\{x_k\}$ uma sequência gerada por um algoritmo de descida. Assuma que $\{d_k\}$ é gradient related e que t_k é escolhido de modo a satisfazer Armijo. Então todo ponto limite de $\{x_k\}$ é estacionário.

Proof.

1. Suponha que uma subsequência $\{x_k\}_{k \in \mathcal{K}}$ convirja a $\bar{x}$ não estacionário.
2. Usando a condição de Armijo e o fato de $\{d_k\}$ ser gradient related, chegamos que $\{t_k\}_{k \in \mathcal{K}} \to 0$.
3. Como $\{d_k\}_{\mathcal{K}}$ é limitada, existe $\{d_k\}_{\bar{\mathcal{K}}}$ convergindo a $\bar{d}$.
4. Daí mostra-se que

$$0 \leq \nabla f(\bar{x})^T \bar{d},$$

que contradiz o fato de $\{d_k\}$ ser gradient related.

Theorem

Seja $\{x_k\}$ uma sequência gerada por um algoritmo de descida. Assuma que $\{d_k\}$ é gradient related e que t_k é escolhido de modo a satisfazer Armijo. Então todo ponto limite de $\{x_k\}$ é estacionário.

Proof.

1. Suponha que uma subsequência $\{x_k\}_{k\in\mathcal{K}}$ convirja a $\bar{x}$ não estacionário.
2. Usando a condição de Armijo e o fato de $\{d_k\}$ ser gradient related, chegamos que $\{t_k\}_{k\in\mathcal{K}} \to 0$.
3. Como $\{d_k\}_{\mathcal{K}}$ é limitada, existe $\{d_k\}_{\bar{\mathcal{K}}}$ convergindo a $\bar{d}$.
4. Daí mostra-se que
$$0 \leq \nabla f(\bar{x})^T \bar{d},$$
que contradiz o fato de $\{d_k\}$ ser gradient related.

Iterações Híbridas

Theorem

Seja $\{x_k\}$ *uma sequência tal que*

$$f(x_{k+1}) \leq f(x_k), \quad k = 0, 1, \ldots$$

Assuma que existe um conjunto infinito de índices $\mathcal{K}$ *tal que*

$$x_{k+1} = x_k + t_k\, d_k, \quad \forall k \in \mathcal{K},$$

onde $\{d_k\}_{\mathcal{K}}$ *é gradient related e* t_k *é escolhido segundo Armijo. Então todo ponto limite de uma subsequência de* $\{x_k\}_{\mathcal{K}}$ *é estacionário.*

- Alternar iterações gradient-based e iterações heurísticas.

Iterações Híbridas

Theorem

Seja $\{x_k\}$ uma sequência tal que

$$f(x_{k+1}) \leq f(x_k), \qquad k = 0, 1, \ldots$$

Assuma que existe um conjunto infinito de índices $\mathcal{K}$ tal que

$$x_{k+1} = x_k + t_k\, d_k, \qquad \forall k \in \mathcal{K},$$

onde $\{d_k\}_{\mathcal{K}}$ é gradient related e t_k é escolhido segundo Armijo. Então todo ponto limite de uma subsequência de $\{x_k\}_{\mathcal{K}}$ é estacionário.

- Alternar iterações gradient-based e iterações heurísticas.

Teorema da Captura

Theorem

Seja $\{x_k\}$ *uma sequência gerada por um algoritmo de descida e globalmente convergente. Assuma que existem* $s > 0$ *e* $c > 0$ *tais que* $\forall k$

$$t_k \leq s, \qquad \|d_k\| \leq c \, \|\nabla f(x_k)\| \, .$$

Seja x_* *um minimizador local estrito de* f*, o qual é o único ponto estacionário de* f *dentro de alguma bola aberta. Então existe um conjunto* S *contendo* x_* *tal que se* $x_{k_0} \in S$ *para algum* $k_0 \geq 0$*, então* $x_k \in S$ *para todo* $k \geq k_0$ *e* $\{x_k\} \to x_*$*.*

