

Método de restrições ativas para minimização com restrições lineares

Marina Andretta

ICMC-USP

27 de outubro de 2018

O problema

O problema a ser considerado é

Minimizar $f(x)$ sujeita a $x \in \Omega$,

onde

- $\Omega = \{x \in \mathbf{R}^n \mid C_i(x) \leq 0, i = 1, \dots, p\}$,
- $C_i(x) = (a^i)^T x - b_i$, $a^i = (a_1^i, \dots, a_n^i)^T$ e
- $f : \mathbf{R}^n \rightarrow \mathbf{R}$ pertence a \mathcal{C}^2 .

Métodos de restrições ativas se baseiam na seguinte **idéia**: se fosse possível conhecer o conjunto de restrições ativas na solução, a solução seria “facilmente” calculada (bastaria “apenas” resolver um problema irrestrito).

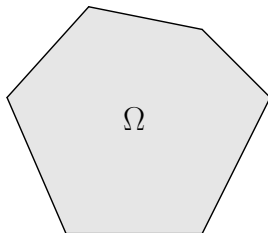
Assim, métodos de restrições ativas tentam, a cada iteração, descobrir qual o conjunto de restrições ativas na solução. Se detectam que o conjunto W de restrições que supunham ser ativas na solução está incorreto, utilizam algum critério para acrescentar ou remover restrições de W .

Método de restrições ativas

Dividimos a região viável Ω em faces abertas disjuntas. Para todo $I \subset \{1, 2, \dots, p\}$ definimos

$$F_I = \{x \in \Omega \mid C_i(x) = 0 \text{ se e somente se } i \in I\}.$$

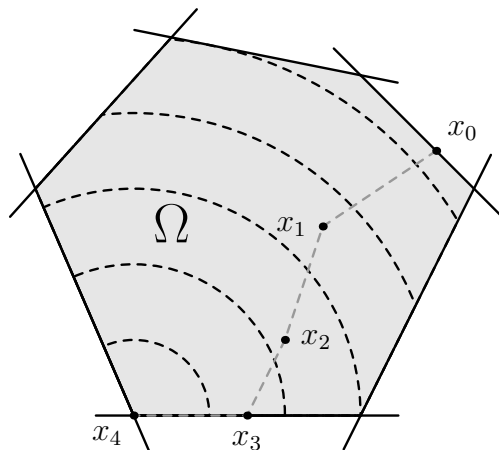
O conjunto Ω é a união das faces abertas.



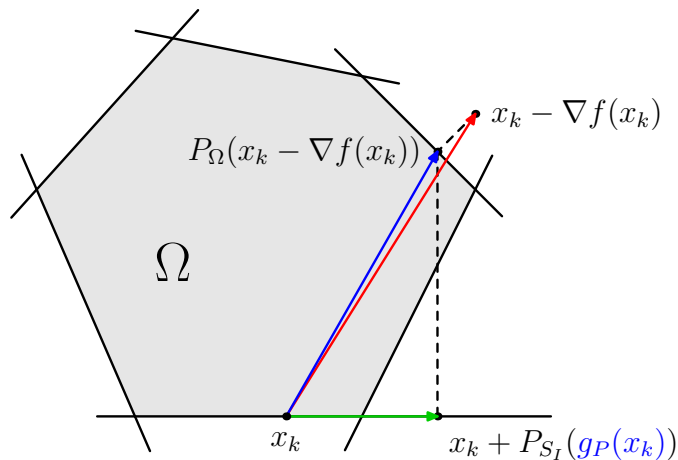
A idéia do método de restrições ativas apresentado aqui é a seguinte:

- A cada **iteração k** , tem-se um ponto x_k pertencente a uma face. No início de cada iteração, **decide-se se a face atual deve ser abandonada** ou não.
- Se o algoritmo decide **abandonar a face atual**, algum método é utilizado para calcular **x_{k+1} em uma face diferente**.
- Se o algoritmo decide **permanecer na face atual**, algum método é utilizado para calcular **x_{k+1} no fecho da face atual**.
- x_{k+1} sempre é calculado de maneira que seu valor de função seja menor do que o valor de função em x_k .

Método de restrições ativas

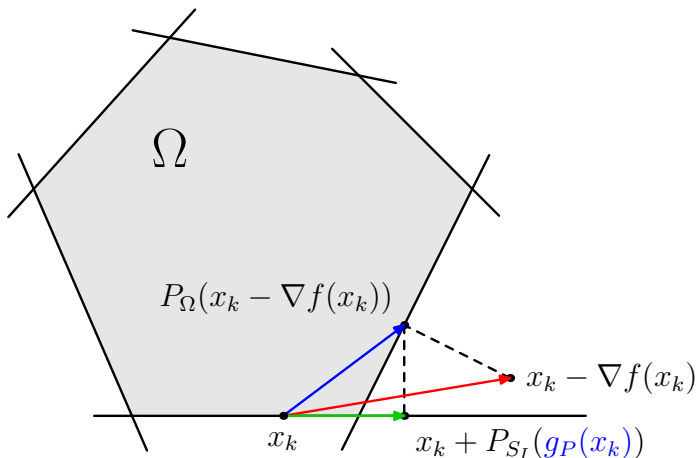


Decisão de permanecer ou não na face



Como $\|\text{blue arrow}\|$ é “muito” maior do que $\|\text{green arrow}\|$, decidimos abandonar a face.

