



Title	構造化低ランク近似アルゴリズムとその初期化戦略に関する研究
Author(s)	吉野, 夏樹
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16986号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16986
Doc URL	https://hdl.handle.net/2115/99830
Type	doctoral thesis
File Information	Natsuki_Yoshino.pdf, 全文



構造化低ランク近似アルゴリズムと
その初期化戦略に関する研究

北海道大学 大学院情報科学院
情報科学専攻情報理工学コース
情報数理学研究室
吉野 夏樹

A Study on Structured Low-Rank Approximation Algorithms and Their Initialization Strategy

In real-world data, prior structural properties often arise from physical laws or system configurations, and these properties are frequently expressed as linear equations involving structured matrices as coefficient matrices. For such equations to hold, the associated matrices must satisfy both structural constraints and prescribed rank conditions. However, due to measurement noise and numerical errors, observed matrices are typically full rank and may not admit an exact solution. In this context, it is essential to adjust the matrices without violating the underlying structural properties. Specifically, finding a nearby low-rank matrix while preserving the original structure is known as Structured Low-Rank Approximation (SLRA), which appears as a fundamental and critical problem in various fields such as signal processing, system identification, and time-series analysis. The simplest approach to low-rank approximation is based on singular value decomposition (SVD), where small singular values are truncated. However, this method does not preserve the structure of the original matrix, resulting in approximations that do not correspond to any valid model or system. The difficulty of structure-preserving low-rank approximation lies in the fact that, except for special cases, the set of structured matrices with a given rank cannot be explicitly described, and this set is non-convex. Consequently, expressing optimization constraints explicitly is challenging, and no direct method exists to compute the global optimum. Even iterative methods applicable to a wide range of cases remain largely unexplored. For these reasons, numerous algorithms have been proposed to address SLRA problems, tailored to the characteristics and requirements of specific applications. This study focuses on SLRA methods based on null-space representations of low-rank structure. Although these approaches have demonstrated practical effectiveness, they lack sufficient theoretical and implementation frameworks, leaving several aspects without clear design guidelines. The objective of this dissertation is to address these gaps by pro-

viding theoretical insights and proposing improvements for two unresolved issues in SLRA using null-space representation. First, we analyze the formulation of the Structured Total Least Norm method (STLN-SLRA) and derive a criterion for selecting initial equations that minimizes computational error. This criterion significantly reduces computational complexity compared to existing methods, achieving substantial efficiency gains. Second, we identify the vulnerability of initialization strategies in SLRA algorithms, particularly their sensitivity to noise due to the phenomenon of singular value swapping. To overcome this issue, we propose new initialization strategies that mitigate this effect. Numerical experiments confirm that the proposed methods maintain accuracy comparable to conventional approaches while improving computational efficiency and robustness. Numerical experiments demonstrate that the proposed approaches improve efficiency and estimation accuracy under noisy conditions.

目次

第 1 章 序論	4
1.1 本研究の背景	4
1.1.1 SLRA 問題の定式化	6
1.1.2 関連研究	7
1.2 本論文の目的	10
1.3 記号の定義	11
第 2 章 数値モデルと構造化行列	13
2.1 低ランク Hankel 行列と線形漸化関係	13
2.1.1 Hankel 行列のランク特性と有限ランクベクトル	15
2.1.2 LRR	15
2.1.3 LRR と変調指数関数の和	16
2.1.4 最小 LRR	18
2.1.5 非最小 LRR と Hankel 行列の右零空間	19
2.1.6 Hankel 行列の SLRA の解釈	20
2.1.7 Hankel 構造より派生する行列構造	22
2.2 近似最大公約多項式計算	23
2.2.1 数学的準備	24
2.2.2 フル・Sylvester 行列	26
2.2.3 数値的な最大公約多項式	28
2.2.4 Sylvester 行列の零空間の性質を利用した近似 GCD 計算	30
2.2.5 近似 GCD 問題の陽表現と変数射影 [1]	31
2.2.6 変数射影法による近似 GCD 問題との関係 [1]	33

第 3 章	非構造化低ランク近似問題 [2]	37
3.1	重み付き低ランク近似問題	39
3.2	特殊なケース: 重みなし低ランク近似	40
3.3	交互射影アルゴリズム	42
3.3.1	交互射影法	43
3.3.2	問題の分離	43
3.3.3	欠損値推定	44
3.3.4	初期近似値の構成	44
3.3.5	アルゴリズム	44
3.4	本章のまとめ	45
第 4 章	構造化低ランク近似	47
4.1	はじめに	47
4.1.1	構造化低ランク近似の定式化	48
4.1.2	本章の構成	51
4.2	低ランク性の零空間表現	51
4.3	STLN-SLRA	52
4.3.1	アルゴリズムの導出	53
4.3.2	初期方程式の選択における問題点	56
4.4	VP-SLRA	57
4.4.1	VP-SLRA の定式化	58
4.4.2	目的関数の勾配及びヘシアン	60
4.4.3	勾配およびヤコビアン の計算効率化に向けた構造の活用	66
4.5	零空間表現以外の低ランク性の記述法	67
4.5.1	像空間表現法 [3]	67
4.5.2	核ノルム緩和法	72
4.5.3	常微分方程式の活用	76
4.5.4	IRLS を適用する手法	77

第 5 章	STLN-SLRA における初期方程式の選択方法	78
5.1	STLN-SLRA における問題点	79
5.1.1	$\mathcal{S}(c)$ が雑音を含まない場合の各列ベクトルの線形従属性	79
5.1.2	$\mathcal{S}(c)$ が雑音を含む場合	82
5.2	従来手法	82
5.3	提案法 [4]	84
5.3.1	本手法の計算量	86
5.4	数値実験	87
5.4.1	近似 GCD 計算への適用例	87
5.4.2	雑音除去への適用例	92
5.5	本章のまとめ	96
第 6 章	低ランク性の零空間表現を使用する SLRA 手法の初期化戦略	97
6.1	はじめに	97
6.2	特異値の交代	97
6.2.1	予備実験	99
6.3	提案法 1：特異ベクトルの選択	102
6.4	提案法 2：特異ベクトルの線型結合	103
6.5	数値実験	103
6.5.1	数値例：雑音除去性能の評価	103
6.5.2	VP-SLRA による近似 GCD 計算での数値例	106
6.5.3	異なる次数の多項式に対する提案法の有効性	108
6.5.4	異なる雑音強度に対する提案法の有効性	110
6.5.5	考察	113
6.6	本章のまとめ	114
第 7 章	結論	115
	謝辞	117

第1章 序論

1.1 本研究の背景

現代社会では、計算機の性能向上とそれに支えられた情報技術の普及により、私たちが日々扱うデータの量と多様性が飛躍的に増大している。その結果、データが持つ価値はかつてないほど高まり、データの解析や処理の重要性も一層増している。実世界のデータには、物理法則やシステム構造、経済法則といった背後の仕組みに由来する規則性が存在することが多い。このような規則性は、しばしばデータを要素とする構造化行列を係数とする斉次線形方程式の形で記述でき、数理モデルとして体系的に取り扱われる。その結果、モデルに基づく解析や処理が可能となり、データから信頼性の高い知見を得たり、計算の安定性や効率を高めたりすることが期待される。しかし、現実には観測データは必然的に雑音や数値計算誤差の影響を受けるため、そのような場合、データから構成される行列はフルランクとなり、上述の方程式が成立しない。このような状況下では、背後にある構造的性質を損なわないように行列を補正する必要がある。特に、もとの構造を保持しつつ、近傍の低ランク行列を求める操作は、構造化低ランク近似 (Structured Low-Rank Approximation: SLRA) と呼ばれる。以下では、SLRA の重要性を自己回帰 (Autoregressive: AR) モデルを例に説明する。

例 1 (AR モデルと Hankel 行列の SLRA). 時系列 $\mathbf{c} \in \mathbb{R}^{n_p}$ が式 (1.1) で表される次数 $d \ll n_p$ の AR モデルによって生成される状況を考える。

$$c_k = -\theta_1 c_{k-1} - \theta_2 c_{k-2} - \cdots - \theta_d c_{k-d} \quad (1.1)$$

$k = d + 1, \dots, n_P$ について上式を連立することで、次式を得る:

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_{d+1} \\ c_2 & c_3 & \cdots & c_{d+2} \\ \vdots & \vdots & & \vdots \\ c_{n_P-d} & c_{n_P-d+1} & \cdots & c_{n_P} \end{bmatrix} \begin{bmatrix} \theta_d \\ \vdots \\ \theta_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (1.2)$$

ここで上式左辺に現れる係数行列は、窓長 $d + 1$ の Hankel 行列と呼ばれ、 $\mathbf{H}_{d+1}(\mathbf{c})$ で表される。この記号により上式は $\mathbf{H}_{d+1}(\mathbf{c})\boldsymbol{\theta} = \mathbf{0}$ と整理され、 $\mathbf{H}_{d+1}(\mathbf{c})$ は少なくとも一次元の零空間を持つことが明らかとなり、モデルに従う時系列の行列はランクが欠損することになる。また、その零空間の基底としてモデル係数 $\boldsymbol{\theta}$ を推定可能である。

しかしながら、 $\mathbf{c}$ に誤差や雑音が含まれる場合、それはもはや式 (1.1) を満たさない。この時 $\text{rank } \mathbf{H}_{d+1}(\mathbf{c}) = d + 1$ 、つまり $\mathbf{H}_{d+1}(\mathbf{c})$ はフルランクである。所望の線形関係を満たさない $\mathbf{c}$ に対して、 $\mathbf{H}_{d+1}(\hat{\mathbf{c}})$ のランクが d となるような修正を加えたベクトル $\hat{\mathbf{c}}$ を求められたとする。そのようなベクトル $\hat{\mathbf{c}}$ は式 (1.1) を満たすため、 $\mathbf{H}_{d+1}(\hat{\mathbf{c}})$ の一次元零空間の基底ベクトルの係数がモデル係数の推定値として妥当となる。

$$\mathbf{H}_{d+1}(\mathbf{c}) : \text{fullrank} \xrightarrow{\text{SLRA}} \text{rank } \mathbf{H}_{d+1}(\hat{\mathbf{c}}) = d$$

そのような近似は、様々存在する。その中でも近似差分 $\mathbf{H}_{d+1}(\mathbf{c}) - \mathbf{H}_{d+1}(\hat{\mathbf{c}})$ の大きさを Frobenius ノルムで評価し、これが最小であるようなものを求めることに価値があるという立場をとることで、次のような Hankel 行列の低ランク近似問題が導かれる。

$$\min_{\hat{\mathbf{c}}} \|\mathbf{c} - \hat{\mathbf{c}}\|_F \quad \text{subject to} \quad \text{rank } \mathbf{H}_{d+1}(\hat{\mathbf{c}}) \leq d \quad (1.3)$$

このように、SLRA によって抽出された低ランクなデータは、そのデータが本来準拠すべきモデルのパラメータを推定することを可能にする。例えば、線形時不変システムのインパルス応答から構築される Hankel 行列のランク欠損は、システムの次数や最小状態表現の次数を決定するために用いられる [5, 6]。また、多項式の係数にノイズが含まれる場合、それらから構成される Sylvester 行列に対して SLRA を適用することで、近似多項式最大公約数 (AGCD) を計算する問題として定式化される [7, 8]。このほかにも、信号処理 [9, 10, 11]、時系列解析 [12, 13, 14] など多くの分野で基本的かつ重要な課題として現れる。このように、SLRA は、雑音を伴う観測からでも、問題の本質的な構造を保ちつつ、ロバストかつ効率的にパラメータ推定を行う能力を提供する。

1.1.1 SLRA 問題の定式化

本研究では、低ランク近似の対象を次の Affine 行列構造に限定する。以降、単に行列構造と述べた場合には、Affine 行列構造を指すものとする。

定義 1 (Affine 行列構造 [2]). 所与の行列 $\mathbf{S}_k \in \mathbb{R}^{K \times L}$ ($k = 1, \dots, n_p$) に対して以下のように定義される、構造化パラメータ空間 $\mathbb{R}^{n_p}$ から行列空間 $\mathbb{R}^{K \times L}$ への Affine 写像

$$\mathcal{S} : \mathbf{c} \in \mathbb{R}^{n_p} \mapsto \mathcal{S}(\mathbf{c}) := \mathbf{S}_0 + \sum_{i=1}^{n_p} c_i \mathbf{S}_i \quad (1.4)$$

を Affine 行列構造と定める。

本論文では、Affine 行列構造を持つ行列を Affine 構造化行列と呼称する。Hankel 行列構造が Affine 行列構造の一種であることは容易に確認可能である。例えば任意の 4×3 Hankel 行列は、次に示すように、行列 $\mathbf{S}_k \in \mathbb{R}^{4 \times 3}$ の線形結合として表現できるため、Affine 構造化行列の一種である。

例 2 (Hankel 行列).

$$\begin{bmatrix} c_1 & c_2 & c_3 \\ c_2 & c_3 & c_4 \\ c_3 & c_4 & c_5 \\ c_4 & c_5 & c_6 \end{bmatrix} = c_1 \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{S}_1} + c_2 \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{S}_2} + c_3 \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{S}_3} \\ + c_4 \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{\mathbf{S}_4} + c_5 \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{\mathbf{S}_5} + c_6 \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}_6}. \quad (1.5)$$

Affine 構造化低ランク近似問題は、次のように定義される。

問題 2 (Affine 構造化低ランク近似 [2]). $\{\mathbf{S}_k\}_{k=1, \dots, n_z}$ で指定される所与の Affine 行列構造, パラメータベクトル $\mathbf{c} \in \mathbb{R}^{n_z}$, ノルム $\|\cdot\|_{\mathbf{W}}$, そして整数 r であって ($0 < r < L$) を満たすものが所与であるとし, 次を求めよ:

$$\arg \min_{\hat{\mathbf{c}} \in \mathbb{R}^{n_z}} \left\| \hat{\mathbf{c}} - \mathbf{c} \right\|_{\mathbf{W}}^2 \quad \text{subject to } \text{rank } \mathcal{S}(\hat{\mathbf{c}}) \leq r. \quad (1.6)$$

ここで、式 (1.6) において目的関数が近似差分 $\mathcal{S}(\mathbf{c}) - \mathcal{S}(\hat{\mathbf{c}})$ の Frobenius ノルムではなく、パラメータベクトル間の重み付きノルム $\|\mathbf{c} - \hat{\mathbf{c}}\|_{\mathbf{W}}$ で定式化されていることには理由がある。実際、 $\mathbf{W}$ を適切に選ぶことで、このノルムは行列表現の差分に関して $\|\mathbf{c} - \hat{\mathbf{c}}\|_{\mathbf{W}} = \|\mathcal{S}(\mathbf{c}) - \mathcal{S}(\hat{\mathbf{c}})\|_{\text{F}}$ を満たすようにすることが可能である。この関係により、元の最小化問題は構造化行列の近似差分の Frobenius ノルムの最小化と同等でありながら、より簡潔で扱いやすい表現としてベクトルノルムを用いて定式化されている。以降、本論文では Affine 行列構造のみを対象とするため、単に構造化低ランク近似又は SLRA と記述する場合は、Affine 構造化低ランク近似を指すものとする。

1.1.2 関連研究

最もよく知られた行列の低ランク近似手法は、所与の行列の特異値分解 (Singular Value Decomposition) を行い、いくつかの小さな特異値を 0 で置き換えるもの (truncated SVD) であるが、一般にこの手法は、行列構造を保たないため、近似結果の行列に対応するモデルやシステムが存在しない点で不適である。

構造を保持した低ランク近似問題の難しさは、一部の特殊な場合を除き、所定のランクを持つ構造化行列の集合の要素を陽に記述できない点と、その集合が凸ではないことに由来する。そのため、最適化の制約条件を明示的に表現することが困難であり、最適解を直接求める方法が存在しないだけでなく、反復計算による方法であっても幅広く適用可能なものは知られていない。こうした背景から、SLRA に帰着させる問題の性質や要求に合わせて、多様なアルゴリズムが提案されてきた。これらの手法は次の二つのクラスに大別される：

1. 制約条件を緩和する手法；
2. 局所最適化手法.

制約条件を緩和する手法は、解析解を持つ問題に緩和するものである。特に行列の低ランク性に関する制約を核ノルム緩和する研究例が多数存在し、例えば行列補完 [15, 16], システム同定 [17, 18] への応用で有効性が示されている。

表 1.1: 構造化低ランク近似問題における低ランク性の定式化方法の比較:
Comparison of Formulations for Low-Rank Properties in SLRA Problems

	像空間表現	零空間表現
行列構造制約	×	○
ランク制約	○	△

他方局所的最適化手法は、問題の構造を保持したまま解を求めることができる点で緩和手法に比べてより良好な解が得られる可能性があるものの、初期値依存性が高い点が課題となる。SLRA 問題を局所最適化手法により解く手法として、像空間表現と零空間表現の二つの代表的な定式化が存在する。表 1.1 では、両者が行列構造の制約とランク制約を扱うのにどちらが適しているかを比較している。

像空間表現では、近似結果となる行列を二つの行列の積として表すことで、低ランク性を自然に課することができる [3]。しかしこの方法では、近似結果の行列が元の行列構造を保持する保証がないという問題がある。さらに、この表現において得られる行列の積は一意に定まらず、最適化変数の探索空間が広がるため、計算効率の面でも不利であることが指摘されている。

零空間表現では、式 (1.6) の制約条件を $\mathcal{S}(\hat{\mathbf{c}})\mathbf{x} = \mathbf{O}$ で置き換えた、

$$\min_{\hat{\mathbf{c}} \in \mathbb{R}^{n_z}} \left\| \hat{\mathbf{c}} - \mathbf{c} \right\|_{\mathbf{W}}^2 \quad \text{subject to} \quad \mathcal{S}(\hat{\mathbf{c}})\mathbf{x} = \mathbf{O} \quad (1.7)$$

という定式化を用いる。ただし、上式において $\mathbf{O}$ は、演算に適合するサイズの行列であって、そのすべての要素が 0 で埋め尽くされているもの（零行列）とする。上式の $\text{rank } \mathcal{S}(\hat{\mathbf{c}}) \leq r$ という制約は、 $\mathcal{S}(\hat{\mathbf{c}})\mathbf{x} = \mathbf{O}$ を満たす最大列階数を有する $\mathbf{x} \in \mathbb{R}^{L \times (L-r)}$ が存在することから従う定式化である。本論文では、このような零空間表現に着目する。

文献 [19] では、勾配降下型の SLRA 手法が提案されており、この手法の反復は二次収束する性質を持つ。また、このアルゴリズムが近似最大公約多項式（近似 GCD）計算、行列補完、雑音除去といった実用的な問題に対して有効に機能することが数値例により示されている。

近年では、変数射影 (Variable Projection: VP) 原理に基づいた SLRA 問題へのアプローチ (VP-SLRA) が盛んに研究されている [20, 21, 1, 22]。この手法では、式 (1.7) を 2 段階の最小化問題に分割する。第 1 段階は変数の数を削減する最小二乗問題であ

り、第2段階は削減後の変数に対する局所最適化問題を解くものである。変数射影法による Hankel 行列の構造化低ランク近似は、主に時系列解析に適用されている [20, 23]。加えて、近似 GCD 問題に対する応用も報告されている [21, 1]。

また、近似 GCD 問題を SLRA 問題に帰着することで求解するものには、構造化全最小ノルム (structured total least norm: STLN) 法に基づくもの (STLN-SLRA) の適用例が報告されている [8, 24, 25, 26, 27, 10]。

これらの手法には計算精度や計算時間の面でそれぞれ長所と短所があり、実用上は SLRA 問題の求解を通して解決すべき課題や、使用する計算機の性能、問題の規模などの条件を総合的に考慮した上で、最も適した SLRA 計算手法を選択することが求められる。

近似 GCD 計算を例にとると、文献 [28] では、複数の SLRA 計算手法を比較した結果、多くの場合において UVGCD [29] が有力な選択肢であることを結論づけている。UVGCD は、2 つの実係数多項式に対して近似 GCD を計算する手法であり、特定の条件下で高い計算精度と安定性を示す上に、計算コストも比較的抑えられるという特徴がある。また、VP-SLRA についても、近似精度や収束性の観点から同等の性能を示すことが文献 [1] で述べられている。この方法の大きな利点は、複数の複素係数多項式に対しても近似 GCD を計算できることであり、より広範な問題設定に適用可能である。

さらに、対象の多項式の次数が大きい場合には、計算効率が重要となる。特に STLN に基づく Sylvester 行列の SLRA は、高次の多項式に対する計算時間の点で、他の手法と比較して有力な選択肢であることが報告されている [28]。この手法は、多項式の次数が増大した場合でも効率的な計算が可能であり、従来手法に比べて実用的な利点があると言える。

また、Hankel 行列の SLRA に関しては、特に VP-SLRA [21, 5, 30] およびその派生手法 [20, 23] が近年盛んに研究され、それらの手法が計算精度や収束性の面で優れた性能を持つことが報告されている。

1.2 本論文の目的

本論文では、STLN-SLRA や VP-SLRA をはじめとする、低ランク性の零空間表現を用いる SLRA 計算手法に着目する。これらの手法については、有効性が多数報告されているものの、現時点では理論・実装の両面で十分に体系化されているとはいえない。特に、数理的な基盤が未整備な部分がいくつか存在し、手法の設計や選択において明確な指針を欠いているのが現状である。

本論文の目的は、こうした現状を踏まえ、低ランク性の零空間表現を用いる SLRA 手法において未解決のまま残されている以下の二点に焦点を当て、その理論的理解と改善を図ることである。

