Prediction-Correction Algorithm for Time-Varying Smooth Optimization (時間変化する平滑最適化問題に対する予測補正アルゴリズム)

数理情報学専攻 48216205　岩切 秀規
指導教員　武田 朗子 教授

1　はじめに

時間変化する最適化問題は工学において広く現れ，重要な研究対象となっている．中でも，時間と共に変化する最適解の追跡精度の向上を目的として，様々な予測補正アルゴリズムと呼ばれる手法が提案されている．しかし，既存手法の理論保証は強凸関数に限定される上，そのほとんどが，大規模な問題へ適用困難である．本研究では，一般の非凸な大規模最適化問題に適用可能な予測補正アルゴリズムを新たに提案した．さらに，強凸でない目的関数に対する勾配降下法及び提案手法の収束誤差の解析を行い，提案手法が勾配降下法を上回る精度を達成することを理論的に示した (表 1 参照).

2　時間変化する平滑最適化問題

以下の時間変化する平滑最適化問題を考える:

$$x_k^* := \operatorname*{argmin}_{x\in\mathbb{R}^d} \ f(x;t_k),\ k\in\mathbb{N} \tag{1}$$

ただし，$t_k := kh$ であり，h はサンプリング周期を表す．また，$f(x;t)$ は x,t に対して滑らかとする．この問題においては，時不変な最適化問題とは異なり，各反復において h 以内に最適化問題を解かなければならないことから，毎反復，一定程度の誤差を許すことになる．そのため，最適解がなす軌道 $\{x_k^*\}$ への収束速度だけでなく，収束後の最適解軌道の追跡精度が重要になる．次章で，この追跡精度の向上を目的に開発された予測補正アルゴリズムについて紹介する．

3　予測補正アルゴリズム

予測補正アルゴリズム (Algorithm 1) は，補正ステップと予測ステップからなる．補正ステップでは，入手した目的関数 $f(\cdot;t_k)$ に対して，時不変な関数に対する最適化手法，例えば勾配降下法を適用する．予測ステップでは，次の時刻における関数値を予測した上で，良い性質を持つ点 x_k を予測する．最適解の追跡速度の向上にあたっては，関数の時間変化の特徴を捉えた優れた予測手法を構築することが鍵となる．

Algorithm 1 予測補正アルゴリズムの手続き

```
1: for k = 0, 1, 2, ... do
2:     目的関数 f(·; t_k) が明らかになる
3:     f(·; t_k) に基づき，予測 x_{k|k−1} を x_k に補正
4:     f(·; t_{k+1}) の近似関数に基づき，x_k を初期点として x_{k+1|k} を予測
5:     損失 f(x_{k+1|k}; t_{k+1}) を被る
6: end for
```

これまでに，勾配降下法より高い精度を達成する予測補正アルゴリズムが様々提案されているが (e.g. [3, 2])，それらは全て，目的関数が強凸関数であるか，強凸関数と良い性質を備えた凸関数の和で書けることを仮定している．さらに，多くの既存アルゴリズムは，ヘッセ行列やその逆行列を用いており，大規模な問題への適用が困難である．

4　提案手法

そこで本研究では，一般の非凸大規模問題に適用可能な予測補正アルゴリズムを新たに提案した．その手続きを Algorithm 2 に示す．4 行目に示したように，提案手法の補正ステップは，勾配降下法そのものである．5 行目が予測ステップに対応しており，これは，次の時刻における関数 $f(\cdot;t_{k+1})$ の 1 次の Taylor 級数近似の制約付き最小化問題

$$\begin{aligned}&\min_{x\in\mathbb{R}^d}\ f(x_k;t_k)+\langle\nabla_x f(x_k;t_k),x-x_k\rangle+h\nabla_t f(x_k;t_k)\\&\text{s.t.}\ \ \|x-x_k\|\le\zeta h\end{aligned}$$

の解となっている．予測ステップでは，勾配及び勾配のノルムしか用いないため，大規模な問題への適用も可能であることが分かる．

5　理論解析

提案手法が勾配降下法と比べて良い精度を達成することを保証するにあたっては，目的関数の滑らかさの仮定に加え，以下の仮定が必要となる:

図 1. 既存手法及び提案手法における収束誤差と必要なオラクルの比較.