Busca não-monótona

$$f(x_k + t\, d_k) \leq \max_{0 \leq j \leq \min\{k, \mathcal{M}-1\}} f(x_{k-j}) + \sigma\, t\, \nabla f(x_k)^T d_k$$

Grippo, Lampariello, Lucidi, 1986

Condições para convergência

1. $\{x \mid f(x) \leq f(x_0)\}$ é compacto
2. Existem $c_1, c_2 > 0$ tais que

$$\nabla f(x_k)^T d_k \leq -c_1 \|\nabla f(x_k)\|^2 ,$$

$$\|d_k\| \leq c_2 \|\nabla f(x_k)\| .$$

Ω convexo

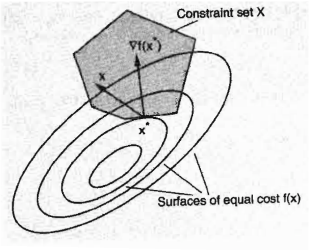
$$\min_{x} \quad f(x)$$
$$\text{s.a} \quad x \in \Omega \subset \mathbb{R}^n$$

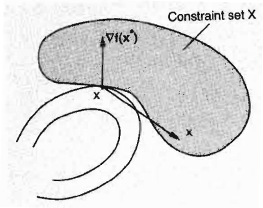
Ω convexo e fechado

1. Se x_* é minimizador local de f em Ω, então

   $$\nabla f(x_*)^T(x - x_*) \geq 0, \quad \forall x \in \Omega$$

2. Se f é convexa em Ω, então a condição anterior é também suficiente.





Bertsekas, 1999

Métodos de direções factíveis

$$x_{k+1} = x_k + t_k\, d_k$$

d_k é uma direção factível se existe $\bar{t} > 0$ tal que

$$x_{k+1} = x_k + t\, d_k \in \Omega, \quad \forall t \in (0, \bar{t}]$$

Podemos expressar direções factíveis como

$$d_k = \bar{x}_k - x_k,$$

onde $\bar{x}_k \in \Omega$, e neste caso

$$x_{k+1} = x_k + t_k\,(\bar{x}_k - x_k), \qquad t_k \in (0, 1].$$

Convergência global

Se x_k é não estacionário, então

$$\exists \bar{x}_k \in \Omega, \quad \nabla f(x_k)^T(\bar{x}_k - x_k) < 0.$$

Theorem

Seja $\{x_k\}$ uma sequência gerada por uma algoritmo de direções factíveis $x_{k+1} = x_k + t_k\, d_k$. Assuma que $\{d_k\}$ é gradient related e que t_k escolhido por Armijo. Então todo ponto limite de $\{x_k\}$ é estacionário.

Gradiente condicional

Resolver o problema

$$\begin{aligned} \min_{x} \quad & \nabla f(x_k)^T (x - x_k) \\ \text{s.a} \quad & x \in \Omega, \end{aligned}$$

e obtendo uma solução $\bar{x}_k$, definir a direção

$$d_k = \bar{x}_k - x_k.$$

Convergência

Suponha Ω compacto. Seja $\{x_k\}_{k \in \mathcal{K}}$ convergindo a $\tilde{x}$ não estacionário.

$$\nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(x_k)(x - x_k), \qquad \forall x \in \Omega,$$

tomando limites

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(\tilde{x})(x - \tilde{x}), \qquad \forall x \in \Omega,$$

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \min_{x \in \Omega} \nabla f(\tilde{x})(x - \tilde{x}) < 0.$$

Gradiente condicional

Resolver o problema

$$\begin{aligned} \min_{x} \quad & \nabla f(x_k)^T (x - x_k) \\ \text{s.a} \quad & x \in \Omega, \end{aligned}$$

e obtendo uma solução $\bar{x}_k$, definir a direção

$$d_k = \bar{x}_k - x_k.$$

Convergência

Suponha Ω compacto. Seja $\{x_k\}_{k \in \mathcal{K}}$ convergindo a $\tilde{x}$ não estacionário.

$$\nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(x_k)(x - x_k), \quad \forall x \in \Omega,$$

tomando limites

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(\tilde{x})(x - \tilde{x}), \quad \forall x \in \Omega,$$

