Variantes do método do gradiente para otimização irrestrita

1 Resumo

Um dos procedimentos mais fundamentais para resolução de problemas de otimização irrestrita é o *método do gradiente*. Este método baseia-se na ideia de que, dada uma função continuamente diferenciável, a direção de maior decréscimo é a contrária ao de seu gradiente. Apesar de seu grande uso em implementações práticas, sabe-se que a minimização iterativa através de passos de gradiente torna-se demasiadamente lenta à medida que o método se aproxima da solução. Estratégias para o aceleramento incluem controle do tamanho de passo ou mesmo ajuste na direção de minimização. Em particular, uma variante do método do gradiente, chamada *q-gradiente*, foi proposta para minimização de funções contínuas gerais, mostrando-se promissora. Outra variante, o *método do gradiente ponderado com atraso*, aplica-se com admirável sucesso à problemas quadráticos estritamente convexos, mostrando vantagens numéricas até mesmo sobre o famoso e largamente usado método de gradientes conjugados de Hestenes e Stiefel. Nesta pesquisa, consideramos as duas variantes citadas. O objetivo é compará-las, do ponto de vista teórico e numérico, entre si e com o método dos gradientes conjugados. Ademais, pretende-se analisar a adaptação da segunda variante a problemas irrestritos gerais.

Palavras-chave: Programação Não-Linear. Método do gradiente e variantes. Gradientes conjugados.

2 Introdução

Nesta pesquisa, consideramos problemas de otimização sem restrições da forma

$$min_{x\in \mathbb{R}} f(x) \ , \tag{1}$$

onde $f:\mathbb{R}^n \rightarrow \mathbb{R}$ é função continuamente diferenciável. Problemas irrestritos são de grande importância, não só porque são utilizados diretamente em aplicações (como *aprendizado de máquina supervisionado*, veja (BOTTOU; CURTIS; NOCEDAL, 2018)), mas também porque subproblemas de métodos iterativos para minimização com restrições, como Lagrangiano aumentado e pontos interiores, são irrestritos. De qualquer modo, problemas na forma (1) são largamente utilizados, e métodos para sua resolução estão presentes em vários pacotes computacionais modernos.

Dentre as ideias para resolução de (1), uma das mais fundamentais (talvez a mais) é caminhar iterativamente em direções que decrescem f localmente a partir de um ponto x (*direções de descida*). Sabemos dos cursos de cálculo que a direção de maior decréscimo para f a partir de x é $-\nabla f(x)$. Isso nos leva ao esquema iterativo do *método do gradiente* (ou *método de máxima descida*), dado por

$$x^{k+1}=x^k-t_k \nabla f(x^k) \ , \tag{2}$$

onde $t_k>0$ é o *tamanho do passo* na direção de $-\nabla f(x)$. Evidentemente, o tamanho do passo deve ser calculado a cada iteração, o que, na prática numérica, é guiado por uma condição que garanta "decréscimo suficiente" de f (usualmente a *condição de Armijo*, veja (RIBEIRO; KARAS, 2013) para detalhes). Tal condição é suficiente para convergência (global) teórica do método, isto é, é possível mostrar que a partir de qualquer ponto inicial x^0, os pontos de acumulação $\bar{x}$ gerados pelo método do gradiente satisfazem $\nabla f(\bar{x})=0$ ($\bar{x}$ é ponto crítico/estacionário). Isso se reflete, por sua vez, no critério de parada prático $\|\nabla f(x^k)\|\leq \varepsilon$.

A despeito de sua simplicidade, o método do gradiente é empregado em pacotes computacionais modernos. No entanto, sabe-se este método converge lentamente próximo a um ponto crítico (fenômeno "zig-zag", veja (RIBEIRO; KARAS, 2013)). Ou seja, são necessários muitos passos próximo à solução, o que leva ao aumento do número de avaliações da função f e seu de gradiente. Uma forma de contornar esse problema é tomar outra direção que não a do gradiente (pelos menos próximo à pontos críticos). Nesse sentido, métodos que fazem uso de derivadas segundas (Newton e Quasi-Newton's) são por vezes empregados. Costumam ser eficazes, possuem boas taxas de convergência, porém são computacionalmente caros. Por outro lado, alternativas com baixo custo computacional funcionam muito bem em problemas específicos, como métodos de gradientes conjugados aplicados à problemas quadráticos estritamente convexos. O mais famoso método de gradientes conjugados foi proposto por Hestenes e Stiefel (1952)[1]. Possui excelentes propriedades teóricas, sendo desde então a referência em implementações computacionais de sucesso. Cabe ressaltarmos que problemas estritamente convexos são importantes; surgem, por exemplo, em (sub)problemas regularizados. Extensões para problemas não lineares gerais foram propostas (RIBEIRO; KARAS, 2013), sem no entanto obter o mesmo sucesso.