アルゴリズム	GD [1]	U-FOPC [2]	AGT [3]	提案手法 (FOA-Min)
必要なオラクル	Grad	Grad, Hessian	Grad, Hessian inv.	Grad, Grad norm
PL	$O(h)$-optimal	-	N/A	$O(h^2)$-optimal
非凸	$O(\sqrt{h})$-stationary	-	N/A	$O(h)$-stationary

Algorithm 2 提案手法

for $k = 0, 1, 2, \ldots$ **do**
  目的関数 $f(\cdot; t_k)$ が明らかになる
  $x_k = x_{k|k-1} - \beta \nabla_x f(x_{k|k-1}; t_k)$
  $x_{k+1|k} = x_k - \zeta h \frac{\nabla_x f(x_k; t_k)}{\|\nabla_x f(x_k; t_k)\|}$
  損失 $f(x_{k+1|k}; t_{k+1})$ を被る
end for

仮定 1. $\exists Z \geq 0,\ \forall x \in \mathbb{R}^d,\ \forall t \geq 0,\ |\nabla_t f(x;t)| - Z\|\nabla_x f(x;t)\| \leq 0$

この仮定は，目的関数の勾配が 0 に近づくとき，時間微分も同程度の速度で 0 に近づいていくことを要請している．これは，大変強い仮定のように思えるが，いくつかの重要な関数クラスがこの仮定を満たすことが示せる．例えば，以下の定理が示すように，強凸関数の最適化問題はこの仮定を満たすように同値変形できる:

定理 5.1. *強凸な目的関数 $f(x;t)$ がある滑らかさの仮定を満たすとする．このとき，目的関数を $\check{f}(x;t) := f(x;t) - \int_0^t \nabla_t f(y;\tau)|_{y=x^*(\tau)} \mathrm{d}\tau$ と再定義すると，この関数は，引き続き滑らかさの仮定を満たす上，仮定 1 を満たす．*

他にも，時間変化と共に平行移動する関数 $f(x+g(t))$ を含むような重要な関数クラスがこの仮定を満たすことが示せる．

この仮定を用いると，非凸関数に対する提案手法の収束誤差を以下のように評価することができる．

定理 5.2. *目的関数がある滑らかさの仮定及び仮定 1 を満たすとする．このとき，アルゴリズムのパラメータを適切に定めることで，任意の $k_0 \in \mathbb{Z}_{\geq 0}$ に対し，$T_{k_0} := \frac{f_{k_0} - f^*_{k_0}}{h^2}$ 回の反復における勾配のノルムの平均は，以下で抑えられる:*

$$\frac{1}{T_{k_0}} \sum_{k=k_0}^{k_0+T_{k_0}-1} \|\nabla_x f(x_{k|k-1}; t_k)\| \leq h\sqrt{2L_1(1+\bar{G}_2)}$$

同様の解析を行うと，勾配降下法に対しては，勾配のノルムの平均は $O(\sqrt{h})$ となることから，提案手法の予測の導入によって，精度が $O(\sqrt{h}) \to O(h)$ と向上していることが分かる．

また，上記の定理は，時間に関して平均を取った結果について評価しており，最悪ケースは評価できていない．そこで，本研究では，アルゴリズムが $h\sqrt{2L_1(1+\bar{G}_2)}$-stationary point に到達した後も，全ての出力が勾配または関数値の意味で良い点になっていることを示した．

6 数値実験

時間変化する線形回帰問題

$$\min_{x \in \mathbb{R}^{10}} \ f(x; t_k) := \frac{1}{2}\|Ax - b(t_k)\|^2,$$

に対して，既存手法及び提案手法を適用した結果を図に示す．提案手法 (黄，緑) は既存手法に比べ，h が小さいときに既存手法を上回る精度を達成し，かつ，勾配のノルムの h への依存が理論保証通り $O(h)$ となっていることが確認できる．

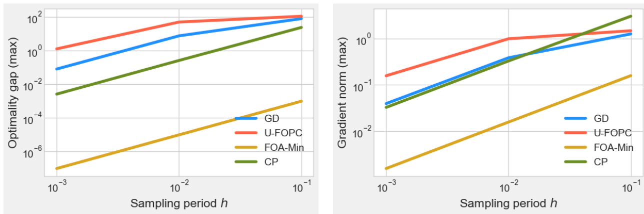

図 2. 横軸にサンプリング周期，縦軸に最適性ギャップ (左) または勾配のノルム (右) を取った両対数プロット．

また，大規模な非凸最適化問題に対する提案手法の性能を確認するため，Netflix のレコメンドデータを用いて，時間変化する行列分解問題を実装した．ここでも，提案手法が勾配降下法を上回る精度を達成した．

参考文献

[1] A. Y. Popkov. Gradient methods for nonstationary unconstrained optimization problems. *Automation and Remote Control*, 66(6):883–891, 2005.

[2] A. Simonetto and E. Dall'Anese. Prediction-correction algorithms for time-varying constrained optimization. *IEEE Transactions on Signal Processing*, 65(20):5481–5494, 2017.

[3] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro. A class of prediction-correction methods for time-varying convex optimization. *IEEE Transactions on Signal Processing*, 64(17):4576–4591, 2016.