

Otimização, derivadas e coloração

Rafael Santos Coelho

UFES

15 de abril de 2009

Sumário

- 1 Introdução e motivação
- 2 Conceitos básicos
 - Matrizes jacobianas e hessianas
- 3 Abordagens para o cálculo de matrizes derivativas
 - Hand-coding
 - Diferenciação finita
 - Diferenciação automática
- 4 Coloração de grafos
 - O que é um grafo e o que significa colorir um?
 - Complexidade algorítmica
 - Detalhes a serem ponderados
 - Relação com a diferenciação finita para o cálculo de jacobianas
- 5 Considerações finais
- 6 Referências

Introdução e motivação

- Via de regra, as soluções para problemas não-lineares de otimização modelados via equações diferenciais ordinárias ou parciais – importantíssimas para o desenvolvimento de várias áreas do conhecimento – requerem o cálculo de matrizes derivativas, especialmente jacobianas e hessianas
- Em aplicações reais, onde existem restrições de tempo de resposta e de consumo de recursos computacionais, propriedades como esparsidade e simetria podem e devem ser exploradas para tornar mais eficiente a determinação dos elementos não-nulos dessas matrizes

Introdução e motivação

- Com o tempo, estudos [3] mostraram que o problema de se calcular otimamente matrizes derivativas pode ser reescrito na forma de um problema de particionamento de matrizes
- Em seguida, provaram que o problema de particionamento de matrizes, na realidade, podia ser convertido em um problema especializado de coloração de grafos
- Desde então, a coloração de grafos tem se revelado muito útil, enquanto estratégia de abstração, na formulação, análise e projeto de algoritmos para a computação de matrizes derivativas

Conceitos básicos – Matrizes jacobianas e hessianas

- Sejam $\mathcal{F} : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma função vetorial “suave” e $z \in \mathcal{D}$.
 Defina-se $\mathcal{J}(\mathcal{F}, z)$ a matriz jacobiana associada a \mathcal{F} e calculada em z

$$\mathcal{J}(\mathcal{F}, z) = \begin{bmatrix} \frac{\partial \mathcal{F}_1(z)}{\partial x_1} & \frac{\partial \mathcal{F}_1(z)}{\partial x_2} & \cdots & \frac{\partial \mathcal{F}_1(z)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{F}_n(z)}{\partial x_1} & \frac{\partial \mathcal{F}_n(z)}{\partial x_2} & \cdots & \frac{\partial \mathcal{F}_n(z)}{\partial x_n} \end{bmatrix}$$

- Sejam $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função escalar “suave” e $z \in \mathcal{D}$.
 Defina-se $\mathcal{H}(f, z)$ a matriz hessiana associada a f e calculada em z

$$\mathcal{H}(f, z) = \begin{bmatrix} \frac{\partial f(z)}{\partial x_1 \partial x_1} & \frac{\partial f(z)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f(z)}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(z)}{\partial x_n \partial x_1} & \frac{\partial f(z)}{\partial x_n \partial x_2} & \cdots & \frac{\partial f(z)}{\partial x_n \partial x_n} \end{bmatrix}$$

Hand-coding – Visão geral

- Definição: consiste em programar separadamente a(s) rotina(s) usada(s) para o cômputo da matriz
- Vantagens: eficiência e alta precisão
- Desvantagens: elevado risco de erros e grande esforço de implementação dependendo do tamanho e da complexidade do problema

Diferenciação finita – Visão geral

- Definição: consiste em utilizar aproximações de diferenças finitas para o cálculo da matriz, dispensando a necessidade de se ter uma rotina estática para fazer isso
- Vantagens: praticidade e economia de memória
- Desvantagens: baixa precisão e dificuldade de predição de um valor adequado para o parâmetro de diferenciação

Diferenciação finita – Formulação

- Sejam $\mathcal{F} : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma função vetorial n vezes continuamente diferenciável, $\varepsilon \in \mathbb{R}_+^*$, $v \in \mathbb{R}^n$ e $\mathcal{J}(\mathcal{F}, x + \varepsilon v)$ a matriz jacobiana associada a \mathcal{F} e calculada em $x + \varepsilon v$, manipulando-se a versão multivariada do teorema de Taylor, pode-se concluir que

$$\mathcal{J}(\mathcal{F}, x + \varepsilon v)v \approx \frac{\mathcal{F}(x + \varepsilon v) - \mathcal{F}(x)}{\varepsilon} + \mathcal{O}(\varepsilon) \quad (1)$$