STLN-SLRA における初期方程式の選択方法

STLN-SLRA では、制約条件

$$\mathbf{S}(\mathbf{c})\mathbf{x} \approx \mathbf{0} \quad (1.8)$$

を、 $\mathbf{S}(\mathbf{c})$ の列ベクトルのうち 1 本を右辺に移動させ、次の線型方程式（初期方程式と呼称）に変換して解を求める：

$$\mathbf{A}(\mathbf{c})\mathbf{y} \approx \mathbf{b}(\mathbf{c}). \quad (1.9)$$

ここで、 $\mathbf{b}(\mathbf{c})$ は移動された列ベクトル、 $\mathbf{A}(\mathbf{c})$ はそれ以外の列ベクトルから構成される行列である。実際の計算においては、精度や収束速度の観点から、どの列ベクトルを右辺に移動させるかという選択が重要となる。しかし、この問題を考慮した既存の選択基準 [25] は計算量が大きく、その理論的妥当性も十分には検証されていない。本論文の第一の目的は、元の制約式 (1.8) と初期方程式 (1.9) との関係を理論的に解析し、従来法よりも計算負荷の少ない新たな選択基準を提案することである。この内容については、第 5 章で詳述する。第 5 章の内容は、著者らによる研究 [4] に基づく。

低ランク性を零空間表現する SLRA における初期値構成問題

VP-SLRA をはじめとする、低ランク性を零空間表現する SLRA では、所与の構造化行列 $\mathbf{S}(\mathbf{c})$ の最小特異値に対応する右特異ベクトルを $\mathbf{x}$ の初期値として用いる手法が

一般的である．しかし，最小特異値は行列の摂動に対して敏感であり，この初期値構成法は雑音に対して脆弱であるという問題がある．本論文の第二の目的は，雑音に対して頑健な新たな初期値構成法を構築することである．これについては，著者らによる研究 [31] 及び [32] に基づき，第 6 章で説明する．

これらを通じて，SLRA 手法の数理的基盤の整備と，より実用的なアルゴリズムの構築を目指す．

1.3 記号の定義

本論文では，ベクトルは太字の斜体を用いて表す．また，多項式はアルファベットの小文字を非斜体の太字で表す．行列を成分表示する際に，空白で表されている部分の要素は 0 で占められるものとする．表 1.2 にて頻出する記号の定義を与える．

表 1.2: 主要な記号 : Notation

v_i	ベクトル $\mathbf{v}$ の第 i 要素
$A_{i,j}$	行列 $\mathbf{A}$ の (i, j) 要素
$A_{i,:}$	行列 $\mathbf{A}$ の第 i 行ベクトル
$\mathbf{A}^\top$	行列 $\mathbf{A}$ の転置
$\mathbf{A}^\dagger$	行列 $\mathbf{A}$ の Moore-Penrose 一般逆行列 [33]
$\text{col}(\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(N)})$	縦ベクトル $\mathbf{c}^{(i)}$ を縦に連結したベクトル
$\text{diag}(a_1, \dots, a_N)$	スカラ $a_1, \dots, a_N$ を対角成分に持つ対角行列
$\text{blkdiag}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)})$	行列 $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}$ により構成されるブロック対角行列
$\mathbf{I}_K$	サイズ $K \times K$ の単位行列
$\mathbf{e}_i$	第 i 要素が 1 であるような単位ベクトル
$\mathbf{O}_{K \times L}$	$K \times L$ の零行列
$\text{rank}(\mathbf{A})$	行列 $\mathbf{A}$ の階数
$\mathbf{A} \succeq 0$	行列 $\mathbf{A}$ が非負定値行列であることを意味する

$\dim(V)$	(部分) 空間 V の次元
$\text{Im}(\mathbf{A})$	行列 $\mathbf{A}$ の像空間
$\text{Ker}(\mathbf{A})$	行列 $\mathbf{A}$ の零空間

$\deg p(t)$	多項式 $p(t)$ の最高次数
$\mathbb{F}[t]$	$\mathbb{F}$ を係数体とし, t を変数とする 1 変数多項式全体の集合

第2章 数理モデルと構造化行列

実際の問題に対して工学的アプローチを行う際、観測データに数理モデルを仮定することは少なくない。そのようなデータの要素間の関係が過決定の線形斉次方程式で表される場合がある。この時、その係数行列はモデルに対応する特定の行列構造を持つ。本章では、その代表例である Hankel 行列と Sylvester 行列に焦点を当て、それらと密接に関連する線形漸化関係および最大公約多項式に関する既知の性質を概説する。なお、第5章および第6章で行う数値実験は、本章で紹介する性質を参照する。

2.1 低ランク Hankel 行列と線形漸化関係

$\mathbf{c} \in \mathbb{R}^{n_p}$ を長さ n_p のベクトルとする。このベクトルから、窓長 L に基づいて構成される Hankel 行列 $\mathbf{H}_L(\mathbf{c}) \in \mathbb{R}^{K \times L}$ を次のように定義する：

$$\mathbf{H}_L(\mathbf{c}) := \begin{bmatrix} c_1 & c_2 & \cdots & c_L \\ c_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ c_K & \cdots & \cdots & c_{K+L-1} \end{bmatrix}. \quad (2.1)$$

ここで、 $K = n_p - L + 1$ とする。このようにして得られる行列は、ベクトル $\mathbf{c}$ の連続する部分系列を行として並べたものであり、特異スペクトル解析（Singular Spectrum Analysis: SSA）の文脈では、 L を「窓長」と呼ぶ [34]。なお、行数 K を明示したい場合には、式 (2.1) の行列を $\mathbf{H}_{K,L}(\mathbf{c})$ と表記することとする。

低ランクな Hankel 行列 $\mathbf{H}_L(\hat{\mathbf{c}})$ のパラメータベクトル $\hat{\mathbf{c}}$ ($\text{rank } \mathbf{H}_L(\hat{\mathbf{c}}) = r$) は、線形漸化式（Linear Recurrence Relation: LRR） [35] と密接に関連している。このような関係式を満たす任意の実ベクトルの各要素は、以下のように多項式と指数減衰する正弦波の積の和として表現できることが知られている：

$$\hat{c}_t = \sum_{j=1}^R p_j(t) \exp(d_j t) \sin(2\pi\omega_j t + \phi_j), \quad t = 1, \dots, n_p. \quad (2.2)$$

ここで、 $p_j(t)$ は実数値多項式であり、 $r = 2R$ である [23].

この定式化は、工学分野で頻繁に現れる 1. (指数的に減衰する) 正弦波 と 2. 純粋な多項式成分 という二つの基本的な信号を含んでいる. これら異なる振る舞いを統一的な枠組みで表現できることから、式 (2.2) は信号処理における一般的なモデリング構造として機能する. その結果、このような信号構造を活用する Hankel 行列は、以下のような問題に対する有効な手法として広く研究されている：

- スペクトル推定 [36, 37]
- スパース信号復元 [11, 38]
- 制御理論 [2, 17]
- モーメントからの形状復元 [39]
- 近似 Prony 法 [40]

さらに、磁気共鳴画像法 (MRI) [41, 13] や耐震性の解析 [42, 14] といった高度な応用においても、Hankel 行列の低ランク構造が活用されている. これらの応用では通常、多次元信号が関与するため、より複雑な依存関係を捉えるためにブロック型 Hankel 行列が用いられる. これらの応用分野において、Hankel 行列の SLRA (1.6) は、所与の信号 $\mathbf{c}$ を式 (2.2) で記述されるモデルへの当てはめあるいは近似として機能し、近似結果を用いて内部変数の推定などが行われる.

Hankel 行列の代数的性質は、Sylvester [43] による二項形式の分解の研究や、Prony [44] の研究にさかのぼる. 本節では、文献 [23] に基づき、Hankel 行列と LRR の関係性について概説する. より包括的な文献については、[45] 及び [35, Ch.5 and 8] を参照されたい.

2.1.1 Hankel 行列のランク特性と有限ランクベクトル

Hankel 行列の重要な性質の一つは、窓長に応じた階数の変化である。パラメータベクトルの長さ n_p を固定し、窓長 L を変化させた場合において、次の補題が知られている：

補題 3 ([35, Corollary 5.1]). 任意のベクトル $\mathbf{c} \in \mathbb{C}^{n_p}$ に対して、 $L' = \lfloor (n_p - 1)/2 \rfloor$ および $r = \text{rank } \mathbf{H}_{L'}(\mathbf{c})$ とすると、以下が成り立つ：

$$\text{rank } \mathbf{H}_L(\mathbf{c}) = \min(L, K, r).$$

この補題は、対応する Hankel 行列 $\mathbf{H}_L(\mathbf{c})$ が、広い範囲の窓長 L に対して一定のランクを保つようなベクトル $\mathbf{c}$ が存在することを保証する。ただし、行列が極端に縦長または横長になる場合、すなわち、 L が非常に小さいか n_p に非常に近い場合は除く。この結果は、そのようなベクトル $\mathbf{c}$ が、極端に短いあるいは長い窓長の場合を除いて、Hankel 行列のランクが不変となるような構造的性質を持つことを示している。Hankel 行列に関する上述の性質に基づき、ベクトル $\mathbf{c}$ のランクは、 $\text{rank } \mathbf{H}_{r+1}(\mathbf{c}) = r$ を満たす整数 r （ただし $r < n_p/2$ ）として定義される。なお、この定義は時系列解析の分野で用いられる「時系列のランク」に対応するものである [20]。

Hankel 行列の縦横比に関する曖昧さを避けるために、ここでは $L \leq n_p - L + 1$ ，すなわち $L \leq \frac{n_p+1}{2}$ の場合に限定して議論を行う。この仮定は、一般性を失うことなく定式化を簡素化する。

特に $r = \frac{n_p+1}{2}$ の場合、Hankel 行列はどの L に対してもフルランクであり、 $L = \frac{n_p+1}{2}$ では正則行列となる。一方、 $r < \frac{n_p+1}{2}$ の場合、行列は特定の構造を持つ。このような $\mathbf{c}$ を文献 [46] に倣い有限ランクであると定める。

2.1.2 LRR

低ランク Hankel 行列は、LRR と密接な関係を持つ。ベクトル $\mathbf{c} \in \mathbb{C}^{n_p}$ が次数 r の LRR を満たすとは、非零ベクトル $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{r+1}) \in \mathbb{C}^{r+1} \setminus \{0\}$ が存在し、次の関係が成り立つことを指す：

$$\theta_1 c_k + \dots + \theta_{r+1} c_{k+r} = 0, \quad \text{for all } k = 1, \dots, n_p - r. \quad (2.3)$$

LRR の特性は, Hankel 行列のランクとして現れる.

注 4. 式 (2.3) は次のように行列表現が可能である:

$$\mathcal{H}_{r+1}(\mathbf{c})\boldsymbol{\theta} = \mathbf{0}. \quad (2.4)$$

なお, ここで考える Hankel 行列 $\mathcal{H}_{r+1}(\mathbf{c})$ は縦長であることから, Hankel 行列がランク欠損している場合, 元のベクトルは LRR を満たすことが分かる. $\diamond$

式 (2.3) に関して $\theta_{r+1} \neq 0$ である時, $\bar{\theta}_k := -\theta_k/\theta_{r+1}$ と定めると, 次の形に変形できる:

$$c_j = \bar{\theta}_r c_{j-1} + \bar{\theta}_{r-1} c_{j-2} + \cdots + \bar{\theta}_1 c_{j-r}. \quad (2.5)$$

式 (2.5) は, ベクトル $\mathbf{c}$ が時系列である場合には, その過去の値から次の値を再帰的に計算する方法を提供する. この性質は, 時系列予測の基礎として利用される.

2.1.3 LRR と変調指数関数の和

本節では, 有限ランクベクトルと, 変調複素指数関数の和の関係について説明する. この関係は無限長の時系列ではよく知られている (詳しくは [47] を参照されたい). 本節及び後続する節では, 有限次元のベクトルに対する挙動について示す.

LRR の性質を理解する上で必要な道具立ては, LRR の **特性多項式** で, これは

$$\theta(z) := \theta_{r+1} z^r + \theta_r z^{r-1} + \cdots + \theta_2 z + \theta_1 \quad (2.6)$$

として定められる. ここに最初と最後の係数が零でないと仮定すると, 代数の基本的な知見により, 次のように特性多項式を因数分解可能である [48, 定理 4.17.1]:

$$\theta(z) = (z - \lambda_1)^{\nu_1} (z - \lambda_2)^{\nu_2} \cdots (z - \lambda_s)^{\nu_s}. \quad (2.7)$$

ただし, ν_i は根 λ_i の重合度であり, $\nu_1 + \cdots + \nu_s = r$ である. さらに, ベクトルの形式は, 次の補題が示すように, その根から決定される.

定理 5. 複素数値のベクトル $\mathbf{c} \in \mathbb{C}^{n_p}$ は、次のように表現できる場合にのみ、 $\theta_1 \neq 0$ および $\theta_s \neq 0$ の LRR を満たす。

$$c_k = \sum_{j=1}^s p^{(j)}(k) \lambda_j^k, \quad k = 1, 2, \dots, n_p \quad (2.8)$$

ここで、 $p^{(j)}(k)$ は、最大次数が $\nu_j - 1$ の多項式で、 $\lambda_j \in \mathbb{C} \setminus \{0\}$ は、(2.7) の多項式の根である。(2.8) の多項式 $p^{(j)}(t)$ の係数は、ベクトルの最初の s 個の値によって決定される。さらに、パラメータの組 $(p^{(j)}(t), \lambda_j)$ が j に対して相異なる場合、ベクトル $\mathbf{c}$ のランク r は

$$r = \sum_{j=1}^s (\deg p^{(j)} + 1) \quad (2.9)$$

で与えられる。

この補題の前半は [23, Theorem 3.1]，後半は [20, Appendix] より直ちに従う。

定理 5 により、低ランクな Hankel 行列のパラメータベクトルについて次の解釈が可能である。Hankel 行列 $\mathbf{H}_L(\hat{\mathbf{c}})$ が特異である場合、注 4 の事実により、ベクトル $\hat{\mathbf{c}}$ は次数 $L - 1$ の LRR を満たし、その形式は (LRR の係数の先頭 θ_1 または末尾 θ_L が 0 である場合を除き) 式 (2.8) によって与えられる。

特性多項式が重複しない根を持つ場合、すなわち (2.7) で $s = r$ および $\nu_j = 1$ である場合に対象を限定することで、次の式が導かれる。

$$c_k = \sum_{j=1}^r a_j \lambda_j^k, \quad k = 1, 2, \dots \quad (2.10)$$

これは、時系列解析の分野では、sum-of-exponential モデルとして知られている [11, 49]。

LRR の係数が実数に限定されている場合、その特性多項式は実多項式である。この時、代数学の基本的な知見により、非実の根は複素共役の対をなす。この事実により、定理 5 の系を得る。

系 6 ([23, Corollary 3.2]). 実数値のベクトル $\mathbf{c}$ が実係数ベクトル $\boldsymbol{\theta} \in \mathbb{R}^{r+1}$, $\theta_1, \theta_{r+1} \neq 0$ により LRR (2.3) を満たすのは, 次のような和として記述できる場合に限られる [50]:

$$c_k = \sum_{j=1}^q p^{(j)}(k) \exp(d_j k) \sin(2\pi\omega_j k + \phi_j), \quad k = 1, \dots, n_p \quad (2.11)$$

ここで $p^{(j)}(k)$ は実多項式であり, 信号処理の文脈においては, d_j は減衰率, ω_j は周波数, ϕ_j は位相に相当する.

項数 q は, $\theta(z)$ の実根の数と複素共役対の数の合計に等しい. 系 6 の詳細については, たとえば [46, 45, 12, 51] を参照されたい.

2.1.4 最小 LRR

定理 5 と系 6 はいくつかのケースのベクトルの形式を提示するが, 表現の一意性については情報を提供しない. 加えて, $\theta_1 = 0$ および $\theta_{r+1} = 0$ の場合を扱えない. 注 4 を踏まえ, 補題 3 の系として得られる次の主張は, 最小 LRR[52] が定義可能であり, さらにそれが一意に定まることを示す.

補題 7 ([23, Lemma 3.2]). 有限ランク r' のベクトル $\mathbf{c}$ が満たす LRR のうち, 最小の次数は r' である. その係数を $\boldsymbol{\theta}' = (\theta'_1, \dots, \theta'_{r'+1})^\top$ とすると, これは定数倍を除いて一意に定まり, 対応する Hankel 行列の右零空間の元として存在する:

$$\mathcal{H}_{r'+1}(\mathbf{c})\boldsymbol{\theta}' = \mathbf{0}.$$

実際, $\mathbf{c}$ に対してより小さな次数の LRR が存在すると仮定すると, 対応する Hankel 行列 $\mathcal{H}_{r'}(\mathbf{c})$ はランク欠損となり, $\mathbf{c}$ のランクが r' であるという事実と矛盾する. さらに, $\mathcal{H}_{r'+1}(\mathbf{c})$ の右零空間の次元が 2 以上であると仮定すると, 本章では Hankel 行列が縦長であることを仮定しているため, $\text{rank}(\mathcal{H}_{r'+1}(\mathbf{c})) \leq r' - 1$ となり, これも $\mathbf{c}$ のランクが r' であることに矛盾する.

さらに, ベクトル $\mathbf{c}$ は, 式 (2.10) に類似する, 一意に定まる最小表現をもつ. 以下では, 係数 θ'_1 や $\theta'_{r'+1}$ が 0 となることも許容する. $\lambda_1, \dots, \lambda_s \in \mathbb{C} \setminus \{0\}$, $\nu_1, \dots, \nu_s$ を

正の整数, $\nu_0, \nu_\infty \in \mathbb{N}$ を負でない整数とし,

$$\nu_0 + \nu_1 + \cdots + \nu_s + \nu_\infty = r' \quad (2.12)$$

とする. 要素が

$$c_k = \underbrace{\sum_{j=1}^{\nu_0} a_j \delta_{j-1}(k) + \sum_{l=1}^{\nu_\infty} b_l \delta_{n_p-l}(k)}_{\text{過渡項}} + \sum_{j=1}^s p^{(j)}(k) \lambda_j^k, \quad (2.13)$$

で表されるベクトル $\mathbf{c}$ を考える. ここで, 各 $p^{(j)}(k)$ は次数 $\nu_k - 1$ の複素係数多項式であり, $\delta_x(k)$ は Kronecker のデルタ

$$\delta_x(k) = \begin{cases} 1, & x = k, \\ 0, & x \neq k; \end{cases}$$

および $a_{\nu_0} \neq 0$, $b_{\nu_\infty} \neq 0$ である. 最後に, $(\theta'_0, \dots, \theta'_{r'})$ を最小の LRR の特性多項式の係数とする.

$$\theta'(z) = \theta'_0 + \theta'_1 z + \cdots + \theta'_{r'} z^{r'} = z^{\nu_0} (z - \lambda_1)^{\nu_1} \cdots (z - \lambda_s)^{\nu_s}; \quad (2.14)$$

ν_0 は根 0 の重複度を表し, ベクトル $\boldsymbol{\theta}'$ の要素列 $\theta'_{r'}, \theta'_{r'-1}, \dots, \theta'_0$ の先頭から連続して現れる 0 の個数を ν_∞ で表す. これは, 文献 [35] の解釈では ∞ 根の重複度に対応する. $\theta'(z)$ は, 補題 7 の最小 LRR $\boldsymbol{\theta}'$ の特性多項式と一致する.

定理 8 ([35, Theorem. 8.2]). ベクトル $\mathbf{c}$ が有限ランク r' であるのは, ベクトル $\mathbf{c}$ が (2.13) の形式で表される場合に限られる. さらに, r' が (2.12) を満たす表現 (以降, 正準表現) (2.13) は一意に定まる.

2.1.5 非最小 LRR と Hankel 行列の右零空間

この節では, 有限ランク r' のベクトル $\mathbf{c}$ が満たす次数 r の LRR を, $r > r'$ の場合に, $\mathbf{c}$ が満たす最小 LRR から構成可能であることを説明する.

命題 9 ([35, Theorem 5.1] の特殊例). $\mathbf{c}$ を有限ランク r' の複素ベクトルとし, $\boldsymbol{\theta}' \in \mathbb{C}^{r'+1}$ をその最小 LRR の係数とする. すると, 任意の $r \in \{r', \dots, n_p - r'\}$ に対して, $\mathbf{c}$ が満たす長さ r のすべての LRR の係数 $(\theta_1 \dots \theta_{r+1})$ は, 次の形式

$$\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{r+1} \end{pmatrix} = \underbrace{\begin{pmatrix} \theta'_1 & 0 & \dots & 0 \\ \theta'_2 & \theta'_1 & \ddots & \vdots \\ \vdots & \theta'_2 & \ddots & 0 \\ \theta'_{r'+1} & \vdots & \ddots & \theta'_1 \\ 0 & \theta'_{r'+1} & & \theta'_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \theta'_{r'+1} \end{pmatrix}}_{\mathcal{M}_{r-r'}(\boldsymbol{\theta}')} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{r-r'+1} \end{pmatrix} \quad (2.15)$$

により表される. ここで, b_k は複素係数である. ($\mathbf{c}$ が実ベクトルの場合, 係数 b_k も実数の範囲内で選択可能である).

注 4 により, このような LRR は, 行列 $\mathcal{H}_{r+1}(\mathbf{c})$ の右零空間内に属するすべてのベクトル (すなわち, $\mathcal{H}_{r+1}(\mathbf{c})\boldsymbol{\theta} = \mathbf{0}$ を満たすすべてのベクトル) に対応することが分かる. したがって, $r' \leq r \leq n_p - r'$ の範囲においては, 式 (2.15) に現れる行列 $\mathcal{M}_{r-r'}(\boldsymbol{\theta}')$ の列空間は, 行列 $\mathcal{H}_{r+1}(\mathbf{c})$ の右零空間と一致する. この右零空間は, $\mathbb{C}^{r+1}$ における次元 $r - r' + 1$ の線形部分空間である.

2.1.6 Hankel 行列の SLRA の解釈

大まかに言えば, Hankel 行列の SLRA

$$\begin{aligned} \hat{\mathbf{c}}_* &:= \arg \min_{\hat{\mathbf{c}}} \|\hat{\mathbf{c}} - \mathbf{c}\|_{\mathbf{w}} \\ \text{subject to } \text{rank } \mathcal{H}_L(\hat{\mathbf{c}}) &\leq L - 1 \end{aligned} \quad (2.16)$$

は, 所与の信号 $\mathbf{c}$ を式 (2.8)

$$c_k = \sum_{j=1}^s p^{(j)}(k) \lambda_j^k, \quad k = 1, 2, \dots, n_p \quad (2.8)$$

で表現される形式の信号で近似する, あるいはそのような形式に当てはめる問題として解釈可能であることを, 定理 2.8 は示す. これにより, Hankel 行列に対する SLRA は次のような応用例が存在する.

