数理解析研究所講究録
第2027巻 2017年 135-143

直交制約つき最適化問題に対する
リーマン多様体上の確率的分散縮小勾配法

東京理科大学・工学部情報工学科　佐藤 寛之
電気通信大学・大学院情報理工学研究科情報ネットワーク工学専攻　笠井 裕之
Amazon Development Centre India, Bamdev Mishra

Hiroyuki Sato[1]
Department of Information and Computer Technology, Tokyo Univerisity of Science
Hiroyuki Kasai
Department of Computer and Network Engineering,
The University of Electro-Communications
Bamdev Mishra
Amazon Development Centre India

概 要

目的関数がサンプルごとに定義される関数の和に分割可能で，そのサンプル数が非常に大きい場合の最小化問題の解法として，サンプルの番号を確率的に選択して対応する関数の勾配を用いる確率的勾配降下法がある．本稿では，確率的勾配を用いて計算される探索方向の分散が小さくなるよう確率的勾配降下法を改良した解法である確率的分散縮小勾配法をリーマン多様体上に拡張する．また，提案アルゴリズムの収束性を議論するとともに，グラスマン多様体上の低ランク行列補完問題への応用についても説明する．

1　はじめに

本稿では損失最小化問題

$$\min f(w) := \frac{1}{N}\sum_{n=1}^{N} f_n(w)$$

を扱う．ここで，w は変数，N はサンプル数，f_n は第 n サンプルについての損失関数である．サンプル数 N は非常に大きく，反復アルゴリズムにおいて f の勾配 $\nabla f(w) = \frac{1}{N}\sum_{n=1}^{N}\nabla f_n(w)$ を毎回の反復で計算するのは計算量が大きく困難である場合を考える．一方，最急降下法や共役勾配法のような通常の勾配アルゴリズムは第 k 反復の点 w_k において勾配 $\nabla f(w_k)$ の情報を用いて探索方向を計算するため，そうした最適化手法をこのような問題に対して適用するのは効率的でない．そこで，第 k 反復の点 w_k において，$\nabla f(w_k)$ の代わりに，ランダムに選んだ n に対する確率的勾配 $\nabla f_n(w_k)$ を用いる確率的勾配降下

[1]e-mail:　hsato[AT]rs.tus.ac.jp

（Stochastic Gradient Descent; SGD）法が提案されている．しかし，確率的勾配の分散が大きくなることで，SDG の収束は最急降下法などのバッチ型のアルゴリズムより遅くなる．そこで，その改良手法として，確率的分散縮小勾配（Stochastic Variance Reduced Gradient; SVRG）法が提案されており，近年多くの注目を集めている [4].

上記のアルゴリズムは基本的には無制約最適化問題に対する手法である．しかし，ユークリッド空間における制約つき最適化問題であっても，その実行可能領域がリーマン多様体 $\mathcal{M}$ をなすとき，その問題は多様体 $\mathcal{M}$ 上の無制約最適化問題と見なせる．また，r, d を $r \leq d$ なる整数とするとき，$\mathbb{R}^d$ の r 次元部分空間全体からなるグラスマン多様体 $\mathrm{Gr}(r, d)$ のように，ユークリッド空間の部分多様体とは限らない，より抽象的な多様体上で定式化される最適化問題もある．こうした最適化問題に対するリーマン多様体上の最適化手法は多くの応用をもつため近年盛んに研究されており，ユークリッド空間における多くの効率的な最適化アルゴリズムがリーマン多様体上に拡張されている [1].

そこで，本稿では，損失最小化問題を制約条件 $w \in \mathcal{M}$ の下で考える．ここで，$\mathcal{M}$ はコンパクトなリーマン多様体とする．すなわち，以降で主に扱う問題は

$$\min_{w \in \mathcal{M}} \frac{1}{N} \sum_{n=1}^{N} f_n(w) \tag{1.1}$$

である．2 節では SGD および SVRG を概説するとともに，リーマン多様体上に拡張された SGD も紹介する．3 節では，とくに $\mathcal{M}$ がグラスマン多様体である場合の最適化問題の例として，主成分分析や行列補完問題を紹介する．また，4 節では SVRG をリーマン多様体 $\mathcal{M}$ 上に拡張した Riemannian SVRG (R-SVRG) を [5] に基づいて紹介し，その収束性や数値的な振る舞いについて議論する．提案手法はリーマン多様体 $\mathcal{M}$ 上のアルゴリズムであり，$\mathcal{M}$ 上の直線探索や，$\mathcal{M}$ 上の異なる点において複数の勾配の平均や和を計算することはユークリッド空間でのようにはできないので，M 上の指数写像およびその逆写像である対数写像や，平行移動と呼ばれる写像を用いてこれらの課題を解決している．

