

Random Subspace Methods for Two Difficult Classes of Optimization Problems

(扱いにくい特徴を持った2つの最適化問題に対するランダム部分空間法)

数理情報学専攻 48226226 野沢 諒太

指導教員 武田 朗子 教授

1 はじめに

近年では、機械学習や統計において最適化問題の変数の数が膨大な大規模最適化問題を解く必要があることが増えている。しかし、大規模問題においては、その変数の多さから計算量が膨大になる課題がある。その課題に対するため、式 (1) で更新するランダム射影を用いた部分空間法がいくつか提案されている:

$$x_{k+1} = x_k + \alpha_k M_k s_k, \quad M_k \in \mathbb{R}^{n \times d}, s_k \in \mathbb{R}^d. \quad (1)$$

例えば, Kozak ら [1] は $s_k = -M_k^\top \nabla f(x_k)$ とする手法を提案した。この手法は、 $\nabla f(x_k)$ を直接計算するのが難しい場合において、方向微分を用いて勾配を計算することを考える。この時に、元の勾配を方向微分で求めるためには n 回の方向微分を必要とするが、 $M_k^\top \nabla f(x_k)$ を計算することで方向微分の計算回数を d 回まで削減することができるという利点がある。

この例のように式 (1) を用いた手法は多く存在するが、そのほとんどが微分可能な無制約最適化問題に対するものであり、制約付き最適化問題や非平滑最適化問題に対するランダム射影を用いたランダム部分空間法はあまり提案されていない。本研究では、制約付き最適化問題、および非平滑最適化問題のそれぞれに対して、ランダム部分空間法を提案し、その収束性を示した。

2 制約付き最適化問題

2.1 既存研究

次の制約付き最適化問題を考える:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) \leq 0.$$

f の勾配はリプシッツ連続であり、 h_i は線形関数であるか、その勾配がリプシッツ連続である。 h_i が全て線形関数であるときに、Rosen [3] によって、現在の解 x_k における有効制約の勾配 $G_k = (\nabla h_i(x_k))_{i \in \mathcal{A}_k}$ を用いて以下のように更新する Gradient projection 法 (GPM) が

提案されている:

$$x_{k+1} = x_k - \alpha_k (I - G_k (G_k^\top G_k)^{-1} G_k^\top) \nabla f(x_k). \quad (2)$$

この手法は、有効制約の勾配の直交する方向に目的関数の勾配を射影することによって有効制約の関数値を変化させないにする手法である。

2.2 提案手法

本研究では、制約付き最適化問題に対して、式 (1) で更新する手法を提案した。ランダム射影と GPM のような手法を組み合わせることによって、線形制約に特化した手法および、非線形制約にも対応できる手法を提案した。線形制約の場合は主に以下の更新式を用いる:

$$x_{k+1} = x_k - \alpha_k M_k (I - Z_k) M_k^\top \nabla f(x_k), \quad (3)$$

$$Z_k := M_k^\top G_k (G_k^\top M_k M_k^\top G_k)^{-1} G_k^\top M_k.$$

M_k は各成分が $\mathcal{N}(0, 1/n^2)$ に従う $n \times d$ のランダム行列である。線形制約においては、次の収束性を示した。

定理. $\varepsilon_0, \varepsilon_1, \varepsilon_2$ は提案手法のパラメータである。

$$\delta(\varepsilon_0, \varepsilon_1, \varepsilon_2) = \min \left(O \left(\frac{d\varepsilon_1^2}{n} \right), O(\varepsilon_0 \varepsilon_1^2), \right. \\ \left. O \left(\frac{d\varepsilon_2^2}{n} \right), O(\varepsilon_0 \varepsilon_2^2) \right)$$

とする。提案手法は $K := \left\lceil \frac{f(x_0) - f^*}{\delta(\varepsilon_0, \varepsilon_1, \varepsilon_2)} \right\rceil$ 回の以内の反復で、高い確率で緩和 KKT 点へ収束する。

この定理から提案手法は、 n が十分大きいときに $O(\frac{n}{d} \max(\varepsilon_1^{-2}, \varepsilon_2^{-2}))$ の反復で収束する。ここで勾配の計算を方向微分によって計算する場合を考えると、提案手法は勾配として $M_k^\top \nabla f(x)$ しか用いられないため、Kozak らの手法 [1] と同様に、勾配の計算量を削減することができる。そのため、方向微分の計算回数として $O(\frac{n}{\varepsilon^2})$ 程度になり、これは勾配を方向微分で計算し、非凸な無制約最適化問題の標準的な手法を適用するときの、方向微分の計算回数 $O(n) \times O(\varepsilon^{-2})$ と一致する結果である。

3 非平滑最適化問題

3.1 既存研究

非平滑かつ凸な関数 f に関して、次の問題を考える:

$$\min_{x \in \mathbb{R}^n} f(x).$$

また f は L_0 -リプシッツ連続であり、関数値 $f(x)$ しか用いることができない0次最適化問題を考える. この問題を解くアルゴリズムに対する理論的な問題点として、 ε -近似解を求めるのに必要な関数の評価回数 (オラクル計算量) が次元数 n に依存する点があげられる. 例えば Nesterov ら [2] 提案した Random gradient-free 法 (RGF) は次のように解 x_k を更新する:

$$x_{k+1} = x_k - \alpha_k g_{\mu_k}(x_k, u_k), \quad u_k \sim \mathcal{N}(0, I_n).$$

$g_{\mu}(x, u)$ は x においての u に沿った方向の勾配の近似であり、中央差分や前進差分が用いられる. 前進差分においてはオラクル計算量が $O(\frac{n^2}{\varepsilon^2})$ であり、中央差分においては $O(\frac{n}{\varepsilon^2})$ であることが示されている.