1. 雑音除去 / パラメータ推定

2. 欠損値推定 / 信号値予測

なお、このような信号のモデル化は一見複雑に見えるかもしれないが、例えば多項式 $p^{(j)}(k)$ を定数とすると、これは信号処理の分野において頻出する、指数的に減衰する正弦波（あるいは単なる正弦波）の和である。

雑音除去 / パラメータ推定 ここでは、式 (2.8) を満たす信号 $\bar{c} \in \mathbb{C}^{n_p}$ に対して、分散共分散行列が $\mathbf{C}$ 、平均が $\mathbf{0}$ の Gauss 性雑音 $\boldsymbol{\xi} \in \mathbb{C}^{n_p}$ が重畳された信号 $\mathbf{c} = \bar{c} + \boldsymbol{\xi}$ が観測される状況を想定する。このような状況において、未知である真の信号 $\bar{c}$ のランク r' が既知であれば、 $\mathbf{W} = \mathbf{C}^{-1}$ とした場合の重み付き最小二乗問題 (2.16) は真の信号 $\bar{c}$ に対して最尤推定となることが知られている [20]。

さらに、真の信号が式 (2.10) の形式で表されるとき、そのパラメータ $\bar{\lambda}_j$ は、最尤推定値 $\hat{c}_*$ を同じく式 (2.10) の形式で表した際のパラメータ λ_{*j} として推定される。 λ_{*j} は、 $\hat{c}_*$ が満たす最小 LRR の係数 $\boldsymbol{\theta}'_*$ の特性多項式の根より計算される。この方法は、Prony の方法 [44, 40] として広く知られているものである。 $\boldsymbol{\theta}'_*$ は、 $\mathcal{H}_{r'+1}(\hat{c}_*)$ の唯一の零特異値に対応する右特異ベクトルとして計算可能である。

欠損値推定 / 信号値予測 信号 $\bar{c} \in \mathbb{C}^{n_p}$ が式 (2.8) の形式で表されることを仮定し、観測に際して、そのうち l 要素が欠損してしまう状況を想定する。欠落している要素の番号を $i_1, \dots, i_l$ とし、欠落した要素は、表記の便宜上、0 として扱うものとする。つまり、

$$c_i = \begin{cases} 0 & \text{if } i \in \{i_1, \dots, i_l\} \\ \bar{c}_i & \text{otherwise} \end{cases} \quad (2.17)$$

が観測であるとする。この時、Hankel 行列の SLRA(2.16) において、重み行列 $\mathbf{W}$ を要素が次で指定される対角行列とすることで、欠損値推定が行われる [23]。

$$W_{ii} = \begin{cases} 1 & \text{if } i \in \{i_1, \dots, i_l\} \\ \infty & \text{otherwise} \end{cases} \quad (2.18)$$

信号予測は、式 (2.8) でモデル化される所与の信号を $\mathbf{c}^0 \in \mathbb{C}^{n_{\text{given}}}$ とし、その先に続く予測したい信号の要素数を n_{forecast} 、 $n_p = n_{\text{given}} + n_{\text{forecast}}$ 、欠損値要素数 $l = n_{\text{forecast}}$ 、

その要素番号を $i_1 = n_{\text{given}} + 1, \dots, i_l = n_{\text{given}} + l = n_p$ で指定した際の欠損値推定問題と見なすことができる [30, 23].

2.1.7 Hankel 構造より派生する行列構造

多変量時系列など複数の時系列を扱う場合は、式 (2.3) の関係が各時系列ごとに成立することを想定することがある。特に、同じ長さの M 個の同一ランクの時系列 $\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)}$ があるとき、各時系列に対応する Hankel 行列を次のように並べた構造 (ストライプ型 Hankel 構造) が意味をなす場合がある。

$$\begin{bmatrix} \mathcal{H}_L(\mathbf{c}^{(1)}) & \dots & \mathcal{H}_L(\mathbf{c}^{(M)}) \end{bmatrix}. \quad (2.19)$$

ストライプ型 Hankel 行列に対する SLRA は、多変量 SSA[53] に用いられる。また、異なる時系列が動的システムの入力および出力である場合、この構造を用いることで、ベヘイビアアプローチに基づく線形システムの理論を展開することができる [54, 17, 55, 56].

一方、異なる長さの場合、(2.19) は、モザイク型 Hankel 行列のランク欠損を考えることになる。[57, 58]. モザイク型 Hankel 行列は近似最大公約多項式計算にも応用があり、第 2.2 節でも取り扱うため、次に定義を掲載する。

定義 10 (モザイク型 Hankel 行列 [21]). $i = 1, \dots, M, j = 1, \dots, T$ に対して $\mathbf{c}^{(i,j)} \in \mathbb{R}^{k_i+l_j-1}$ が定められているとき、モザイク型 Hankel 行列 $\mathcal{H}_{k,l}(\mathbf{c}^{(k,l)}) \in \mathbb{R}^{K \times L}$ を次式で定める:

$$\mathcal{H}_{k,l}(\mathbf{c}^{(k,l)}) := \begin{bmatrix} \mathcal{H}_{k_1,l_1}(\mathbf{c}^{(1,1)}) & \dots & \mathcal{H}_{k_1,l_T}(\mathbf{c}^{(1,T)}) \\ \vdots & \ddots & \vdots \\ \mathcal{H}_{k_M,l_1}(\mathbf{c}^{(M,1)}) & \dots & \mathcal{H}_{k_M,l_T}(\mathbf{c}^{(M,T)}) \end{bmatrix}. \quad (2.20)$$

ただし、ここで $\mathbf{c}^{(k,l)}$ は $\mathbf{c}^{(i,j)}$ を結合したベクトル

$$\mathbf{c}^{(k,l)} := \text{col}(\mathbf{c}^{(1,1)}, \dots, \mathbf{c}^{(1,T)}, \dots, \mathbf{c}^{(M,T)})$$

を表し、 $n_p = \sum_{i=1}^M \sum_{j=1}^T (k_i + l_j - 1)$ である。

以降、 $\mathbf{c}^{(k,l)}$ が $\mathcal{H}_{k,l}(\mathbf{c}^{(k,l)})$ のようにモザイク型 Hankel 行列の構造化パラメータとなっている場合に単に $\mathbf{c}$ と略記し ($\mathcal{H}_{k,l}(\mathbf{c}^{(k,l)}) \mapsto \mathcal{H}_{k,l}(\mathbf{c})$)、また同一数式内に現れる $\mathbf{c}^{(k,l)}$ も同様に $\mathbf{c}$ と略す。

特殊なモザイク型 Hankel 行列に、次の形式を持つブロック型 Hankel 行列がある：

$$\begin{pmatrix} \mathcal{H}_L(\mathbf{c}^{(1)}) & \mathcal{H}_L(\mathbf{c}^{(2)}) & \mathcal{H}_L(\mathbf{c}^{(3)}) & \cdots & \mathcal{H}_L(\mathbf{c}^{(T)}) \\ \mathcal{H}_L(\mathbf{c}^{(2)}) & \mathcal{H}_L(\mathbf{c}^{(3)}) & \mathcal{H}_L(\mathbf{c}^{(4)}) & \cdots & \mathcal{H}_L(\mathbf{c}^{(T+1)}) \\ \vdots & \vdots & & \ddots & \\ \mathcal{H}_L(\mathbf{c}^{(M)}) & \mathcal{H}_L(\mathbf{c}^{(M+1)}) & \mathcal{H}_L(\mathbf{c}^{(M+2)}) & \cdots & \mathcal{H}_L(\mathbf{c}^{(T+M+1)}) \end{pmatrix}$$

ブロック型 Hankel 行列の低ランク近似は、線形システム理論において広く使用されており、線形システム実現 (system realization) やモデル縮約 (model reduction) の問題と関連がある [59, 60, 17]. また、ブロック型 Hankel 行列は短時間 Fourier 変換 (STFT: Short-Time Fourier Transform) とも関係がある [61]. 有限サイズのブロック型 Hankel 行列に関する代数的理論も存在するが、この場合には複数の線形漸化式を考慮する必要があるため、陽的なパラメータ化がより複雑になる [62].

2.2 近似最大公約多項式計算

最大公約多項式 (greatest common divisor: GCD) を求めることは、代数的計算における基本的な問題である. この問題は通常、実数または複素数係数をもつ 2 つの多項式が与えられたとき、それらの最大公約数の係数を求める問題として定式化される.

この問題の応用範囲は非常に広く、以下にいくつかの例を挙げる.

多項式の根の計算 重解をもつ多項式 $p(t)$ の根を求めることは悪条件問題である. しかし、ロバストな GCD 計算手法が利用可能であれば、 $p(t)$ そのものとその導関数の GCD $g(t)$ を計算することにより、この問題の解決に役立つ可能性がある. これは、 $p(t)/g(t)$ の根のほうが、数値的に良好な条件をもつことが多いためである [63].

有理関数の簡約化 有理関数 $g(t) = a(t)/b(t)$ に対して、計算や表現を行う際に、 $a(t)$ と $b(t)$ が互いに素である必要がある場合がある. このとき、 $w(t) = \gcd(a(t), b(t))$ を計算すれば、 $w(t)$ を $\bar{g}(t) = \bar{a}(t)/\bar{b}(t)$ に置き換えることができる. ここで、 $\bar{a}(t) = a(t)/w(t)$ 、 $\bar{b}(t) = b(t)/w(t)$ である. この考え方は、Bézier 曲線などの有理曲線の次数削減などに応用される. この応用例については、例えば、文献 [64] を参照されたい.

制御理論 多項式の互いに素である性質は，線形制御システムの可制御性と関係がある [65, 6]. また，システム同定 [66, 67, 68] の分野にも応用可能である．

画像復元 多項式の GCD 計算は，ブラインド画像ボケ除去 (blind image deblurring) にも利用される [10].

本節では主に，一変数多項式の GCD を求める問題を扱う．この問題は，厳密な状況（すなわち，多項式の係数が誤差を含まないと仮定した場合）については，Euclid の互助法をはじめとした古くからの研究により深く理解されている．しかしながら，実際の多くの応用では，入力データは浮動小数点数として与えられたり，物理実験や以前の計算の結果に由来したりするため，何らかの誤差を含むのが一般的である．このような多項式に対して通常が多項式計算を適用する場合，さらなる工夫が必要となる．しかし，多項式の組が次数 1 以上の GCD を持つ場合でも，係数にわずかな摂動が加わることで互いに素となることがある．このため，数値表現の精度限界や観測誤差が存在するこれらの応用分野では，多項式の GCD を近似的に計算する手法が必要となる．

本節では，多項式についての基本的な性質を説明し，その後に近似 GCD の概念を導入する．

2.2.1 数学的準備

多項式ベクトル空間

整数 n に対して，次数が高々 n の多項式全体の集合はベクトル空間を形成する．このベクトル空間を

$$\mathcal{P}_n = \{p \in \mathbb{C}[t] \mid \deg(p) \leq n\}$$

と表す．この時， $\mathcal{P}_n$ は次に定める $\Phi: \mathcal{P}_n \rightarrow \mathbb{R}^{n+1}$ によって $\mathbb{R}^{n+1}$ と全単射の関係にある．

$$\mathbf{p} := \Phi(p_1 + p_2 t + \cdots + p_{n+1} t^n). \quad (2.21)$$

$p(t) \in \mathcal{P}_n$ に対してノルムを $\|p(t)\| := \|\Phi(p(t))\| = \|\mathbf{p}\|$ と定める．

したがって， $\mathcal{P}_n$ の次元は次のように与えられる．

$$\dim(\mathcal{P}_n) = \begin{cases} 0 & \text{for } n < 0 \\ n + 1 & \text{for } n \geq 0 \end{cases}$$

本論文では、 $\mathcal{P}_n$ の基底として単項式基底 $\{1, t, t^2, \dots, t^n\}$ を用いる．この基底のもとで、任意の多項式 $p(t)$ は $p(t) = \sum_{i=0}^n p_{i+1} t^i$ の形で表され、その係数ベクトルは次のように表現される．

$$\mathbf{p} = (p_0 \ \cdots \ p_n)^\top \in \mathbb{C}^{n+1}$$

ここで、 $p_n = 0$ となることもありうる．また、同じ多項式 $p(t)$ は、より高次元のベクトル空間 $\mathcal{P}_m \supset \mathcal{P}_n$ に埋め込むことが可能であり、その際にはより高次元の係数ベクトルを持つ．

二つの多項式 $p^{(1)}(t) \in \mathcal{P}_{n_1}$ と $p^{(2)}(t) \in \mathcal{P}_{n_2}$ の乗算は $\mathbb{R}^{n_1+1}$ の元と $\mathbb{R}^{n_2+1}$ の元の畳み込みと等価であり、次のように表される：

$$\Phi(p^{(1)}(t) \cdot p^{(2)}(t)) = \mathbf{M}_{n_2+1}(\mathbf{p}^{(1)})\mathbf{p}^{(2)} = \mathbf{M}_{n_1+1}(\mathbf{p}^{(2)})\mathbf{p}^{(1)}. \quad (2.22)$$

ここで、 $\mathbf{M}_m(\mathbf{h})$ は次式で定める $\mathbf{h} \in \mathbb{R}^{d+1}$ の畳み込み行列である．

$$\mathbf{M}_m(\mathbf{h}) := \begin{bmatrix} h_1 & & & \\ & \ddots & & \\ & & h_1 & \\ & & & \ddots \\ & & & & h_{d+1} \end{bmatrix} \in \mathbb{R}^{(m+d) \times (m)}. \quad (2.23)$$

$p(t) \in \mathcal{P}_n$ と $h(t) \in \mathcal{P}_d$, $(0 \leq d \leq n)$ について、 $p(t) = q(t) \cdot h(t)$ を満たす $q(t) \in \mathcal{P}_{n-d}$ が存在する時、 $h(t)$ は $p(t)$ の除数であるという．この定義のもとでは、全ての $h(t) \in \mathcal{P}_d \setminus \{0\}$, $(0 \leq d \leq n)$ は $0 \in \mathcal{P}_n$ の除数であり、また全ての $h(t) \in \mathcal{P}_0 \setminus \{0\}$ は任意の多項式の除数である．

本論文では、2つの多項式だけでなく、3つ以上の多項式の組に対する GCD の計算も議論の対象とする．そこで、表記の煩雑化を避けるため、 N 個の多項式の組の略記法を定める [1]． $\mathbf{n} = [n_1, n_2, \dots, n_N]$ により、多項式たちの次数を並べたベクトルを表す． $p^{(i)}(t) \in \mathcal{P}_{n_i}$ ($i = 1, \dots, N$) に対し、 $p^{(\mathbf{n})}(t) := (p^{(1)}(t), \dots, p^{(N)}(t)) \in \mathcal{P}_{\mathbf{n}}$ と表すこととする．ここで $\mathcal{P}_{\mathbf{n}} := \mathcal{P}_{n_1} \times \cdots \times \mathcal{P}_{n_N}$ と定めた．また、 $p^{(\mathbf{n})}(t)$ の係数を縦ベクトルとして $\mathbf{p}^{(\mathbf{n})} := \text{col}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(N)}) \in \mathbb{R}^{n_p}$ と表すこととする．ただし $n_p := \sum_{i=1}^N (n_i + 1)$ とした． N 個の多項式の組 $g^{(\mathbf{n})}(t) \in \mathcal{P}_{\mathbf{n}-d} := \mathcal{P}_{n_1-d} \times \cdots \times \mathcal{P}_{n_N-d}$ と多項式 $h(t) \in \mathcal{P}_d$ との積を次のように定義する：

$$(g^{(\mathbf{n})} \cdot h)(t) := ((g^{(1)} \cdot h)(t), \dots, (g^{(N)} \cdot h)(t)).$$

定義 11 (Greatest Common Divisor of polynomials [1]). 多項式の組 $p^{(n)}(t) \in \mathcal{P}_n$ に対して $h(t) \in \mathcal{P}_d \setminus \{0\}$ が $p^{(1)}(t), \dots, p^{(N)}(t)$ 全ての除数であるとき, $h(t)$ はそれらの公約多項式であるという. 特に, $d' > d$ なる $\mathcal{P}_{d'}$ に公約多項式が存在しない場合, $h(t)$ は最大公約多項式 (Greatest Common Divisor of polynomials: GCD) であるという.

$1 \in \mathcal{P}_0$ は任意の多項式の組の公約多項式であるため, 任意の多項式の組に対し GCD は存在する. ここでは計算上の理由から, 多項式に対してモニック (monic, 最高次係数が 1) の条件を課さない. この設定では, GCD は一意ではなく, 同じ多項式対の GCD は非零定数倍だけ異なる. このことから, 我々は二つの多項式 $p(t), q(t)$ の間に次の同値関係を定義する.

$$p(t) \sim q(t) \iff p(t) = \alpha q(t), \quad \alpha \neq 0$$

このとき, ある多項式対 $p^{(n)}(t)$ の GCD 全体の集合は, この同値関係 $\sim$ のもとでの同値類を形成する. これを

$$\gcd(p^{(n)}(t))$$

と表す. なお, これは剰余環 $\mathbb{C}[t]/\sim$ のもとで一意的に定まる.

特定の次数 k を持つ GCD を持つ多項式対の集合を次のように表記する.

$$\mathcal{P}_n^k = \{p^{(n)}(t) \in \mathcal{P}_n \mid \deg(\gcd(p^{(n)})) = k\}$$

2.2.2 フル・Sylvester 行列

$p^{(n)}(t) \in \mathcal{P}_n^d$ について, $h(t) = \gcd(p^{(n)}(t))$ とすると, 全ての $i = 1, 2, \dots, N$ に対して

$$p^{(i)}(t) = h(t) \cdot g^{(i)}(t) \tag{2.24}$$

を満たす商多項式 $g^{(i)}(t) \in \mathcal{P}_{n_i-d}$ が存在する. この時,

$$h(t) \cdot g^{(i)}(t) \cdot g^{(j)}(t) = p^{(i)}(t) \cdot g^{(j)}(t) = p^{(j)}(t) \cdot g^{(i)}(t)$$

が任意の i, j , $(1 \leq i < j \leq N)$ に対して成立するため, 連立方程式

$$\mathcal{M}_{n_j-d+1}(p^{(i)})g^{(j)} - \mathcal{M}_{n_i-d+1}(p^{(j)})g^{(i)} = 0 \tag{2.25}$$

を得る．一般化 Sylvester 行列を式 (2.26) で定める．

$$\text{Syl}_d(\mathbf{p}^{(n)}) := \begin{pmatrix} \mathcal{M}_{f_1}(\mathbf{p}^{(2)}) & -\mathcal{M}_{f_2}(\mathbf{p}^{(1)}) & & \\ \vdots & & \ddots & \\ \mathcal{M}_{f_1}(\mathbf{p}^{(N)}) & & & -\mathcal{M}_{f_N}(\mathbf{p}^{(1)}) \end{pmatrix} \in \mathbb{R}^{E \times F}. \quad (2.26)$$

ここで $f_i := n_i - d + 1$ ，および

$$F := \sum_{i=1}^N f_i, \quad E := (N-1)f_1 + \sum_{i=2}^N n_i \quad (2.27)$$

である．一般化 Sylvester 行列に対して

$$\text{Syl}_d(\mathbf{p}^{(n)}) \begin{pmatrix} \mathbf{g}^{(1)} \\ \vdots \\ \mathbf{g}^{(N)} \end{pmatrix} = \mathbf{0} \quad (2.28)$$

が成立することは容易に確認できる．この事実より， $\mathbf{p}^{(1)}$ が零多項式でない場合に¹，式 (2.28) を利用して商多項式の係数 $\mathbf{g}^{(n)}$ を計算可能である．ところが，一般化 Sylvester 行列をさらに拡張することで， $\mathbf{p}^{(1)}$ に対する条件を取り除くことができる． N 個の多項式の組に対して次の形で再帰的に定義される，フル・Sylvester 行列

$$\text{fSyl}_d^N(\mathbf{p}^{(n)}) := \begin{cases} \begin{bmatrix} \text{Syl}_d(\mathbf{p}^{(n)}) \\ \mathbf{0} \text{ fSyl}_d^{N-1}(\mathbf{p}'^{(n')}) \end{bmatrix} & \text{if } N > 2 \\ \text{Syl}_d(\mathbf{p}^{(n)}) & \text{if } N = 2 \end{cases} \in \mathbb{R}^{E^{\text{full}} \times F}, \quad (2.29)$$

を導入することで，式 (2.25) は次の形で表される [1]:

$$\text{fSyl}_d(\mathbf{p}^{(n)}) \begin{bmatrix} \mathbf{g}^{(1)} \\ \vdots \\ \mathbf{g}^{(N)} \end{bmatrix} = \mathbf{0}. \quad (2.30)$$

ただし， $\mathbf{n}' := [n_2, \dots, n_N] \in \mathbb{N}^{N-1}$ ， $\mathbf{p}'^{(n')} := \text{col}(\mathbf{p}^{(2)}, \dots, \mathbf{p}^{(N)})$ ， $E^{\text{full}} := (N-1)(n_p - \frac{1}{2}N(d+1))$ である．例えば $N = 3$ の場合，

$$\text{fSyl}_d^3(\mathbf{p}^{(n)}) = \begin{pmatrix} \mathcal{M}_{f_1}(\mathbf{p}^{(2)}) & -\mathcal{M}_{f_2}(\mathbf{p}^{(1)}) & & \\ \mathcal{M}_{f_1}(\mathbf{p}^{(3)}) & & -\mathcal{M}_{f_3}(\mathbf{p}^{(1)}) & \\ & \mathcal{M}_{f_2}(\mathbf{p}^{(3)}) & -\mathcal{M}_{f_3}(\mathbf{p}^{(2)}) & \end{pmatrix}. \quad (2.31)$$

