Resolução de problemas com apenas restrições lineares de igualdade

Marina Andretta

ICMC-USP

14 de outubro de 2014

Problema com restrições lineares de igualdade

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

$$\boxed{\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeita a} & Ax = b, \end{array}} \qquad (1)$$

onde

- $x \in \boldsymbol{R}^n$,
- $A \in \boldsymbol{R}^{m \times n}$, $b \in \boldsymbol{R}^m$,
- $f : \boldsymbol{R}^n \to \boldsymbol{R}$ função suave.

Ao trabalhar com o problema (1), trataremos apenas do caso em que a matriz A possui linhas linearmente independentes.

Isso é razoável na prática, pois quando alguma linha de A depende linearmente de outra, ou a linha em questão pode ser eliminada sem nenhuma alteração da solução de (1) ou a região viável é vazia, não havendo solução para o problema.

Problema com restrições lineares de igualdade

Vejamos algumas definições:

- Z é uma matriz de $\boldsymbol{R}^{n\times(n-m)}$ cujas colunas formam uma base para núcleo do espaço gerado pelas linhas da matriz A.
- Y é uma matriz de $\boldsymbol{R}^{n\times m}$ cujas colunas formam uma base para a imagem do espaço formado pelas linhas de A.

Note que que $AZ = 0$ e AY é não-singular.

Uma vez escolhidas as matrizes Z e Y, todo vetor x em $\boldsymbol{R}^n$ possui uma única representação da forma

$$\boxed{x = Yx_y + Zx_z,}$$

com $x_y \in \boldsymbol{R}^m$ e $x_z \in \boldsymbol{R}^{(n-m)}$.

Em particular, tomando x^* solução de (1), temos

$$\boxed{x^* = Yx_y^* + Zx_z^*.}$$

Problema com restrições lineares de igualdade

Como x^* é viável, vale que

$$\boxed{Ax^* = A(Yx_y^* + Zx_z^*) = b.}$$

Como $AZ = 0$ e AY é não-singular, temos

$$\boxed{x_y^* = (AY)^{-1}b,}$$

Assim, a solução x^* de (1) (assim como qualquer ponto viável de (1)) pode ser vista como a soma de uma solução particular para o sistema linear $Ax = b$ (tomando, por exemplo, $x_z^* = 0$) e de um ponto no núcleo do espaço gerado pelas linhas de A.

Desta forma, reduzimos o problema restrito original (1) a um problema irrestrito de $n - m$ variáveis, dado por

$$\boxed{\text{Minimizar} \quad \bar{f}(x_z) = f(Y(AY)^{-1}b + Zx_z) = f(x).} \tag{2}$$

Problema com restrições lineares de igualdade

Note que o gradiente e a Hessiana de $\bar{f}$ são dados por

$$\boxed{\nabla \bar{f}(x_z) = Z^T \nabla f(x)}$$

e

$$\boxed{\nabla^2 \bar{f}(x_z) = Z^T \nabla^2 f(x) Z.}$$

Problema com restrições lineares de igualdade

Para resolver o problema (2), podemos usar qualquer método para resolução de problemas irrestritos.

Formularemos, então, um algoritmo para resolver (2) baseado em busca linear que gera uma sequência de pontos viáveis.

Para manter a viabilidade de cada ponto gerado pelo algoritmo, a cada iteração k, partimos de um ponto x_k viável e calculamos um passo p_k tal que

$$\boxed{Ap_k = 0.}$$

Problema com restrições lineares de igualdade

Como x_k é viável e $Ap_k = 0$, temos

$$\boxed{Ax_{k+1} = A(x_k + \alpha_k p_k) = Ax_k + \alpha_k Ap_k = Ax_k = b.}$$

Ou seja, o ponto $x_{k+1} = x_k + \alpha_k p_k$ é viável.

Problema com restrições lineares de igualdade

Note que

$$\boxed{Ap_k = 0 \Leftrightarrow p_k = Zp_z,}$$

para algum vetor $p_z \in \boldsymbol{R}^{(n-m)}$.

Resolução de problemas com restrições lineares de igualdade

Método de busca linear para minimização com restrições lineares de igualdade: Dados uma matriz $A \in \mathbf{R}^{m \times n}$, $m < n$, de posto m, um vetor $b \in \mathbf{R}^m$, um ponto inicial x_0 viável e um escalar $\epsilon > 0$.

Passo 1: Faça $k \leftarrow 0$.

Passo 2: Calcule Z tal que $AZ = 0$.

Passo 3: Se $\|Z^T \nabla f(x_k)\| \leq \epsilon$, pare com x_k como solução.