3.2 提案手法

本研究では、ランダム射影を用いることによってオラクル計算量の次元に対する依存性を削減することを目的とする. ランダム射影と RGF を組み合わせた次の手法を提案する:

$$x_{k+1} = x_k - \alpha_k \frac{h_k(\mu_k u_k) - h_k(-\mu_k u_k)}{2\mu_k} P_k u_k$$

$$h_k(u) := f(x_k + \frac{1}{\sqrt{n}} P_k u), \quad u_k \sim \mathcal{N}(0, I_d)$$

ここで、 P_k は各成分が $\mathcal{N}(0, 1)$ に従う $n \times d$ ($d < n$) のランダム行列である. この手法は、目的関数 f の部分空間関数 $h_k(u)$ に対して、RGF を適用したのものであり、本研究では、提案手法が非平滑かつ凸な目的関数に対して、大域的に $O(\frac{n}{\varepsilon^2})$ のオラクル計算量を達成することを示した. また局所的に目的関数に対して、いくつかの仮定を追加することによって次の定理を示した.

定理. α_k, μ_k は提案手法のパラメータである. f に関する適切な仮定の下で、以下のうち少なくとも1つが成立する.

1. $\sum_{k=0}^{N-1} \alpha_k (\mathbb{E} u_k [f(x_k)] - f(x^*)) \leq \frac{1}{2} r_0^2 + L_0 \sqrt{d} \sum_{k=0}^{N-1} \alpha_k \mu_k + \frac{L_0^2 (n+4)(d+4)^2}{cn} \sum_{k=0}^{N-1} \alpha_k^2,$
2. $\min_{0 \leq k \leq N-1} f(x_k) - f(x^*) \leq L_0 C_1 \sqrt{\frac{10d+41}{d^2}} + \frac{L_0 C_2}{n\sqrt{d}}$

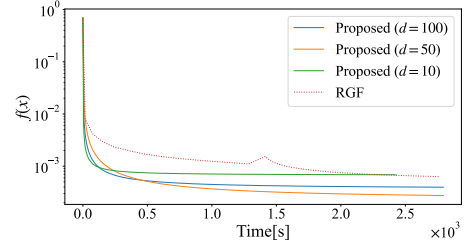


図 1. 提案手法における数値実験

C_1, C_2, c は n には依存しない定数, $\mathbb{E} u_k$ は $u_0, P_0, \dots, u_{k-1}, P_{k-1}$ に関する期待値であり, $r_0 = \|x_0 - x^*\|$, x^* は最適解の1つである.

この定理から、パラメータ α_k, μ_k を適切に決定することで、提案手法が $O(\frac{d^2}{\varepsilon^2})$ のオラクル計算量を達成することを示した.

3.3 数値実験

目的関数が以下のような形をしている場合を考える:

$$f(x) := \max_{\|y\| \leq \delta} g(x^\top y) \quad (4)$$

提案手法において、 $f(x_k + \frac{1}{\sqrt{n}} P_k u_k)$ を計算する場合に、以下の n 次元の最適化問題を解くことになる.

$$\max_{\|y\|^2 \leq \delta^2} g(x_k^\top y + \frac{1}{\sqrt{n}} u_k^\top P_k^\top y). \quad (5)$$

ここで問題 (5) を解く代わりに、以下の $(d+1)$ 次元の最適化問題を解く:

$$\max_{\alpha_k^2 / \|x_k\|^2 + \|\beta\|^2 \leq \delta^2} g(\alpha + u_k^\top \beta). \quad (6)$$

$d = o(n)$ で元の次元 n が十分大きければ、問題 (6) は問題 (5) の近似であることを示せることができ、次元のより小さい問題を解くことができるため $f(x)$ の評価時間を削減することができる. 図 1 は提案手法と RGF を目的関数が (4) の構造を持つ問題に適用した結果であり、収束までの時間が短くなっていることが確認できる.

参考文献

- [1] D. Kozak, S. Becker, A. Doostan, and L. Tenorio. A stochastic subspace approach to gradient-free optimization in high dimensions. *Computational Optimization and Applications*, 79(2):339–368, 2021.
- [2] Y. Nesterov and V. Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17:527–566, 2017.
- [3] J. B. Rosen. The gradient projection method for nonlinear programming. part i. linear constraints. *Journal of the society for industrial and applied mathematics*, 8(1):181–217, 1960.