である．

¹このような場合には一般化 Sylvester 行列は自明にランク欠損となってしまう

フル・Sylvester 行列は 2 つ以上の多項式の組に対して定義される ($N \geq 2$)。また、通常自明な公約多項式は考察の対象にならないため $d > 0$ が暗黙の条件となる。これら 2 つの条件下で、 $E^{\text{full}} > F$ であり、フル・Sylvester 行列は縦長であると仮定できる。

次の補題は、フル・Sylvester 行列を用いて、 $p^{(n)}(t)$ を $\gcd(p^{(n)}(t))$ で除した際の商多項式 $g^{(n)}(t)$ を式 (2.30) の形式で計算可能であることを示している。

補題 12 ([1, Lemma 4.6]). $\deg \gcd(p^{(n)}(t)) = d$ かつ $\text{fSyl}_d(\mathbf{p}^{(n)})\mathbf{g}^{(n)} = \mathbf{0}$ (ただし $\mathbf{g}^{(n)} \neq \mathbf{0}$) である時、 $p^{(k)}(t) = g^{(k)}(t) \cdot h(t)$, $h(t) = \gcd(p^{(n)}(t))$ である。

この補題を証明するために、先にいくつかの補題を準備する。

補題 13. N 個の非零の多項式の組 $p^{(n)}(t) \in \mathcal{P}_n$ の GCD の次数が d であるとする。この時

$$\text{rank Syl}_{d'}(\mathbf{p}^{(n)}) < L, \text{ for } d' \leq d \quad (2.32)$$

$$\text{rank Syl}_{d'}(\mathbf{p}^{(n)}) = L, \text{ for } d' > d, \quad (2.33)$$

特に $\text{Syl}_d(\mathbf{p}^{(n)})$ は 1 次元の零空間を持つ。ここで L は $\text{Syl}_{d'}(\mathbf{p}^{(n)})$ の列数である。

証明. [63, Lemma 2.1] 及び [1, Lemma 4.6] から容易に証明可能である。 □

注 14. 2 つの多項式の組を考えると、すなわち $N = 2$ である時、特に

$$\text{Syl}_d(\mathbf{p}) := [\mathcal{M}_{n_2-d+1}(p^{(1)}) \quad \mathcal{M}_{n_1-d+1}(-p^{(2)})] \quad (2.34)$$

は単に Sylvester 行列と呼ばれる。 ◇

2.2.3 数値的な最大公約多項式

現実の問題では、多項式の組は係数の数値誤差や観測雑音の影響により互いに素となり、 $\text{fSyl}_d(\mathbf{p}^{(n)})$ がフルランクとなる場合が少なくない。このような場合には、所与の多項式に対して何らかの修正差分²を加えることで得る「近似多項式」たちに対して厳密に GCD が存在するように調整するという戦略を立てることができる。そのような数値的な最大公約多項式の定式化は、大きく 2 つに分類 [28] される。1 つ目は ϵ -GCD[69, 70, 71, 29] と呼ばれるもので、修正差分の大きさがある基準以下に収まる範

²本論文では、所与の多項式と、それに対する近似多項式の差のことを修正差分と呼称する。

囲の中で、近似多項式の GCD の次数を最大化するものである。もう 1 つは、次数に関する制約付きの近似 GCD (approximate GCD with fixed degree, 以降単に近似 GCD と呼称) [69, 70, 7] で、近似多項式の GCD の次数が d 以上になる条件の下で、その修正差分の大きさを最小化するものである。これら 2 つの近似 GCD の定式化は双対の関係にあり、片方の解がもう一方の部分問題の解として機能する。後者の近似は、GCD の次数が事前に判明している場合に有用である。

ϵ -GCD [69]

2 つの多項式の組に対する ϵ -GCD の定義を次に与える。

定義 15 (ϵ -GCD). $p^{(1)}(t) \in \mathcal{P}_{n_1}$ と $p^{(2)}(t) \in \mathcal{P}_{n_2}$ 多項式, 及び正の実数 ϵ を所与とする. 修正を加えた多項式 $\hat{p}^{(1)}(t) \in \mathcal{P}_{n_1}$ と $\hat{p}^{(2)}(t) \in \mathcal{P}_{n_2}$ に対して多項式 $h(t)$ が次の式を満たすとき, $p^{(1)}(t)$ と $p^{(2)}(t)$ の ϵ -divisor と定める.

$$\begin{aligned} \|\hat{p}^{(1)}(t) - p^{(1)}(t)\| &\leq \epsilon, \\ \|\hat{p}^{(2)}(t) - p^{(2)}(t)\| &\leq \epsilon \\ h(t) &= \gcd(\hat{p}^{(1)}(t), \hat{p}^{(2)}(t)) \end{aligned} \tag{2.35}$$

ϵ -GCD は, 次で定められる.

$$\arg \max_{h(t) \in \mathcal{P}_n} \{ \deg h(t) : h(t) \text{ は } p^{(1)}(t) \text{ と } p^{(2)}(t) \text{ の } \epsilon\text{-divisor} \} \tag{2.36}$$

近似 GCD

二つ目の戦略は, $\deg \gcd(\hat{p}^{(n)}(t)) \geq d$ を満たすように修正が加えられた多項式 $\hat{p}^{(n)}(t)$ のうち, 修正差分の大きさが最小となるものを求めるものである。これは次の最小化問題に帰着する。

問題 16 (近似 GCD[1]). $p^{(n)}(t) \in \mathcal{P}_n$ 及び整数 $d : 1 \leq d \leq \min(n_1, \dots, n_N)$ に対して, 次に定めるノルム

$$\|p^{(n)}(t) - q^{(n)}(t)\| := \sqrt{\sum_{i=1}^N \|p^{(i)} - q^{(i)}\|^2} \quad (2.37)$$

に関する以下の最小化問題を, 近似 GCD 問題と呼ぶ:

$$\begin{aligned} & \underset{\hat{p}^{(n)}(t) \in \mathcal{P}_n}{\text{minimize}} \quad \|p^{(n)}(t) - \hat{p}^{(n)}(t)\|^2 \\ & \text{subject to} \quad \deg \gcd(\hat{p}^{(n)}(t)) \geq d. \end{aligned} \quad (2.38)$$

注 17. 次に定義する「真の GCD の推定問題」において, 摂動項 $\eta^{(n)}$ が平均 0 の白色 Gauss 性雑音であると仮定すると, 式 (2.38) により定義される近似 GCD (近似 GCD) 問題は, 真の GCD に対する最尤推定に一致することが知られている [67, 68]. $\diamond$

問題 18 (真の GCD の推定). 未知の「真の」多項式 $\bar{g}^{(n)}(x) \in \mathcal{P}_{n-d}$ 及び $\bar{h}(x) \in \mathcal{P}_d$ により構成される多項式の組

$$\bar{p}^{(n)}(x) = (\bar{g}^{(n)} \cdot \bar{h})(x) \in \mathcal{P}_n$$

を考える. ここで, $\deg \text{GCD}(\bar{g}^{(n)}(x)) = 0$ であるとする. また, 観測された多項式の組 $p^{(n)}(x)$ は, 雑音項 $\eta^{(n)}(x)$ を通じて

$$p^{(n)}(x) = \bar{p}^{(n)}(x) + \eta^{(n)}(x)$$

と表されるものとする. このような状況下で, 観測 $p^{(n)}(x)$ から未知の「真の GCD」 $\bar{h}(x)$ を推定する.

2.2.4 Sylvester 行列の零空間の性質を利用した近似 GCD 計算

近似 GCD 問題 (2.38) に対する解法の一つに, フル・Sylvester 行列の性質を応用するものがある [8, 25, 10, 7, 63]. この手法では, 問題 (2.38) 内の GCD に関する次数の制約を定理 12 を援用して書き換えた次の最適化問題

$$\begin{aligned} & \underset{\hat{p}^{(n)}(t) \in \mathcal{P}_n}{\text{minimize}} \quad \|p^{(n)} - \hat{p}^{(n)}\|^2 \\ & \text{subject to} \quad \text{rank fSyl}_d(\hat{p}^{(n)}) \leq F - 1. \end{aligned} \quad (2.39)$$

に取り組む。式 (2.39) は、フル・Sylvester 行列の SLRA 問題であり、その解法は第 4 章で説明する。

2.2.5 近似 GCD 問題の陽表現と変数射影 [1]

問題 16 に対し、近似 GCD と商多項式に対応する新たな変数 $\hat{h}(t) \in \mathcal{P}_d, \hat{g}^{(n)}(t) \in \mathcal{P}_{n-d}$ を導入することで、式 (2.38) の制約条件を満たす $\hat{p}^{(n)}(t)$ は $\hat{p}^{(n)}(t) = \hat{g}^{(n)}(t) \cdot \hat{h}(t)$ と表現される。この表現方法に基づき、問題 16 に類似する問題が定式化される。

問題 19. 問題 16 と同様の $p^{(n)}(t)$ 及び d に対し、次の最小化問題の解となる $\hat{h}(t)$ 及び $\hat{g}^{(n)}(t) \in \mathcal{P}_{n-d}$ を計算する。

$$\underset{\substack{\hat{h}(t) \in \mathcal{P}_d, \\ \hat{g}^{(n)}(t) \in \mathcal{P}_{n-d}}}{\text{minimize}} \quad \|\hat{g}^{(n)}(t) \cdot \hat{h}(t) - p^{(n)}(t)\|^2 \quad (2.40)$$

式 (2.40) の目的関数は双線形であるため、式 (2.40) は $\hat{h}(t)$ 、または $\hat{g}^{(n)}(t)$ の一方を固定した際にもう一方についての線形最小二乗問題となり、その解は陽に表される。このような性質を持つ最適化問題については、変数射影原理 [72] の適用が可能である。

$\hat{h}$ に関する変数射影により、式 (2.40) は次の最適化問題として書き換えられる。

$\hat{h}$ に関する変数射影：VP $_{\hat{h}}$

$$\hat{\mathbf{h}}_* := \arg \min_{\hat{\mathbf{h}} \in \mathbb{R}^{d+1}} \mathcal{L}_1(\hat{\mathbf{h}}) \quad (2.41)$$

ここで

$$\begin{aligned} \mathcal{L}_1(\hat{\mathbf{h}}) &:= \min_{\hat{\mathbf{g}}^{(n)} \in \mathbb{R}^F} \sum_{i=1}^N \left\| \mathcal{M}_{f_i}(\hat{\mathbf{h}}) \hat{\mathbf{g}}^{(i)} - \mathbf{p}^{(i)} \right\|^2 \\ &= \min_{\hat{\mathbf{g}}^{(n)} \in \mathbb{R}^F} \left\| \mathcal{M}_{\mathbf{f}, [d+1]}(\hat{\mathbf{h}}) \hat{\mathbf{g}}^{(n)} - \mathbf{p}^{(n)} \right\|^2 \\ &= \left\| (\mathbf{\Pi}_1 - \mathbf{I}_{n_p}) \mathbf{p}^{(n)} \right\|^2 \end{aligned} \quad (2.42)$$

であり、

$$\begin{aligned} \mathcal{M}_{\mathbf{f}, [d+1]}(\hat{\mathbf{h}}) &:= \begin{bmatrix} \mathcal{M}_{f_1}(\hat{\mathbf{h}}) & & \\ & \ddots & \\ & & \mathcal{M}_{f_N}(\hat{\mathbf{h}}) \end{bmatrix} \in \mathbb{R}^{n_p \times F}, \\ \mathbf{\Pi}_1 &:= \mathcal{M}_{\mathbf{f}, [d+1]}(\hat{\mathbf{h}}) (\mathcal{M}_{\mathbf{f}, [d+1]}(\hat{\mathbf{h}}))^\dagger \end{aligned}$$

表 2.1: 変数の対応関係

 Variable correspondence between VP_P , $\text{VP}_{\hat{\mathbf{h}}}$ and $\text{VP}_{\hat{\mathbf{g}}^{(n)}}$.

$\text{VP}_{\boldsymbol{\xi}}$	$\boldsymbol{\xi}$	L	K	$\mathbf{y}$	$\mathbf{k}$	$\mathcal{M}_{\mathbf{k},l}(\boldsymbol{\xi})$
$\text{VP}_{\hat{\mathbf{h}}}$	$\hat{\mathbf{h}}$	$d+1$	F	$\hat{\mathbf{g}}^{(n)}$	$\mathbf{f}$	$\mathcal{M}_{\mathbf{f},[d+1]}(\hat{\mathbf{h}})$
$\text{VP}_{\hat{\mathbf{g}}^{(n)}}$	$\hat{\mathbf{g}}^{(n)}$	F	$d+1$	$\hat{\mathbf{h}}$	$[d+1]$	$\mathcal{M}_{[d+1],\mathbf{f}}(\hat{\mathbf{g}}^{(n)})$

とした。また、 $\hat{\mathbf{g}}^{(n)}$ に関する変数射影により、次の最適化問題を得る。

$\hat{\mathbf{g}}^{(n)}$ に関する変数射影： $\text{VP}_{\hat{\mathbf{g}}^{(n)}}$

$$\hat{\mathbf{g}}_*^{(n)} := \arg \min_{\hat{\mathbf{g}}^{(n)} \in \mathbb{R}^F} \mathcal{L}_2(\hat{\mathbf{g}}^{(n)}) \quad (2.43)$$

ここで

$$\begin{aligned} \mathcal{L}_2(\hat{\mathbf{g}}^{(n)}) &:= \min_{\hat{\mathbf{h}} \in \mathbb{R}^{d+1}} \sum_{i=1}^N \left\| \mathcal{M}_{d+1}(\hat{\mathbf{g}}^{(i)}) \hat{\mathbf{h}} - \mathbf{p}^{(i)} \right\|^2 \\ &= \min_{\hat{\mathbf{h}} \in \mathbb{R}^{d+1}} \left\| \mathcal{M}_{[d+1],\mathbf{f}}(\hat{\mathbf{g}}^{(n)}) \hat{\mathbf{h}} - \mathbf{p}^{(n)} \right\|^2 \\ &= \left\| (\mathbf{\Pi}_2 - \mathbf{I}_{n_p}) \mathbf{p}^{(n)} \right\|^2 \end{aligned} \quad (2.44)$$

であり、

$$\begin{aligned} \mathcal{M}_{[d+1],\mathbf{f}}(\mathbf{g}^{(n)}) &:= \begin{bmatrix} \mathcal{M}_{d+1}(\hat{\mathbf{g}}^{(1)}) \\ \vdots \\ \mathcal{M}_{d+1}(\hat{\mathbf{g}}^{(N)}) \end{bmatrix} \in \mathbb{R}^{n_p \times (d+1)}, \\ \mathbf{\Pi}_2 &:= \mathcal{M}_{[d+1],\mathbf{f}}(\hat{\mathbf{g}}^{(n)}) (\mathcal{M}_{[d+1],\mathbf{f}}(\hat{\mathbf{g}}^{(n)}))^\dagger \end{aligned}$$

とした。

$\text{VP}_{\hat{\mathbf{h}}}$ と $\text{VP}_{\hat{\mathbf{g}}^{(n)}}$ のどちらの変数射影を用いるかは、問題に含まれるベクトルの長さに依存する。例えば、 $\text{VP}_{\hat{\mathbf{h}}}$ は、GCD の次数が小さい場合に用いられ [73]、 $\text{VP}_{\hat{\mathbf{g}}^{(n)}}$ は商多項式の次数が小さい場合に用いられる [66, 67]。この使い分けは、要素数が少ない変数について変数射影を行うことで、最適化問題全体の計算量が小さくなるという事実に基づく [1]。

$\text{VP}_{\hat{\mathbf{h}}}$ あるいは $\text{VP}_{\hat{\mathbf{g}}^{(n)}}$ を、 $\boldsymbol{\xi} \in \mathbb{R}^L$ に関する変数射影として式 (2.45) の形式で統一的に表すこととする。その際の変数の対応関係を表 2.1 に示す。

ξ に関する変数射影：VP $_{\xi}$

$$\xi_* := \arg \min_{\xi \in \mathbb{R}^L} \mathcal{L}^{(LS)}(\xi), \quad (2.45)$$

ただし

$$\mathcal{L}^{(LS)}(\xi) := \min_{\mathbf{y} \in \mathbb{R}^K} \|\mathcal{M}_{k,l}(\xi)\mathbf{y} - \mathbf{p}^{(n)}\|^2 \quad (2.46)$$

であり、 $\mathcal{M}_{k,l}(\xi) \in \mathbb{R}^{n_p \times K}$ である。

さらに Usevich らは、式 (2.45), (2.46) により定められる最適化問題を、後述のモザイク型 Hankel 行列の低ランク近似問題に帰着させた [1]。この形式の問題は、高速かつ高精度な計算パッケージ [57] によって求解可能であるため、近似 GCD 問題を求解するにあたり、数値計算の観点で有利である。

2.2.6 変数射影法による近似 GCD 問題との関係 [1]

天下りの的であるが、次の二段階最適化問題を考える：

$$\mathbf{x}_* := \arg \min_{\mathbf{x} \in \mathbb{R}^{L \times 1} \setminus \{\mathbf{0}\}} \mathcal{L}^{(LN)}(\mathbf{x}) \quad (2.47)$$

ただし

$$\mathcal{L}^{(LN)}(\mathbf{x}) := \min_{\hat{\mathbf{c}} \in \mathbb{R}^{n_p}} \|\mathbf{c} - \hat{\mathbf{c}}\|^2 \quad \text{subject to} \quad \mathcal{H}_{k,l}(\hat{\mathbf{c}})\mathbf{x} = \mathbf{0}. \quad (2.48)$$

ここで式 (2.48) の問題は、線形最小ノルム問題であり、閉形式解を持つことに注意されたい。式 (2.46) と式 (2.48) との間に成立する関係について次の命題 20 が知られている。この命題は、式 (2.46) における最小化が式 (2.48) における最小化を介して計算可能であることを意味する。後に第 4.4 節で説明するように、上記の二段階最適化問題は、モザイク型 Hankel 行列に対する SLRA 問題を変数射影原理により変数の分離を施したものと同一の形式である。そのため $\mathcal{L}^{(LS)}$ の ξ に関する最適化をモザイク型 Hankel 行列の構造化低ランク近似の形式で計算可能となる。また、モザイク型 Hankel 行列の SLRA 問題については、先に述べたように計算パッケージ [57] が利用可能である。

命題 20 ([1, Proposition 6.5]). 式 (2.46), 及び式 (2.48) による最小化問題の解をそれぞれ $\mathbf{y}_*$, $\hat{\mathbf{c}}_*^{(k,l)}$ と表し, $\mathbf{c}^{(k,l)} = \mathbf{p}^{(n)}$, $t = 1$, $\mathbf{x} = \boldsymbol{\xi}$ とすると,

$$\mathcal{L}^{(LS)}(\boldsymbol{\xi}) = \|\mathbf{c}^{(k,l)}\|^2 - \mathcal{L}^{(LN)}(\mathbf{x}), \quad (2.49)$$

$$\mathcal{M}_{k,l}(\boldsymbol{\xi})\mathbf{y}_* = \mathbf{c}^{(k,l)} - \hat{\mathbf{c}}_*^{(k,l)}. \quad (2.50)$$

証明. $\boldsymbol{\xi} = \hat{\mathbf{h}}$ の場合について証明する. この時 $k = f$, $l = [d+1]$, $\mathcal{L}^{(LS)}(\boldsymbol{\xi}) = \mathcal{L}_1(\hat{\mathbf{h}})$ である. よく知られているように, $\epsilon := \min_{\mathbf{y}} \|\mathbf{A}\mathbf{y} - \mathbf{b}\|^2$ と $\epsilon' := \min_{\mathbf{z}} \|\mathbf{b} - \mathbf{z}\|^2$ sub.to $\mathbf{A}^\top \mathbf{z} = \mathbf{0}$ の間には図 2.1 のように $\epsilon + \epsilon' = \|\mathbf{b}\|^2$ なる関係があり, それらの最適解は,

$$\mathbf{y}_* = \mathbf{A}^\dagger \mathbf{b}, \quad (2.51)$$

$$\mathbf{z}_* = \mathbf{b} - \mathbf{A}\mathbf{y}_* \quad (2.52)$$

となる. この関係を利用するために, $\mathcal{L}_1(\hat{\mathbf{h}})$ に対して

$$\begin{aligned} \mathcal{L}'(\hat{\mathbf{h}}) &:= \min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{p}^{(n)} - \mathbf{z}\|^2 \\ \text{subject to } \mathcal{M}_{f,[d+1]}^\top(\hat{\mathbf{h}})\mathbf{z} &= \mathbf{0} \end{aligned} \quad (2.53)$$

とおくと, $\mathcal{L}_1(\hat{\mathbf{h}}) = \|\mathbf{p}^{(n)}\|^2 - \mathcal{L}'(\hat{\mathbf{h}})$ となる. ここで, Hankel 行列 $\mathcal{H}_{k,l}(\mathbf{c})$ と, $\mathbf{h} \in \mathbb{R}^l$ の畳み込み行列 $\mathcal{M}_k(\mathbf{h}) \in \mathbb{R}^{(k+l-1) \times k}$ の間に存在する,

$$\mathcal{H}_{k,l}(\mathbf{c})\mathbf{h} = \mathcal{M}_k^\top(\mathbf{h})\mathbf{c} \quad (2.54)$$

なる関係 [2] を用いることで

$$\mathcal{M}_{f,[d+1]}^\top(\hat{\mathbf{h}})\mathbf{z} = \mathcal{H}_{f,[d+1]}(\mathbf{z})\hat{\mathbf{h}} \quad (2.55)$$

を得る. 以上より, 式 (2.53) は

$$\begin{aligned} \mathcal{L}'(\hat{\mathbf{h}}) &:= \min_{\mathbf{z} \in \mathbb{R}^{np}} \|\mathbf{p}^{(n)} - \mathbf{z}\|^2 \\ \text{subject to } \mathcal{H}_{f,[d+1]}(\mathbf{z})\hat{\mathbf{h}} &= \mathbf{0} \end{aligned} \quad (2.56)$$

と表され $\mathcal{L}^{(LN)}(\hat{\mathbf{h}})$ に一致し, 式 (2.49) が成立する. また, 式 (2.52) を変形すると, $\mathbf{b} - \mathbf{z}_* = \mathbf{A}\mathbf{y}_*$ となるため, 命題の記号を当てはめると直ちに式 (2.50) が従う.