- ε é chamado de parâmetro de diferenciação e v é chamado de vetor direcional
- A relação (1) é uma fórmula de diferença finita adiantada de passo real que implica em erros de truncamento e erros de cancelamento

Diferenciação finita – Erros de truncamento

- Erros de truncamento surgem a partir do truncamento da série de Taylor e geralmente são proporcionais à ordem de grandeza do parâmetro de diferenciação. Tais erros podem ser atenuados diminuindo-se o valor do parâmetro de diferenciação ou aumentando-se a ordem de aproximação da fórmula de diferença finita. Entretanto, a última opção provoca um aumento do número de avaliações da função $\mathcal{F}(x)$, o que conseqüentemente implica em um maior custo computacional

Diferenciação finita – Erros de cancelamento

- Erros de cancelamento surgem quando dois números muito próximos são subtraídos, restando apenas os dígitos mais à direita. Isso acontece quando se escolhe um valor muito diminuto para o parâmetro de diferenciação. Tais erros possuem uma natureza acumulativa, o que significa que quanto maior for o número de operações propagadoras, maiores serão as chances de eles “contaminarem” os dígitos à esquerda mais significativos, comprometendo, portanto, a precisão dos resultados
- Uma forma de se livrar por completo do erro de cancelamento é a estratégia da derivada de passo complexo (*complex step derivative*) [4], que incorpora variáveis complexas na fórmula de diferenças finitas

Diferenciação finita – Parâmetro de diferenciação

- Se o parâmetro de diferenciação ε for muito grande, a derivada é pobremente aproximada, porém se ele for muito pequeno o resultado da aproximação pode ser gravemente afetado por erros de cancelamento
- A melhor escolha para o parâmetro de diferenciação deve ser feita de maneira a balancear esses fatores contrários
- “Estimar devidamente o valor para o parâmetro de diferenciação é tanto uma ciência quanto é uma arte.”

Diferenciação finita – Calculando a matriz jacobiana

- Sejam $\mathcal{F} : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma função vetorial “suave”, $z \in \mathcal{D}$, $\mathcal{J}(\mathcal{F}, z)$ a matriz jacobiana associada a \mathcal{F} e calculada em z e j_{ij} o elemento da i -ésima linha e j -ésima coluna de $\mathcal{J}(\mathcal{F}, z)$. Analisando-se a relação (1), sabe-se que

$$\begin{bmatrix} \sum_{i=1}^n j_{1i} v_i \\ \vdots \\ \sum_{i=1}^n j_{ni} v_i \end{bmatrix} = \mathcal{J}(\mathcal{F}, z) v \approx \frac{\mathcal{F}(x + \varepsilon v) - \mathcal{F}(x)}{\varepsilon}$$

- Sejam j_i a i -ésima coluna de $\mathcal{J}(\mathcal{F}, z)$ e $e = \{e_1, \dots, e_n\}$ a base canônica de \mathbb{R}^n . Então, é natural esperar que

$$j_i = \mathcal{J}(\mathcal{F}, z) e_i \quad i = 1, \dots, n$$

Diferenciação finita – Calculando a matriz jacobiana

- Sejam $\mathcal{F} : \mathcal{D} \subset \mathbb{R}^4 \rightarrow \mathbb{R}^4$ uma função vetorial “suave”, $z \in \mathcal{D}$ e $\mathcal{J}(\mathcal{F}, z)$ a matriz jacobiana associada a \mathcal{F} e calculada em z cujo padrão de esparsidade é conhecido a priori por

$$\mathcal{J}(\mathcal{F}, z) = \begin{bmatrix} j_{11} & 0 & 0 & j_{14} \\ 0 & j_{22} & j_{23} & j_{24} \\ j_{31} & j_{32} & 0 & 0 \\ 0 & j_{42} & j_{43} & 0 \end{bmatrix}$$

- Para $e = \{e_1, e_2, e_3, e_4\}$, com $e_2 = (0, 1, 0, 0)$, fica claro que

$$j_2 = \begin{bmatrix} 0 \\ j_{22} \\ j_{32} \\ j_{42} \end{bmatrix} = \mathcal{J}(\mathcal{F}, z)e_2 \approx \frac{\mathcal{F}(x + \varepsilon e_2) - \mathcal{F}(x)}{\varepsilon}$$

Diferenciação finita – Calculando a matriz jacobiana

- Por indução, pode-se concluir que o cálculo de uma matriz jacobiana de ordem n requer, no máximo, n produtos matriz-vetor. Porém, como cada produto matriz-vetor é aproximado por diferenças finitas, isso corresponde fundamentalmente a, no máximo, n avaliações da função $\mathcal{F}(x)$
- Para problemas complexos e de grande escala, repetidas avaliações da função $\mathcal{F}(x)$ podem ser computacionalmente custosas. Logo, o ideal seria que o número de avaliações pudesse ser, de alguma forma, reduzido: é justamente esse o *insight* da coloração de grafos!

