Método de restrições ativas para minimização em caixas

Marina Andretta

ICMC-USP

16 de novembro de 2010

Problema com restrições de caixa

Estamos interessados em resolver o problema

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeita a} & x \in \Omega = \{x \in \boldsymbol{R}^n \mid \ell \leq x \leq u\}, \end{array}} \qquad (1)$$

com

- $x, \ell, u \in \boldsymbol{R}^n$, $\ell < u$, e
- $f : \boldsymbol{R}^n \to \boldsymbol{R}$ suave.

Denotaremos o gradiente da função f por g.

Método de restrições ativas

Se fosse possível saber o conjunto de restrições ativas na solução, o ponto x^* solução do problema (1) seria "facilmente" calculado (pois o problema se tornaria irrestrito).

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

O método de restrições ativas que estudaremos parte de um conjunto de restrições ativas W.

A cada iteração, fixa as variáveis de W e resolve um subproblema irrestrito, usando apenas as variáveis livres.

Se a solução do subproblema for também a solução do problema original (1), o conjunto W está correto e o método para.

Se a solução do subproblema não for solução do problema original (1), o conjunto W é modificado e inicia-se uma nova iteração.

Note que, ao fixar algumas variáveis do conjunto Ω, estamos definindo faces deste conjunto.

Podemos então dividir a região viável Ω em faces abertas disjuntas.

Para todo $I \subset \{1, 2, \ldots, n, n+1, n+2, \ldots, 2n\}$ definimos

$$\boxed{F_I = \left\{ x \in \Omega \;\middle|\; \begin{array}{ll} x_i = \ell_i, & \text{se } i \in I, \\ x_i = u_i, & \text{se } n+i \in I, \\ \ell_i < x_i < u_i & \text{caso contrário} \end{array} \right\}.}$$

Método de restrições ativas

O conjunto Ω é a união das faces abertas.

Definimos V_I o menor espaço afim que contém F_I e S_I o subespaço vetorial paralelo a V_I.

Definimos o gradiente projetado contínuo como

$$\boxed{g_P(x) = P_\Omega(x - g(x)) - x.}$$

Lembre-se que um ponto viável x_k satisfaz as condições KKT se e somente se $g_P(x_k) = 0$.

Método de restrições ativas

Para todo $x \in F_I$, definimos o gradiente interno como

$$\boxed{g_I(x) = P_{S_I}[g_P(x)].}$$

Ou seja,

$$\boxed{[g_I(x)]_i = \begin{cases} [g_P(x)]_i & \text{, se } x_i \text{ é livre,} \\ 0 & \text{, caso contrário} \end{cases}}$$

Note que a norma de $g_I(x)$ sempre é menor ou igual à norma de $g_P(x)$.

Método de restrições ativas

A idéia do método de restrições ativas é a seguinte:

- O conjunto viável é dividido em faces abertas e, a cada iteração k, tem-se um ponto x_k pertencente a uma dessas faces.
- No início de cada iteração, verifica-se qual a relação entre a norma do gradiente interno e a norma do gradiente projetado contínuo.
- Quando a norma do gradiente projetado contínuo g_P é "muito maior" que a norma do gradiente interno g_I, decide-se mudar de face.
- Quando a norma do gradiente projetado contínuo g_P não é "muito maior", decide-se continuar na mesma face.

Teste para decidir se permanece na face

O teste principal, usado para decidir se permanecemos ou não na face atual, é

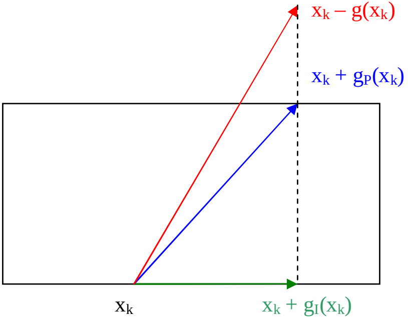
$$\boxed{\|g_I(x_k)\| \geq \eta \|g_P(x_k)\|,}$$

para $\eta \in (0, 1)$.

A ideia deste teste é: se a função objetivo parece diminuir mais saindo da face (norma de g_P "muito maior" do que a norma de g_I), decidimos abandonar a face corrente.

Se ficar na face atual parece promissor, decidimos permanecer na face corrente.

Teste para decidir se permanece na face



Como $\|\nearrow\|$ é “maior” do que $0.1\|\nearrow\|$, decidimos permanecer na face.

O que fazer para permanecer ou abadonar uma face

Para permanecer na face atual, precisamos resolver um problema de minimização com algumas variáveis do problema original fixas em seus limitantes.

Em vez de encontrar o minimizador na face, realizamos apenas uma iteração de um método para minimização "irrestrita" para calcular um ponto x_{k+1} no fecho da face atual, no qual há decréscimo do valor da função objetivo.

O que fazer para permanecer ou abadonar uma face

Para abandonar a face, fazemos uma iteração do método SPG (Gradiente Espectral Projetado), que fornecerá um novo ponto x_{k+1} viável em uma nova face, com valor de função objetivo menor.