$\boldsymbol{\xi} = \hat{\mathbf{g}}^{(n)}$ の場合も, 同様に証明可能である. □

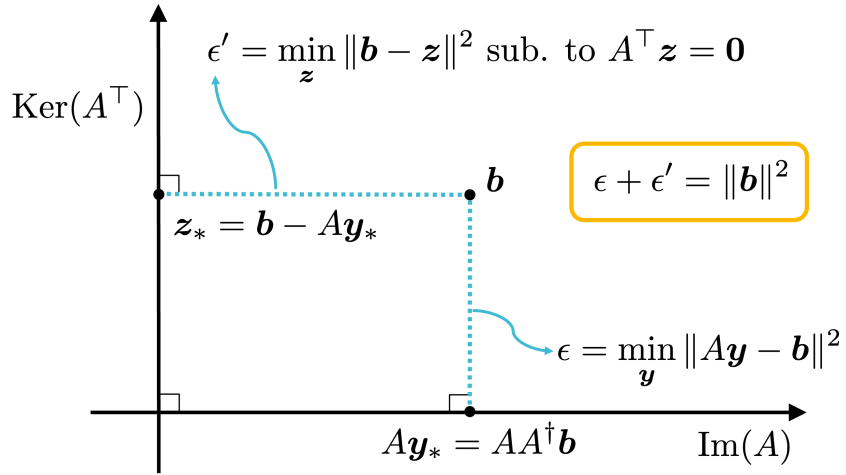


図 2.1: 線形最小二乗問題と線形最小ノルム問題の最小値同士の関係: Relationship between the Minimum Values of the Linear Least Squares Problem and the Linear Minimum Norm Problem

注 21. 近似 GCD 問題の像空間表現では、最適化変数は $\hat{g}^{(n)}$ と $\hat{h}$ であり、 $p^{(n)}$ の近似多項式はそれらの畳み込みとして表される:

$$\hat{p}^{(n)} = \mathcal{M}_{k,l}(\xi)y. \quad (2.57)$$

一方、 ξ ($\hat{h}$ あるいは $\hat{g}^{(n)}$) に関する変数射影を適用し、命題 20 によりモザイク型 Hankel 行列の SLRA 問題として解く場合、 y ($\hat{g}^{(n)}$ あるいは $\hat{h}$) は陽には求まらないため、式 (2.57) を y についての線型方程式とみなして求解する必要がある。しかしながら、[1, § 6.5] でも指摘がなされているように、実用上は $\mathcal{M}_{k,l}(\xi)$ の条件数が大きく、そのような y の計算精度が悪い場合があることに注意が必要である。◇

勾配ベースの手法により最適化問題 (2.47) の最適化にあたっては、変数 ξ に対して初期推定値を与える必要がある。近似 GCD 問題を $VP_{\hat{g}^{(n)}}$ により解く場合、 $\xi = \hat{g}^{(n)}$ であるので、商多項式の初期推定値を必要とする。GCD の次数に関する条件を満たす近似多項式 $\hat{p}^{(n)}$ について $\text{fSyl}_d(\hat{p}^{(n)})\hat{g}^{(n)} = \mathbf{0}$ が成立することから、商多項式の初期推定値としては、 $\text{fSyl}_d(p^{(n)})\hat{g}^{(n)} \approx \mathbf{0}$ の近似解を用いる。具体的には、 $\text{fSyl}_d(p^{(n)})$ の最小特異値に対応する右特異ベクトル $v_1 \in \mathbb{R}^F$ を用いる。 $VP_{\hat{h}}$ を用いる場合は、 $\xi = \hat{h}$ の

Algorithm 1 変数射影法に基づく近似 GCD 計算手法

(Conventional Approximate GCD Calculation Method based on Variable Projection Approach)

Input: $p^{(n)}(t) \in \mathcal{P}_n$, $d \in \mathbb{N}$ ($1 \leq d \leq \min_i n_i$), $\mathbf{v} \in \mathbb{R}^F$

Output: $\boldsymbol{\xi}_*(\mathbf{v}) \in \mathbb{R}^L$, $\mathbf{y}_*(\mathbf{v}) \in \mathbb{R}^K$, $\hat{\mathbf{p}}_*^{(n)}(\mathbf{v}) \in \mathbb{R}^{n_p}$

- 1: 式 (2.58) を計算し, $\hat{\mathbf{h}}$ の初期推定値とする.
 - 2: 命題 20 を利用し式 (2.45) を解き, その解を $\boldsymbol{\xi}_*(\mathbf{v})$ とする.
 - 3: 手順 2 で付随して計算される $\mathbf{p}^{(n)}$ の近似値を $\hat{\mathbf{p}}_*^{(n)}(\mathbf{v})$ とおく.
 - 4: 式 (2.57) の解 $\mathbf{y}_*(\mathbf{v}) = \mathcal{M}_{k,l}^\dagger(\boldsymbol{\xi}_*(\mathbf{v}))\hat{\mathbf{p}}_*^{(n)}(\mathbf{v})$ を計算する.
-

初期値として

$$\hat{\mathbf{h}}_{\text{ini}} = \arg \min_{\hat{\mathbf{h}} \in \mathbb{R}^{d+1}} \sum_{i=1}^N \|\mathcal{M}_{d+1}(\hat{\mathbf{g}}_{\text{ini}}^{(i)})\hat{\mathbf{h}} - \mathbf{p}^{(i)}\|^2 \quad (2.58)$$

を与える. ここで $\hat{\mathbf{g}}_{\text{ini}}^{(n)}$ は, 何らかの方法によって与えた $\hat{\mathbf{g}}^{(n)}$ の初期値である. このように, どちらの変数に関して変数射影を行う場合でも, 商多項式の初期推定値を与える必要がある.

$\hat{\mathbf{g}}^{(n)}$ の初期値を $\mathbf{v}$ とした際に計算される式 (2.45) の局所最適解を $\boldsymbol{\xi}_*(\mathbf{v})$ と表し, 対応する $\mathbf{p}^{(n)}$ の近似多項式を $\hat{\mathbf{p}}_*^{(n)}(\mathbf{v})$ と表す. $\boldsymbol{\xi} = \boldsymbol{\xi}_*(\mathbf{v})$, $\hat{\mathbf{p}}^{(n)} = \hat{\mathbf{p}}_*^{(n)}(\mathbf{v})$ に対し定まる式 (2.57) の, $\mathbf{y}$ についての解を $\mathbf{y}_*(\mathbf{v})$ と表す.

以上により説明される, 商多項式の初期推定値を $\mathbf{v}$ で与えた際の $\text{VP}_{\boldsymbol{\xi}}$ による近似 GCD 計算アルゴリズムを, Algorithm 1 に示した.

第3章 非構造化低ランク近似問題 [2]

本章では、行列に対する構造を加味しない場合の実行列の重みつき低ランク近似について、既存の手法を概説する。これらの手法は、それ自体で工学的な応用が存在するだけでなく、本論文の主論点である「構造化」低ランク近似問題に対する局所最適化手法に対して、妥当な初期推定値を与えるという点で有用である。

低ランク近似問題は、画像補完 [74, 75] や欠損値推定 [76] に代表されるように、工学の様々な場面に応用が存在する [77, 78, 79]。文献 [80] ではSVD (Singular Value Decomposition) の線形代数及び線形システムにおける応用例の包括的な説明を行っている。また、文献 [81, 82, 83, 84] では、複素行列の低ランク近似による二次元フィルタ設計手法が提案されている。

例えばデジタルフィルタの設計 [82, 83, 84] においては、まず所望の周波数応答を標準化して得られる行列 $\mathbf{D}$ を用意し、この $\mathbf{D}$ に対してSVDを行うことで、二次元設計問題を単純な一次元設計問題へと分解する。このとき活用されるSVDの重要な性質は、行列 $\mathbf{D}$ に対してSVDが、Euclid ノルムおよび Frobenius ノルムの両方においてランク r の最適な低ランク近似を提供するという点である。つまり、次に示す行列の低ランク近似問題

問題 22 (低ランク近似問題).

$$\hat{\mathbf{D}}_* = \arg \min_{\hat{\mathbf{D}}} \left\| \mathbf{D} - \hat{\mathbf{D}} \right\|_{\text{F}} \text{ subject to } \text{rank} \left(\hat{\mathbf{D}} \right) = r \quad (3.1)$$

に対して $\mathbf{D}$ のSVDを

$$\mathbf{D} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\top} = \sum_{i=1}^L \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top} \quad (3.2)$$

と表す。ここで

$$\mathbf{D}_r := \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top} \quad (3.3)$$

と定めると、任意の $1 \leq r \leq L$ に対して、

$$\|\mathbf{D} - \mathbf{D}_r\|_F = \min_{\hat{\mathbf{D}}} \|\mathbf{D} - \hat{\mathbf{D}}\|_F \quad \text{subject to } \text{rank}(\hat{\mathbf{D}}) = r \quad (3.4)$$

が成り立つ。

SVD を使用した低ランク近似はフィルタ設計手法として高い成功を収めているものの、その弱点として、標本化された周波数応答行列のすべての要素を等しく扱ってしまうという点がある。このため、場合によっては設計の質が低下する可能性がある。このように、データ行列に対して低ランク近似を行う際に、取り組んでいる課題の性質を考慮し、データ行列中の要素の重要度を加味した近似を行いたいという動機が存在する。行列の要素に対して重要度を導入するために、重み付き Frobenius ノルムを定義する。

定義 23 (重み付き Frobenius ノルム). 行列 $\mathbf{D} \in \mathbb{R}^{K \times L}$ 及び非負定値の重み行列 $\mathbf{W} \in \mathbb{R}^{(KL) \times (KL)}$ に対し、次のように定める。

$$\|\mathbf{D}\|_{\mathbf{W}} := \sqrt{(\text{vec}(\mathbf{D}))^\top \mathbf{W} \text{vec}(\mathbf{D})} \quad (3.5)$$

$\mathbf{W}$ は非負定値行列であるから、 $\mathbf{W} = \mathbf{L}_{\mathbf{W}}^\top \mathbf{L}_{\mathbf{W}}$ を満たす $\mathbf{W}$ の平方根 $\mathbf{L}_{\mathbf{W}}$ が存在する。このことから、

$$\|\mathbf{D}\|_{\mathbf{W}}^2 = (\text{vec}(\mathbf{D}))^\top \mathbf{W} \text{vec}(\mathbf{D}) = \|\mathbf{L}_{\mathbf{W}} \text{vec}(\mathbf{D})\|_2^2 \quad (3.6)$$

と書き換えることができるため、 $\|\cdot\|_{\mathbf{W}}$ が行列のノルムになっていることは容易に確認される。また、 $\mathbf{W} = \mathbf{I}_{KL}$ とすることで $\|\mathbf{D}\|_{\mathbf{W}} = \sqrt{\text{vec}^\top(\mathbf{D}) \text{vec}(\mathbf{D})} = \sqrt{\sum_{i=1}^K \sum_{j=1}^L D_{i,j}^2} = \|\mathbf{D}\|_F$ を得、このことから $\|\cdot\|_{\mathbf{W}}$ は Frobenius ノルムに対する一般化である。

本章では、文献 [2] の内容をもとに、次に定義する (実行列の) 非構造化重み付き低ランク近似問題に対する求解手法を紹介する。

問題 24.

$$\hat{\mathbf{D}}_* = \arg \min_{\hat{\mathbf{D}}} \|\mathbf{D} - \hat{\mathbf{D}}\|_{\mathbf{W}} \quad \text{subject to } \text{rank } \hat{\mathbf{D}} = r \quad (3.7)$$

ここで $\mathbf{W}$ は サイズ $(KL) \times (KL)$ の非負定値の重み行列である。

表記の簡略化のため、式 (3.1) の目的関数内に現れる差分を表す記号 $\Delta \mathbf{D}$ を導入する:

$$\Delta \mathbf{D} := \mathbf{D} - \hat{\mathbf{D}}. \quad (3.8)$$

3.1 重み付き低ランク近似問題

近似基準に導入する重み行列 $\mathbf{W}$ の分類を行う。これに先立ち、次式により $\mathbf{W}$ と関連づけられる重み行列 $\bar{\mathbf{W}}$ を導入する：

$$\|\Delta \mathbf{D}\|_{\mathbf{W}} = \|\Delta \mathbf{D}^\top\|_{\bar{\mathbf{W}}}, \quad \text{for all } \Delta \mathbf{D} \in \mathbb{R}^{K \times L}. \quad (3.9)$$

重み付き低ランク近似では、近似の評価に用いる重み行列が次の3種類のブロック対角構造をとることがある。

列ごとの重み付け

$$\mathbf{W} = \text{blkdiag}(\mathbf{W}_1, \dots, \mathbf{W}_L), \quad \mathbf{W}_j \in \mathbb{R}^{K \times K}, \quad \mathbf{W}_j \succeq 0 \quad (3.10)$$

行ごとの重み付け

$$\bar{\mathbf{W}} = \text{blkdiag}(\mathbf{W}_1, \dots, \mathbf{W}_K), \quad \mathbf{W}_i \in \mathbb{R}^{L \times L}, \quad \mathbf{W}_i \succeq 0 \quad (3.11)$$

要素ごとの重み付け

$$\mathbf{W} = \text{diag}(\mathbf{w}), \quad w_i \geq 0 \text{ for } i = 1, \dots, KL \quad (3.12)$$

要素ごとの重み付けでは、近似基準は新たに導入する行列 Ω 、及び Hadamard 積 $\odot$ を用いて

$$\|\Delta \mathbf{D}\|_{\mathbf{W}} = \|\mathbf{D} - \hat{\mathbf{D}}\|_{\mathbf{W}} = \|\Omega \odot (\mathbf{D} - \hat{\mathbf{D}})\|_{\text{F}} = \sqrt{\sum_{i=1}^K \sum_{j=1}^L \Omega_{ij} (D_{ij} - \hat{D}_{ij})^2} \quad (3.13)$$

と表せる。ここで、 $\Omega \in \mathbb{R}^{K \times L}$ は、式 (3.12) 中の $\mathbf{w}$ に対して $\text{vec}(\Omega) = \sqrt{\mathbf{w}}$ を満たすことに注意されたい。ただし $\sqrt{\mathbf{w}}$ は、非負ベクトル $\mathbf{w}$ の各要素の平方根を要素に持つベクトルである。なお、 $\Omega_{ij} = 0$ とすれば、対応する $\mathbf{D}$ の要素 D_{ij} は近似基準に影響を与えず、欠損値として扱われる。

3.2 特殊なケース: 重みなし低ランク近似

重み行列が単位行列のスカラー倍（すなわち、すべての要素が同じ重みを持つ場合）となる場合を考える。このとき、ノルム $\|\cdot\|_{\mathbf{W}}$ は Frobenius ノルム $\|\cdot\|_F$ に比例し、低ランク近似問題は行列 $\mathbf{D}$ の特異値分解を用いて解析的に解くことができる。この結果は Eckart-Young-Mirsky の定理 [85, 86, 87] として知られている。

定理 25 (Eckart-Young-Mirsky). $\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^\top$ を $\mathbf{D}$ の特異値分解とし、以下のよう
にブロック分割を行う。

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \in \mathbb{R}^{K \times K}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \in \mathbb{R}^{K \times K}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \in \mathbb{R}^{K \times L}$$

ここで、 $\mathbf{U}_1 \in \mathbb{R}^{K \times r}$, $\mathbf{U}_2 \in \mathbb{R}^{K \times (K-r)}$, $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\Sigma_2 = \text{diag}(\sigma_{r+1}, \dots, \sigma_K)$, $\mathbf{V}_1 \in \mathbb{R}^{r \times L}$, $\mathbf{V}_2 \in \mathbb{R}^{(K-r) \times L}$ である。このとき、打ち切り特異値分解 (truncated SVD)

$$\hat{\mathbf{D}}^* = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^\top \quad (3.14)$$

によりランク r の行列を計算すると、これは Frobenius ノルムに関する低ランク近似問題 (3.1) の解である。この時の近似誤差は

$$\|\mathbf{D} - \hat{\mathbf{D}}^*\|_F = \min_{\text{rank } \hat{\mathbf{D}} \leq r} \|\mathbf{D} - \hat{\mathbf{D}}\|_F = \sqrt{\sigma_{r+1}^2 + \dots + \sigma_K^2} \quad (3.15)$$

となる。なお特異値が $\sigma_r \neq \sigma_{r+1}$ である場合に限り、 $\hat{\mathbf{D}}^*$ が一意に定まる。

注 26 (ユニタリ不変ノルム). 定理 25 は、任意のユニタリ不変ノルム $\|\cdot\|$ に対しても成り立つ。すなわち、任意の直交行列 $\mathbf{U}, \mathbf{V}$ に対して

$$\|\mathbf{U}\Delta\mathbf{D}\mathbf{V}\| = \|\Delta\mathbf{D}\|$$

が成り立つ場合、定理の結論がそのまま適用可能である。◇

注 27 (スペクトルノルムでの近似). 行列 $\Delta\mathbf{D}$ のスペクトルノルム $\|\Delta\mathbf{D}\|_2$ は $\Delta\mathbf{D}$ の最大特異値で定義される。この場合、最適なランク r のスペクトルノルム誤差は、最初
に無視された特異値と等しくなる。すなわち、

$$\min_{\text{rank } \hat{\mathbf{D}}=r} \|\mathbf{D} - \hat{\mathbf{D}}\|_2 = \sigma_{r+1}.$$

この場合も, truncated SVD によって最適な近似が得られるが, Frobenius ノルムの場合とは異なり, $\sigma_r \neq \sigma_{r+1}$ であっても最適解は一意ではないことに注意を要する. $\diamond$

一般の重み付き低ランク近似問題に対する解析的解

定理 25 に示されたような解析解は, 一般の重み付き低ランク近似問題 ($\|\cdot\|_{\mathbf{W}}$) に対しては知られていない. 現時点で, 解析解が得られる最も広いクラスの重み付き低ランク近似問題は, 重み行列 $\mathbf{W}$ が正定値で, 以下の形を持つ場合であることが示されている [88, Theorem 3].

$$\mathbf{W} = \mathbf{W}_2 \otimes \mathbf{W}_1, \quad \text{where } \mathbf{W}_1 \in \mathbb{R}^{K \times K}, \quad \mathbf{W}_2 \in \mathbb{R}^{L \times L} \quad (3.16)$$

ここで, 記号 ' $\otimes$ ' は Kronecker 積を表し, $\mathbf{A} \in \mathbb{R}^{K \times L}$ に対して,

$$\mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} A_{11}\mathbf{B} & \cdots & A_{1L}\mathbf{B} \\ \vdots & \ddots & \vdots \\ A_{K1}\mathbf{B} & \cdots & A_{KL}\mathbf{B} \end{bmatrix} \quad (3.17)$$

で定義される. このような形式重み行列の設定では, 近似基準が次の簡単な形に整理される (導出は次段落):

$$\|\sqrt{\mathbf{W}_1} \Delta \mathbf{D} \sqrt{\mathbf{W}_2}\|_{\text{F}}. \quad (3.18)$$

近似基準の形式より, このような重み行列を採用した近似基準に基づく近似問題は両側重み付き低ランク近似 (two-side-weighted low rank approximation) と呼ばれる [2]. この設定は, 次の特殊ケースを含む. ただし, $\sqrt{\mathbf{W}}$ は半正定値行列 $\mathbf{W}$ の主平方根を表す.

- 式 (3.10) で全列に対して同じ重み行列 $\mathbf{W}_1$ を用いる場合:

$\mathbf{W}_1 = \mathbf{W}_2 = \cdots = \mathbf{W}_L$ とすることに相当し, 近似基準は $\|\sqrt{\mathbf{W}_1} \Delta \mathbf{D}\|_{\text{F}}$ である.

- 式 (3.11) で全行に対して同じ重み行列 $\mathbf{W}_1$ を用いる場合:

$\mathbf{W}_1 = \mathbf{W}_2 = \cdots = \mathbf{W}_K$ とすることに相当し, 近似基準は $\|\Delta \mathbf{D} \sqrt{\mathbf{W}_1}\|_{\text{F}}$ となる.

また, 対角行列 $\mathbf{W}$, すなわち要素ごとの重み付けを行う場合, 解析解を持つ特別なケース ($\mathbf{W}_2 \otimes \mathbf{W}_1$) は, Ω によって重みが要素ごとに与えられる場合に対応する.

重み行列を式 (3.16) の形式に設定した重み付き低ランク近似問題 (3.7) の目的関数が式 (3.18) と等しいことは、次の恒等式

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X}) \quad (3.19)$$

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = (\mathbf{A}_1 \mathbf{A}_2) \otimes (\mathbf{B}_1 \mathbf{B}_2) \quad (3.20)$$

を利用することで、以下のように導出できる [2, 88].

$$\begin{aligned} \|\mathbf{D} - \hat{\mathbf{D}}\|_{\mathbf{W}_2 \otimes \mathbf{W}_1} &= \sqrt{\text{vec}^\top(\Delta \mathbf{D})(\mathbf{W}_2 \otimes \mathbf{W}_1) \text{vec}(\Delta \mathbf{D})} \\ &= \left\| (\sqrt{\mathbf{W}_2} \otimes \sqrt{\mathbf{W}_1}) \text{vec}(\Delta \mathbf{D}) \right\|_2 \\ &= \left\| \text{vec}(\sqrt{\mathbf{W}_1} \Delta \mathbf{D} \sqrt{\mathbf{W}_2}) \right\|_2 \\ &= \left\| \sqrt{\mathbf{W}_1} \Delta \mathbf{D} \sqrt{\mathbf{W}_2} \right\|_F \end{aligned} \quad (3.21)$$

したがって、ノルム $(\|\cdot\|_{\mathbf{W}_2 \otimes \mathbf{W}_1})$ を持つ重み付き低ランク近似問題 (3.7) は、次に定める両側重み付き低ランク近似問題と等価である [2, Chapter 4].