2 確率的最適化手法について

2.1 ユークリッド空間における確率的勾配降下法

ユークリッド空間における目的関数 f の無制約最小化問題において，最急降下法では更新式

$$w_{k+1} = w_k - \alpha_k \nabla f(w_k)$$

によってユークリッド空間内の点列 $\{w_k\}$ を生成する．ここで，スカラー $\alpha_k > 0$ はステップ幅である．

ところが，目的関数が $f(w) := \frac{1}{N} \sum_{n=1}^{N} f_n(w)$ で N が非常に大きいときは $\nabla f(w)$ の計算が困難になる．そこで，確率的勾配降下法（SGD）では点 w_k の探索方向を $-\nabla f(w_k)$ とする代わりにランダムに選んだ番号 n に対して $-\nabla f_n(w_k)$ を探索方向とする．すなわ

ち，f を構成する項を一つランダムにサンプリングし，その勾配の逆方向を探索方向とするのである．したがって，SGD における点列の更新式は

$$w_{k+1} = w_k - \alpha_k \nabla f_n(w_k)$$

である．ここで，$\nabla f_n(w)$ を w における確率的勾配という．任意の $n \in \{1, 2, \ldots, N\}$ が選ばれる確率が一様に $1/N$ である場合，確率的勾配の n についての期待値を計算すると

$$\mathbb{E}_n[\nabla f_n(w)] = \frac{1}{N} \sum_{n=1}^{N} \nabla f_n(w) = \nabla f(w)$$

となり，f の勾配 $\nabla f(w)$ と一致するので，この意味で確率的勾配は元の関数 f の勾配の近似になっていることがわかる．

しかしながら，確率的勾配の分散の影響で，SGD の収束を保証するには一般に $\alpha_k = O(1/k)$ とする必要があり，その結果として収束率は劣線形で $O(1/t)$ となる．すなわち，SGD は最急降下法を確率的に近似したアルゴリズムであり，1 反復あたりの計算量は最急降下法より小さくなるが，その一方で収束は最急降下法より遅くなる．

2.2 ユークリッド空間における確率的分散縮小勾配法

前節で述べたように SGD は線形収束せず，その収束率を向上させるために SGD の多くの改良版が研究されている．本節では，そのうちの一つである，[4] で提案された確率的分散縮小勾配法（SVRG）を紹介する．

SVRG では，SGD の確率的勾配 $\nabla f_n(w)$ にその分散を小さくするための補正項を付け加えることで，修正された確率的勾配を計算し，その逆方向を探索方向とする．そのための工夫として，SVRG はアルゴリズム 1 のように二重のループ構造からなる．

外部反復の第 s 反復において最初に，一つ前の外部反復終了時に得られている点 $\tilde{w}^{s-1}$ を w_0^s として保存しておく．内部反復の第 t 反復においては，w_{t-1}^s から次の点 w_t^s を計算することになる．このとき $\{1, 2, \ldots, N\}$ からランダムに選択した番号を i_t^s とすると，SVRG では，SGD で用いられる確率的勾配 $\nabla f_{i_t^s}(w_{t-1}^s)$ に修正を加えた

$$v_t^s = \nabla f_{i_t^s}(w_{t-1}^s) - \nabla f_{i_t^s}(\tilde{w}^{s-1}) + \nabla f(\tilde{w}^{s-1})$$

を計算する．ここで，i_t^s が任意の値をとる確率は一様に $1/N$ であるとすると，

$$\mathbb{E}_{i_t^s}[\nabla f_{i_t^s}(\tilde{w}^{s-1})] = \nabla f(\tilde{w}^{s-1})$$