Decisão de permanecer ou não na face



Como $\|\nearrow\|$ NÃO é “muito” maior do que $\|\searrow\|$, decidimos permanecer na face.

Cálculo da projeção P_Ω

Projetar um ponto $\bar{x} \in \mathbf{R}^n$ em Ω significa encontrar um ponto \bar{x}^* que seja solução do seguinte problema:

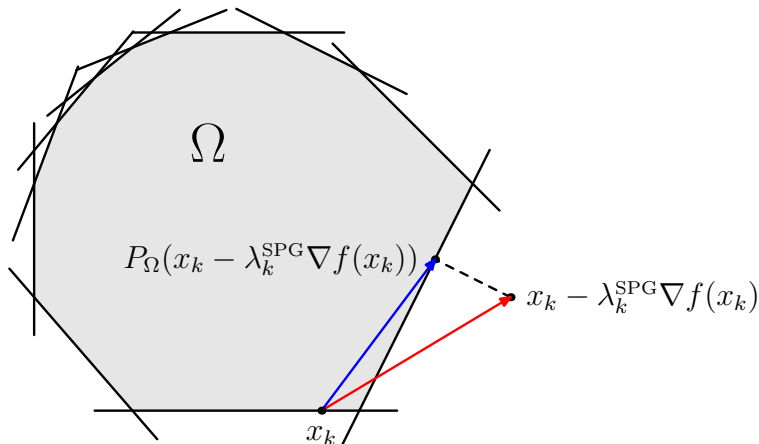
$$\text{Minimizar } \frac{1}{2} \|\bar{x} - x\|_2^2 \text{ sujeita a } x \in \Omega.$$

Note que, quando há apenas **restrições de caixa**, a **projeção é trivial**. Na presença de **restrições lineares**, a **projeção** passou a ser um **problema de minimização**.

Vejamos agora algumas idéias para diminuir o número de projeções e o custo de cada uma delas.

- Como, na presença de restrições lineares, o custo computacional da projeção é alto, decidimos **permanecer na mesma face** até que seja **encontrado o minimizador** desta face **ou** até que uma “borda” seja atingida.
- Seguindo este critério, podemos acrescentar restrições ao conjunto de restrições ativas por várias iterações seguidas para, no fim, chegar a uma face que não contém a solução do problema. Para evitar que isso aconteça, **se acrescentamos restrições ao conjunto de restrições ativas por 20 iterações consecutivas, abandonamos a face.**

Sair da face: Método SPG parcial



O passo espectral é dado por

$$\lambda_k^{spg} = \frac{s_k^T s_k}{s_k^T y_k},$$

onde

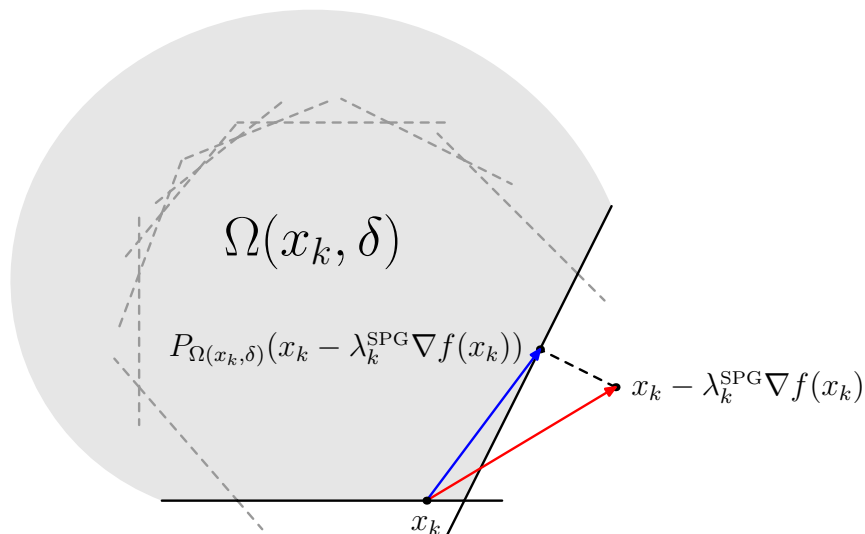
$$s_k = x_k - x_{k-1}, \quad y_k = \nabla f(x_k) - \nabla f(x_{k-1}).$$

A interpretação para λ_k^{spg} é a seguinte:

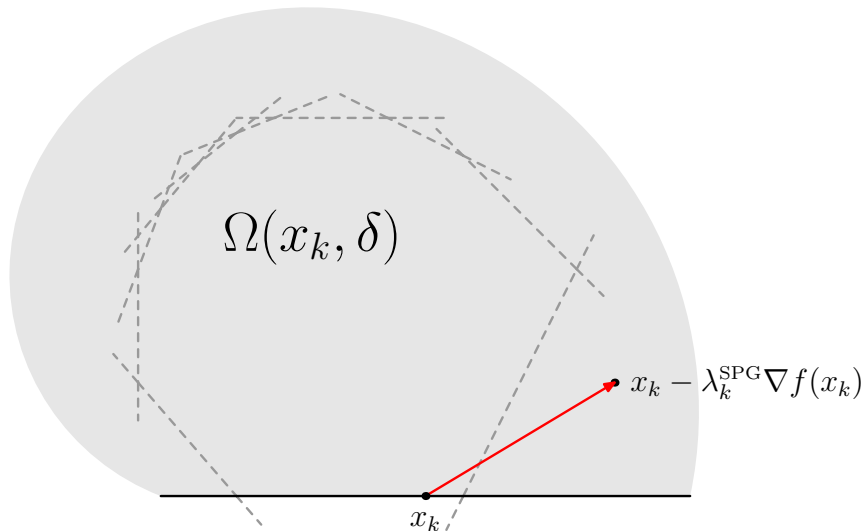
$$\lambda_k^{spg} = \underset{D}{\operatorname{argmin}} \quad \frac{1}{2} \|Ds_k - y_k\|_2^2$$

sujeita a $D = \frac{1}{\lambda} I$.

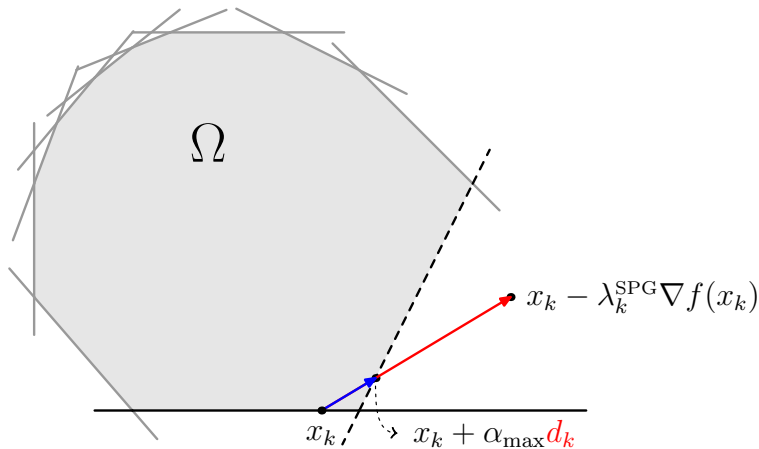
Método SPG: δ ideal



Método PSPG: δ pequeno



Método PSPG: δ pequeno



Permanecer na face: minimização nas faces

Em cada face deseja-se encontrar solução para o problema

$$\text{Minimizar } f(x) \text{ sujeita a } \bar{A}x = \bar{b},$$

com

- $x \in \mathbf{R}^n$, $\bar{A} \in \mathbf{R}^{t \times n}$ e $\bar{b} \in \mathbf{R}^t$,
- \bar{A} e \bar{b} formados pelas t restrições ativas na face atual.

Ao resolver este problema, devemos tomar o cuidado de **não violar as restrições restantes** do problema original.

Este problema pode ser escrito como um problema irrestrito no espaço nulo do espaço gerado pelas linhas \bar{A} .

O problema a ser resolvido em cada face pode ser escrito como

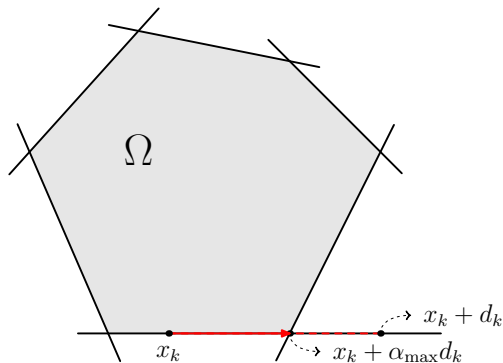
$$\text{Minimizar } \bar{f}(x^z) = f(Y(\bar{A}Y)^{-1}\bar{b} + Zx^z),$$

onde

- \bar{A} e \bar{b} como definidos anteriormente,
- t número das restrições ativas,
- $x^z \in \mathbf{R}^{(n-t)}$,
- $Z \in \mathbf{R}^{n \times (n-t)}$ tal que $\bar{A}Z = 0$ e
- $Y \in \mathbf{R}^{n \times t}$ tal que $Z^T Y = 0$.

Minimização nas faces

O que fazer quando o ponto calculado pelo algoritmo irrestrito não é viável?



Critério de convergência

Se $z \in \Omega$ satisfaz as condições KKT do problema original, então z também satisfaz as condições KKT de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{sujeita a} & x \in \Omega(z, \delta), \end{array} \quad (P_\delta)$$

para todo $\delta \geq 0$.

Reciprocamente, se z satisfaz as condições KKT de (P_δ) para algum $\delta \geq 0$, então z satisfaz as condições KKT do problema original.

Portanto, usaremos o critério de convergência $\|g_P(x_k, \delta^k)\| = 0$ (ao invés de $\|g_P(x_k)\| = 0$).