問題 28 (両側重み付き低ランク近似).

$$\begin{aligned} &\underset{\hat{\mathbf{D}}}{\text{minimize}} && \left\| \sqrt{\mathbf{W}_1}(\mathbf{D} - \hat{\mathbf{D}})\sqrt{\mathbf{W}_2} \right\|_F \\ &\text{subject to} && \text{rank}(\hat{\mathbf{D}}) \leq r \end{aligned} \quad (3.22)$$

上述の最小化問題 (3.22) は解析解を持つ.

定理 29 (両側重み付き低ランク近似 [2, Theorem 4.12]). 修正されたデータ行列を

$$\mathbf{D}_m := \sqrt{\mathbf{W}_1} \mathbf{D} \sqrt{\mathbf{W}_2} \quad (3.23)$$

と定義し、 $\mathbf{D}_m$ の最適（非重み付き）低ランク近似を $\hat{\mathbf{D}}_m^*$ とすると、

$$\hat{\mathbf{D}}^* := \sqrt{\mathbf{W}_1}^{-1} \hat{\mathbf{D}}_m^* \sqrt{\mathbf{W}_2}^{-1} \quad (3.24)$$

が式 (3.22) の解となる. この解は常に存在し、 $\hat{\mathbf{D}}_m^*$ が一意である場合に限り、一意に定まる.

3.3 交互射影アルゴリズム

この節では、要素ごとに重み付けされた近似基準を用いた低ランク近似問題 (3.7) を扱う. 解法は局所最適化に基づき、交互射影アルゴリズム (alternating projections

algorithm) を利用する.

まず, ランク制約の像表現を用いて低ランク性の制約を満たす行列を表現する. すなわち,

$$\text{rank } \hat{\mathbf{D}} \leq r \iff \exists \mathbf{Q} \in \mathbb{R}^{K \times r}, \mathbf{P} \in \mathbb{R}^{r \times L} \text{ s.t. } \hat{\mathbf{D}} = \mathbf{Q}\mathbf{P} \quad (3.25)$$

この表現 (3.25) を用いることで, 等価なパラメータ最適化問題を導出する.

$$\underset{\mathbf{Q} \in \mathbb{R}^{K \times r}, \mathbf{P} \in \mathbb{R}^{r \times L}}{\text{minimize}} \quad \|\Omega \odot (\mathbf{D} - \mathbf{Q}\mathbf{P})\|_{\text{F}} \quad (3.26)$$

3.3.1 交互射影法

この方法は, 問題 (3.26) が双線形であることを利用する. すなわち, $\hat{\mathbf{D}} = \mathbf{Q}\mathbf{P}$ という構造のため, $\mathbf{Q}$ または $\mathbf{P}$ のいずれかを固定すれば, 残りは線形最小二乗問題となる. この性質に基づいて問題 (3.26) を以下のように $\mathbf{Q}$ と $\mathbf{P}$ に関する2つの部分問題に分解する.

$\mathbf{Q}$ に関する最適化問題:

$$\underset{\mathbf{Q} \in \mathbb{R}^{K \times r}}{\text{minimize}} \quad \|\Omega \odot (\mathbf{D} - \mathbf{Q}\mathbf{P})\|_{\text{F}} \quad (3.27)$$

$\mathbf{P}$ に関する最適化問題:

$$\underset{\mathbf{P} \in \mathbb{R}^{r \times L}}{\text{minimize}} \quad \|\Omega \odot (\mathbf{D} - \mathbf{Q}\mathbf{P})\|_{\text{F}} \quad (3.28)$$

交互射影アルゴリズムではこれら二つを交互に更新することで収束を目指す.

3.3.2 問題の分離

問題 (3.28) は, 行列 $\mathbf{P}$ の各列 $\mathbf{p}_1, \dots, \mathbf{p}_L$ に対する独立した最小化問題に分離可能であり,

$$\underset{\mathbf{p}_i \in \mathbb{R}^r}{\text{minimize}} \quad \|\Omega_{:,i} \odot (\mathbf{d}_i - \mathbf{Q}\mathbf{p}_i)\|_{\text{F}} \quad (3.29)$$

となる. 同様に問題 (3.27) も K 個の独立した最小化問題に分離可能である. この分離性により, 計算効率の良い実装が可能となる.

3.3.3 欠損値推定

行列 $\mathbf{D}$ に欠損値が含まれる場合も、交互射影法を用いた低ランク近似を適用できる。ここで、欠損値として与えられた要素の添字集合を $\mathcal{J}_g$ とすることで、問題 (3.29) は以下のように記述できる。

$$\underset{\mathbf{p}_i \in \mathbb{R}^r}{\text{minimize}} \|\Omega_{\mathcal{J}_i, i} \odot (\mathbf{D}_{\mathcal{J}_i, i} - \mathbf{Q}_{\mathcal{J}_i, :} \mathbf{p}_i)\|_F \quad (3.30)$$

この修正を導入することで、要素ごとに重み付けされた低ランク近似に対する交互射影法が、欠損値を含む場合にも適用可能となる。

3.3.4 初期近似値の構成

交互射影アルゴリズムは逐次計算アルゴリズムの一種であり、最適化変数 $\mathbf{Q}$ に対して初期値を与える必要がある。初期値の与え方次第では、アルゴリズムの収束が遅くなるなど望ましくない事象が生じるため、初期値の構成は重要である。ここでは、初期値の構成方法として、Algorithm 2 に示す、すべての欠損要素を 0 で補完等加重低ランク近似問題を解く方法を紹介する。

Algorithm 2 欠損値を含むデータ行列に対する（非構造化）低ランク近似

:(Unstructured) Low-Rank Approximation with Missing Data.

Input: データ行列 $\mathbf{D} \in \mathbb{R}^{K \times L}$, 階数 $r \leq L$, $\mathbf{D}$ の各列 $D_{:,i}$ 中の要素が与えられている添字集合 $\mathcal{J}_i^1$, $\mathbf{D}$ の各行 $D_{i,:}$ 中の要素が与えられている添字集合 $\mathcal{J}_i^2$,

Output: $\mathbf{Q} \in \mathbb{R}^{K \times r}$, $\mathbf{P} \in \mathbb{R}^{r \times L}$ であって, $\mathbf{QP}$ が (3.30) の最適解となっているもの

- 1: $\mathbf{D}$ 中の欠損値を 0 で置き換える.
 - 2: 特異値分解により $\mathbf{D}$ の最も大きい r 個の特異値に対応する左特異ベクトルを計算し, $\mathbf{Q}$ とする.
 - 3: $\mathbf{P} := \mathbf{Q}^\top \mathbf{D}$.
-

3.3.5 アルゴリズム

以上の議論により導出される、交互射影法に基づく欠損値を含む要素ごとの重み付き低ランク近似のアルゴリズムを、Algorithm 3 に掲載する。なお、 j 回目の反復計算

において手順4及び手順4で計算される近似誤差ベクトル $\mathbf{r}^{[j]}$ は、次式で定められる2乗誤差を表す.

$$\mathbf{r}^{[j]} = \left\| \Omega \odot (\mathbf{D} - \hat{\mathbf{D}}^{[j]}) \right\|_F^2$$

Algorithm 3に示したアルゴリズム内での反復計算のたびに、近似基準値は単調減少する．また局所最適解への収束が保証され、局所的な収束速度は線形である．収束判定は、更新後の相対誤差の減少率に基づいて行う．これは、目的関数値の相対減少に対する許容誤差 ε を設定することで実現される．より計算コストの高い収束判定の方法としては、以下のようなものがある．

- 近似行列 $\hat{\mathbf{D}}^{[j]}$ の変化量の直接評価
- 目的関数の勾配の大きさを計算し、モデルパラメータに関する収束性を評価

3.4 本章のまとめ

本章では、行列に構造が与えられていない、あるいは構造を加味しない場合の重みつき低ランク近似手法について説明し、重みに対する考え方を拡張することでこれを欠損値を含むデータ行列に対する低ランク近似へと拡張可能であることを概説した．本章で紹介した低ランク近似アルゴリズムは、近似前の行列が構造化されていたとしても、それを全く考慮しないで近似を行う．しかしながら、次章で説明する構造化低ランク近似アルゴリズム（非凸な問題に対して凸最適化手法を適用しているため、解の初期値依存性が高い）に対してある意味で適切な初期値を与えるアルゴリズムとして応用が可能である点を付け加えて説明する．

Algorithm 3 交互射影法に基づく重み付き低ランク近似アルゴリズム

: Alternating Projections for Weighted Low-Rank Approximation.

Input: データ行列 $\mathbf{D} \in \mathbb{R}^{K \times L}$, $\mathbf{D}$ の各列 $D_{:,i}$ 中の要素が与えられている添字集合 $\mathcal{J}_i^1$, $\mathbf{D}$ の各行 $D_{i,:}$ 中の要素が与えられている添字集合 $\mathcal{J}_i^2$, ランク制約 r , 重み行列 $\Omega \in \mathbb{R}^{K \times L} \geq 0$, 相対許容誤差 ε .

Output: 最適化問題 (3.26) に対する局所最適解 $\hat{\mathbf{D}} = \hat{\mathbf{D}}^{[j]} := \mathbf{Q}^{[j]} \mathbf{P}^{[j]}$.

- 1: Algorithm 2 により $\mathbf{Q}$ の初期近似値を計算し, $\mathbf{Q}^{[0]}$ とする.
 - 2: $j := 0$.
 - 3: **repeat**
 - 4: $\mathbf{r}^{[j]} := 0$. { 近似誤差ベクトルの初期化 }
 - 5: **for** $i = 1, \dots, L$ **do**
 - 6: $\mathbf{c} := \text{diag}(\Omega_{\mathcal{J}_i^1, i}) D_{\mathcal{J}_i^1, i} = \Omega_{\mathcal{J}_i^1, i} \odot D_{\mathcal{J}_i^1, i}$.
 - 7: $\mathbf{Q} := \text{diag}(\Omega_{\mathcal{J}_i^1, i}) \mathbf{Q}_{\mathcal{J}_i^1, :}^{[j]} = (\Omega_{\mathcal{J}_i^1, i} \mathbf{1}_r) \odot \mathbf{Q}_{\mathcal{J}_i^1, :}^{[j]}$.
 - 8: $\mathbf{p}_i^{[j]} := (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{c}$. { $\mathbf{p}_i$ の更新 }
 - 9: $\mathbf{r}^{[j]} := \mathbf{r}^{[j]} + \|\mathbf{c} - \mathbf{Q} \mathbf{p}_i^{[j]}\|^2$. { 近似誤差ベクトルの更新 }
 - 10: **end for**
 - 11: $\mathbf{P}^{[j]} := \begin{bmatrix} \mathbf{p}_1^{[j]} & \dots & \mathbf{p}_L^{[j]} \end{bmatrix}$. { $\mathbf{P}$ の構成 }
 - 12: $\mathbf{r}^{[j+1]} := 0$. { 近似誤差ベクトルの初期化 }
 - 13: **for** $i = 1, \dots, K$ **do**
 - 14: $\gamma := \mathbf{D}_{i, \mathcal{J}_i^2} \text{diag}(\Omega_{i, \mathcal{J}_i^2}) = \mathbf{D}_{i, \mathcal{J}_i^2} \odot \Omega_{i, \mathcal{J}_i^2}$
 - 15: $\mathbf{P} := \mathbf{P}_{\mathcal{J}_i^2, :}^{[j]} \text{diag}(\Omega_{i, \mathcal{J}_i^2}) = \mathbf{P}_{\mathcal{J}_i^2, :}^{[j]} \odot (\mathbf{1}_r^T \Omega_{i, \mathcal{J}_i^2})$.
 - 16: $\mathbf{q}_i^{[j+1]} := \gamma \mathbf{P}^T (\mathbf{P} \mathbf{P}^T)^{-1}$. { $\mathbf{q}_i$ の更新 }
 - 17: $\mathbf{r}^{[j+1]} := \mathbf{r}^{[j+1]} + \|\gamma - \mathbf{q}_i^{[j+1]} \mathbf{p}\|^2$. { 近似誤差ベクトルの更新 }
 - 18: **end for**
 - 19: $\mathbf{Q}^{[j+1]} := \begin{bmatrix} \mathbf{q}_1^{[j+1]} \\ \vdots \\ \mathbf{q}_K^{[j+1]} \end{bmatrix}$. { 更新された $\mathbf{q}_i$ より $\mathbf{Q}$ を構成 }
 - 20: $j := j + 1$.
 - 21: **until** $|\mathbf{r}^{[j]} - \mathbf{r}^{[j-1]}| / \mathbf{r}^{[j]} < \varepsilon$.
-

第4章 構造化低ランク近似

4.1 はじめに

構造化低ランク近似 (Structured Low-Rank Approximation, SLRA) 問題は、信号処理や線形システム論といった応用分野において、データの背後にある構造やモデルを抽出・推定するための有力な手法として近年注目されている [17, 55]。この問題を考える主な動機は、「観測されたデータが、理想的には何らかのモデル（例えば再帰関係やシステムのインパルス応答）に従って生成されているはずだが、ノイズや誤差によりそのモデルが完全には満たされていない」という問題が実用上様々な場面で発生するためである [17]。SLRA は、このような不完全なデータに対して、元の構造に最も近い形で低ランクな行列を復元することを目指す。

SLRA の具体的な応用例とその動機

信号処理：スペクトル解析・ノイズ除去 ある信号が有限個の正弦波の線形結合として表されると仮定した場合、対応する Hankel 行列のランクはその正弦波の個数に等しい。しかし実際には観測ノイズの影響で Hankel 行列のランクは増大する。このとき、Hankel 構造を持つ低ランク行列に近似することで、ノイズを除去し、信号の本質的な成分（周波数成分など）を抽出できる。これは例えば耐震性の解析 [89, 14] に応用されている。

システム同定：モデル推定 SLRA は、離散時間線形システムの応答データを用いてシステムの状態空間モデルを推定する問題においても、重要な役割を果たす。理論的には、観測データから作成される Hankel 行列が低ランクであることが期待される（システムの次数がランクに対応）。しかし、実際のデータには外乱や測定誤差が含まれるため、得られた行列はフルランクとなってしまう。このような場

合, SLRA によりランク制約を満たす構造化行列を復元し, システムモデルをより正確に推定することが可能である [90, 91, 92].

制御理論：モデル次数削減と予測制御 モデル予測制御などでは, 状態や出力の予測に基づいて制御入力を最適化するために, システムモデルの簡略化が求められることがある. ここでも SLRA により, 高次のシステムをランクの低い近似モデルに変換することで, 制御の計算コストを削減しつつ, 必要な精度を確保することが期待できる [59, 11].

計算機代数：近似 GCD 問題の求解 数値計算において, 丸め誤差のために厳密な GCD が存在しないが, 理論的には共通因子があると仮定されるときや, システム制御・信号処理で, システムの共通極・共通零点などの構造を抽出したいときなど, 近似的に GCD を計算する必要がある場合がある. この計算問題に対して, Sylvester 行列の SLRA により解が提供される [7, 6].

このように, SLRA はノイズにより満たすべき性質を満たさなくなったデータを修正し, その背後にある物理的・数学的モデルを復元するための強力な枠組みとして, 多くの実問題に対して応用されている. 問題の性質や構造に応じた適切な近似を行うことで, 信号処理・制御・システム同定といったさまざまな分野において, より信頼性の高い解析や設計が可能である.

4.1.1 構造化低ランク近似の定式化

Affine 行列構造の定義を再掲する.

定義 30 (Affine 行列構造 [2]). 体 $\mathbb{F}$ に対して, 以下のように定義される, 構造化パラメータ空間 $\mathbb{F}^{n_p}$ から行列空間 $\mathbb{F}^{K \times L}$ への Affine 写像

$$\mathcal{S}(\mathbf{c}) := \mathbf{S}_0 + \sum_{i=1}^{n_p} c_i \mathbf{S}_i, \quad \mathbf{c} \in \mathbb{F}^{n_p}, \quad \mathbf{S}_i \in \mathbb{F}^{K \times L} \quad (4.1)$$

を Affine 行列構造と定める.

ほとんどの応用例において, 体 $\mathbb{F}$ は実数体 $\mathbb{R}$ または複素数体 $\mathbb{C}$ である. 本論文の主目

的は, SLRA 手法の定式化とその初期化戦略の妥当性を改善することである. この目的と本質的に関係しない部分で議論が煩雑になるのを避けるため, 以下では $\mathbb{F} = \mathbb{R}$ とし議論を進めるが, ほとんどの場合, これらの議論はそのまま $\mathbb{C}$ へ拡張可能である. また, 多くの応用例では, $\mathbf{S}_0$ が零行列で, $1 \leq k \leq n_p$ に対しては, 要素が 0 または 1 の行列 $\mathbf{S}_k$ で指定されるような構造行列に関心がある. 本論文では, 次の仮定をおく.

仮定 31. $\mathbf{S}_i$ ($i = 1, \dots, n_p$) の成分は 0 または 1 のいずれかであり, 行列

$$\begin{bmatrix} \text{vec}(\mathbf{S}_0) & \dots & \text{vec}(\mathbf{S}_{n_p}) \end{bmatrix}$$

の各行には高々 1 つの非零要素しか含まれないものとする. すなわち, 構造化行列 $\mathcal{S}(\mathbf{p})$ の各要素は, $\mathbf{p}$ の要素のいずれか (高々 1 つ) に対応するか, あるいは固定値として定まっている.

SLRA 問題についても再掲する.

問題 32 (構造化低ランク近似 (SLRA) [2]). $\{\mathbf{S}_k\}_{k=1, \dots, n_z}$ で指定される行列構造 (4.1), パラメータベクトル $\mathbf{c} \in \mathbb{R}^{n_z}$, ノルム $\|\cdot\|_{\mathbf{W}}$, そして整数 r であって ($0 < r < L$) を満たすものが所与であるとする. (Affine) 構造化低ランク近似問題を次の最小化問題として定める:

$$\begin{aligned} & \underset{\hat{\mathbf{c}} \in \mathbb{R}^{n_z}}{\text{minimize}} && \left\| \hat{\mathbf{c}} - \mathbf{c} \right\|_{\mathbf{W}}^2 \\ & \text{subject to} && \text{rank } \mathcal{S}(\hat{\mathbf{c}}) \leq r. \end{aligned} \tag{4.2}$$

本章においては, $\mathbf{W}$ が正定値対称行列であることを要請する.

例えば, Frobenius 計量における $\mathcal{S}(\mathbf{c})$ の SLRA を求める場合, 次の最適化問題を考える:

$$\begin{aligned} & \underset{\hat{\mathbf{c}} \in \mathbb{R}^{n_p}}{\text{minimize}} && \left\| \mathcal{S}(\hat{\mathbf{c}}) - \mathcal{S}(\mathbf{c}) \right\|_{\text{F}}^2 \\ & \text{subject to} && \text{rank } \mathcal{S}(\hat{\mathbf{c}}) \leq r. \end{aligned} \tag{4.3}$$

式 (4.3) の目的関数は、次のように展開される:

$$\begin{aligned}
 \|\mathcal{S}(\hat{\mathbf{c}}) - \mathcal{S}(\mathbf{c})\|_F^2 &= \left\| \left(\sum_{k=1}^{n_p} \mathbf{S}_k \hat{c}_k + \mathbf{S}_0 \right) - \left(\sum_{k=1}^{n_p} \mathbf{S}_k c_k + \mathbf{S}_0 \right) \right\|_F^2 \\
 &= \left\| \sum_{k=1}^{n_p} \mathbf{S}_k (\hat{c}_k - c_k) \right\|_F^2 \\
 &= \left\| \sum_{k=1}^{n_p} \text{vec}(\mathbf{S}_k) (\hat{c}_k - c_k) \right\|_2^2 \\
 &= \left\| \begin{bmatrix} \text{vec}(\mathbf{S}_1) & \text{vec}(\mathbf{S}_2) & \dots & \text{vec}(\mathbf{S}_{n_p}) \end{bmatrix} (\hat{\mathbf{c}} - \mathbf{c}) \right\|_2^2 \\
 &= \|\hat{\mathbf{c}} - \mathbf{c}\|_{\mathbf{W}}^2.
 \end{aligned}$$

ここで重み行列 $\mathbf{W}$ は, $\mathbf{S} \in \mathbb{R}^{KL \times n_p}$ を

$$\mathbf{S} := [\text{vec}(\mathbf{S}_1) \text{vec}(\mathbf{S}_2) \dots \text{vec}(\mathbf{S}_{n_p})] \quad (4.4)$$

として,

$$\mathbf{W} = \mathbf{S}^\top \mathbf{S}$$

とした. 以上に説明したように, 式 (4.3) は式 (4.2) の形に帰着する. 先の例のように, $\mathcal{S}$ が Hankel 構造である場合, $\mathbf{W}$ を, $\mathcal{S}(\mathbf{c})$ における c_i の出現回数を対角要素に持つ対角行列として選択することにより, 適切な重み付けが行われる.