であるから，修正項 $-\nabla f_{i_t^s}(\tilde{w}^{s-1}) + \nabla f(\tilde{w}^{s-1})$ の期待値は 0 であり，修正された確率的勾配 v_t^s の期待値は確率的勾配 $\nabla f_{i_t^s}(w_{t-1}^s)$ の期待値と一致する．すなわち，v_t^s の期待値も目的関数の勾配 $\nabla f(w_{t-1}^s)$ と等しい．しかし，修正項の影響により v_t^s の分散は $\nabla f_{i_t^s}(w_{t-1}^s)$ の分散より小さくなり，その結果として SGD より速い収束性が保証される．より具体的には，たとえば次の定理が証明されている [4]．

アルゴリズム 1 SVRG [4]

Require: $m_s > 0$ とステップ幅 $\alpha > 0$.

```
w^0 を初期化する.
for s = 1, 2, ... do
  勾配 ∇f(w̃^{s-1}) を計算する.
  w_0^s = w̃^{s-1} を保存する.
  for t = 1, 2, ..., m_s do
    i_t^s ∈ {1, 2, ..., N} を一様にランダムに選択する.
    修正された確率的勾配 v_t^s を
        v_t^s = ∇f_{i_t^s}(w_{t-1}^s) − ∇f_{i_t^s}(w̃^{s-1}) + ∇f(w̃^{s-1})
    により計算する.
    w_t^s を w_t^s = w_{t-1}^s − αξ_t^s と更新する.
  end for
  option I: w̃^s を w_1^s, ..., w_{m_s}^s の平均 (1/m_s) Σ_{t=1}^{m_s} w_t とするか, または, ランダムに選んだ t ∈ {1, 2, ..., m_s} に対して w̃^s = w_t^s とする.
  option II: w̃^s = w_{m_s}^s.
end for
```

定理 2.1. アルゴリズム 1 において m_s が s によらず一定の値 m をとり, さらに option II を用いる場合を考える. $f_n, n = 1, 2, \ldots, N$ はすべて凸関数であるとし, さらにその滑らかさについて $L > 0$ が存在して

$$f_n(w) - f_n(w') - \frac{1}{2}L\|w - w'\|_2^2 \leq \nabla f_n(w')^T(w - w')$$

が成り立つと仮定する. さらに f は, $\gamma > 0$ が存在して

$$f(w) - f(w') - \frac{1}{2}\gamma\|w - w'\|_2^2 \geq \nabla f(w')^T(w - w')$$

を満たす強凸関数とする. また, m は十分大きく

$$\alpha := \frac{1}{\gamma\alpha(1 - 2L\alpha)m} + \frac{2L\alpha}{1 - 2L\alpha}$$

が 1 より小さくなるとする. このとき, f の最小点を w^* として,

$$\mathbb{E}[f(\tilde{w}^s)] \leq \mathbb{E}[f(w^*)] + \alpha^s(f(\tilde{w}^0) - f(w^*))$$

が成り立つ.

2.3 リーマン多様体における確率的勾配降下法

ユークリッド空間における SGD はすでにリーマン多様体上に拡張されている [3]. 以下では, リーマン多様体上の SGD を R-SGD と表記する. リーマン計量 $\langle \cdot, \cdot \rangle$ を備えたリーマ

ン多様体 $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ においては，一般の滑らかな関数 f の $w \in \mathcal{M}$ における勾配 $\mathrm{grad} f(w)$ が

$$\mathrm{D}f(w)[\xi] = \langle \mathrm{grad} f(w), \xi \rangle_w, \quad \xi \in T_w\mathcal{M}$$

を満たす w における接ベクトルとして定義される．そこで，R-SGD における確率的勾配は，勾配をリーマン多様体 $\mathcal{M}$ 上のものに拡張して，$\mathrm{grad} f_n(w)$ とする．n は $\{1, 2, \ldots, N\}$ からランダムに選択される番号である．

