Método do Gradiente Espectral Projetado para minimização em caixas

Marina Andretta

ICMC-USP

14 de novembro de 2018

Problema com restrições de caixa

Vamos nos concentrar em problemas apenas com restrições de caixa. Ou seja, estamos interessados em resolver o seguinte problema:

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

com

- $x, \ell, u \in (\boldsymbol{R} \cup \{-\infty, +\infty\})^n$,
- $f : \boldsymbol{R}^n \to \boldsymbol{R}$ função suave.

Problema com restrições de caixa

Como vimos, as condições necessárias de primeira ordem para que um ponto x^* seja solução para um problema do tipo

$$\boxed{\begin{array}{lll} \text{Minimizar} & f(x) & \\ \text{Sujeita a} & c_i(x) = 0, & i \in \mathcal{E}, \\ & c_i(x) \geq 0, & i \in \mathcal{I}, \end{array}} \qquad (2)$$

com restrições satisfazendo LICQ, é que exista um vetor λ^* tal que as condições KKT sejam satisfeitas. Ou seja,

$$\boxed{\begin{array}{rcll} \nabla_x \mathcal{L}(x^*, \lambda^*) & = & 0, & \\ c_i(x^*) & = & 0, & \forall i \in \mathcal{E}, \\ c_i(x^*) & \geq & 0, & \forall i \in \mathcal{I}, \\ \lambda_i^* & \geq & 0, & \forall i \in \mathcal{I}, \\ \lambda_i^* c_i(x^*) & = & 0, & \forall i \in \mathcal{E} \cup \mathcal{I}. \end{array}} \qquad (3)$$

Problema com restrições de caixa

A condição LICQ é a seguinte: dados um ponto x^* e o conjunto de restrições ativas $\mathcal{A}(x^*)$, dizemos que a condição de qualificação de restrição de independência linear (LICQ) vale se o conjunto dos gradientes das restrições ativas $\{\nabla c_i(x^*),\ i \in \mathcal{A}(x^*)\}$ é linearmente independente.

Note que, no caso do problema (1), que tem apenas restrições de caixa, a condição LICQ sempre pode ser satisfeita.

O gradiente de cada restrição é uma coluna da matriz identidade multiplicada por 1 ou -1.

Dois gradientes de restrições ativas seriam linearmente dependentes apenas quando uma variável x_i fosse igual tanto a seu limitante inferior como ao seu limitante superior. Ou seja, apenas quando $\ell_i = x_i = u_i$.

Neste caso, a variável x_i é fixa e a restrição $\ell_i \leq x_i \leq u_i$ pode ser eliminada do problema.

Consideraremos, daqui em diante, apenas problemas (1) com $\ell_i < u_i$, para todo i.

Problema com restrições de caixa

Assim, as condições necessárias de primeira ordem para que um ponto x^* seja solução de (1) são

$$\boxed{\begin{array}{rcll} \nabla f(x^*) - \sum_{i \in \mathcal{I}_\ell} \lambda_i^* e_i + \sum_{i \in \mathcal{I}_u} \mu_i^* e_i & = & 0, & \\ x_i - \ell_i & \geq & 0, & \forall i \in \mathcal{I}_\ell, \\ u_i - x_i & \geq & 0, & \forall i \in \mathcal{I}_u, \\ \lambda_i^* & \geq & 0, & \forall i \in \mathcal{I}_\ell, \\ \mu_i^* & \geq & 0, & \forall i \in \mathcal{I}_u, \\ \lambda_i^*(x_i - \ell_i) & = & 0, & \forall i \in \mathcal{I}_\ell, \\ \mu_i^*(u_i - x_i) & = & 0, & \forall i \in \mathcal{I}_u, \end{array}} \tag{4}$$

com $\mathcal{I}_\ell$ e $\mathcal{I}_u$ conjunto dos índices das restrições $\ell_i \leq x_i$, $\ell_i \in \boldsymbol{R}$ e $x_i \leq u_i$, $u_i \in \boldsymbol{R}$, respectivamente. e_i i-ésima coluna da matriz identidade.

Problema com restrições de caixa

Antes de prosseguir, vamos definir a projeção ortogonal de um ponto x no conjunto viável Ω. Esta será denotada por $P_\Omega(x)$.

$$\boxed{P_\Omega(x) = \begin{cases} x_i, & \text{se } \ell_i \leq x_i \leq u_i, \\ \ell_i, & \text{se } x_i < \ell_i, \\ u_i, & \text{se } x_i > u_i. \end{cases}}$$

Problema com restrições de caixa

Note que um ponto viável x^* satisfaz

$$\boxed{P_\Omega(x^* - \nabla f(x^*)) - x^* = 0,}$$

se e somente se este ponto x^* também satisfaz as condições KKT (exercício).

Método do Gradiente Espectral Projetado

O método do Gradiente Espectral Projetado (SPG) é usado para resolver o seguinte problema:

$$\boxed{\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeita a} & x \in \Omega, \end{array}}$$

com f uma função em C^1, $x \in \boldsymbol{R}^n$ e Ω um conjunto convexo fechado.

Método do Gradiente Espectral Projetado

Estamos interessados no caso em que Ω é uma caixa.

A idéia do algoritmo é a seguinte: na iteração k, calculamos uma direção de descida $p_k \in \Omega$ e fazemos uma busca linear para que $x_k + \alpha_k p_k$ satisfaça a condição de Armijo.

Quando isso acontece, temos o novo ponto $x_{k+1} = x_k + \alpha_k p_k \in \Omega$.