構造化行列 $\mathcal{S}(\mathbf{c})$ と式 (4.4) で導入した行列 $\mathbf{S}$ の関係を vec 作用素とその逆作用素 vec^{-1} により記述する:

$$\text{vec}(\mathcal{S}(\mathbf{c})) = \text{vec}(\mathbf{S}_0) + \mathbf{S}\mathbf{c}, \quad (4.5)$$

$$\mathcal{S}(\mathbf{c}) = \mathbf{S}_0 + \text{vec}^{-1}(\mathbf{S}\mathbf{c}). \quad (4.6)$$

目的関数の 2 ノルムでの表現

任意の問題 (4.2) は計量を含まない単純な 2 ノルムの最小化の形式に表すことが可能である. まず, 重み行列の最大階数分解を行う:

$$\mathbf{W} = \mathbf{L}_W^\top \mathbf{L}_W. \quad (4.7)$$

ここで、 $\mathbf{W}$ の正定値性より、 $\mathbf{L}_{\mathbf{W}}$ は正則であるため、その逆行列 $\mathbf{L}_{\mathbf{W}}^{-1}$ が存在することを注記する。次に、

$$\Delta \mathbf{c} := \mathbf{L}_{\mathbf{W}}(\hat{\mathbf{c}} - \mathbf{c}) \quad (4.8)$$

で問題 (4.2) の変数を置き換えると、

$$\begin{aligned} \|\hat{\mathbf{c}} - \mathbf{c}\|_{\mathbf{W}}^2 &= (\hat{\mathbf{c}} - \mathbf{c})^\top \mathbf{W}(\hat{\mathbf{c}} - \mathbf{c}) \\ &= (\hat{\mathbf{c}} - \mathbf{c})^\top \mathbf{L}_{\mathbf{W}}^\top \mathbf{L}_{\mathbf{W}}(\hat{\mathbf{c}} - \mathbf{c}) \\ &= \|\Delta \mathbf{c}\|_2^2 \end{aligned} \quad (4.9)$$

であるため、最小化問題 (4.2) は次の形に縮約される:

$$\begin{aligned} &\underset{\Delta \mathbf{c} \in \mathbb{R}^{np}}{\text{minimize}} && \|\Delta \mathbf{c}\|_2^2 \\ &\text{subject to} && \text{rank}(\mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c})) \leq r. \end{aligned} \quad (\text{SLRA})$$

4.1.2 本章の構成

第 4.2 節では、零空間表現に基づく SLRA 問題の定式化について説明する。第 4.3 節では、構造化全最小ノルム法 (STLN) に基づく Affine 構造に対する SLRA 手法 [93, 8] (STLN-SLRA) を紹介する。この手法は、対象の行列がほぼ正方に近い場合に適している。第 4.4 節では、Affine 構造を持つ行列に対して変数射影法を適用する手法 [21] (VP-SLRA) について、詳細に説明する。この手法は、一方の次元が小さく、もう一方が大きいような行列に適しており、特にシステム同定のように、小さい次元がモデルの複雑さによって、大きい次元がサンプル数によって決まる場面において有効である。

これらに加えて、構造化低ランク近似には、他にも低ランク制約の凸緩和に基づくアプローチや、像空間表現に基づくアプローチなどが提案されている。本論文の主な検討対象ではないものの、関連性が高いため、第 4.5 節で簡単に紹介する。

4.2 低ランク性の零空間表現

行列のランクが欠損しているという条件 $\text{rank } \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}) \leq r$ は同値な条件

$$\exists \mathbf{X} \in \mathbb{R}^{L \times (L-r)} \quad \text{such that} \quad \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{X} = \mathbf{O}_{K, L-r}, \text{rank } \mathbf{X} = L - r \quad (4.10)$$

で置き換えられる．これは低ランク性の零空間表現と呼ばれる [88, 2]

本論文では，議論が煩雑になることを防ぎ，本質的な問題点に焦点を当てることを目的とし，

$$r = L - 1 \quad (4.11)$$

つまり，ランク 1 落ちの場合のみを扱うこととする．この時対応する零空間の次元も 1 であり，零空間の基底行列はベクトルとみなすことができるため， $\mathbf{X}$ の代わりに $\mathbf{x}$ を記号として使用する．以後 SLRA 問題 (SLRA) を次の形で取り扱う．

$$\begin{aligned} & \underset{\Delta \mathbf{c} \in \mathbb{R}^{np}, \mathbf{x} \in \mathbb{R}^L}{\text{minimize}} && \|\Delta \mathbf{c}\|_2^2 \\ & \text{subject to} && \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{x} = \mathbf{0} \text{ with } \mathbf{x} \neq \mathbf{0}. \end{aligned} \quad (4.12)$$

低ランク性を零空間表現する SLRA 計算手法として

- STLN-SLRA : 構造化全最小ノルム (Structured Total Least Norm: STLN) に基づく手法 (第 4.3 節で説明)
- VP-SLRA : 変数射影 (Variable Projection) に基づく手法 (第 4.4 節で説明)

の二つが知られている．

4.3 STLN-SLRA

Affine 構造に対して，STLN (Structured Total Least Norm : 構造化全最小ノルム) 法に基づいた SLRA [7] が知られている．STLN は，変動が想定されるベクトル $\mathbf{c}$ を構造化パラメータに持つ構造化行列 $\mathbf{A}(\mathbf{c})$ 及び $\mathbf{b}(\mathbf{c})$ によって定められる，必ずしも解を持たない過決定線形方程式

$$\mathbf{A}(\mathbf{c})\mathbf{X} = \mathbf{B}(\mathbf{c}) \quad (4.13)$$

に対して $\mathbf{c}$ に微小な修正を加えるとともに， $\mathbf{X}$ を求解する手法である．STLN に基づく SLRA では，問題 (4.12) の制約条件を (4.13) の形式に変形する．この考えを実現す

るために、まず単位行列の i 列目を消去した行列 $\tilde{\mathbf{I}}_i$ を導入する:

$$\tilde{\mathbf{I}}_i := \begin{matrix} & \begin{matrix} 1 \\ \vdots \\ \vdots \\ i \\ \vdots \\ \vdots \\ n \end{matrix} \end{matrix} \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & 0 & 0 & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix} \in \mathbb{R}^{L \times (L-1)}. \quad (4.14)$$

明らかな事実として $\tilde{\mathbf{I}}_i \tilde{\mathbf{I}}_i^\top + \mathbf{e}_i \mathbf{e}_i^\top = \mathbf{I}_L$ が成立する. この関係に基づき次のように問題 (4.12) の制約条件に対して同値変形を行う:

$$\mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{x} = \mathbf{0} \iff \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) (\tilde{\mathbf{I}}_i \tilde{\mathbf{I}}_i^\top + \mathbf{e}_i \mathbf{e}_i^\top) \mathbf{x} = \mathbf{0} \quad (4.15)$$

$$\iff \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \tilde{\mathbf{I}}_i \tilde{\mathbf{I}}_i^\top \mathbf{x} = -x_i \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{e}_i. \quad (4.16)$$

次に $x_i \neq 0$ を仮定し, (4.16) の両辺に $-1/x_i$ を乗算することで, 問題 (4.12) の制約条件 $\mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{x} = \mathbf{0}$ と等価な方程式

$$\mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \tilde{\mathbf{I}}_i \mathbf{x}_{(i)} = \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{e}_i \quad (4.17)$$

を得る. ここで

$$\mathbf{x}_{(i)} := \frac{-1}{x_i} \tilde{\mathbf{I}}_i^\top \mathbf{x}$$

である. 以降式 (4.17) を第 i 初期方程式と呼称する. 制約条件を初期方程式で置き換えることにより新たな問題

$$\begin{aligned} & \underset{\mathbf{x}_{(i)}, \Delta \mathbf{c}}{\text{minimize}} && \|\Delta \mathbf{c}\|_2^2 \\ & \text{subject to} && \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \tilde{\mathbf{I}}_i \mathbf{x}_{(i)} = \mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c}) \mathbf{e}_i \end{aligned} \quad (4.18)$$

を得る.

4.3.1 アルゴリズムの導出

新たな変数

$$\Delta \mathbf{c}' := \mathbf{L}_{\mathbf{W}}^{-1} \Delta \mathbf{c} \quad (4.19)$$

を導入し，式 (4.18) の（非線形）制約に対する残差 $\mathbf{r}$ を定める：

$$\mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}) := \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}')(\tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i). \quad (4.20)$$

この時 SLRA 問題 (4.12) の制約条件は次の形となる：

$$\mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}) = \mathbf{0}. \quad (4.21)$$

$\delta \mathbf{c}'$ と $\delta \mathbf{x}_{(i)}$ をそれぞれ $\Delta \mathbf{c}'$ と $\mathbf{x}_{(i)}$ の微小変化とし，式 (4.21) を線形近似する：

$$\begin{aligned} & \mathbf{r}(\Delta \mathbf{c}' + \delta \mathbf{c}', \mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}) \\ &= \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}' + \delta \mathbf{c}')(\tilde{\mathbf{I}}_i(\mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}) - \mathbf{e}_i) \\ &= (\mathcal{S}(\mathbf{c} + \Delta \mathbf{c}') + \mathcal{S}(\delta \mathbf{c}') - \mathbf{S}_0)(\tilde{\mathbf{I}}_i(\mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}) - \mathbf{e}_i) \\ &= \underbrace{\mathcal{S}(\mathbf{c} + \Delta \mathbf{c}')(\tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i)}_{\mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)})} + (\mathcal{S}(\delta \mathbf{c}') - \mathbf{S}_0)(\tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i) \end{aligned} \quad (4.22)$$

$$\begin{aligned} & \quad + \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}')\tilde{\mathbf{I}}_i \delta \mathbf{x}_{(i)} + (\mathcal{S}(\delta \mathbf{c}') - \mathbf{S}_0)\tilde{\mathbf{I}}_i \delta \mathbf{x}_{(i)} \\ & \approx \mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}) + (\mathcal{S}(\delta \mathbf{c}') - \mathbf{S}_0)(\tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i) + \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}')\tilde{\mathbf{I}}_i \delta \mathbf{x}_{(i)}. \end{aligned} \quad (4.23)$$

ここで，

$$(\mathcal{S}(\delta \mathbf{c}') - \mathbf{S}_0)\tilde{\mathbf{I}}_i \delta \mathbf{x}_{(i)} = \left(\sum_{k=1}^{n_p} \mathbf{S}_k \delta c_k \right) \tilde{\mathbf{I}}_i \delta \mathbf{x}_{(i)} \quad (4.24)$$

$$= \left(\sum_{k=1}^{n_p} \mathbf{S}_k \tilde{\mathbf{I}}_i \underbrace{\delta \mathbf{x}_{(i)} \delta c_k}_{\approx \mathbf{0}} \right) \quad (4.25)$$

$$\approx \mathbf{0} \quad (4.26)$$

とした． $\phi_i : \mathbb{R}^{n-1} \rightarrow \mathbb{R}^n$ を $\mathbf{x}_{(i)}$ の i 番目の要素に -1 を挿入する写像と定義する．すなわち，

$$\phi_i(\mathbf{x}_{(i)}) := \tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i \quad (4.27)$$

次に式 (4.23) の右辺 2 番目の項を展開する：

$$\begin{aligned} (\mathcal{S}(\delta \mathbf{c}) - \mathbf{S}_0)(\tilde{\mathbf{I}}_i \mathbf{x}_{(i)} - \mathbf{e}_i) &= \left(\sum_{k=1}^{n_p} \mathbf{S}_k \delta c_k \right) \phi_i(\mathbf{x}_{(i)}) \\ &= \sum_{k=1}^{n_p} (\mathbf{S}_k \phi_i(\mathbf{x}_{(i)})) \delta c_k \\ &= [\mathbf{S}_1 \phi_i(\mathbf{x}_{(i)}) \ \dots \ \mathbf{S}_{n_p} \phi_i(\mathbf{x}_{(i)})] \delta \mathbf{c}. \end{aligned}$$

さらに

$$\mathbf{T}_{\mathcal{S}}(\mathbf{x}) := \begin{bmatrix} \mathbf{S}_1 \mathbf{x} & \cdots & \mathbf{S}_{n_p} \mathbf{x} \end{bmatrix} \quad (4.28)$$

$$= \begin{bmatrix} \text{vec}(\mathbf{S}_1 \mathbf{x}) & \cdots & \text{vec}(\mathbf{S}_{n_p} \mathbf{x}) \end{bmatrix} \quad (4.29)$$

$$= \begin{bmatrix} \text{vec}(\mathbf{x}^\top \mathbf{S}_1^\top) & \cdots & \text{vec}(\mathbf{x}^\top \mathbf{S}_{n_p}^\top) \end{bmatrix} \quad (4.30)$$

$$= \begin{bmatrix} \text{vec}(\mathbf{x}^\top \mathbf{S}_1^\top \mathbf{I}_K) & \cdots & \text{vec}(\mathbf{x}^\top \mathbf{S}_{n_p}^\top \mathbf{I}_K) \end{bmatrix} \quad (4.31)$$

$$= \begin{bmatrix} (\mathbf{I}_K \otimes \mathbf{x}^\top) \text{vec}(\mathbf{S}_1^\top) & \cdots & (\mathbf{I}_K \otimes \mathbf{x}^\top) \text{vec}(\mathbf{S}_{n_p}^\top) \end{bmatrix} \quad (4.32)$$

$$= (\mathbf{I}_K \otimes \mathbf{x}^\top) \underbrace{\begin{bmatrix} \text{vec}(\mathbf{S}_1^\top) & \cdots & \text{vec}(\mathbf{S}_{n_p}^\top) \end{bmatrix}}_{=:\bar{\mathbf{S}}} \quad (4.33)$$

$$= (\mathbf{I}_K \otimes \mathbf{x}^\top) \bar{\mathbf{S}} \quad (4.34)$$

と定め、式 (4.23) を整理することにより、式 (4.21) の近似を行う:

$$\left[\mathbf{T}_{\mathcal{S}}(\phi_i(\mathbf{x}_{(i)})) \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}') \tilde{\mathbf{I}}_i \right] \begin{pmatrix} \delta \mathbf{c}' \\ \delta \mathbf{x}_{(i)} \end{pmatrix} \approx -\mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}). \quad (4.35)$$

$\delta \mathbf{c}', \delta \mathbf{x}_{(i)}$ を固定された $\Delta \mathbf{c}'$ と $\mathbf{x}_{(i)}$ に対する代替変数として考えると、最適化問題 (4.12) は、線形制約最小二乗問題となる [8]:

$$\begin{aligned} & \underset{\delta \mathbf{c}', \mathbf{x}_{(i)}}{\text{minimize}} \quad \left\| \begin{bmatrix} \mathbf{L}_{\mathbf{W}} & \mathbf{O} \end{bmatrix} \begin{pmatrix} \delta \mathbf{c}' \\ \delta \mathbf{x}_{(i)} \end{pmatrix} - (-\mathbf{L}_{\mathbf{W}} \Delta \mathbf{c}') \right\|^2 \\ & \text{subject to} \quad \left[\mathbf{T}_{\mathcal{S}} \phi_i((\mathbf{x}_{(i)})) \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}') \tilde{\mathbf{I}}_i \right] \begin{pmatrix} \delta \mathbf{c}' \\ \delta \mathbf{x}_{(i)} \end{pmatrix} = -\mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}). \end{aligned} \quad (4.36)$$

なお、もともとの最適化問題は $\|\Delta \mathbf{c} + \delta \mathbf{c}\|_2^2$ を目的関数として定式化されていたが、式 (4.36) においてはこれと恒等的に等価な表現 $\|\mathbf{L}_{\mathbf{W}} \Delta \mathbf{c}' + \mathbf{L}_{\mathbf{W}} \delta \mathbf{c}'\|_2^2$ を用いている。式 (4.36) は線形な制約を持つ最小二乗問題であるため、零空間法 [87] によって最適解を計算可能である。その具体的な計算手順は、Algorithm 4 に示す。

式 (4.36) の最適化手法として、ペナルティ法を用いたアプローチが報告されている [93, 94, 95]。この手法では、次のような最適化問題 (4.37) を解くことになる:

$$\underset{\delta \mathbf{c}', \mathbf{x}_{(i)}}{\text{minimize}} \quad \left\| \begin{bmatrix} \mathbf{L}_{\mathbf{W}} & \mathbf{O} \\ \lambda \mathbf{T}_{\mathcal{S}} \phi_i((\mathbf{x}_{(i)})) & \lambda \mathcal{S}(\mathbf{c} + \Delta \mathbf{c}') \tilde{\mathbf{I}}_i \end{bmatrix} \begin{pmatrix} \delta \mathbf{c}' \\ \delta \mathbf{x}_{(i)} \end{pmatrix} - \begin{bmatrix} -\mathbf{L}_{\mathbf{W}} \Delta \mathbf{c}' \\ -\lambda \mathbf{r}(\Delta \mathbf{c}', \mathbf{x}_{(i)}) \end{bmatrix} \right\|^2. \quad (4.37)$$

この問題は単純な線形最小二乗問題であり、効率的に解くことが可能である。

式 (4.36) (あるいは式 (4.37)) を解くことで得た解 $\Delta \mathbf{c}' + \delta \mathbf{c}'$ 及び $\mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}$ は、制約条件を一次近似したものを満たすのみであり、本来の条件

$$\mathbf{r}(\Delta \mathbf{c}' + \delta \mathbf{c}', \mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}) = \mathbf{0}$$

は厳密には満たさない。したがって、許容誤差 ϵ_c および ϵ_x を与え、更新量 $\delta \mathbf{c}$ および $\delta \mathbf{x}_{(i)}$ がそれぞれ $\delta \mathbf{c} < \epsilon_c$ および $\delta \mathbf{x}_{(i)} < \epsilon_x$ を満たすようになるまで、式 (4.36) (あるいは式 (4.37)) を繰り返し解き、近似誤差を最小化する。ただし、反復が行われるたびに、 $\Delta \mathbf{c}$ 及び $\mathbf{x}_{(i)}$ を次の更新式

$$\begin{aligned}\Delta \mathbf{c}' &\leftarrow \Delta \mathbf{c}' + \delta \mathbf{c}', \\ \mathbf{x}_{(i)} &\leftarrow \mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}\end{aligned}\tag{4.38}$$

に従って更新する。以上の手順を Algorithm 5 に掲載する。

式 (4.37) を基にした STLN-SLRA は、準 Newton 法と等価なアルゴリズムであることが知られており [93]、理論的な収束性を備えている点が大きな利点である。しかしながら、ペナルティ法では制約条件の充足を促すために、ペナルティ係数 λ を十分大きく（たとえば 10^{12} 程度）設定する必要がある。その結果として、式 (4.37) は悪条件化し、求められる解が数値的に不安定になるという新たな問題が生じる [8]。加えて、実際には式 (4.36) を最小化する手法の方が収束が速いことの方が多いという側面もあり、ペナルティ法を用いるメリットよりもデメリットの方が多いことが指摘されている [8, §3.1]。

4.3.2 初期方程式の選択における問題点

原理的には、 $x_i \neq 0$ である任意の整数 i に対して第 i 初期方程式を使用可能である。しかし、 $\mathbf{x}$ 自体が未知の変数であるため、Algorithm 5 の手順 3 においては、この条件を初期方程式選択の判断材料とすることができない。また、限定的な状況を除いて、適切な初期方程式の選択が STLN の収束速度や近似精度に影響を及ぼすことが文献 [25] により示されている。すなわち、選択する初期方程式によって STLN-SLRA の性能を向上させることができる可能性がある。このように、初期方程式の選択は STLN-SLRA の実用的な応用において重要な論点の一つである。具体的な選択手法については、第 5 章で詳述する。

Algorithm 4 零空間法（線形制約 $\mathbf{C}\mathbf{x} = \mathbf{d}$ を伴う線形最小二乗問題 $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ を最小化する最小ノルム解を計算するアルゴリズム）: nullspace method [87]

Input: $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{C} \in \mathbb{R}^{p \times n}, \mathbf{d} \in \mathbb{R}^p$ ($m + p \geq n \geq p$), $\text{rank } \mathbf{C} = p$,
 $\text{rank} \begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix} = n$

Output: 線形制約 $\mathbf{C}\mathbf{x} = \mathbf{d}$ を伴う線形最小二乗問題 $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ に対する最小ノルム解

1: $\mathbf{C}^\top$ の QR 分解を計算:

$$\mathbf{Q}_\mathbf{C}^\top \mathbf{C}^\top = \begin{pmatrix} \mathbf{R}_\mathbf{C} \\ \mathbf{O}_{(n-p) \times p} \end{pmatrix}, \quad (4.39)$$

ただし, $\mathbf{R}_\mathbf{C} \in \mathbb{R}^{p \times p}$.

2: 行列 $\mathbf{Q}_\mathbf{C}$ を次形式に分割:

$$\mathbf{Q}_\mathbf{C} = \begin{pmatrix} \mathbf{Q}_s & \mathbf{Q}_n \end{pmatrix}, \quad (4.40)$$

ただし, $\mathbf{Q}_s \in \mathbb{R}^{n \times p}, \mathbf{Q}_n \in \mathbb{R}^{n \times (n-p)}$.

3: 制約条件を満たす $\mathbf{x}$ の特殊解を計算:

$$\mathbf{x}_1 \leftarrow \mathbf{C}^\dagger \mathbf{d} = \mathbf{Q}_p \mathbf{R}_\mathbf{C}^{-\top} \mathbf{d}. \quad (4.41)$$

4: 上記特殊解に対し, 一般解を計算するためのベクトルを計算:

$$\mathbf{x}_2 \leftarrow (\mathbf{A}\mathbf{Q}_n)^\dagger (\mathbf{b} - \mathbf{A}\mathbf{x}_1). \quad (4.42)$$

5: 出力を構成:

$$\mathbf{x} \leftarrow \mathbf{x}_1 + \mathbf{Q}_n \mathbf{x}_2. \quad (4.43)$$

4.4 VP-SLRA

Affine 行列構造の SLRA に対し, 変数射影 (Variable Projection: VP) 法の適用により, SLRA 問題を変数 $\hat{\mathbf{c}}$ と $\mathbf{x}$ のそれぞれに関する最適化の二段階に分離する手法（以後, VP-SLRA）が報告されている [2, 21]. 本節では [21] に基づき, 一般の Affine 構造 $\mathcal{S}$ および重み付き 2 ノルム $\|\cdot\|_{\mathbf{W}}^2$ に対して, 最適化問題の明示的な表現を導出したのち, その問題を解くためのアルゴリズムである VP-SLRA の高効率な実装に必要な勾配の計算方法について紹介する.

Algorithm 5 STLN-SLRA

Input: $\mathbf{c}, \{\mathbf{S}_k \in \mathbb{R}^{m \times n}\}_{k=1, \dots, n_p}, \mathbf{W} \in \mathbb{R}^{n_p \times n_p}$, 及び閾値 ϵ_x, ϵ_z .