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \min_{x \in \Omega} \nabla f(\tilde{x})(x - \tilde{x}) < 0.$$

Gradiente condicional

Resolver o problema

$$\begin{aligned} \min_{x} \quad & \nabla f(x_k)^T (x - x_k) \\ \text{s.a} \quad & x \in \Omega, \end{aligned}$$

e obtendo uma solução $\bar{x}_k$, definir a direção

$$d_k = \bar{x}_k - x_k.$$

Convergência

Suponha Ω compacto. Seja $\{x_k\}_{k \in \mathcal{K}}$ convergindo a $\tilde{x}$ não estacionário.

$$\nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(x_k)(x - x_k), \qquad \forall x \in \Omega,$$

tomando limites

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \nabla f(\tilde{x})(x - \tilde{x}), \qquad \forall x \in \Omega,$$

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \nabla f(x_k)(\bar{x}_k - x_k) \leq \min_{x \in \Omega} \nabla f(\tilde{x})(x - \tilde{x}) < 0.$$

Gradiente projetado

$$x_{k+1} = x_k + t_k \left(\bar{x}_k - x_k \right),$$

onde

$$\bar{x}_k = \mathcal{P}_\Omega(x_k - s_k \nabla f(x_k))$$

$\mathcal{P}_\Omega(.)$ denota a projeção em Ω, $s_k > 0$ $t_k \in (0, 1]$.

Projection theorem

Seja $\Omega \subset \mathbb{R}^n$ convexo, não vazio e fechado.

1. $\forall z \in \mathbb{R}^n$ existe um único $x_* \in \Omega$ que minimiza $\|z - x\|$ sobre Ω. x_* é chamado de projeção de z em Ω e denotado por $\mathcal{P}_\Omega(z)$.
2. Dado $z \in \mathbb{R}^n$, $x_* = \mathcal{P}_\Omega(z)$ se e somente se

$$(z - x_*)^T (x - x_*) \leq 0, \quad \forall x \in \Omega$$

3. A funcão $\mathcal{P}_\Omega : \mathbb{R}^n \to \Omega$ é contínua e não-expansiva

$$\|\mathcal{P}_\Omega(x) - \mathcal{P}_\Omega(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

Gradiente projetado

$$x_{k+1} = x_k + t_k\,(\bar{x}_k - x_k),$$

onde

$$\bar{x}_k = \mathcal{P}_\Omega(x_k - s_k\,\nabla f(x_k))$$

$\mathcal{P}_\Omega(.)$ denota a projeção em Ω, $s_k > 0$ $t_k \in (0, 1]$.

Projection theorem

Seja $\Omega \subset \mathbb{R}^n$ convexo, não vazio e fechado.

1. $\forall z \in \mathbb{R}^n$ existe um único $x_* \in \Omega$ que minimiza $\|z - x\|$ sobre Ω. x_* é chamado de projeção de z em Ω e denotado por $\mathcal{P}_\Omega(z)$.
2. Dado $z \in \mathbb{R}^n$, $x_* = \mathcal{P}_\Omega(z)$ se e somente se
$$(z - x_*)^T(x - x_*) \leq 0, \quad \forall x \in \Omega$$
3. A funcão $\mathcal{P}_\Omega : \mathbb{R}^n \to \Omega$ é contínua e não-expansiva
$$\|\mathcal{P}_\Omega(x) - \mathcal{P}_\Omega(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

Convergência

Suponha que $\{x_k\}_{k\in\mathcal{K}}$ converge a um ponto não estacionário $\tilde{x}$. A sequência de direções $\{d_k\}_{k\in\mathcal{K}}$, $d_k = \bar{x}_k - x_k$, é limitada, já que $\|\bar{x}_k - x_k\| \to \|\mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x})) - \tilde{x}\|$. Como $\bar{x}_k$ é solução do problema de projeção , temos que

$$(x_k - s\nabla f(x_k) - \bar{x}_k)^T(x - \bar{x}_k) \leq 0, \quad \forall x \in \Omega.$$