O método do gradiente ainda é objeto de estudo no meio científico. Em particular, um método correlato foi proposto por Soterroni, Galski e Ramos (2011) para minimização de funções contínuas gerais (veja também (GOUVÊA; REGIS; SOTERRONI; SCARABELLO; RAMOS, 2016)). Este método consiste em trocar o gradiente clássico de f por outra noção de gradiente, que os autores chamam de *q-gradiente*. Resultados numéricos mostraram que, para uma certa classe de funções de teste, o *método do q-gradiente* se comportou muito melhor que seu par clássico. Outra recente variante do método do gradiente, proposta por Leon (2019), é o *método do gradiente ponderado com atraso*. Este método aplica-se com admirável sucesso à problemas quadráticos estritamente convexos. Nele, a minimização é feita em duas fases: a primeira consiste em dar um passo na direção de $-\nabla f(x)$, tal como em (2); na segunda fase, o ponto x^{k+1} é "corrigido" com uma ponderação usando as iterações atual a anterior, no seguinte sentido:

$$x^{k+1} \leftarrow x^{k-1} + \beta_k (x^{k+1} - x^{k-1}) ,$$

onde β é dado por

$$\beta_k = arg\,min_{\beta \in \mathbb{R}} \| \nabla f(x^{k-1} + \beta(x^{k+1} - x^{k-1})) \|_2 .$$

Notemos que o último problema tem solução fechada, e logo não há custo computacional em resolvê-lo. A análise teórica, estendida somente muito recentemente por Andreani e Raydan (2020) (artigo ainda não publicado), indica que tal método possui algumas das excelentes propriedade teóricas de gradientes conjugados, mas sendo mais estável numericamente em problemas mal condicionados. Isso abre possibilidade para agregar este método em pacotes computacionais, em substituição ao gradientes conjugados de Hestenes e Stiefel, pelo menos quando nos deparamos com problemas mal condicionados.

Nesta pesquisa, consideramos as duas variantes do método do gradiente citadas anteriormente, a saber, os métodos do q-gradiente e do gradiente ponderado com atraso. **O objetivo é compará-los, do ponto de vista teórico e numérico, entre si e com o método dos gradientes conjugados, bem como estudar uma possível adaptação da segunda variante a problemas gerais.**

1 Neste trabalho seminal, o método é apresentado para resolver sistemas lineares quadrados com matriz dos coeficientes simétrica e definida positiva. Cabe notarmos que encontrar pontos estacionários de problemas quadráticos estritamente convexos é equivalente à resolver sistemas do tipo.

3 Referências

ANDREANI, R.; RAYDAN, M. Properties of the delayed weighted gradient method. **Relatório Técnico**, 2020. Acesso em 23/04/20. Disponível em http://www.optimization-online.org/DB_FILE/2020/03/7655.pdf

BOTTOU, L.; CURTIS, F. E.; NOCEDAL, J. Optimization Methods for Large-Scale Machine Learning. **SIAM Review**, v. 60, n. 2, p. 223-311, 2018.

GOUVÊA, E. J. C.; REGIS, R. G.; SOTERRONI, A. C.; SCARABELLO, M. C.; RAMOS, F. M. Global optimization using q-gradients. **European Journal of Operational Research**, v. 251, n. 3, p. 727-738, 2016.

HESTENES, M. R., STIEFEL, E. Methods of conjugate gradients for solving linear systems. **J. Res. Natl. Bur. Stand.**, v. 49, n. 1, p. 409-436, 1952.

LEON, H. F. O. A delayed weighted gradient method for strictly convex quadratic minimization. **Computational Optimization and Applications**, v. 74, n. 1, p. 729-746, 2019.

RIBEIRO, A. A.; KARAS, E. W. **Otimização Contínua.** São Paulo: Cengage, 2013.

SOTERRONI A. C.; GALSKI R. L.; RAMOS F. M. The q-gradient vector for unconstrained continuous optimization problems. In: Hu B., Morasch K., Pickl S., Siegle M. (eds) Operations Research Proceedings 2010. **Operations Research Proceedings**. Berlin: Springer, 2011.