

リーマン多様体上の最適化に基づく 離散時間線形システム同定アルゴリズム

京都大学・大学院情報学研究科・数理工学専攻/白眉センター 佐藤 寛之

北見工業大学・工学部・地域未来デザイン工学科 佐藤 一宏

HiroYuki Sato¹

Department of Applied Mathematics and Physics/

The Hakubi Center for Advanced Research,

Kyoto University

Kazuhiro Sato

School of Regional Innovation and Social Design Engineering,

Kitami Institute of Technology

概要

本稿では、著者らの論文 [10] に基づいて、離散時間線形状態空間モデルのシステム同定問題に対する、リーマン多様体上の最適化に基づくアルゴリズムについて議論する。決定すべき変数であるシステムのパラメータ行列を、その成分の組をベクトルと見なすのではなく行列変数のまま扱うことで、効率的なアルゴリズムを提案する。さらに、入出力等価なシステムを実現するパラメータ間に同値関係を導入することで、あるリーマン商多様体を導き、その上の最適化問題および解法アルゴリズムを議論する。また、この商多様体の幾何学を調べることで、導出された最適化問題に対する共役勾配法を導き、多様体上の共役勾配法とユークリッド空間における共役勾配法との間に差異が生じることを説明する。

1 はじめに

制御器を設計する上で、システム同定は重要なプロセスである。本稿ではシステム同定法の一つである予測誤差法に基づいた新しいアルゴリズムを提案する。とくに、入出力等価なシステムに対しては最小化すべき目的関数値が等しくなることに着目し、そのようなシステムを実現するパラメータを互いに同値であると定義することによって商多様体を構成する。さらに、そこにリーマン計量を与えることで、システム同定問題をリーマン商多様体上の最適化問題として定式化する。こうすることで、互いに入出力等価なシステム全体からなる集合を商多様体上の1つの点と見なして最適化を行うことができる。さらにその最適化問題の幾何学を詳しく調べることで、近年盛んに研究されているリーマン多様体上の幾何学的な最適化アルゴリズムのうち、最急降下法および共役勾配法における具体的

¹e-mail: [hsato\[AT\]amp.i.kyoto-u.ac.jp](mailto:hsato[AT]amp.i.kyoto-u.ac.jp)

な計算式を求める．とくに，商多様体の水平空間への直交射影を導くことで，考えている多様体の vector transport を定義して，幾何学的な共役勾配法において用いる．この操作によって，この問題に対する幾何学的な共役勾配法が，ユークリッド空間における従来の共役勾配法と本質的に異なる計算を行うことになることを示す．

本稿の構成は次の通りである．2 節ではリーマン多様体上の最適化の一般論を概説し，関数の勾配やレトラクションなど，最適化において必要となる幾何学的な量を説明する．3 節では，予測誤差法に基づく，[7, 11] で提案されたベクトルを決定変数にもつ最適化問題を紹介し，行列を変数とする問題として定式化し直す．次に 4 節では，システムの入出力等価性に注目して商多様体を導入し，その上の最適化問題を提案する．また，その商多様体について，垂直空間や水平空間などを求める．これらを用いて 5 節でユークリッド空間およびリーマン多様体上の最急降下法と共役勾配法を導出する．6 節では本稿のまとめを行うとともに，数値実験結果についても言及する．

2 リーマン多様体上の最適化について

$\mathcal{M}$ を多様体とする． $\mathcal{M}$ 上の任意の点 x における接空間 $T_x\mathcal{M}$ はベクトル空間であるが，ここに内積 $g_x(\cdot, \cdot)$ が与えられているとする．点 x に対して g_x を対応させる g を考え， g_x が点 $x \in \mathcal{M}$ に対して滑らかであるとき， g を $\mathcal{M}$ 上のリーマン計量といい， $(\mathcal{M}, g)$ または単に $\mathcal{M}$ をリーマン多様体という．

リーマン多様体 $\mathcal{M}$ 上の制約なし最適化問題を考えよう：

問題 2.1.

$$\begin{aligned} &\text{minimize } f(x), \\ &\text{subject to } x \in \mathcal{M}. \end{aligned}$$

$\mathcal{M}$ がユークリッド空間 $\mathbb{R}^n$ であるとき，問題 2.1 に対する直線探索に基づく反復アルゴリズムでは，初期点 $x_0 \in \mathbb{R}^n$ を与えた上で，

$$x_{k+1} = x_k + \alpha_k \eta_k \tag{2.1}$$

によって点列 $\{x_k\}$ を生成する． $\alpha_k > 0$ はステップ幅， $\eta_k \in \mathbb{R}^n$ は探索方向である．

$\mathcal{M}$ が一般のリーマン多様体であるときには，(2.1) の右辺の和が必ずしも定義されないため，直線探索の代わりに曲線上の探索を行う必要がある．そこで，現在得られている点 x_k から次の点 x_{k+1} を計算する際に，点 x_k から探索方向，すなわち $\eta_k \in T_{x_k}\mathcal{M}$ の方向に伸びる曲線 γ を考える．つまり， $\gamma(0) = x_k$ かつ $\dot{\gamma}(0) = \eta_k$ が成り立つとする．この曲線 γ 上で探索を行うことで適切なステップ幅 $\alpha_k > 0$ を求め，

$$x_{k+1} = \gamma(\alpha_k)$$

によって x_{k+1} を計算する. このような曲線を定義する上で, 次のレトラクションと呼ばれる写像があれば便利である [1]. ここで, $TM := \bigcup_{x \in M} (\{x\} \times T_x M)$ は M の接バンドルである.