Diferenciação automática - Visão geral

- Definição: consiste em decompor a função a ser derivada em uma sequência de funções elementares e então utilizar a regra da cadeia para realizar a derivação de fato, dispensando a necessidade de se ter uma rotina estática para fazer isso
- Vantagens: praticidade, alta precisão e eficiência
- Desvantagem: elevado gasto de memória

O que é um grafo e o que significa colorir um?

- Um grafo $\mathcal{G} = (V, E)$ pode ser definido como um par ordenado (V, E) , onde V é um conjunto de vértices e E é um conjunto de arestas do tipo $\{v_1, v_2\}$, com $v_1, v_2 \in V$
- Dado um grafo $\mathcal{G} = (V, E)$, em geral, o problema da coloração ótima de grafos se resume em encontrar uma coloração de vértices própria $\mathcal{C} : V \rightarrow K$, tal que para todo par de vértices (v_1, v_2) , com $v_1, v_2 \in V$, se $\{v_1, v_2\} \in E$, então $\mathcal{C}(v_1) \neq \mathcal{C}(v_2)$ e a cardinalidade $|K| = \chi(\mathcal{G})$, onde $\chi(\mathcal{G})$ é o número cromático de \mathcal{G}

Complexidade algorítmica

- Foi provado que o problema da coloração de grafos é NP-Completo e, portanto, é improvável que exista um algoritmo que resolva otimamente qualquer instância do problema em um tempo de execução expresso como uma função polinomial do tamanho do grafo
- No entanto, em 1984, Goldfarb e Toint [2] mostraram que para determinadas classes de grafos, a saber grafos que modelam malhas estruturadas oriundas de discretizações de diferenças finitas, é possível, sim, desenvolver um algoritmo polinomial ótimo

Detalhes importantes a serem ponderados

Ao se projetar um algoritmo heurístico de coloração de grafos para otimizar o cálculo aproximado de matrizes derivativas é preciso levar em consideração vários fatores:

- A abordagem a ser usada para a execução do cálculo, isto é, diferenciação finita ou diferenciação automática
- A natureza da matriz derivativa em si, ou seja, se se trata de uma matriz jacobiana (não-simétrica) ou hessiana (simétrica), quão esparsa é a matriz, qual é a estrutura da esparsidade da matriz e qual é o tamanho da parte não-constante da matriz
- A natureza do cálculo, a saber, se se trata do cálculo integral ou parcial da matriz derivativa

Detalhes importantes a serem ponderados

- Que tipo de partição será realizada, isto é, unidirecional ou bidirecional
- O modelo de programação do algoritmo heurístico, isto é, sequencial ou paralelo
- Se é possível relaxar as exigências da coloração para tornar o algoritmo menos custoso

Relação com a diferenciação finita para o cálculo de jacobianas

Sejam $D = [d_{ij}]_{l \times c}$ uma matriz qualquer e o conjunto de suas colunas $\mathcal{C} = \{k_1, k_2, \dots, k_c\}$

- Define-se $\sigma(k_i) = \{j \in \{1, 2, \dots, l\} : d_{ji} \neq 0\}$ a estrutura de k_i
- Duas colunas k_i e k_j são ditas estruturalmente ortogonais se $\sigma(k_i) \cap \sigma(k_j) = \emptyset$
- Define-se que $\mathfrak{B}(D) = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\alpha\}$ é uma partição de D em colunas estruturalmente ortogonais, se

$$\textcircled{1} \bigcup_{i=1}^{\alpha} \mathcal{C}_i = \mathcal{C}$$

$\textcircled{2}$ Para todo par $(\mathcal{C}_i, \mathcal{C}_w)$, com $i \neq w$ e $i, w \in \{1, 2, \dots, \alpha\}$, tem-se $\mathcal{C}_i \cap \mathcal{C}_w = \emptyset$

$\textcircled{3}$ Para todo $i \in \{1, 2, \dots, \alpha\}$, os elementos de \mathcal{C}_i são colunas estruturalmente ortogonais entre si

Relação com a diferenciação finita para o cálculo de jacobianas

Sejam $D = [d_{ij}]_{I \times c}$ uma matriz derivativa qualquer e o conjunto de suas colunas $\mathcal{C} = \{k_1, k_2, \dots, k_c\}$

