Otimização: Método Gradiente Descendente

Jayme Augusto Avelino de Paiva – 2210289
João Gabriel da Cunha Vasconcellos – 2211302
Luiz Felipe da Silva Seibel – 2110799

Introdução

Este projeto visa estudar o método do gradiente descendente (método do máximo declive) para encontrar mínimos de funções quaisquer, em 'n' dimensões. O objetivo é comparar os resultados obtidos pelo método, utilizando o algoritmo da Interpolação Parabólica Sucessiva para determinação do passo durante as iterações, com os resultados obtidos utilizando passos constantes. Além disso, analisamos as diferenças nos resultados entre diferentes implementações do método da Interpolação Parabólica Sucessiva.

Desenvolvimento

O método gradiente descendente consiste em calcular o vetor gradiente da função no ponto da iteração atual para dar um passo na direção contrária do gradiente, atualizando o valor do ponto atual, para dessa forma, após 'n' iterações, se aproximar do mínimo da função.

Para iniciar o algoritmo é necessário passar um valor inicial, esse valor pode ser um valor qualquer do domínio da função, mas quanto mais perto do mínimo da função esse valor estiver, mais rápido o método irá convergir para o mínimo, além disso, dependendo do comportamento da função na região do ponto inicial passado, o método pode acabar convergindo para um mínimo local ao invés do mínimo global. Também deve-se notar que dependendo da função e forma como o passo é calculado, pode ocorrer cenários que o método não convirja. Isso mostra a importância de calcular o passo de maneira que o método consiga achar um valor e que esse valor seja o mínimo global.

Embora seja impossível garantir que o método converge para o mínimo global, existem diversas heurísticas para tentar fazer com que ele convirja e para fazer com que caso ele convirja para um mínimo local, ele consiga sair desse mínimo local e achar o verdadeiro mínimo global. Neste trabalho iremos ver uma heurística que tenta fazer com que o método convirja, sem se preocupar se está achando um mínimo global ou local, essa heurística se resume a calcular o passo de cada iteração de forma a minimizar o valor de $f(X - aG)$, onde 'f' é a função que estamos minimizando, X é o vetor que guarda as coordenadas do ponto da iteração atual, 'a' é o valor do passo que queremos encontrar e G o valor do vetor gradiente da função no ponto X. A heurística é usar esse valor de 'a' que minimiza $f(X - aG)$como o passo que a função dará em direção contrária ao gradiente, fazendo com que Xi (novo valor de X após a iteração) seja igual à $X - AG$.

Para achar o valor de 'a' que minimiza f(X-aG) iremos usar o método da Interpolação Parabólica Sucessiva, esse método consiste em passar 3 valores iniciais para achar o mínimo da função, após isso achamos a parábola que passa por esses 3 valores e usamos o mínimo dessa parábola para para substituir um dos 3 valores iniciais, depois repetimos o processo 'n' vezes, até que o valor mínimo achado não esteja melhorando em relação ao valor mínimo achado na iteração anterior ou até que cheguemos a um limite de iterações passados no código, para evitar ciclos infinitos, já que não é garantido que esse método sempre convirja. A escolha de qual dos 3 valores da iteração será substituído pelo novo valor achado também pode variar, iremos analisar a diferença entre substituir o valor mais antigo e substituir o pior valor.

Para calcular o vetor gradiente usamos o método das diferenças finitas de segunda ordem :

$$\mathbf{f}'(\mathbf{x}) = \frac{\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}-\mathbf{h})}{\mathbf{2h}} - \frac{h^2}{6}f'''(c), \text{ onde } c \in [x-h, x+h]$$

Figura 1: Método de segunda ordem para calcular derivadas.

Resultados e Análise

Para estudar o método iremos usar 3 funções diferentes, sendo duas de 2 dimensões e uma de 3 dimensões.

Inicialmente, avaliamos $f(x, y) = x^4 + y^4 + 2x^2y^2 + 6xy - 4x - 4y + 1$, resolvendo de forma analitica achamos que o mínimo local dessa função é (½, ½) onde a função vale -1.25, enquanto os mínimos globais são (1.13264, -0.465972) e (-0.465972, 1.13264) onde a função vale -2.58333, todos valores aproximados.

Em seguida usamos o método do gradiente descendente para achar o mínimo da função usando passo constante, iremos fazer usando o passo sendo 0.0001, 0.01 e 1 para analisar o resultado para cada um dos casos, todos os casos começando no ponto (0, 0).

- Passo 0.0001: Obtemos valores muito próximos do mínimo local, aproximadamente (0.494, 0.494) que possui o valor de aproximadamente -1.249, contudo para isso demorou 4121 iterações ao todo.
- Passo 0.01: Novamente o método convergiu para valores muito próximos do mínimo local, nesse caso foi aproximadamente (0.4996, 0.4996) cujo valor da função é aproximadamente -1.249, dessa vez demorou apenas 59 iterações. Essa queda no número de iterações acontece devido ao passo ser maior, assim é necessário menos iterações para avançar a mesma "distância" em direção ao mínimo da função.
- Passo 1: Devido ao passo ser muito grande, o método divergiu completamente de qualquer mínimo, achando valores como (1177585836, 1177585836) em 4 iterações que deu 7.691840778749311e+36 de valor "mínimo". Isso mostra uma das desvantagens de usar passos constantes, dependendo do valor do passo o método pode divergir completamente do valor esperado. O motivo do método parar em apenas 4 iterações é porque chega em um ponto em que a função é praticamente vertical, fazendo o valor tender a infinito, onde na fórmula de calcular o gradiente esse infinito-infinito acaba dando 0 e o valor de X não avança mais, fazendo com que o método pare.