定義 2.1. $R: TM \rightarrow M$ が次の性質を満たすとき, R を M のレトラクションという. ここで, $R_x := R|_{T_x M}: T_x M \rightarrow M$ であり, 0_x は $T_x M$ の零ベクトルを表す.

1. $R_x(0_x) = x$.
2. $DR_x(0_x)[\eta] = \eta, \quad \eta \in T_x M$.

レトラクションの定義から, $\gamma(t) := R_{x_k}(t\eta_k)$ によって曲線 γ を定めれば, γ は上記の性質を満たす曲線となる. したがって, x_k から x_{k+1} を計算するための更新式は

$$x_{k+1} = R_{x_k}(\alpha_k \eta_k)$$

となる.

また, 探索方向を計算する際には, ユークリッド空間の場合と同様に目的関数の勾配やヘシアンを用いるが, これらはリーマン計量 g に依存した量として定義される. たとえば f の (M, g) 上の点 x における勾配 $\text{grad } f(x)$ は,

$$g_x(\text{grad } f(x), \xi) = Df(x)[\xi], \quad \xi \in T_x M$$

を満たすように定義される. このような $\text{grad } f(x)$ は一意的存在することが示される [1].

最急降下法では探索方向を $\eta_k = -\text{grad } f(x_k)$ として反復を行う. 共役勾配法では初期の探索方向是最急降下法と同じく $\eta_0 = -\text{grad } f(x_0)$ とするが, その次以降の反復では, η_k を最急降下方向 $-\text{grad } f(x_k)$ と 1 つ前の探索方向 η_{k-1} の β_{k-1} 倍の和として計算する. ここで, β_{k-1} はスカラーである. ところが, リーマン多様体 M では $\text{grad } f(x_k) \in T_{x_k} M$ かつ $\eta_{k-1} \in T_{x_{k-1}} M$ であるため, これらをそのまま足し合わせることができないので, η_{k-1} を $T_{x_k} M$ に移す必要がある. そこで重要な役割を果たすのが次の vector transport と呼ばれる写像である [1].

定義 2.2. $\mathcal{T}: \bigcup_{x \in M} (T_x M \times T_x M) \rightarrow TM$ が次の性質を満たすとき, $\mathcal{T}$ を M の vector transport という. すなわち, 任意の $x \in M$ および任意の $\xi, \eta, \zeta \in T_x M$ に対して,

1. レトラクション R が存在して, $\mathcal{T}_\xi(\eta) \in T_{R_x(\xi)} M$.
2. $\mathcal{T}_{0_x}(\eta) = \eta$.
3. 任意の $a, b \in \mathbb{R}$ に対して, $\mathcal{T}_\xi(a\eta + b\zeta) = a\mathcal{T}_\xi(\eta) + b\mathcal{T}_\xi(\zeta)$.

たとえば, レトラクション R に対して,

$$\mathcal{T}_\xi(\eta) = DR_x(\xi)[\eta], \quad \xi, \eta \in T_x M, x \in M$$

によって $\mathcal{T}$ を定義すると, $\mathcal{T}$ は vector transport となる. あるいは接空間への適当な射影を用いて vector transport を定義することもできる. このようにして, vector transport が得られているときには, 共役勾配法における探索方向を

$$\eta_k = -\text{grad } f(x_k) + \beta_k \mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1}) \quad (2.2)$$

と計算することができる. ここで, $R_{x_{k-1}}(\alpha_{k-1}\eta_{k-1}) = x_k$ であるから, $\mathcal{T}_{\alpha_{k-1}\eta_{k-1}}(\eta_{k-1}) \in T_{x_k}\mathcal{M}$ であることに注意する.

3 予測誤差法と対応する最適化問題

次の離散時間状態空間モデルを考える:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + Kv_t, \\ y_t &= Cx_t + Du_t + v_t. \end{aligned} \quad (3.1)$$

ここで, $u_t \in \mathbb{R}^m$ はシステムの入力, $y_t \in \mathbb{R}^p$ は出力, $x_t \in \mathbb{R}^n$ は状態を表す. また, $v_t \in \mathbb{R}^p$ は平均 0 で独立同一分布に従う確率過程であり, 観測誤差を表す. 行列 $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$, $K \in \mathbb{R}^{n \times p}$ は定数行列であり, システムのパラメータである. システム (3.1) はただ一つの雑音源 v_t をもち, このようなシステムをイノベーション形式という [4]. プロセス雑音と観測雑音が無相関であるようなシステムの同定も, イノベーション形式 (3.1) のシステム同定に帰着できることが知られている [5].