Passo 4: Calcule uma direção de descida $p_z \in \mathbf{R}^{(n-m)}$ e faça $p_k \leftarrow Zp_z$.

Passo 5: Faça $x_{k+1} \leftarrow x_k + \alpha_k p_k$, α_k satisfaz condições de Wolfe ou é calculado usando *backtracking* com Armijo.

Passo 6: Faça $k \leftarrow k + 1$ e volte para o Passo 3.

Cálculo da matriz Z

A primeira idéia para o cálculo da matriz Z é o uso da fatoração LQ da matriz A.

Nesta fatoração, que gasta $O(nm^2)$ operações, temos uma matriz ortogonal $Q \in \boldsymbol{R}^{n \times n}$ tal que

$$\boxed{AQ = (L \quad 0),}$$

onde $L \in \boldsymbol{R}^{m \times m}$ é triangular inferior.

Cálculo da matriz Z

Como estamos supondo que as linhas da matriz A são linearmente independentes, a matriz L é não-singular.

Escrevendo a matriz Q como $Q = [Q_1; Q_2]$, com $Q_1 \in \boldsymbol{R}^{n \times m}$ e $Q_2 \in \boldsymbol{R}^{n \times (n-m)}$, temos que

$$\boxed{AQ_1 = L \text{ e } AQ_2 = 0.}$$

Cálculo da matriz Z

É fácil ver que as colunas de Q_1 e Q_2 formam, respectivamente, uma base ortogonal para a imagem e o núcleo do espaço gerado pelas linhas de A.

Assim, uma escolha natural para a matriz Z é Q_2.

Cálculo da matriz Z

Note que, usando, por exemplo, técnicas do tipo Newton para calcular o passo p_z (Passo 4), temos que

$$\boxed{\text{cond}(Z^T \nabla^2 f(x_k) Z) \leq \text{cond}(\nabla^2 f(x_k))(\text{cond}(Z))^2.}$$

Assim, como este limitante é justo (pode valer a igualdade), quando o número de condição da matriz Z é grande, a matriz Hessiana projetada $Z^T \nabla^2 f(x_k) Z$ poderia ser mal-condicionada mesmo quando a matriz Hessiana $\nabla^2 f(x_k)$ é bem-condicionada.

Cálculo da matriz Z

No entanto, utilizando a fatoração LQ para determinar Z, temos $\text{cond}(Z) = 1$. Ou seja,

$$\boxed{\text{cond}(Z^T \nabla^2 f(x_k) Z) \leq \text{cond}(\nabla^2 f(x_k)),}$$

o que mostra que, neste caso, o condicionamento do problema (2) não é pior que o condicionamento do problema (1).

Portanto, uma grande vantagem do uso desta técnica é o fato de que esta escolha de Z não piora o condicionamento do problema de encontrar uma solução para (1).

Cálculo da matriz Z

Para formar a matriz Q usando a fatoração LQ é necessário calcular o produto de transformações ortogonais usadas para triangularizar a matriz A.

Na implementação de métodos que usam a matriz Z somente em produtos matriz-vetor, pode não ser eficiente calcular a matriz Q (e, consequentemente, Z) explicitamente, principalmente quando o número de restrições é pequeno.

Neste caso é mais interessante armazenar as transformações ortogonais de forma compacta e aplicá-las ao vetor apropriado para calcular o produto matriz-vetor sempre que necessário.

Cálculo da matriz Z

Apesar das vantagens do uso da fatoração LQ para o cálculo da matriz Z, esta é uma operação computacionalmente custosa.

Uma alternativa é a técnica de redução de variáveis.

Tomando a matriz A com linhas linearmente independentes, podemos particioná-la em

$$\boxed{A = (V \quad U),}$$

onde $V \in R^{m \times m}$ é não-singular e $U \in R^{m \times (n-m)}$.

Cálculo da matriz Z

Usando esta partição de A, a matriz Z pode ser definida como

$$\boxed{Z = \begin{pmatrix} -V^{-1}U \\ I \end{pmatrix},}$$

onde I é a matriz identidade $(n-m) \times (n-m)$.

Cálculo da matriz Z

Note que

$$\boxed{AZ = \begin{pmatrix} V & U \end{pmatrix} \begin{pmatrix} -V^{-1}U \\ I \end{pmatrix} = -VV^{-1}U + U = 0,}$$

como desejamos.

Cálculo da matriz Z

Geralmente a matriz Z, tal como descrita acima, não é armazenada explicitamente.

Produtos matriz-vetor envolvendo Z e Z^T podem ser obtidos resolvendo sistemas lineares envolvendo V e V^T.

A grande dificuldade desta técnica é encontrar uma maneira eficiente de particionar A de forma que a matriz V escolhida fique relativamente bem-condicionada.