Calculado o novo ponto, verificamos se encontramos a solução do problema original. Em caso negativo, verificamos novamente se vale a pena ou não permanecer na face atual e repetimos o procedimento.

Método de restrições ativas: Dados $x_0 \in \Omega$ e $\eta \in (0,1)$.

Passo 1: Faça $k \leftarrow 0$.

Passo 2: Se $\|g_P(x_k)\| \leq \epsilon$, pare com x_k como solução.

Passo 3: Se $\|g_I(x_k)\| \geq \eta \|g_P(x_k)\|$,

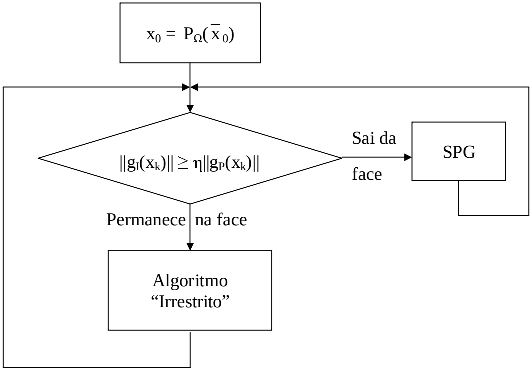
suponha, sem perda de generalidade, que as variáveis livres em F_I são $x_1, ..., x_m$ e que as variáveis restantes estão fixas em $\bar{x}_{m+1}, ..., \bar{x}_n$ ($\bar{x}_i \in \{l_i, u_i\}$, para $m+1 \leq i \leq n$). Defina $\bar{f}(x_1, ..., x_m) = f(x_1, ..., x_m, \bar{x}_{m+1}, ..., \bar{x}_n)$, $\forall x_1, ..., x_m \in \boldsymbol{R}$. Use uma iteração de um algoritmo "irrestrito" para calcular $x_{k+1} \in \bar{F}_I$ minimizador de $\bar{f}$.

Senão

faça uma iteração do SPG para calcular x_{k+1}.

Passo 4: Faça $k \leftarrow k+1$ e volte para o Passo 2.

Método de restrições ativas

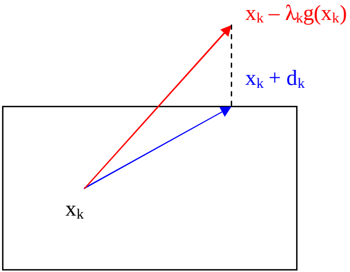


Método de restrições ativas

Note que, como o valor da função objetivo sempre diminui, quando uma face é abandonada ela nunca é visitada novamente.

Como o número de faces é finito e tanto o SPG como o algoritmo "irrestrito" possuem convergência garantida, este método de restrições ativas converge em um número finito de iterações para a solução do problema (1).

Como sair da face: método SPG



- $s_k = x_k - x_{k-1}$, $y_k = g(x_k) - g(x_{k-1})$
- $\lambda_k = \frac{{s_k}^T s_k}{{s_k}^T y_k}$
- $d_k = P_\Omega(x_k - \lambda_k g(x_k)) - x_k$

Permanecer na face: minimização na face

Em cada face F_I estamos interessados em resolver o seguinte problema "irrestrito"

$$\boxed{\text{Minimizar } \bar{f}(x_1, \ldots, x_m),} \tag{2}$$

- $x_1, \ldots, x_m$ variáveis livres e $\bar{x}_{m+1}, \ldots, \bar{x}_n$ variáveis fixas em F_I
- $\bar{f}(x_1, \ldots, x_m) = f(x_1, \ldots, x_m, \bar{x}_{m+1}, \ldots, \bar{x}_n)$, para todo $x_1, \ldots, x_m \in \boldsymbol{R}$

Permanecer na face: minimização na face

Para resolver o problema (2), qualquer algoritmo irrestrito pode ser usado.

No entanto, é preciso tomar cuidado para que as restrições do problema original (1) nas variáveis livres $x_1, \ldots, x_m$ sejam respeitadas, mantendo os iterandos do método de restrições ativas sempre viáveis.

Minimização na face com busca linear

Se o método escolhido para fazer a minimização na face for de busca linear, a viabilidade do ponto $x_{k+1} = x_k + \alpha_k p_k$ é garantida pelo seguinte procedimento:

- Calcula-se $\alpha_{\max}$ o maior escalar entre 0 e 1 tal que $x_k + \alpha_{\max} p_k \in \Omega$.

$$\boxed{\alpha_{\max} = \max\{\alpha \in [0, 1] \mid x_k + \alpha p_k \in \Omega\}.}$$

- O tamanho de passo inicial para a busca linear é tomado como $\alpha_{\max}$.

Minimização na face com busca linear

Iteração do algoritmo "irrestrito" com busca linear: Seja k a iteração atual. Dados x_k viável, $\epsilon > 0$ e $c \in (0, 1)$.