また，ユークリッド空間における一般の直線探索アルゴリズムをリーマン多様体に拡張する際と同様に，直線探索の代わりに曲線上の探索を行う必要がある．レトラクションと呼ばれる写像 $R : T\mathcal{M} \to \mathcal{M}$ を用いて

$$w_{k+1} = R_{w_k}(-\alpha_k \mathrm{grad} f_n(w_k))$$

とするのが一般的な多様体上での更新式であるが，今回はとくにレトラクション R として指数写像 Exp を用いる場合を考える．

このように，SGD のユークリッド空間における改良版として SVRG が提案されており，SGD のリーマン多様体への拡張として R-SGD が提案されている．そこで本稿では，SVRG をリーマン多様体に拡張した R-SVRG を提案する．

3　グラスマン多様体上の最適化問題の例

本節では，グラスマン多様体上の損失最小化問題の例として，主成分分析と低ランク行列補完問題への応用について述べる．

3.1　主成分分析への応用

N 個の d 次元のデータベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ に対する主成分分析においては，次の最適化問題を解くことになる．

$$\min_{\mathbf{U} \in \mathrm{St}(r,d)} \frac{1}{N} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \mathbf{U}\mathbf{U}^T \boldsymbol{x}_n\|_2^2. \tag{3.1}$$

ここで，$\mathbf{U}\mathbf{U}^T$ はデータを射影するための行列であり，$\mathbf{U}$ には各列が正規直交ベクトルであるという制約が課せられる．そこで，問題 (3.1) は $d \times r$ の列直交行列全体からなるシュティーフェル多様体 $\mathrm{St}(r,d)$ 上の問題として定式化されている．

問題 (3.1) は，最大化問題

$$\max_{\mathbf{U} \in \mathrm{St}(r,d)} \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n^T \mathbf{U}\mathbf{U}^T \boldsymbol{x}_n$$

と等価である．ここで，目的関数は r 次直交群 $\mathcal{O}(r)$ による作用に対して不変である．すなわち，任意の r 次直交行列 $\mathbf{U}$ に対して $\mathbf{U} \mapsto \mathbf{U}\mathbf{O}$ なる変換を施しても目的関数値は一

定であるので，$\mathrm{St}(r,d)$ において f の臨界点は孤立点でない．このことから，問題 (3.1) を商多様体 $\mathrm{St}(r,d)/\mathcal{O}(r)$ 上の最適化問題として最定式化することができる．この商多様体はグラスマン多様体 $\mathrm{Gr}(r,d)$ と同相であり，こうしてグラスマン多様体上の最適化問題が得られた．

3.2 低ランク行列補完問題への応用

少数の要素だけが既知である行列 $\mathbf{X} \in \mathbb{R}^{d\times N}$ に対して，$\mathbf{X}$ の低ランク性を仮定して補完を行う．既知である $\mathbf{X}$ の成分の添字集合を Ω とすると，ランク r の行列補完問題は次の最適化問題として定式化される．

$$\min_{\mathbf{U}\in\mathbb{R}^{d\times r},\mathbf{A}\in\mathbb{R}^{r\times N}} \|\mathcal{P}_\Omega(\mathbf{UA}) - \mathcal{P}_\Omega(\mathbf{X})\|_F^2.$$

ここで，$(i,j)\in\Omega$ のとき $\mathcal{P}_\Omega(\mathbf{X}_{ij}) = \mathbf{X}_{ij}$ であり，そうでないとき $\mathcal{P}_\Omega(\mathbf{X}_{ij}) = 0$ である．また，$\|\cdot\|_F$ はフロベニウスノルムを表す．$\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$ と列ベクトルごとに分割すると，この問題は

$$\min_{\mathbf{U}\in\mathbb{R}^{d\times r},\boldsymbol{a}_n\in\mathbb{R}^r} \frac{1}{N}\sum_{n=1}^{N} \|\mathcal{P}_{\Omega_n}(\mathbf{U}\boldsymbol{a}_n) - \mathcal{P}_{\Omega_n}(\boldsymbol{x}_n)\|_2^2 \tag{3.2}$$

と等価である．ここで，$\boldsymbol{x}_n \in \mathbb{R}^d$ であり，写像 $\mathcal{P}_{\Omega_n}$ は $\mathcal{P}_\Omega$ の第 n 列への作用を表す．$\mathbf{U}$ が定まれば最適な $\boldsymbol{a}_n$ は $\mathbf{U}$ を用いて書くことができ，この問題は $\mathbf{U}$ の列空間のみに依存するため，グラスマン多様体上の最小化問題と見なせる [2].