- Define-se $\mathcal{G}_c(D) = (V, E)$ o grafo de intersecção de colunas de D , com $V = \mathcal{C}$ e $E = \{(k_i, k_j) : \sigma(k_i) \cap \sigma(k_j) \neq \emptyset\}$
- Em 1984, Curtis, Powell e Reid [1] demonstraram que o problema de se otimizar a computação de D via diferenciação finita é equivalente a obter otimamente uma coloração própria de vértices para o grafo $\mathcal{G}_c(D)$, isso significa encontrar uma partição de colunas $\mathfrak{P}(\mathcal{C})$ estruturalmente ortogonais para D , com $|\mathfrak{P}(\mathcal{C})| = \chi(\mathcal{G}_c(D))$
- A idéia é que uma vez encontrada uma partição $\mathfrak{P}(\mathcal{C}) = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k\}$ de colunas estruturalmente ortogonais para D , com $k < c$, então para $i \in \{1, 2, \dots, k\}$, tem-se que as colunas de \mathcal{C}_i podem ser computadas concomitantemente em uma única avaliação da função $\mathcal{F}(x)$

Relação com a diferenciação finita para o cálculo de jacobianas

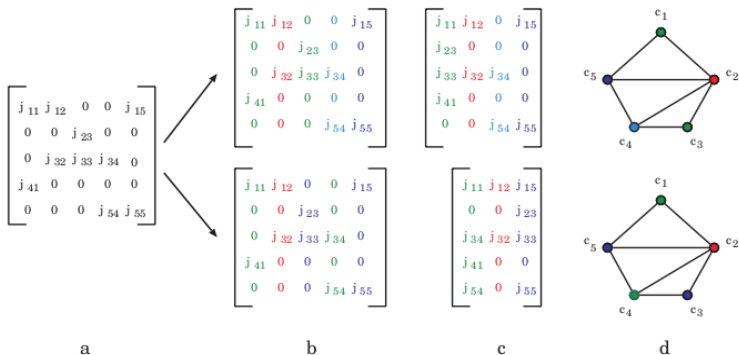


Figura: (a) Matriz jacobiana $\mathcal{J}(\mathcal{F}, z)$ associada a $\mathcal{F} : \mathcal{D} \subset \mathbb{R}^5 \rightarrow \mathbb{R}^5$, com $z \in \mathcal{D}$
 – (b) Partições coloridas de $\mathcal{J}(\mathcal{F}, z)$ em colunas estruturalmente ortogonais – (c)
 Versão comprimida da matriz jacobiana $\mathcal{J}(\mathcal{F}, z)$ – (d) Grafo de intersecção de
 colunas $\mathcal{G}_c(\mathcal{J}(\mathcal{F}, z))$ da matriz jacobiana





Considerações finais

- Diversos problemas em Computação Científica possuem, direta ou indiretamente, implicações de cunho combinatorial, e, na maioria dos casos, são passíveis de serem modelados em termos da Teoria de Grafos
- Nesse contexto, um nicho de problemas que possuem exigências de particionamento de objetos segundo regras pré-estipuladas acabam naturalmente sendo tratados como variações de problemas de colorações de grafos

Considerações finais

- Em particular, a coloração de grafos tem se mostrado imprescindível no estudo de estratégias algorítmicas para o cálculo aproximado de matrizes derivativas, recorrentemente presentes em problemas de otimização baseados em equações diferenciais
- Todavia, dado que o problema de coloração de grafos, em sua formulação mais abrangente, se enquadra na categoria NP-Completo, atualmente a única saída de propósito geral se resume em algoritmos heurísticos, o que configura tal problema ainda como um grande desafio para a comunidade científica, dada a sua grande relevância nos meios acadêmico, comercial e industrial

Referências

-  A. R. Curtis, M. J. D. Powell, and J. K. Reid.
On the estimation of sparse Jacobian matrices.
J. Inst. Math. Appl., 13:117–119, 1974.
-  D. Goldfarb and P. L. Toint.
Optimal estimation of Jacobian and Hessian matrices that arise in finite difference calculations.
Math. Comp., 43:69–88, 1984.
-  B. Hendrickson and A. Pothen.
Combinatorial Scientific Computing: The Enabling Power of Discrete Algorithms in Computational Science.
Lecture Notes in Computer Science, 4395:260–280, 2007.
-  J. R. R. A. Martins, P. Sturdza, and J. J. Alonso.
The complex-step derivative approximation.
ACM Trans. Math. Softw., 29:245–262, 2003.