ここで, パラメータ全体を行列として扱うために, 行列空間の直積

$$M := \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times m} \times \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times m} \times \mathbb{R}^{n \times p}$$

を定義し, システム行列 A, B, C, D, K の組を $\Theta := (A, B, C, D, K) \in M$ とする. Θ によってパラメータ付けられたモデル (3.1) の 1 段先予測 $\hat{y}_{t|t-1}(\Theta)$ は次の式を満たす [6, 11]:

$$\begin{aligned} \hat{x}_{t+1|t} &= (A - KC)\hat{x}_{t|t-1} + (B - KD)u_t + Ky_t, \\ \hat{y}_{t|t-1}(\Theta) &= C\hat{x}_{t|t-1} + Du_t. \end{aligned} \quad (3.2)$$

モデル (3.2) はカルマンフィルタと密接な関係を持つことが知られている [5]. [11] では, N をデータ長として, 予測誤差ベクトル

$$e(\Theta) := \left(y_1^\top - \hat{y}_{1|0}^\top(\Theta) \quad \dots \quad y_N^\top - \hat{y}_{N|N-1}^\top(\Theta) \right)^\top$$

の 2 ノルムの最小化が議論されている. ここで, $\cdot^\top$ はベクトルや行列の転置を表す. ただし, [11] では $\Theta = (A, B, C, D, K)$ を行列の組と見るのではなく, $n^2 + nm + mp + 2np$ 次元の列ベクトル

$$\theta = \left(\text{vec}(A)^\top \quad \text{vec}(B)^\top \quad \text{vec}(C)^\top \quad \text{vec}(D)^\top \quad \text{vec}(K)^\top \right)^\top$$

を変数としている。ここで、 $\text{vec}(\cdot)$ は行列の各列ベクトルを 1 列に並べた列ベクトルであり、たとえば

$$\text{vec}(A) = (A_{11}, \dots, A_{n1}, \dots, A_{1n}, \dots, A_{nn})^\top \in \mathbb{R}^{n^2}$$

である。

未知パラメータを推定するための最適化問題を導出する際には、列ベクトル θ と行列の組 $\Theta \in M$ は、理論的にはどちらの表記を用いて議論しても本質的に同じことである。そこで、[11] で用いられている θ の代わりに、以降の議論では Θ を用いることにし、行列変数 A, B, C, D, K についての最適化アルゴリズムを提案する。最小化したい目的関数を $\bar{f}$ とおき、次の最適化問題を考える：

問題 3.1.

$$\begin{aligned} & \text{minimize} \quad \bar{f}(\Theta) := \|e(\Theta)\|_2^2, \\ & \text{subject to} \quad \Theta \in M. \end{aligned}$$

ここで、 $\|\cdot\|_2$ はベクトルの 2-ノルムを表す。

最急降下法や共役勾配法などを用いるには、目的関数 $\bar{f}(\Theta) = \|e(\Theta)\|_2^2$ の勾配が必要である。以下ではこの勾配を、列ベクトル θ ではなく行列の組 $\Theta = (A, B, C, D, K)$ に関して求める。まず、

$$\bar{f}(\Theta) = \|e(\Theta)\|_2^2 = \sum_{t=1}^N \|y_t - \hat{y}_{t|t-1}(\Theta)\|_2^2$$

であるから、 $\bar{f}$ の点 $\Theta = (A, B, C, D, K) \in M$ における $\Theta' = (A', B', C', D', K') \in M$ への方向微分は、 $\hat{y}_{t|t-1}$ の点 Θ における Θ' への方向微分 $D\hat{y}_{t|t-1}(\Theta)[\Theta']$ を用いて

$$D\bar{f}(\Theta)[\Theta'] = -2 \sum_{t=1}^N (y_t - \hat{y}_{t|t-1}(\Theta))^\top D\hat{y}_{t|t-1}(\Theta)[\Theta'] \quad (3.3)$$

と表される。一方、 $\hat{y}_{t|t-1}$ の方向微分は (3.2) より、 $\hat{x}_{t|t-1}$ の点 Θ における Θ' への方向微分 $D\hat{x}_{t|t-1}(\Theta)[\Theta']$ を用いて

$$D\hat{y}_{t|t-1}(\Theta)[\Theta'] = C D\hat{x}_{t|t-1}(\Theta)[\Theta'] + C' \hat{x}_{t|t-1} + D' u_t \quad (3.4)$$

となる。再び (3.2) により、 $D\hat{x}_{t+1|t}(\Theta)[\Theta']$ と $D\hat{x}_{t|t-1}(\Theta)[\Theta']$ の間の関係式

$$\begin{aligned} D\hat{x}_{t+1|t}(\Theta)[\Theta'] &= (A - KC) D\hat{x}_{t|t-1}(\Theta)[\Theta'] \\ &\quad + (A' - K'C - KC') \hat{x}_{t|t-1} + (B' - K'D - KD') u_t + K' y_t \end{aligned} \quad (3.5)$$

が得られる。初期状態の推定 $\hat{x}_{0|0}$ を x_0 と定義すると、 $\hat{x}_{1|0}$ は x_0, u_0, y_0 を用いて

$$\hat{x}_{1|0} = (A - KC)x_0 + (B - KD)u_0 + Ky_0$$

となるから、その方向微分は

$$D\hat{x}_{1|0}(\Theta)[\Theta'] = (A' - K'C - KC')x_0 + (B' - K'D - KD')u_0 + K'y_0 \quad (3.6)$$

