Report for 2024/07/30

Hiroki Hamaguchi

July 24, 2024

1 CG-Newton method

On the last time meeting, I have implemented the CG-Newton (I named it) method, non-linear conjugate gradient method with Newton's method based line search, and showed the result of the numerical experiment. Although I'm quite interested in this topic, since I couldn't find the corresponding method in literature, I temporarily suspended the research on this topic.

According to Takahashi-san, this method might be related to A coordinate gradient descent method for nonsmooth separable minimization, which adopt the following line search method (I modified little bit):

> Choose α_{init} and let α be the largest element of $\{\alpha_{\text{init}}\beta^i\}_i$ satisfying
> $$f(x+\alpha d) \le f(x) + \alpha\sigma\left(\boldsymbol{\nabla} f(x)^\top d + \frac{1}{2} d^\top \boldsymbol{\nabla}^2 f(x) d\right).$$

Although this is not completely the same as the line search method I implemented, I think this is quite suggestive. I implemented as $\alpha_{\text{init}} = -\frac{d^\top \boldsymbol{\nabla} f(x)}{d^\top \boldsymbol{\nabla}^2 f(x) d}$ with normal Armijo rule, but in analysis α_{init} doesn't affect the convergence rate. Thus, we need to prove that the obtained α satisfies stronger condition than the normal Armijo rule like above one. Let us only consider the case of f is convex and backtracking will not occur. Since

$$\begin{aligned} f(x+\alpha d) - f(x) &\simeq \alpha \boldsymbol{\nabla} f(x)^\top d + \frac{\alpha^2}{2} d^\top \boldsymbol{\nabla}^2 f(x) d \\ &= -\frac{(d^\top \boldsymbol{\nabla} f(x))^2}{2 d^\top \boldsymbol{\nabla}^2 f(x) d}, \end{aligned}$$

we can expect that

$$f(x_{k+1}) - f(x_k) \le -\frac{(d_k^\top \boldsymbol{\nabla} f(x_k))^2}{2 d_k^\top \boldsymbol{\nabla}^2 f(x_k) d_k}.$$

However, I don't have any idea what it means and how to utilize it in the analysis. It might necessary to consider the smoothness of $\nabla^2 f$ or something like that.

2 Graph Drawing

I'm sorry for sudden topic change, but I'm currently quite interested in the graph drawing problem. Firstly, let me explain why I'm interested in this topic.

On the last time meeting, I said that Random Subspace Method does work well for a narrow valley, and thus I proposed the CG-Newton method. But based on Prof.Takeda's advise, I considered what kind of problems can be well solved by Random Subspace Method, and realized that Random Subspace Method is very suitable for the problem with separable variable $X = (x_1, x_2, \ldots, x_n) \quad (x_i \in \mathbb{R}^2)$, i.e., the graph drawing problem, which is I'm very interested in from years ago. (Actually, one of the candidates for my graduation thesis theme was about graph drawing.)

$$\underset{X}{\text{minimize}} \quad \sum_{i<j} f(x_i, x_j)$$

As you know, stochastic gradient decent works well for the separable problem ($f = \sum_i f_i$), and separableness of $X = (x_1, x_2, \ldots, x_n)(x_i \in \mathbb{R}^2)$ is corresponding to this property.

Let us consider the case a random matrix P is a permutation matrix like, which is a practical situation. Then, the separableness of X is necessary to terminate the algorithm in a few step.