Em particular para $x = x_k$

$$\nabla f(x_k)^T(\bar{x}_k - x_k) \leq -\frac{1}{s}\|x_k - \bar{x}_k\|^2,$$

e tomando limites

$$\lim_{k\to\infty}\sup_{k\in\mathcal{K}} \nabla f(x_k)^T(\bar{x}_k - x_k) \leq= -\frac{1}{s}\|\tilde{x} - \mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x}))\|^2 < 0.$$

FIM

Convergência

Suponha que $\{x_k\}_{k\in\mathcal{K}}$ converge a um ponto não estacionário $\tilde{x}$. A sequência de direções $\{d_k\}_{k\in\mathcal{K}}$, $d_k = \bar{x}_k - x_k$, é limitada, já que $\|\bar{x}_k - x_k\| \to \|\mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x})) - \tilde{x}\|$. Como $\bar{x}_k$ é solução do problema de projeção , temos que

$$(x_k - s\nabla f(x_k) - \bar{x}_k)^T(x - \bar{x}_k) \leq 0, \qquad \forall x \in \Omega.$$

Em particular para $x = x_k$

$$\nabla f(x_k)^T(\bar{x}_k - x_k) \leq -\frac{1}{s}\|x_k - \bar{x}_k\|^2,$$

e tomando limites

$$\lim_{k\to\infty}\sup_{k\in\mathcal{K}} \nabla f(x_k)^T(\bar{x}_k - x_k) \leq= -\frac{1}{s}\|\tilde{x} - \mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x}))\|^2 < 0.$$

FIM

Convergência

Suponha que $\{x_k\}_{k\in\mathcal{K}}$ converge a um ponto não estacionário $\tilde{x}$. A sequência de direções $\{d_k\}_{k\in\mathcal{K}}$, $d_k = \bar{x}_k - x_k$, é limitada, já que $\|\bar{x}_k - x_k\| \to \|\mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x})) - \tilde{x}\|$. Como $\bar{x}_k$ é solução do problema de projeção , temos que

$$(x_k - s\nabla f(x_k) - \bar{x}_k)^T(x - \bar{x}_k) \leq 0, \qquad \forall x \in \Omega.$$

Em particular para $x = x_k$

$$\nabla f(x_k)^T(\bar{x}_k - x_k) \leq -\frac{1}{s}\left\|x_k - \bar{x}_k\right\|^2,$$

e tomando limites

$$\lim_{k\to\infty}\sup_{k\in\mathcal{K}} \nabla f(x_k)^T(\bar{x}_k - x_k) \leq= -\frac{1}{s}\left\|\tilde{x} - \mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x}))\right\|^2 < 0.$$

FIM

Convergência

Suponha que $\{x_k\}_{k\in\mathcal{K}}$ converge a um ponto não estacionário $\tilde{x}$. A sequência de direções $\{d_k\}_{k\in\mathcal{K}}$, $d_k = \bar{x}_k - x_k$, é limitada, já que $\|\bar{x}_k - x_k\| \to \|\mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x})) - \tilde{x}\|$. Como $\bar{x}_k$ é solução do problema de projeção , temos que

$$(x_k - s\nabla f(x_k) - \bar{x}_k)^T(x - \bar{x}_k) \leq 0, \qquad \forall x \in \Omega.$$

Em particular para $x = x_k$

$$\nabla f(x_k)^T(\bar{x}_k - x_k) \leq -\frac{1}{s}\|x_k - \bar{x}_k\|^2,$$

e tomando limites

$$\lim_{k\to\infty} \sup_{k\in\mathcal{K}} \nabla f(x_k)^T(\bar{x}_k - x_k) \leq= -\frac{1}{s}\|\tilde{x} - \mathcal{P}_\Omega(\tilde{x} - s\nabla f(\tilde{x}))\|^2 < 0.$$

FIM

Referências

- D. P. Bertsekas, Nonlinear programming: 2nd edition, Athena Scientific, 1999.
- J. Nocedal & S. J. Wright, *Numerical Optimization*, Springer, 1999.
- J. M. Martínez & S. A. Santos, *Métodos Computacionais de Otimização*, IMPA, 20º Colóquio Brasileiro de Matemática, Rio de Janeiro, SBM, 1995.