INF1608 – Análise Numérica
Projeto: Método Gradiente Descendente

Leonardo Seperuelo Duarte
lduarte@tecgraf.puc-rio.br
Departamento de Informática, PUC-Rio

Descrição

O método do Gradiente Descendente (método do Máximo Declive) é um método iterativo para determinação de mínimos de funções de n variáveis, $f(\mathbf{x})$. Esse método pode ser expresso pelo seguinte pseudo-código:

$$
\begin{array}{l}
\textbf{for } k = 0, 1, 2, 3, \cdots \\
\quad \mathbf{v} = \nabla f(\mathbf{x}_k) \\
\quad \text{Minimize} \quad f(\mathbf{x} - \alpha \mathbf{v}) \quad \text{for scalar} \quad \alpha = \alpha^* \\
\quad \mathbf{x}_{k+1} = \mathbf{x}_k - \alpha^* \mathbf{v}
\end{array}
$$

Isto é, o método parte de uma estimativa inicial, $\mathbf{x}_0$, e avança na direção contrária à do gradiente da função:

$$
\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_0} \\ \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_{n-a}} \end{bmatrix}
$$

Existem diferentes heurísticas para determinar o tamanho do passo, no sentido inverso do gradiente. No pseudo-código acima, o passo é tomado de tal forma que minimiza o valor da função $f(\mathbf{x})$ na direção de busca. Portanto, a cada iteração, recai-se num problema de otimização de uma função escalar (achar o mínimo da função ao longo de uma direção). Para isso, podemos empregar o método da Interpolação Parabólica Sucessiva, que parte de três estimativas iniciais, r, s e t:

$$
u = \frac{r+s}{2} - \frac{(f(s) - f(r))(t-r)(t-s)}{2[(s-r)(f(t) - f(s)) - (f(s) - f(r))(t-s)]}
$$

O valor de u substitui a estimativa menos recente r, ciclando as estimativas, ou a pior estimativa entre r, s e t.

Tarefa

O objetivo deste projeto é implementar o método do Gradiente Descendente, usando passos constantes e passos determinados pela Interpolação Parabólica Sucessiva (IPS), como discutido acima. Para testar a implementação, considere determinar o mínimo das seguintes funções:

- $f(x, y) = x^4 + y^4 + 2x^2y^2 + 6xy - 4x - 4y + 1$
- $f(x, y) = 100(y - x^2)^2 + (x - 1)^2$ (função de Rosenbrock em 2D)

Em seguida, experimente sua implementação para tratar a função de Rosenbrock em 3D:

- $f(x, y, z) = 100(y - x^2)^2 + (x - 1)^2 + 100(z - y^2)^2 + (y - 1)^2$

A função de Rosenbrock é popularmente utilizada para testar implementações de métodos de otimização baseados em gradientes. A função é unimodal, tendo o mínimo global dentro de um vale estreito; alcançar o vale é rápido, mas convergir para o mínimo dentro do vale tende a ser difícil.

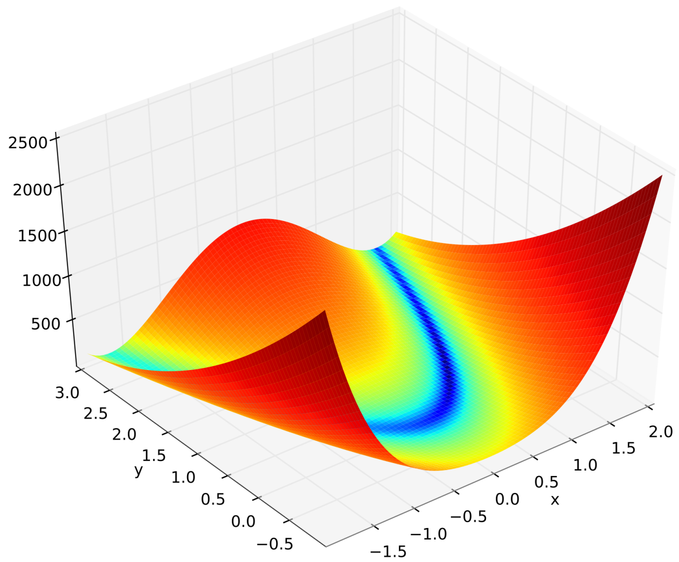

Figura 1: Função de Rosenbrock em 2D; o mínimo ocorre em $f(1, 1) = 0$ (imagem da Wikipedia).

Análise

Ao desenvolver seu trabalho e testá-lo, procure, baseado em experimentos computacionais, responder as seguintes perguntas:

- Para as funções 2D, como a convergência do método varia com as estimativas iniciais? Quantas iterações foram necessárias para uso de passos constantes e passos dada pela IPS?
- No uso do método de IPS, qual estratégia apresentou melhor resultado: u substituir a estimativa menos recente ou substituir a pior estimativa?
- Sua implementação se estende para n dimensões? Como foi o comportamento do método para a função 3D?