A direção p_k é calculada com base na projeção ortogonal de $-\lambda_k \nabla f(x_k)$ no conjunto Ω, em que λ_k (chamado de passo espectral) é um escalar calculado a partir de informações fornecidas pelo ponto atual x_k e pelo ponto x_{k-1} da iteração anterior.

Método do Gradiente Espectral Projetado

Método SPG: Dados $x_0 \in \Omega$, $\epsilon > 0$, $0 < \lambda_{min} < \lambda_{max} < \infty$, $c \in \left(0, \frac{1}{2}\right)$, $0 < \rho_1 < \rho_2 < 1$.

Passo 1: Faça $k \leftarrow 0$.

Passo 2: Se $\| P_\Omega(x_k - \nabla f(x_k)) - x_k \| \leq \epsilon$, pare com x_k como solução.

Passo 3: Calcule $s_k = x_k - x_{k-1}$ e $y_k = \nabla f(x_k) - \nabla f(x_{k-1})$.

Passo 4: Se ${s_k}^T y_k \leq 0$ então

faça $\lambda_k \leftarrow \lambda_{max}$.

Senão

faça $\lambda_k \leftarrow \min \left\{ \lambda_{max}, \max \left\{ \lambda_{min}, \frac{{s_k}^T s_k}{{s_k}^T y_k} \right\} \right\}$.

Passo 5: Calcule $p_k = P_\Omega(x_k - \lambda_k \nabla f(x_k)) - x_k$.

Método do Gradiente Espectral Projetado

Passo 6: Faça $\alpha \leftarrow 1$.

Passo 7: Se $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k$ então

faça $\alpha_k \leftarrow \alpha$.

Senão

faça $\alpha \leftarrow \alpha_{novo} \in [\rho_1 \alpha, \rho_2 \alpha]$ e repita o Passo 7.

Passo 8: Faça $x_{k+1} \leftarrow x_k + \alpha_k p_k$.

Passo 9: Faça $k \leftarrow k + 1$.

Passo 10: Volte para o Passo 2.

O valor obtido para λ_k possui alguma informação de segunda ordem da função objetivo. Na verdade, temos que

$$\boxed{\begin{array}{rll} \left(\frac{1}{\lambda_k}\right) I = & \text{argmin}_D & \|Ds_k - y_k\| \\ & \text{sujeita a} & D = \gamma I, \end{array}}$$

em que I é a matriz identidade.

Ou seja, $\left(\frac{1}{\lambda_k}\right) I$ é a matriz diagonal que melhor aproxima a equação secante (equação esta que é uma aproximação da hessiana da função objetivo).

Método do Gradiente Espectral Projetado

Quanto à convergência do algoritmo do Gradiente Espectral Projetado, temos que todo ponto de acumulação da sequência $\{x_k\}$ gerada por ele é um ponto estacionário restrito.

Cálculo do passo espectral inicial

Uma maneira de calcular λ_0 é calcular um $\bar{x}$ dando um passo muito pequeno a partir de x_0 na direção de $-\nabla f(x_0)$.

Neste caso, esperamos que a função objetivo decresça e satisfaça a condição de Armijo.

Por isso utilizamos, para calcular λ_0, os mesmos passos usados para calcular λ_k, considerando $\bar{x}$ como x_k e x_0 como x_{k-1}.

Cálculo do passo espectral inicial: Dados $\epsilon_{rel}, \epsilon_{abs} > 0$ e $x_o \in \Omega$.

Passo 1: Calcule $\alpha_{peq} = \max\{\epsilon_{rel}\|x_0\|_\infty, \epsilon_{abs}\}$.

Passo 2: Calcule $\bar{x} = x_0 - \alpha_{peq}\nabla f(x_0)$.

Passo 3: Calcule $\bar{s} = \bar{x} - x_0$.

Passo 4: Calcule $\bar{y} = \nabla f(\bar{x}) - \nabla f(x_0)$.

Passo 5: Se $\bar{s}^T\bar{y} \leq 0$ então

faça $\lambda_0 \leftarrow \lambda_{max}$.

Senão

faça $\lambda_0 \leftarrow \min\left\{\lambda_{max}, \max\left\{\lambda_{min}, \frac{\bar{s}^T\bar{s}}{\bar{s}^T\bar{y}}\right\}\right\}$.

Cálculo do tamanho de passo

Uma forma simples é calcular $\alpha \leftarrow \frac{\alpha_k}{2}$.

Uma forma mais interessante é utilizar a interpolação quadrática uni-dimensional $q(w)$ tal que $q(0) = f(x_k)$, $q(\alpha) = f(x_k + \alpha p_k)$ e $\nabla q(0) = \nabla f(x_k)^T p_k$.

Para obter α_{novo}, calculamos o mínimo de $q(w)$ e fazemos a salvaguarda (lembre-se que estamos interressados em $\alpha_{novo} \in [\rho_1 \alpha, \rho_2 \alpha]$).

A seguir estão os passos utilizados para a escolha de α_{novo} utilizando a interpolação quadrática e salvaguarda:

Cálculo do tamanho de passo

Cálculo do tamanho de passo: Dados $p_k \in \boldsymbol{R}^n$ direção de descida, $\alpha > 0$ e $x_k \in \Omega$.

Passo 1: Calcule $\alpha_{tent} = -\frac{\alpha^2 \nabla f(x_k)^T p_k}{2(f(x_k+\alpha p_k) - f(x_k) - \alpha \nabla f(x_k)^T p_k)}$.

Passo 2: Se $(\alpha_{tent} \geq \rho_1 \alpha)$ e $(\alpha_{tent} \leq \rho_2 \alpha)$ então

faça $\alpha_{novo} \leftarrow \alpha_{tent}$.

Senão

faça $\alpha_{novo} \leftarrow \frac{\alpha}{2}$.