Questões de Computação Científica no contexto da Otimização

Leonardo D. Secchin

Universidade Federal do Espírito Santo
Mestrado em Informática

Abril de 2009

Sumário

Sumário

Introdução e Motivação

Dada uma função $F : D \subset \mathbb{R}^n \to \mathbb{R}^m$,
$F(x) = (F^1(x), \cdots, F^m(x))$, a *jacobiana de F em x* é a matriz

$$F'(x) = \left[\frac{\partial F^i}{\partial x_j}(x)\right]_{m\times n}$$

onde supomos que tais derivadas existam. Também, dada uma função $f : D \subset \mathbb{R}^n \to \mathbb{R}$ que possui todas derivadas parciais até segunda ordem, a *hessiana de f em x* é dada por

$$f''(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]_{n\times n}.$$

Em otimização, alguns métodos para problemas não lineares exigem o cálculo de hessianas. No atual contexto, é conveniente usar informações de esparsidade para cálculo dessas matrizes num dado x.

Um exemplo são os *Métodos de Newton*. Considere o problema

$$\min_x f(x) \quad \text{s.a.} \quad x \in \mathbb{R}^n, \tag{1}$$

com $f : \mathbb{R}^n \to \mathbb{R}$ duas vezes continuamente diferenciável e tal que sua hessiana seja semidefinida positiva $\forall x$. Os pontos estacionários deste problema são caracterizados como soluções de

$$f'(x) = 0.$$

Agora, seja x^k uma aproximação a um ponto estacionário $\overline{x}$ do problema (1). A aproximação x^{k+1} seguinte é uma solução de

$$f'(x^k) + f''(x^k)(x - x^k) = 0.$$

Supondo agora que $f''(x^k)$ seja não singular, temos

$$x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k), \quad k = 0, 1, 2, \cdots$$

Este é o esquema iterativo geral para Métodos de Newton.

Sumário

Cálculo de jacobianas

O método para o cálculo de jacobianas apresentado aqui é o das *diferenças finitas*. Dada $F : D \subset \mathbb{R}^n \to \mathbb{R}^m$, podemos estimar a k-ésima coluna de $F'(x)$ por

$$\frac{1}{t}[F(x + te_k) - F(x)] \approx F'(x)e_k$$

fazendo medições para $|t| > 0$ pequeno, onde e_k é o k-ésimo vetor canônico de $\mathbb{R}^n$. Aqui são necessários n processos deste, um para cada coluna de $F'(x)$.

Explorando esparsidade de $F'(x)$

Podemos tentar diminuir o custo para o cálculo de $F'(x)$ explorando a esparsidade dessa matriz, caso seja conhecida. Dizemos que duas colunas de uma matriz são *estruturalmente ortogonais* se elas não possuem elementos não nulos na mesma linha. Agora, supondo que $F'(x)$ esteja particionada em p grupos de colunas estruturalmente ortogonais, "somamos" as colunas em cada grupo j e então $F'(x)$ pode ser determinada por p avaliações da forma $F'(x)v^j$, onde para o grupo j temos

$$v_i^j = \begin{cases} 1 & \text{se a } i\text{-ésima coluna pertence ao grupo } j \\ 0 & \text{caso contrário} \end{cases} .$$

Para isso, utiliza-se por exemplo o método de diferenças finitas exposto anteriormente.

Vejamos um exemplo. Seja dada $F : \mathbb{R}^4 \to \mathbb{R}^3$, $F(x) = (x_1 x_2 - x_1,\ x_3^2 + 2x_2,\ x_1^3 x_4)$, e queiramos $F'(x^0 = (1,1,1,1))$, cuja estrutura de esparsidade é

$$\begin{bmatrix} 0 & \cdot & 0 & 0 \\ 0 & \cdot & \cdot & 0 \\ \cdot & 0 & 0 & \cdot \end{bmatrix}.$$

Os grupos $J_1 = \{1,2\}$ e $J_2 = \{3,4\}$ são de colunas estruturalmente ortogonais e daí $v^1 = (1,1,0,0)$ e $v^2 = (0,0,1,1)$. Podemos então estimar $F'(x^0)$ pelo método das diferenças finitas na matriz compacta 3x2 $\tilde{F}'(x^0)$ cujas colunas são as "somas" das colunas em cada grupo J_i:

$$\frac{1}{t}[F(x^0 + tv^i) - F(x^0)] \approx \tilde{F}'(x^0)[1\ \ 1]^t, \quad i = 1, 2.$$

Naturalmente, a matriz $\tilde{F}'(1,1,1,1)$ é menos esparsa que sua original. Sua estrutura de esparsidade é

$$\begin{bmatrix} \cdot & 0 \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}.$$

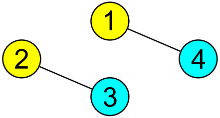
Por fim, o processo de obtenção de $F'(1,1,1,1)$ a partir de $\tilde{F}'(1,1,1,1)$ é trivial, visto que não perdemos informações quando da soma das colunas estruturalmente ortogonais de $F'(1,1,1,1)$.