問題 (3.1), (3.2) はいずれも (1.1) の形になっており，確率的最適化手法が有効な問題である．

4 グラスマン多様体上の確率的分散縮小勾配法

以降では，グラスマン多様体上の問題を考え，変数 w を $\mathbf{U}\in\mathrm{Gr}(r,d)$ と書くことにする．しかしながら，提案するアルゴリズムとその収束性解析は他のコンパクトなリーマン多様体にも同様に適用できる．さて，グラスマン多様体 $\mathrm{Gr}(r,d)$ 上の点は $\mathbb{R}^d$ の r 次元部分空間であり，$d\times r$ の列直交行列 $\mathbf{U}$ によって代表させることで，同値類 $[\mathbf{U}] := \{\mathbf{UO}_r : \mathbf{O}\in\mathcal{O}(r)\}$ と同一視できる．つまり，$\mathrm{Gr}(r,d) = \mathrm{St}(r,d)/\mathcal{O}(r)$ である．ここで，$\mathrm{St}(r,d)$ は $d\times r$ の列直交行列全体であり，シュティーフェル多様体と呼ばれる．こうして，グラスマン多様体は商多様体の構造をもつ [1].

提案する R-SVRG では，$\mathrm{Gr}(r,d)$ 上の点列を生成するにあたり，

$$\mathrm{Exp}_{\mathbf{U}(0)}(\xi) = [\mathbf{U}(0)\mathbf{V} \ \ \mathbf{W}] \begin{bmatrix} \cos\Sigma \\ \sin\Sigma \end{bmatrix} \mathbf{V}^T$$

で与えられる指数写像を用いる [1]．ここで，$\xi = \mathbf{W}\Sigma\mathbf{V}^T$ は $\mathbf{U}(0)$ における接ベクトル ξ のランク r の特異値分解であり，$\cos(\cdot)$ および $\sin(\cdot)$ は対角成分のみに作用する．指数写像により点 $\mathbf{U}(0)$ から接ベクトル ξ の方向に延びる測地線 $\mathbf{U}(t)$ が

$$\mathbf{U}(t) = \mathrm{Exp}_{\mathbf{U}(0)}(t\xi) = [\mathbf{U}(0)\mathbf{V} \;\; \mathbf{W}] \begin{bmatrix} \cos t\Sigma \\ \sin t\Sigma \end{bmatrix} \mathbf{V}^T$$

と定義され，この曲線上の点が更新後の点として計算される．

提案する R-SVRG をアルゴリズム 2 に示す．ここで，リーマン多様体 $\mathcal{M}$ において $w_1, w_2, \ldots, w_N$ の Karcher 平均は最小化問題

$$\min_{w\in\mathcal{M}} \frac{1}{2N} \sum_{n=1}^{N} (\mathrm{dist}(w, w_n))^2,$$

の解として定義される．ただし，dist は測地線距離を表す．

アルゴリズム 2 R-SVRG [5]

Require: $m_s > 0$ とステップ幅 $\alpha > 0$.

$\tilde{\mathbf{U}}^0$ を初期化する．
for $s = 1, 2, \ldots$ **do**
  勾配 $\mathrm{grad} f(\tilde{\mathbf{U}}^{s-1})$ を計算する．
  $\mathbf{U}_0^s = \tilde{\mathbf{U}}^{s-1}$ を保存する．
  for $t = 1, 2, \ldots, m_s$ **do**
    $i_t^s \in \{1, 2, \ldots, N\}$ を一様にランダムに選択する．
    $\tilde{\mathbf{U}}^{s-1}$ から $\mathbf{U}_{t-1}^s$ へ向かう接ベクトル ζ を対数写像により計算する．
    確率的勾配 ξ_t^s を (4.1) により計算する．
    指数写像 Exp を用いて $\mathbf{U}_t^s$ を $\mathbf{U}_t^s = \mathrm{Exp}_{\mathbf{U}_{t-1}^s}(-\alpha\xi_t^s)$ と更新する．
  end for
  option I: $\tilde{\mathbf{U}}^s$ を $\mathbf{U}_1^s, \ldots, \mathbf{U}_{m_s}^s$ の Karcher 平均とするか，または，ランダムに選んだ $t \in \{1, 2, \ldots, m_s\}$ に対して $\tilde{\mathbf{U}}^s = \mathbf{U}_t^s$ とする．
  option II: $\tilde{\mathbf{U}}^s = \mathbf{U}_{m_s}^s$.
end for