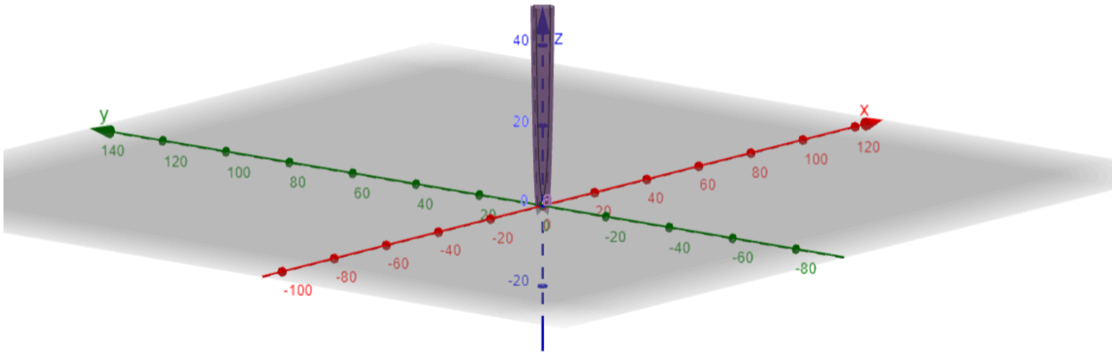

Figura 2: Plot da função que mostra como o gradiente se aproxima do infinito ao se afastar do mínimo

Na tentativa de evitar a divergência, repetimos o método começando no ponto (-1,1), um ponto mais próximo de um dos mínimos globais.

- Passo 0.0001: O método achou valores muito próximos do mínimo global em 4393 iterações.
- Passo 0.01: O método novamente achou valores muito próximos do mínimo global, mas desta vez em apenas 73 iterações.
- Passo 1: Mais uma vez o método divergiu completamente de algum mínimo devido ao passo ser muito grande.

Ao realizar mais testes com pontos mais distantes foi observado que o método diverge completamente, mesmo com passos pequenos, por exemplo, no ponto (-50,50) o método ainda consegue convergir para um dos mínimos globais, agora no ponto (-100,100) o método diverge completamente. O motivo disso é que nesses pontos mais distantes, a função é muito íngreme, o que causa o vetor gradiente possuir valores muito grandes, dessa forma o método acaba quebrando, pois mesmo com passos pequenos, o método dá saltos na função, passando por cima do mínimo e divergindo cada vez mais.

Em seguida analisamos as funções usando substituindo o pior valor e substituindo o valor mais antigo

- Substituindo valor mais antigo e iniciando no ponto (0, 0): O método demora apenas 2 iterações, mas gasta ao todo 11 iterações na IPS, e converge para o mínimo local.
- Substituindo pior valor e iniciando no ponto (0, 0): Resultados iguais ao teste anterior.
- Substituindo valor mais antigo e iniciando no ponto (-1, 1): O método demora 7 iterações, gastando 27 na IPS e acha um dos mínimos globais
- Substituindo pior valor e iniciando no ponto (-1, 1): Semelhante ao teste anterior o método acha o mínimo global em apenas 7 iterações, mas desta vez gasta 25 iterações na IPS.

Por fim realizamos mais alguns testes com a função de Rosenbrock em 3D, uma função famosa pela dificuldade em achar o seu mínimo. Isso ocorre por ela possuir um vale de gradiente muito pequeno. A função possui mínimo (1, 1, 1) onde vale 0.

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

Figura 3: Função de Rosenbrock 3D

- Passo constante = 0.01, ponto inicial (-2,4,5): Devido ao ponto inicial distante do ponto mínimo correto e o passo ser da grandeza de 0.01 o método diverge.
- Passo constante = 0.0001, ponto inicial (-2,4,5): Graças à diminuição do passo o método passa convergir, no entanto, demora tanto para convergir que atinge o limite de 50000 iterações colocado no código, chegando até o ponto (-1, 1, 1) aproximadamente, onde a função vale 4.
- Passo via IPS substituindo valor mais antigo, ponto inicial (-2,4,5): Devido ao ponto inicial o método diverge completamente.
- Passo via IPS substituindo pior valor, ponto inicial (-2,4,5): O método do IPS não consegue convergir para minimizar o passo logo na primeira iteração.
- Passo via IPS substituindo pior valor, ponto inicial (0,0,0): O método converge no mínimo em 1768 iterações, gastando o total de 7168 no IPS.
- Passo via IPS substituindo valor mais antigo, ponto inicial (0,0,0): O método converge no mínimo em 1285 iterações, gastando o total de 7027 no IPS.

Conclusão

Após analisar todos os resultados do trabalho que estudou o método do gradiente descendente utilizando o método da Interpolação Parabólica Sucessiva (IPS), conclui-se que, na maioria dos casos, o IPS é superior ao uso de um passo constante. O IPS ajusta dinamicamente o tamanho do passo, permitindo passos maiores ou menores até chegar ao ponto em que o gradiente é nulo (na direção analisada da iteração). No entanto, apesar de ser superior na maioria dos casos, existem situações em que o IPS pode divergir, enquanto um passo constante pequeno converge, embora necessite de muito mais iterações.

Além disso, ao comparar o IPS substituindo o pior valor com o IPS substituindo o valor mais antigo, verificou-se que nenhum dos métodos é completamente superior ao outro. Existem casos em que o IPS substituindo o pior valor é mais rápido, casos em que o IPS substituindo o valor mais antigo é mais rápido, e casos em que ambos não convergem. Sendo assim, é necessário testar os métodos para o tipo de problema específico que está sendo resolvido e escolher o melhor para resolvê-lo.