Conexão com o problema de coloração de grafos

É interessante que o número p de grupos de colunas estruturalmente ortogonais seja pequeno. Minimizar p pode ser visto como um problema de coloração de grafos: as colunas são os vértices, e as arestas dizem se duas colunas não são estruturalmente ortogonais. No exemplo anterior, a estrutura de esparsidade de $F'(1,1,1,1)$ é

$$\begin{bmatrix} 0 & \cdot & 0 & 0 \\ 0 & \cdot & \cdot & 0 \\ \cdot & 0 & 0 & \cdot \end{bmatrix}.$$

Associamos o *grafo de interseção de colunas* não orientado $G(V,E)$ onde $V = \{1,2,3,4\}$ e $E = \{\{1,4\}, \{2,3\}\}$.



Note que as partições $J_1 = \{1,2\}$ e $J_2 = \{3,4\}$ correspondem a uma coloração ótima.

Em geral vale o resultado:

Teorema

$J_1, \cdots, J_p$ é uma coloração para $G(V, E)$ se, e somente se, $J_1, \cdots, J_p$ induz uma partição em colunas estruturalmente ortogonais de $F'(x)$

Há ainda problemas de coloração de grafos especializados. Por exemplo, para o cálculo de hessianas, Gebremedhin e outros (2006) discutem problemas de coloração

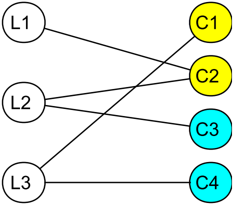
- acíclica, onde além de dois vértices adjacentes terem cores distintas (coloração de distância 1), todo ciclo tem no mínimo 3 cores;
- e de estrela, onde além da coloração de distância 1, todo caminho com 4 vértices tem no mínimo 3 cores.

Vejamos detalhadamente a seguir uma outra abordagem.

Uma forma mais eficiente de tratar o problema é associar um grafo bipartido para representar a matriz jacobiana (Gebremedhin, 2005). No exemplo anterior, a estrutura de esparsidade de $F'(1,1,1,1)$ é

$$\begin{bmatrix} 0 & \cdot & 0 & 0 \\ 0 & \cdot & \cdot & 0 \\ \cdot & 0 & 0 & \cdot \end{bmatrix}.$$

Associamos o grafo bipartido $G = (V_1 \cup V_2, E)$, $V_1 \cap V_2 = \emptyset$, onde V_1 e V_2 representam as linhas e colunas de $F'(1,1,1,1)$, respectivamente, e $(i,j) \in E \Leftrightarrow$ a entrada da posição (i,j) em $F'(1,1,1,1)$ for não nula.



Aqui, duas colunas i, j são estruturalmente ortogonais se, e só se, seus vértices Ci, Cj estão a uma distância ≥ 2. Por isso, a coloração usada é a de distância 2, e somente nos vértices de V_2.

Sumário

Referências

Hendrickson B., Pothen A. *Combinatorial scientific computing: the enabling power of discrete algorithms in computacional science,* Em Lecture Notes in Computer Science 4395:260-280, 2007.

Gebremedhin A. H., Tarafdar A. Manne F., Pothen A. *New acyclic and star coloring algorithms with application to computing hessians.* ???, 2008.

Birgin, E. J. G. *Diferenciação computacional e aplicações.* Tese de Doutorado, Universidade Estadual de Campinas, 1998.

Steihaug T., Hossain A. K. M. S. *Graph coloring and the estimation of sparse jacobian matrices with segment columns.* ???, 1997.

Coleman T. F., Moré J. J. *Estimation of sparse hessian matrices and graph coloring problems.* Mathematical Programming 28:243-270, 1984.

Solodov M., Izmailov. *Otimização volume 2. Métodos Computacionais.* IMPA, 2007.