修正された確率的勾配 ξ_t^s を計算するにあたり，確率的勾配 $\mathrm{grad} f(\mathbf{U}_{t-1}^s)$ は $\mathbf{U}_{t-1}^s$ における接ベクトルである一方で，修正項 $\mathrm{grad} f(\tilde{\mathbf{U}}^{s-1})$ および $\mathrm{grad} f_{i_t^s}(\tilde{\mathbf{U}}^{s-1})$ は $\tilde{\mathbf{U}}^{s-1}$ における接ベクトルであるため，これらをそのまま加減することはできない．そこで，提案アルゴリズムでは修正項のベクトルを $\log_{\tilde{\mathbf{U}}^{s-1}}(\mathbf{U}_{t-1}^s)$ の方向に測地線 γ 上で平行移動することによって，

$$\xi_t^s = \mathrm{grad} f_{i_t^s}(\mathbf{U}_{t-1}^s) - P_\gamma^{\mathbf{U}_{t-1}^s \leftarrow \tilde{\mathbf{U}}^{s-1}} \left(\mathrm{grad} f_{i_t^s}(\tilde{\mathbf{U}}^{s-1}) - \mathrm{grad} f(\tilde{\mathbf{U}}^{s-1}) \right) \tag{4.1}$$

を修正された確率的勾配として用いる．ここで，対数写像 log は指数写像の逆写像であり，

$\mathrm{Gr}(r,d)$ 上の $\mathbf{U}(0)$ において，$\mathbf{U}(t)$ に次のように作用する．

$$\log_{\mathbf{U}(0)}(\mathbf{U}(t)) = \mathbf{W}\arctan(\Sigma)\mathbf{V}^T.$$

$\mathbf{W}\Sigma\mathbf{V}^T$ は $(\mathbf{U}(t)-\mathbf{U}(0)\mathbf{U}(0)^T\mathbf{U}(t))(\mathbf{U}(0)^T\mathbf{U}(t))^{-1}$ のランク r の特異値分解である．また，ξ に沿った $\zeta \in T_{\mathbf{U}(0)}\mathrm{Gr}(r,d)$ の平行移動は

$$P_\gamma^{\mathbf{U}(t)\leftarrow\mathbf{U}(0)} = \left([\mathbf{U}(0)\mathbf{V}\ \ \mathbf{W}]\begin{bmatrix} -\sin t\Sigma \\ \cos t\Sigma \end{bmatrix}\mathbf{W}^T + (\mathbf{I}-\mathbf{W}\mathbf{W}^T)\right)\zeta$$

と計算される [1].

アルゴリズム 2 の大域的収束性および局所的収束性についてそれぞれ次の結果を得た [5].

定理 4.1. 連結で単射半径が一様に下に有界（その下限を $I>0$ とする）であるようなリーマン多様体 $\mathcal{M}$ 上で**アルゴリズム 1** を考える．ステップ幅の列 $(\alpha_t^s)_{m_s\geq t\geq 1, s\geq 1}$ が条件 $\sum(\alpha_t^s)^2<\infty$ および $\sum\alpha_t^s=+\infty$ を満たし，さらにコンパクト集合 K が存在して任意の $t\geq 0$ に対して $\mathbf{U}_t^s\in K$ が成り立つとする．このとき，$f(\mathbf{U}_t^s)$ は概収束し，$\mathrm{grad}f(\mathbf{U}_t^s)$ は 0 に概収束する．