となる。

方向微分 $D\hat{x}_{t|t-1}(\Theta)[\Theta']$ は漸化式 (3.5) および (3.6) により、

$$\begin{aligned} D\hat{x}_{t|t-1}(\Theta)[\Theta'] &= \sum_{i=0}^{t-1} (A - KC)^{t-i-1} (A' - K'C - KC') \hat{x}_{i|i-1} \\ &\quad + \sum_{i=0}^{t-1} (A - KC)^{t-i-1} ((B' - K'D - KD')u_i + K'y_i) \end{aligned}$$

と計算される。これを (3.4) に代入すると $D\hat{y}_{t|t-1}(\Theta)$ が求まり、さらに (3.3) に代入すると $D\bar{f}(\Theta)[\Theta']$ が求まる。 $e_t(\Theta) := y_t - \hat{y}_{t|t-1}(\Theta)$ とし、これを単に e_t と書くことにすると、

$$\begin{aligned} D\bar{f}(\Theta)[\Theta'] &= -2 \sum_{t=1}^N e_t^\top (CD\hat{x}_{t|t-1}(\Theta)[\Theta'] + C'\hat{x}_{t|t-1} + D'u_t) \\ &= -2 \sum_{t=1}^N \text{tr} (C'^\top e_t \hat{x}_{t|t-1}^\top + D'^\top e_t u_t^\top) \\ &\quad - 2 \sum_{t=1}^N \sum_{i=0}^{t-1} e_t^\top C (A - KC)^{t-i-1} ((A' - K'C - KC') \hat{x}_{i|i-1} + (B' - K'D - KD')u_i + K'y_i) \\ &= \text{tr}(A'^\top G_A) + \text{tr}(B'^\top G_B) + \text{tr}(C'^\top G_C) + \text{tr}(D'^\top G_D) + \text{tr}(K'^\top G_K) \end{aligned}$$

となる。ここで、

$$\begin{aligned} G_A &= -2 \sum_{t=1}^N \sum_{i=0}^{t-1} W_{t-i-1}^\top e_t \hat{x}_{i|i-1}^\top, \quad G_B = -2 \sum_{t=1}^N \sum_{i=0}^{t-1} W_{t-i-1}^\top e_t u_i^\top, \\ G_C &= -2 \sum_{t=1}^N \left(e_t \hat{x}_{t|t-1}^\top - \sum_{i=0}^{t-1} K^\top W_{t-i-1}^\top e_t \hat{x}_{i|i-1}^\top \right), \\ G_D &= -2 \sum_{t=1}^N \left(e_t u_t^\top - \sum_{i=0}^{t-1} K^\top W_{t-i-1}^\top e_t u_i^\top \right), \\ G_K &= -2 \sum_{t=1}^N \sum_{i=0}^{t-1} W_{t-i-1}^\top e_t (y_i - C\hat{x}_{i|i-1} - Du_i)^\top, \\ W_j &= C(A - KC)^j, \quad j = 0, 1, 2, \dots, N-1 \end{aligned}$$

とおいた。したがって、ユークリッド勾配 $\nabla \bar{f}(\Theta)$ は上記の G_A, G_B, G_C, G_D, G_K を用いて

$$\nabla \bar{f}(\Theta) = (G_A, G_B, G_C, G_D, G_K)$$

と表される。

4 商多様体上の最適化問題

4.1 入出力等価性に基づく同値関係から得られる商多様体

本節では、以下の仮定を置く：

1. ランクに関する条件：

$$\text{rank} \begin{pmatrix} A & B & K \end{pmatrix} = \text{rank} \left(\begin{pmatrix} C^\top & (CA)^\top & \cdots & (CA^{n-1})^\top \end{pmatrix}^\top \right) = n.$$

2. $A - KC$ は安定である [3, 9].

3. 初期状態 x_0 は $x_0 = 0$ である.

仮定 1. の条件は、[8] の定理 2.2.60 の条件を今回の設定において書き直したものである。また、 $\hat{x}_{t|t-1}$ は (3.2) により

$$\hat{x}_{t|t-1}(\Theta) = (A - KC)^t x_0 + \sum_{i=0}^{t-1} (A - KC)^{t-i-1} ((B - KD)u_i + Ky_i)$$

と書けるので、仮定 2. の下で第 1 項 $(A - KC)^t x_0$ は t が十分大きければ 0 に近付き、 $\hat{x}$ は定常状態で初期状態 x_0 に依存しないため、仮定 3. は自然である。

本節では、次の Θ の探索領域を考える：

$$\bar{\mathcal{M}} := \{\Theta \in M \mid A, B, C, K \text{ は仮定 1. を満たす}\}.$$

$\bar{\mathcal{M}}$ にさらに商構造を導入する。 $GL(n)$ を n 次一般化線形群とすると、任意の $T \in GL(n)$ に対して $\Theta = (A, B, C, D, K)$ と $\Theta_T = (T^{-1}AT, T^{-1}B, CT, D, T^{-1}K)$ は入出力等価なシステムを実現する。実際、 Θ でパラメータ付けられた (3.1) の伝達関数を $G_\Theta(z)$ とすると、

$$G_\Theta(z) = C(zI_n - A)^{-1} \begin{pmatrix} B & K \end{pmatrix} + \begin{pmatrix} D & I_p \end{pmatrix}$$