Passo 1: Se $\|\nabla \bar{f}(x_k)\| \leq \epsilon$ então pare e com x_k minimizador na face.

Passo 2: Calcule uma direção de descida p_k.

Passo 3: Calcule $\alpha_{\max} = \max\{\alpha \in [0, 1] \mid x_k + \alpha p_k \in \Omega\}$.

Passo 4: Se $\alpha_{\max} < 1$ e $\bar{f}(x_k + \alpha_{\max} p_k) < \bar{f}(x_k)$ então

faça $x_{k+1} = x_k + \alpha_{\max} p_k$ e pare.

Passo 5: Calcule $\alpha_k \leq \alpha_{\max}$ que satisfaz condições de Wolfe ou é calculado usando *backtracking* com Armijo.

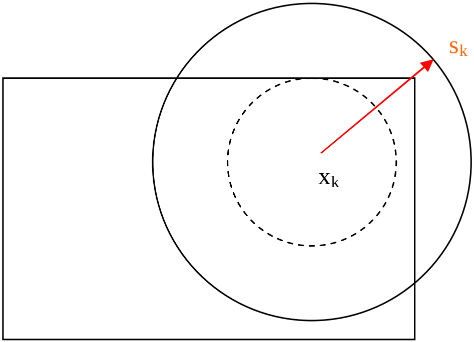
Passo 6: Faça $x_{k+1} \leftarrow x_k + \alpha_k p_k$.

Minimização na face com regiões de confiança

Se o método escolhido para fazer a minimização na face for de regiões de confiança, a viabilidade do ponto $x_{k+1} = x_k + p_k$ é garantida pelo seguinte procedimento:

- Se a solução do subproblema de regiões de confiança gera um ponto $x_k + p_k \notin \Omega$, calcula-se o escalar $\alpha_{\max}$ que faz com que $x_k + \alpha_{\max}$ fique na borda de Ω.
- Se a função $\bar{f}$ tem decréscimo simples neste ponto, ele é tomado como x_{k+1}.
- Se o ponto na borda não produz decréscimo da função, diminui-se o raio da região de confiança de modo que a solução do subproblema de regiões de confiança gere $x_k + p_k \in \Omega$.

Minimização na face com regiões de confiança



Minimização na face com regiões de confiança

Além disso, para garantir convergência deste método de restrições ativas, quando um ponto está muito próximo da borda de Ω, não se deve usar uma iteração de regiões de confiança.

Neste caso, utilizamos uma iteração do método de SPG para resolver o seguinte problema

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeita a} & x \in F_I. \end{array}}$$

Minimização na face com regiões de confiança

Iteração do algoritmo "irrestrito" com regiões de confiança: Seja k a iteração atual. Dados x_k viável, $\epsilon > 0$, $\Delta_{\min} > 0$, $\Delta_k \geq \Delta_{\min}$ e $\sigma > 0$.

Passo 1: Se $\|\nabla \bar{f}(x_k)\| \leq \epsilon$ então pare com x_k minimizador na face.

Passo 2: Calcule Δ_{borda} a distância de x_k à borda da face.

Passo 3: Se $\Delta_{borda} < 2\Delta_{\min}$ então

faça uma iteração de SPG restrito a F_I (face atual) para calcular x_{k+1} e pare.

Passo 4: Calcule p_k uma solução aproximada de

$$\begin{array}{ll} \text{Minimizar} & \varphi(w) \equiv \frac{1}{2} w^T \nabla^2 \bar{f}(x_k) w + (\nabla \bar{f}(x_k))^T w \\ \text{sujeita a} & \|w\| \leq \Delta_k. \end{array}$$

Passo 5: Se $\varphi(p_k) = 0$ então pare com x_k minimizador na face.

Minimização na face com regiões de confiança

Passo 6: Calcule

$$\alpha_{\max} = \max\{\alpha \in [0,1] \mid [x_k, x_k + \alpha p_k] \subset \Omega\}.$$

Passo 7: Se $\alpha_{\max} < 1$ e $\bar{f}(x_k + \alpha_{\max} p_k) < \bar{f}(x_k)$ então

faça $x_{k+1} = x_k + \alpha_{\max} p_k$ e pare.

Passo 8: Se $\alpha_{\max} < 1$ e $\bar{f}(x_k + \alpha_{\max} p_k) \geq \bar{f}(x_k)$ então

faça $\Delta_k = 0.9\Delta_{borda}$ e volte para o Passo 4.

Passo 9: Se $\frac{\bar{f}(x_k+p_k)-\bar{f}(x_k)}{\varphi(p_k)} \leq 0.1$ então

escolha $\Delta_k = \Delta \in [0.1\|p_k\|, 0.9\|p_k\|]$ e volte para o Passo 4.

Passo 10: Faça $x_{k+1} \leftarrow x_k + p_k$ e escolha $\Delta_{k+1} \geq \Delta_{\min}$.