定理 4.2. $\mathcal{M}$ をグラスマン多様体とし，$\mathbf{U}^*\in\mathcal{M}$ を f の非退化な局所的最小点とする．$\mathbf{U}^*$ の凸な近傍 $\mathcal{U}$ および $\sigma>0$ が存在し，各点 $\mathbf{U}\in\mathcal{U}$ での $\mathrm{Hess}f(\mathbf{U})$ の最小固有値は σ 以上とする．また，各 $\mathrm{grad}f_n$ が連続的微分可能で，導関数はリプシッツ連続（リプシッツ定数を β とする）であるとし，$\alpha>0$ は $0<\alpha(\sigma-14\alpha\beta^2)<1$ を満たすとする．このとき，アルゴリズム 2 により生成される，$\mathbf{U}^*$ に収束する任意の点列 $\{\tilde{\mathbf{U}}^s\}$ に対して，十分大きい $s>K$ について

$$\mathbb{E}[(\mathrm{dist}(\tilde{\mathbf{U}}^s,\mathbf{U}^*))^2] \leq \frac{4(1+8m\alpha^2\beta^2)}{\alpha m(\sigma-14\alpha\beta^2)}\mathbb{E}[(\mathrm{dist}(\tilde{\mathbf{U}}^{s-1},\mathbf{U}^*))^2]$$

が成り立つ．

主成分分析や行列補完問題などに対して提案手法 R-SVRG と既存手法（リーマン多様体上の最急降下法 R-SD および確率的勾配降下法 R-SGD [3]）の性能を比較した数値実験の結果については，[5] を参照されたい．それらの結果から，提案アルゴリズムが既存の R-SD や R-SGD より優れていることが実証される．

5　結論と今後の課題

本稿では，[5] に基づいてリーマン多様体上の確率的分散縮小勾配法 R-SVRG を提案した．これは，確率的勾配降下法 SGD の改良アルゴリズムである SVRG をリーマン多様体上に拡張したものであると同時に，リーマン多様体上のアルゴリズムである R-SGD の改良アルゴリズムにもなっている．数値的にも実際に速い収束性を示すことを確認している．

本研究ではグラスマン多様体の場合に特化して議論を行った．グラスマン多様体の断面曲率は非負であり，そのことにより議論を行いやすい．しかしながら，他のリーマン多様

体でも断面曲率が下に有界なものであれば，より複雑な議論を経て，断面曲率の下限を用いて定理 4.2 と同様の収束率の結果を得ることができる．また，アルゴリズム自体は他の多様体にもそのまま適用可能である．ただし，異なるリーマン多様体においては，指数写像，対数写像，平行移動の公式が本稿で紹介したものとは異なる．さらに，グラスマン多様体においても，指数写像と平行移動の代わりにそれらを一般化した概念であるレトラクションや vector transport を用い，対数写像の代わりにレトラクションの逆写像を用いると，計算量が小さくなり結果として高速なアルゴリズムが得られると期待される．そこで，今後の課題として，本稿で紹介した結果の様々な観点からの一般化が挙げられる．すなわち，一般の断面曲率の下限をもつリーマン多様体に対する R-SVRG の，レトラクションや vector transport を用いた場合の収束性の解析を行いたい．

また，本研究により，リーマン多様体上においても様々な確率的最適化アルゴリズムが考えられる可能性が示唆される．そこで，ユークリッド空間ですでに提案されている様々な確率的最適化アルゴリズムをリーマン多様体に拡張することも今後の課題である．

参考文献

[1] P.-A. Absil, R. Mahony and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.

[2] L. Balzano, R. Nowak, and B. Recht, Online identification and tracking of subspaces from highly incomplete information, *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 704–711, 2010.

[3] S. Bonnabel, Stochastic gradient descent on Riemannian manifolds, *IEEE Transactions on Automatic Control*, **58** (2013), 2217–2229.

[4] R. Johnson and T. Zhang, Accelerating stochastic gradient descent using predictive variance reduction, Advances in Neural Information Processing Systems (NIPS), **26** (2013), 315–323.

[5] H. Sato, H. Kasai, and B. Mishra, Riemannian stochastic variance reduced gradient, arXiv preprint, arXiv:1702.05594 (2017).