となり、

$$G_{\Theta_T}(z) = CT(zI_n - T^{-1}AT)^{-1} \begin{pmatrix} T^{-1}B & T^{-1}K \end{pmatrix} + \begin{pmatrix} D & I_p \end{pmatrix} = G_\Theta(z)$$

が導かれる。ここで、 I_n は n 次単位行列である。したがって、 Θ と Θ_T についての予測誤差も等しくなり、 $\bar{f}(\Theta) = \bar{f}(\Theta_T)$ が成り立つ。そこで、 Θ と任意の $T \in GL(n)$ に対する Θ_T を同一視するために、 $\bar{\mathcal{M}}$ に同値関係 $\sim$ を、

$$\begin{aligned} (A_1, B_1, C_1, D_1, K_1) &\sim (A_2, B_2, C_2, D_2, K_2) \\ \iff T \in GL(n) \text{ が存在して } (A_2, B_2, C_2, D_2, K_2) &= (T^{-1}A_1T, T^{-1}B_1, C_1T, D_1, T^{-1}K_1) \end{aligned}$$

と定義する. この同値関係によって同値類 $[\Theta] := \{\Theta_1 \in \bar{\mathcal{M}} \mid \Theta_1 \sim \Theta\}$ を定義すると, $\mathcal{M} := \{[\Theta] \mid \Theta \in \bar{\mathcal{M}}\}$ は商多様体となる. さらに, 群 $GL(n)$ の $\bar{\mathcal{M}}$ への作用 $\circ$ を

$$T \circ (A, B, C, D, K) = (T^{-1}AT, T^{-1}B, CT, D, T^{-1}K)$$

と定義すると, $\mathcal{M}$ は $\bar{\mathcal{M}}/GL(n)$ と表すことができる. こうして次の最適化問題 4.1 が導かれる:

問題 4.1.

$$\begin{aligned} & \text{minimize} \quad f([\Theta]) := \|e(\Theta)\|_2^2, \\ & \text{subject to} \quad [\Theta] \in \mathcal{M}. \end{aligned}$$

4.2 問題 4.1 の幾何学

多様体 $\bar{\mathcal{M}}$ にリーマン計量 $\bar{g}$ を, 任意の $(a_1, b_1, c_1, d_1, k_1), (a_2, b_2, c_2, d_2, k_2) \in T_{\Theta}\bar{\mathcal{M}}$ に対して

$$\begin{aligned} & \bar{g}_{\Theta}((a_1, b_1, c_1, d_1, k_1), (a_2, b_2, c_2, d_2, k_2)) \\ &= \text{tr}(a_1^{\top} a_2) + \text{tr}(b_1^{\top} b_2) + \text{tr}(c_1^{\top} c_2) + \text{tr}(d_1^{\top} d_2) + \text{tr}(k_1^{\top} k_2) \end{aligned}$$

と定める. この $\bar{g}$ に基づいて商多様体 $\mathcal{M}$ 上のリーマン計量を定めたい. そのためには $\mathcal{M}$ の接ベクトルを $\bar{\mathcal{M}}$ の接ベクトルと対応づけられれば良いが, そこで $\mathcal{M}$ の接ベクトルの水平リフトという概念が重要となる. 水平リフトを定義するには $\bar{\mathcal{M}}$ の水平空間が必要となるが, これは垂直空間の直交補空間として定義される. 以下では垂直空間, 水平空間, 水平リフトの順に $\bar{\mathcal{M}}$ および $\mathcal{M}$ の幾何学を調べ, 最後に $\mathcal{M}$ のリーマン計量を導入しよう. これらの概念図を図 4.1 に示しておく.

まず, $\pi: \bar{\mathcal{M}} \rightarrow \mathcal{M}$ を任意の $\Theta \in \bar{\mathcal{M}}$ に対して $\pi(\Theta) = [\Theta]$ と定義する. このとき, $\Theta \in \bar{\mathcal{M}}$ における垂直空間 $\mathcal{V}_{\Theta}$ は $\mathcal{V}_{\Theta} := T_{\Theta}\pi^{-1}([\Theta])$ と定義される. $\mathcal{V}_{\Theta}$ の具体的な表示を求めよう. $\pi^{-1}([\Theta]) \subset \bar{\mathcal{M}}$ 上の任意の曲線 $\Theta(t)$ で $\Theta(0) = \Theta$ となるものを考えると, これは $T(0) = I_n$ となる $GL(n)$ 上の曲線 $T(t)$ を用いて

$$\Theta(t) = (T(t)^{-1}AT(t), T(t)^{-1}B, CT(t), D, T(t)^{-1}K) \quad (4.1)$$

と表される. ここで, $\dot{T}(t) := \frac{d}{dt}(T(t))$ とおくと, $T(t)T^{-1}(t) = I_n$ の両辺を t で微分して $t=0$ とおくことで,

$$\dot{T}(0)T^{-1}(0) + T(0)\frac{d}{dt}(T^{-1}(t))\big|_{t=0} = 0$$

となるから,

$$\frac{d}{dt}(T^{-1}(t))\big|_{t=0} = -T(0)^{-1}\dot{T}(0)T(0)^{-1} = -\dot{T}(0).$$

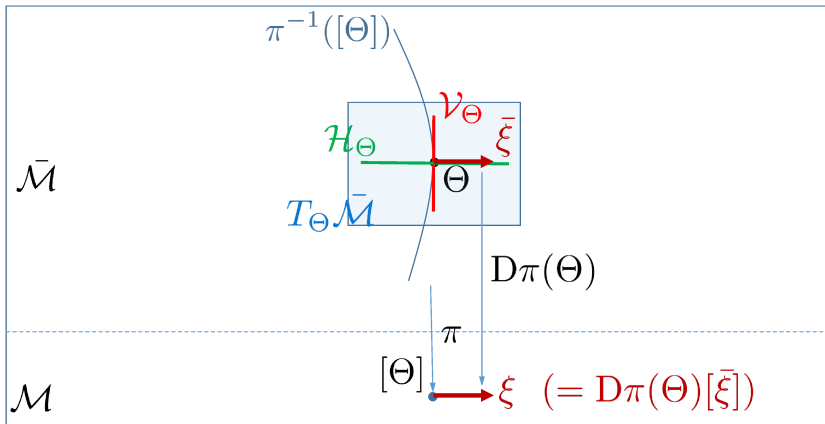


図 4.1: 接空間 $T_{\Theta}\bar{\mathcal{M}}$ における垂直空間 $\mathcal{V}_{\Theta}$ と水平空間 $\mathcal{H}_{\Theta}$ および、商多様体 $\mathcal{M}$ 上の接ベクトル ξ の水平リフト $\bar{\xi}$ の概念図。

よって、(4.1) の両辺を t で微分して $t=0$ とおくことで、

$$\dot{\Theta}(0) = (-\dot{T}(0)A + A\dot{T}(0), -\dot{T}(0)B, C\dot{T}(0), 0, -\dot{T}(0)K)$$

を得る。したがって、垂直空間の定義から、

$$\mathcal{V}_{\Theta} = \{(-T'A + AT', -T'B, CT', 0, -T'K) \mid T' \in \mathbb{R}^{n \times n}\} \quad (4.2)$$

となる。

続いて、 Θ における水平空間 $\mathcal{H}_{\Theta}$ を、 $\bar{\mathcal{M}}$ の接空間 $T_{\Theta}\bar{\mathcal{M}}$ における垂直空間 $\mathcal{V}_{\Theta}$ の、内積 $\bar{g}_{\Theta}$ に関する直交補空間として定義する。 $\Theta' = (A', B', C', D', K')$ を水平空間 $\mathcal{H}_{\Theta}$ 上の点とすると、 Θ' は $T_{\Theta}\bar{\mathcal{M}}$ において垂直空間 $\mathcal{V}_{\Theta}$ の任意の元と内積 $\bar{g}_{\Theta}$ に関して直交するから、

$$\bar{g}_{\Theta}((-T'A + AT', -T'B, CT', 0, -T'K), \Theta') = 0$$

が任意の $T' \in \mathbb{R}^{n \times n}$ に対して成り立つ。ゆえに、

$$\text{tr}(A'^{\top}(-T'A + AT')) - \text{tr}(B'^{\top}T'B) + \text{tr}(C'^{\top}CT') - \text{tr}(K'^{\top}T'K) = 0$$

であり、これを整理すると

$$\text{tr}(T'^{\top}(A^{\top}A' - A'A^{\top} - B'B^{\top} + C^{\top}C' - K'K^{\top})) = 0$$

となる。 T' は任意だから、

$$A^{\top}A' - A'A^{\top} - B'B^{\top} + C^{\top}C' - K'K^{\top} = 0$$

を得る。したがって、水平空間は

$$\mathcal{H}_\Theta = \{(A', B', C', D', K') \mid A'^\top A' - A'A'^\top - B'B'^\top + C'C'^\top - K'K'^\top = 0\} \quad (4.3)$$

となる。

さて、商多様体 $\mathcal{M}$ の接ベクトル $\xi \in T_\Theta \mathcal{M}$ を $\bar{\mathcal{M}}$ の接ベクトルとして表現するために、 ξ の水平リフト $\bar{\xi} \in T_\Theta \bar{\mathcal{M}}$ を $D\pi(\Theta)[\bar{\xi}] = \xi$ を満たすものとして定義する。 ξ に対して、このような $\bar{\xi}$ は一意に存在することが示される [1]。

以上の議論を踏まえて、 $\mathcal{M}$ 上のリーマン計量 g を

$$g_{[\Theta]}(\xi, \zeta) := \bar{g}_\Theta(\bar{\xi}, \bar{\zeta}), \quad [\Theta] \in \mathcal{M}, \quad \xi, \zeta \in T_{[\Theta]} \mathcal{M}$$

と定める。ここで、 $\bar{\xi}$ および $\bar{\zeta}$ はそれぞれ ξ および ζ の水平リフトである。 $\bar{\mathcal{M}}$ 上のリーマン計量 $\bar{g}$ は行列空間 M の標準内積から誘導されたものであるから、リーマン多様体 $\bar{\mathcal{M}}$ 上の $\bar{f}|_{\bar{\mathcal{M}}}$ のリーマン計量 $\bar{g}$ に関する勾配は 3 節で求めた $\bar{f}$ のユークリッド勾配と一致する：

$$\text{grad } \bar{f}(\Theta) = \nabla \bar{f}(\Theta), \quad \Theta \in \bar{\mathcal{M}}.$$

商多様体 $\mathcal{M}$ 上の実際の数値計算においては、抽象的な量である $\text{grad } f([\Theta])$ をそのまま用いることはできないので、 Θ への水平リフト $\overline{\text{grad } f}_\Theta$ を用いる。このとき、

$$\overline{\text{grad } f}_\Theta = \text{grad } \bar{f}(\Theta)$$

となる。同様の性質は一般のリーマン商多様体についても成り立つ [1]。まとめると、

$$\overline{\text{grad } f}_\Theta = \nabla \bar{f}(\Theta) = (G_A, G_B, G_C, G_D, G_K)$$

となる。

最後に、水平空間への射影に基づく vector transport について考察する。 $P_\Theta^h : T_\Theta \bar{\mathcal{M}} \rightarrow \mathcal{H}_\Theta$ を水平空間 $\mathcal{H}_\Theta$ への直交射影とすると、

$$\overline{\mathcal{T}_\xi(\eta)}_{\Theta+\bar{\xi}} = P_{\Theta+\bar{\xi}}^h(\bar{\eta}), \quad \xi, \eta \in T_{[\Theta]} \mathcal{M}, \quad \Theta \in \bar{\mathcal{M}} \quad (4.4)$$

と定義したとき、この $\mathcal{T}$ は $\mathcal{M}$ 上の vector transport になることが知られている [1]。ここで、上式の左辺は $\mathcal{T}_\xi(\eta)$ の $\Theta + \bar{\xi}$ への水平リフトである。そこで、直交射影 P_Θ^h の具体的な表示を求めよう。 $\mathcal{M}$ の接ベクトル $\eta = (a, b, c, d, k) \in T_\Theta \bar{\mathcal{M}}$ を任意に選ぶと、水平空間が垂直空間の $T_\Theta \bar{\mathcal{M}}$ の直交補空間であることから、

$$\eta = \eta^h + \eta^v, \quad \eta^h \in \mathcal{H}_\Theta, \quad \eta^v \in \mathcal{V}_\Theta$$

と一意的に分解することができる。このとき (4.2) より、 $T' \in \mathbb{R}^{n \times n}$ が存在して

$$\eta^v = (-T'A + AT', -T'B, CT', 0, -T'K)$$

となるから、 η^h は

$$\eta^h = (a + T'A - AT', b + T'B, c - CT', d, k + T'K)$$

と書くことができる。 $\eta^h \in \mathcal{H}_\Theta$ より (4.3) から

$$\begin{aligned} & A^\top(a + T'A - AT') - (a + T'A - AT')A^\top \\ & - (b + T'B)B^\top + C^\top(c - CT') - (k + T'K)K^\top = 0 \end{aligned}$$

となる。これを整理すると、 T' の線形方程式

$$\begin{aligned} & A^\top T'A + AT'A^\top - T'(AA^\top + BB^\top + KK^\top) - (A^\top A + C^\top C)T' \\ & = aA^\top - A^\top a + bB^\top - C^\top c + kK^\top \end{aligned}$$

が得られる。この方程式の解を $T' = T'_\Theta(\eta)$ と書くと、直交射影 P_Θ^h は $\eta \in T_\Theta \bar{\mathcal{M}}$ に対して

$$\begin{aligned} P_\Theta^h(\eta) &= (a + T'A - AT', b + T'B, c - CT', d, k + T'K) \\ &= \eta + (T'_\Theta(\eta)A - AT'_\Theta(\eta), T'_\Theta(\eta)B, -CT'_\Theta(\eta), 0, T'_\Theta(\eta)K) \end{aligned}$$

と作用することになる。

5 最急降下法と共役勾配法

問題 3.1 はユークリッド空間 M 上の最適化問題であるから、通常の制約なし問題に対する最適化アルゴリズムを適用することができる。具体的には、最急降下法では Θ_k における探索方向を $\eta_k = -\nabla \bar{f}(\Theta_k)$ と計算し、共役勾配法では $\eta_0 = -\nabla \bar{f}(\Theta_0)$ 、また $k \geq 0$ のときは $\eta_k = -\nabla \bar{f}(\Theta_k) + \beta_k \eta_{k-1}$ と計算する。スカラー β_k の計算方法は多くのものが提案されている [2]。

一方、問題 4.1 は商多様体 $\mathcal{M}$ 上の問題であるから、レトラクションを用いる必要がある。そこで、

$$R_{[\Theta]}(\eta) = \pi(\Theta + \bar{\eta})$$

とすれば、 R は $\mathcal{M}$ 上のレトラクションであることがわかる。すなわち、 $\mathcal{M}$ 上の点 $\Phi_k = [\Theta_k]$ と探索方向 ζ_k から $\Phi_{k+1} \in \mathcal{M}$ を

$$\Phi_{k+1} = R_{\Phi_k}(\alpha_k \zeta_k) = \pi(\Theta_k + \alpha_k \bar{\zeta}_{k\Theta_k})$$

によって生成する。ここで、 $\bar{\zeta}_{k\Theta_k} \in T_{\Theta_k} \bar{\mathcal{M}}$ は $\zeta_k \in T_{\Phi_k} \mathcal{M}$ の Θ_k への水平リフトである。数値的には代表元を用いて計算することになるので、実際の数値計算では

$$\Theta_{k+1} = \Theta_k + \alpha_k \bar{\zeta}_{k\Theta_k}$$

を用いることになる。

最急降下法における探索方向 ζ_k の水平リフトは

$$\bar{\zeta}_{k\Theta_k} = -\overline{\text{grad } f}_{\Theta_k} = -\nabla \bar{f}(\Theta_k)$$

となるので、結局、問題 3.1 および問題 4.1 に対する最急降下法は本質的に同じアルゴリズムであることがわかる。

次に、共役勾配法について考察する。 $\mathcal{M}$ 上の共役勾配法においては、(2.2) のように、探索方向の計算において vector transport を用いる必要がある。水平空間への直交射影から定まる (4.4) による $\mathcal{T}$ を vector transport として用いることにすると、 $\Phi_k = [\Theta_k] \in \mathcal{M}$ における探索方向 $\zeta_k \in T_{\Phi_k} \mathcal{M}$ は、

$$\zeta_k = -\text{grad } f(\Phi_k) + \beta_k \mathcal{T}_{\alpha_{k-1}\zeta_{k-1}}(\zeta_{k-1})$$

によって計算される。実際の数値計算では両辺の水平リフトを考え、

$$\overline{\zeta_{k\Theta_k}} = -\overline{\text{grad } f_{\Theta_k}} + \beta_k P_{\Theta_k}^h(\overline{\zeta_{k-1\Theta_{k-1}}}).$$

を用いることになる。商多様体 $\mathcal{M}$ 上の共役勾配法では、ユークリッド空間における共役勾配法にはなかった水平空間への直交射影が現れており、これによって効率化が期待される。

6 結論

本稿では、予測誤差法に基づくシステム同定に対応する最適化問題を行列変数の問題として扱うとともに、システムの入出力等価性に着目することで商多様体を構成し、リーマン多様体上の問題を導出した。ユークリッド空間および多様体上の両問題に対して、目的関数の勾配や幾何学的な量を求めることで、最急降下法および共役勾配法を導出した。とくに、今回の問題に対する多様体上の共役勾配法は、ユークリッド空間における共役勾配法と本質的に異なるアルゴリズムであることを示した。

提案アルゴリズムは数値的にも良い結果を示すことが実証されている。たとえば、有名なシステム同定法の一つである N4SID で適切に解けない例として [3] で紹介されている問題に対して提案アルゴリズムを適用すると、目的関数の勾配が 0 に収束する様子を確認できる。また、同じ例に対しては、リーマン商多様体上の共役勾配法がユークリッド空間における共役勾配法よりも速い収束を達成することも観察されている。これらの結果の詳細については [10] を参照されたい。我々の提案法は、[5] で紹介されているような既存の部分空間法で得られた同定結果をアルゴリズムの初期点として採用することで、既存法の結果を改善できることにも注意しておく。

謝辞

本研究は日本学術振興会科学研究費補助金(課題番号: 16K17647)の助成を受けている。

参考文献

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

- [2] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific Journal on Optimization*, **2** (2006) 35–58.
- [3] M. Jansson and B. Wahlberg, On consistency of subspace methods for system identification, *Automatica*, **34** (1998), 1507–1519.
- [4] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Pearson, 2000.
- [5] T. Katayama, *Subspace Methods for System Identification*, Springer, 2006.
- [6] L. Ljung, *System Identification: Theory for the User*, 2nd ed., Prentice-Hall, 1999.
- [7] T. McKelvey, A. Helmersson, and T. Ribarits, Data driven local coordinates for multivariable linear systems and their application to system identification, *Automatica*, **40** (2004), 1629–1635.
- [8] R. Peeters, *System Identification Based on Riemannian Geometry: Theory and Algorithms*, Thesis Publ., 1994.
- [9] S. J. Qin, W. Lin, and L. Ljung, A novel subspace identification approach with enforced causal models, *Automatica*, **41** (2005), 2043–2053.
- [10] H. Sato and K. Sato, Riemannian optimal system identification algorithm for linear MIMO systems, *IEEE Control Systems Letters*, **1** (2017), 376–381.
- [11] A. Wills and B. Ninness, On gradient-based search for multivariable system estimates, *IEEE Transactions on Automatic Control*, **53** (2008), 298–306.