Output: ランクが $n - 1$ であるような行列 $\mathbf{S}(\mathbf{c} + \Delta \mathbf{c}) \in \mathbb{R}^{m \times n}$ 及び $\mathbf{x}_{(i)} \in \mathbb{R}^{n-1}$

- 1: 最大階数分解により $\mathbf{L}_\mathbf{W}$ を計算
- 2: $\Delta \mathbf{c} \leftarrow \mathbf{0}$ (最適化変数の初期化)
- 3: いずれかの列選択基準により初期方程式 (4.17) を構成
- 4: 最小二乗法により $\|\mathbf{r}(\mathbf{0}, \mathbf{x}_{(i)})\|^2$ を最小化する $\mathbf{x}_{(i)}$ を計算:

$$\mathbf{x}_{(i)} \leftarrow (\mathbf{S}(\mathbf{c})\mathbf{Q}_i)^\dagger \mathbf{S}(\mathbf{c})\mathbf{e}_i.$$

- 5: **while** $\frac{\|\delta \mathbf{x}_{(i)}\|}{\|\mathbf{x}_{(i)}\|} > \epsilon_x$ OR $\frac{\|\delta \mathbf{c}\|}{\|\mathbf{c}\|} > \epsilon_c$ **do**
- 6: Algorithm 4 により, 式 (4.36) を求解.
- 7: 最適化変数の更新:

$$\mathbf{x}_{(i)} \leftarrow \mathbf{x}_{(i)} + \delta \mathbf{x}_{(i)}, \quad (4.44)$$

$$\Delta \mathbf{c} \leftarrow \Delta \mathbf{c} + \delta \mathbf{c}. \quad (4.45)$$

8: **end while**

4.4.1 VP-SLRA の定式化

最適化問題の書き換え

問題の制約を $\Delta \mathbf{c}$ の線形方程式として記述するために, ベクトル化演算の性質を用いて方程式に対し同値な変形を行う. これに先立ち, $\mathbf{S}(\mathbf{c})\mathbf{x}$ を, 式 (4.34) で導入した行列 $\mathbf{T}_\mathbf{S}(\mathbf{x})$ を用いて表す.

$$\mathbf{S}(\mathbf{c})\mathbf{x} = \left(\sum_{i=1}^{n_p} c_i \mathbf{S}_i + \mathbf{S}_0 \right) \mathbf{x} \quad (4.46)$$

$$= \sum_{i=1}^{n_p} (c_i \mathbf{S}_i \mathbf{x}) + \mathbf{S}_0 \mathbf{x} \quad (4.47)$$

$$= \begin{bmatrix} \mathbf{S}_1 \mathbf{x} & \cdots & \mathbf{S}_{n_p} \mathbf{x} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_{n_p} \end{bmatrix} + \mathbf{S}_0 \mathbf{x} \quad (4.48)$$

$$= \mathbf{T}_\mathbf{S}(\mathbf{x})\mathbf{c} + \mathbf{S}_0 \mathbf{x} \quad (4.49)$$

この等式を利用することで、問題 (4.12) の等式制約部分は

$$\begin{aligned}\mathcal{S}(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1}\Delta\mathbf{c})\mathbf{x} = \mathbf{0} &\iff \mathbf{T}_{\mathcal{S}}(\mathbf{x})(\mathbf{c} + \mathbf{L}_{\mathbf{W}}^{-1}\Delta\mathbf{c}) + \mathbf{S}_0\mathbf{x} = \mathbf{0} \\ &\iff \mathbf{T}_{\mathcal{S}}(\mathbf{x})\mathbf{L}_{\mathbf{W}}^{-1}\Delta\mathbf{c} = -\mathbf{T}_{\mathcal{S}}(\mathbf{x})\mathbf{c} - \mathbf{S}_0\mathbf{x}\end{aligned}\quad (4.50)$$

と同値変形できる。新たに、

$$\begin{aligned}\mathbf{G}(\mathbf{x}) &:= \mathbf{T}_{\mathcal{S}}(\mathbf{x})\mathbf{L}_{\mathbf{W}}^{-1} = (\mathbf{I}_K \otimes \mathbf{x}^\top)\bar{\mathbf{S}}\mathbf{L}_{\mathbf{W}}^{-1}, \\ \mathbf{h}(\mathbf{x}) &:= -\mathbf{T}_{\mathcal{S}}(\mathbf{x})\mathbf{c} - \mathbf{S}_0\mathbf{x} = \mathcal{S}(\mathbf{c})\mathbf{x}\end{aligned}$$

とおくことで、式 (4.50) と同値な等式

$$\mathbf{G}(\mathbf{x})\Delta\mathbf{c} = \mathbf{h}(\mathbf{x}) \quad (4.51)$$

を得、最適化問題 (4.12) は次のように書き換えられる:

$$\begin{aligned}\underset{\Delta\mathbf{c} \in \mathbb{R}^{n_p}}{\text{minimize}} \quad & \|\Delta\mathbf{c}\|_2^2 \\ \text{subject to} \quad & \mathbf{G}(\mathbf{x})\Delta\mathbf{c} = \mathbf{h}(\mathbf{x}).\end{aligned}\quad (4.52)$$

この最適化問題 (4.52) は、固定された $\mathbf{x}$ に対しては標準的な形の線形最小ノルム問題であり、閉形式の解を持つことが知られている [96, Chapter 6]. もし $\mathbf{G}(\mathbf{x})$ が行フルランクであれば、(4.52) の解は存在し、以下で与えられる:

$$\Delta\mathbf{c}_*(\mathbf{x}) := \mathbf{G}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x}) \quad (4.53)$$

$$\mathcal{L}(\mathbf{x}) := \|\Delta\mathbf{c}_*(\mathbf{x})\|_2^2 = \mathbf{h}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x}) \quad (4.54)$$

ただし

$$\mathcal{G}(\mathbf{x}) := \mathbf{G}(\mathbf{x})\mathbf{G}^\top(\mathbf{x}) \in \mathbb{R}^{K \times K} \quad (4.55)$$

である。

注 33. $\mathbf{G}(\mathbf{x})$ が行フルランクでない場合であっても、最適化問題 (4.52) に実行可能解が存在すれば、解は上記の式において $\mathcal{G}(\mathbf{x})$ の逆行列をその擬似逆行列で置き換えることによって得られる. $\diamond$

以降では、 $\mathbf{G}(\mathbf{x})$ が行フルランクであると仮定する。これは、 $\mathcal{G}(\mathbf{x})$ が可逆であることと同値であり、 $\mathbf{G}(\mathbf{x})$ の行数が列数以下でなければならない。つまり $n_p \geq K$ 。

注 34. $\mathcal{G}(\mathbf{x})$ は $\mathbf{x}$ の成分に依存する多項式行列であるため、ある $\mathbf{x}$ に対して非特異であれば、記号的に逆行列を計算することができる。この場合、 $\mathcal{G}(\mathbf{x})$ は $\mathcal{G}(\mathbf{x})$ が非特異であるようなすべての $\mathbf{x}$ に対して無限回微分可能である [97]. $\diamond$

以上の議論により、式 (2.47) は次の最小化問題に帰着される:

$$\mathbf{x}_* = \arg \min_{\mathbf{x} \in \mathbb{R}^L} \mathbf{h}^\top(\mathbf{x}) \mathcal{G}^{-1}(\mathbf{x}) \mathbf{h}(\mathbf{x}). \quad (4.56)$$

文献 [21] では、式 (4.56) を Levenberg–Marquardt アルゴリズムを用いて逐次的に求解する手法が提案された。

$\mathcal{G}(\mathbf{x})$ の計算に関する計算過程の簡素化を目的として、新たに行列 $\mathbf{\Gamma}$ を次式で定める。

$$\mathbf{\Gamma} := \bar{\mathbf{S}} \mathbf{W}^{-1} \bar{\mathbf{S}}^\top \in \mathbb{R}^{(KL) \times (KL)}. \quad (4.57)$$

また、次のように $\mathbf{\Gamma}$ に対して区分分けを行う。

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{(1,1)} & \cdots & \mathbf{\Gamma}_{(1,K)} \\ \vdots & & \\ \mathbf{\Gamma}_{(K,1)} & \cdots & \mathbf{\Gamma}_{(K,K)} \end{bmatrix} \quad (4.58)$$

ただし、 $\mathbf{\Gamma}_{(i,j)} \in \mathbb{R}^{L \times L}$ である。ここで、

$$\mathcal{G}(\mathbf{x}) = \mathbf{G}(\mathbf{x}) \mathbf{G}^\top(\mathbf{x}) \quad (4.59)$$

$$= \mathbf{T}_{\mathcal{S}}(\mathbf{x}) \mathbf{L}_{\mathbf{W}}^{-1} \mathbf{L}_{\mathbf{W}}^{-\top} \mathbf{T}_{\mathcal{S}}^\top(\mathbf{x}) \quad (4.60)$$

$$= (\mathbf{I}_K \otimes \mathbf{x}^\top) \bar{\mathbf{S}} \mathbf{W}^{-1} \bar{\mathbf{S}}^\top (\mathbf{I}_K \otimes \mathbf{x}) \quad (4.61)$$

$$= (\mathbf{I}_K \otimes \mathbf{x}^\top) \mathbf{\Gamma} (\mathbf{I}_K \otimes \mathbf{x}) \quad (4.62)$$

であるため、 $\mathcal{G}(\mathbf{x})$ の第 (i, j) 要素 $\mathcal{G}_{i,j}(\mathbf{x})$ は

$$\mathcal{G}_{i,j}(\mathbf{x}) = \mathbf{x} \mathbf{\Gamma}_{(i,j)} \mathbf{x}^\top \quad (4.63)$$

である。

4.4.2 目的関数の勾配及びヘシアン

この節では、頻繁に $\mathcal{G}^{-1}(\mathbf{x}) \mathbf{h}(\mathbf{x})$ という因子が現れるため、次の表記を定める:

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) := \mathcal{G}^{-1}(\mathbf{x}) \mathbf{h}(\mathbf{x}) \quad (4.64)$$

この記号を用いると、目的関数は $\mathcal{L}(\mathbf{x}) = \mathbf{y}^\top(\mathbf{x})\mathbf{h}(\mathbf{x})$ と表せる。 $\mathcal{L}(\mathbf{x})$ 及び $\Delta\mathbf{c}_*(\mathbf{x})$ は、Algorithm 6 により計算可能である。

Algorithm 6 コスト関数の評価: Cost Function Evaluation

Input: $\mathcal{S}, \mathbf{W}, \Gamma_{(i,j)}, \mathbf{G}, \mathbf{x}$

Output: $\mathcal{L}(\mathbf{x})$

- 1: $\mathbf{h} \leftarrow \mathcal{S}(\mathbf{c})\mathbf{x}$.
 - 2: $\mathcal{G}_{(i,j)} \leftarrow \mathbf{x}^\top \Gamma_{(i,j)} \mathbf{x}^\top$.
 - 3: $\mathbf{y}$ についての方程式 $\mathcal{G}\mathbf{y} = \mathbf{h}$ を解く.
 - 4: $\mathcal{L}(\mathbf{x}) \leftarrow \mathbf{y}^\top \mathbf{h}$.
 - 5: $\Delta\mathbf{c}_* \leftarrow \mathbf{G}^\top \mathbf{y}$.
-

命題 35. 目的関数 $\mathcal{L}(\mathbf{x})$ の勾配 $\nabla\mathcal{L}(\mathbf{x})$ は以下で与えられる：

$$\nabla\mathcal{L}(\mathbf{x}) = -2\mathcal{S}^\top(\mathbf{c})\mathbf{y} - 2\left(\sum_{i,j=1}^K y_i y_j \Gamma_{(i,j)}^\top\right)\mathbf{x} \quad (4.65)$$

証明. 目的関数 $\mathcal{L}(\mathbf{x})$ の微分 $d\mathcal{L}(\mathbf{x})$ が

$$d\mathcal{L}(\mathbf{x}) = -2\mathbf{y}^\top \mathcal{S}(\mathbf{c})d\mathbf{x} - 2\mathbf{x}^\top \left(\sum_{i,j=1}^K y_i y_j \Gamma_{(i,j)}\right) d\mathbf{x} \quad (4.66)$$

で与えられることを示す。 $\mathcal{L}(\mathbf{x}) = \mathbf{y}^\top(\mathbf{x})\mathbf{h}(\mathbf{x})$ より、先に $d\mathbf{h}(\mathbf{x})$ 及び $d\mathbf{y}(\mathbf{x})$ の展開を行う。前者は容易に計算可能である：

$$d\mathbf{h}(\mathbf{x}) = \mathcal{S}(\mathbf{c})d\mathbf{x}. \quad (4.67)$$

一方、 $\mathbf{y}$ は $\mathbf{x}$ は複数の関数の合成として定められているため、最後まで展開してしまうと表記が煩雑となる。よって $d\mathbf{y}(\mathbf{x})$ を途中まで展開する：

$$d\mathbf{y}(\mathbf{x}) = d(\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x})) \quad (4.68)$$

$$= (d\mathcal{G}^{-1}(\mathbf{x}))\mathbf{h}(\mathbf{x}) + \mathcal{G}^{-1}(\mathbf{x})d\mathbf{h}(\mathbf{x}) \quad (4.69)$$

$$= -\mathcal{G}^{-1}(\mathbf{x})(d\mathcal{G}(\mathbf{x}))\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x}) + \mathcal{G}^{-1}(\mathbf{x})d\mathbf{h}(\mathbf{x}) \quad (4.70)$$

$$= \mathcal{G}^{-1}(\mathbf{x})(-(d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x}) + d\mathbf{h}(\mathbf{x})). \quad (4.71)$$

目的関数の微分は

$$d\mathcal{L}(\mathbf{x}) = \mathbf{y}^\top(\mathbf{x})d\mathbf{h}(\mathbf{x}) + d\mathbf{y}^\top(\mathbf{x})\mathbf{h}(\mathbf{x}) \quad (4.72)$$

$$= \mathbf{y}^\top(\mathbf{x})d\mathbf{h}(\mathbf{x}) + \mathbf{h}^\top(\mathbf{x})d\mathbf{y}(\mathbf{x}) \quad (4.73)$$

$$= \mathbf{y}^\top(\mathbf{x})d\mathbf{h}(\mathbf{x}) + \mathbf{h}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x})((d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x}) + d\mathbf{h}(\mathbf{x})) \quad (4.74)$$

$$= 2\mathbf{y}^\top(\mathbf{x})d\mathbf{h}(\mathbf{x}) - \mathbf{y}^\top(\mathbf{x})(d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x}) \quad (4.75)$$

として計算される．なお， $\mathcal{G}(\mathbf{x})$ は対称行列であるため

$$\mathbf{h}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x}) = \mathbf{h}^\top(\mathbf{x})\mathcal{G}^{-\top}(\mathbf{x}) \quad (4.76)$$

$$= (\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x}))^\top = \mathbf{y}^\top(\mathbf{x}) \quad (4.77)$$

と表せることを利用した．式 (4.75) の右辺第二項を展開するために，先んじて $d\mathcal{G}(\mathbf{x})$ の計算を行う． $d\mathbf{G}(\mathbf{x}) = (\mathbf{I}_K \otimes d\mathbf{x}^\top)\bar{\mathbf{S}}\mathbf{L}_\mathbf{W}^{-1}$ より

$$d\mathcal{G}(\mathbf{x}) = (d\mathbf{G}(\mathbf{x}))\mathbf{G}^\top(\mathbf{x}) + \mathbf{G}(\mathbf{x})(d\mathbf{G}(\mathbf{x}))^\top \quad (4.78)$$

$$= (\mathbf{I}_K \otimes d\mathbf{x}^\top)\mathbf{\Gamma}(\mathbf{I}_K \otimes \mathbf{x}) + (\mathbf{I}_K \otimes \mathbf{x}^\top)\mathbf{\Gamma}(\mathbf{I}_K \otimes d\mathbf{x}) \quad (4.79)$$

である．ここで

$$(\mathbf{I}_K \otimes d\mathbf{x}^\top)\mathbf{\Gamma}(\mathbf{I}_K \otimes \mathbf{x}) = \begin{bmatrix} d\mathbf{x}^\top\mathbf{\Gamma}_{(1,1)}\mathbf{x} & \cdots & d\mathbf{x}^\top\mathbf{\Gamma}_{(1,K)}\mathbf{x} \\ \vdots & & \\ d\mathbf{x}^\top\mathbf{\Gamma}_{(K,1)}\mathbf{x} & \cdots & d\mathbf{x}^\top\mathbf{\Gamma}_{(K,K)}\mathbf{x} \end{bmatrix} \quad (4.80)$$

$$= \begin{bmatrix} \mathbf{x}^\top\mathbf{\Gamma}_{(1,1)}d\mathbf{x} & \cdots & \mathbf{x}^\top\mathbf{\Gamma}_{(1,K)}d\mathbf{x} \\ \vdots & & \\ \mathbf{x}^\top\mathbf{\Gamma}_{(K,1)}d\mathbf{x} & \cdots & \mathbf{x}^\top\mathbf{\Gamma}_{(K,K)}d\mathbf{x} \end{bmatrix} \quad (4.81)$$

$$= (\mathbf{I}_K \otimes \mathbf{x}^\top)\mathbf{\Gamma}(\mathbf{I}_K \otimes d\mathbf{x}) \quad (4.82)$$

であることに留意すると，

$$d\mathcal{G}(\mathbf{x}) = 2(\mathbf{I}_K \otimes \mathbf{x}^\top)\mathbf{\Gamma}(\mathbf{I}_K \otimes d\mathbf{x}) \quad (4.83)$$

であり, その第 (i, j) 要素は $2\mathbf{x}^\top \mathbf{\Gamma}_{i,j} d\mathbf{x}$ である. 最終的に

$$\mathbf{y}^\top d\mathcal{G}(\mathbf{x})\mathbf{y} = 2\mathbf{y}^\top (\mathbf{I}_K \otimes \mathbf{x}^\top) \mathbf{\Gamma} (\mathbf{I}_K \otimes d\mathbf{x})\mathbf{y} \quad (4.84)$$

$$= \sum_{i,j=1}^K y_i (2\mathbf{x}^\top \mathbf{\Gamma}_{i,j} d\mathbf{x}) y_j \quad (4.85)$$

$$= 2\mathbf{x}^\top \left(\sum_{i,j=1}^K y_i y_j \mathbf{\Gamma}_{i,j} \right) d\mathbf{x}. \quad (4.86)$$

と展開でき, 以上により主張が従う. $\square$

$\Delta \mathbf{c}_*(\mathbf{x})$ のヤコビ行列

$\Delta \mathbf{c}_*(\mathbf{x})$ のヤコビ行列を計算する. $\Delta \mathbf{c}_* = \mathbf{G}^\top(\mathbf{x})\mathbf{y}(\mathbf{x})$ の微分を考える:

$$d\Delta \mathbf{c}_* = d\mathbf{G}^\top(\mathbf{x})\mathbf{y}(\mathbf{x}) + \mathbf{G}^\top(\mathbf{x})d\mathbf{y}(\mathbf{x}) \quad (4.87)$$

$$= d\mathbf{G}^\top(\mathbf{x})\mathbf{y}(\mathbf{x}) + \mathbf{G}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x}) (d\mathbf{h}(\mathbf{x}) - (d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x})) \quad (4.88)$$

を得る. 式 (4.88) の右辺の各項はそれぞれ次のように展開できる:

1. 第一項について $\mathbf{G}(\mathbf{x})$ の定義 $\mathbf{G}(\mathbf{x}) = \mathbf{T}_{\mathcal{S}}(\mathbf{x})\mathbf{L}_{\mathbf{W}}^{-1} = [\mathbf{S}_1\mathbf{x} \cdots \mathbf{S}_{n_p}\mathbf{x}]\mathbf{L}_{\mathbf{W}}^{-1}$ より

$$d\mathbf{G}(\mathbf{x}) = [\mathbf{S}_1 d\mathbf{x} \cdots \mathbf{S}_{n_p} d\mathbf{x}]\mathbf{L}_{\mathbf{W}}^{-1}$$

と計算でき,

$$(d\mathbf{G}^\top(\mathbf{x}))\mathbf{y} = \mathbf{L}_{\mathbf{W}}^{-\top} \begin{bmatrix} d\mathbf{x}^\top \mathbf{S}_1^\top \\ \vdots \\ d\mathbf{x}^\top \mathbf{S}_{n_c}^\top \end{bmatrix} \mathbf{y} = \mathbf{L}_{\mathbf{W}}^{-\top} \begin{bmatrix} d\mathbf{x}^\top \mathbf{S}_1^\top \mathbf{y} \\ \vdots \\ d\mathbf{x}^\top \mathbf{S}_{n_c}^\top \mathbf{y} \end{bmatrix} = \mathbf{L}_{\mathbf{W}}^{-\top} \begin{bmatrix} \mathbf{y}^\top \mathbf{S}_1 \\ \vdots \\ \mathbf{y}^\top \mathbf{S}_{n_c} \end{bmatrix} d\mathbf{x}. \quad (4.89)$$

2. 第二項のうち, $(d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x})$ は, 次のように展開できる:

$$\begin{aligned}
 (d\mathcal{G}(\mathbf{x}))\mathbf{y}(\mathbf{x}) &= (\mathbf{I}_K \otimes \mathbf{x}^\top) \mathbf{\Gamma} (\mathbf{I}_K \otimes d\mathbf{x}) \mathbf{y} \\
 &= \begin{bmatrix} \mathbf{x}^\top & & \\ & \ddots & \\ & & \mathbf{x}^\top \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_{(1,1)} & \cdots & \mathbf{\Gamma}_{(1,K)} \\ \vdots & & \\ \mathbf{\Gamma}_{(K,1)} & \cdots & \mathbf{\Gamma}_{(K,K)} \end{bmatrix} \begin{bmatrix} d\mathbf{x} & & \\ & \ddots & \\ & & d\mathbf{x} \end{bmatrix} \mathbf{y} \\
 &= \begin{bmatrix} \mathbf{x}^\top \mathbf{\Gamma}_{(1,1)} d\mathbf{x} & \cdots & \mathbf{x}^\top \mathbf{\Gamma}_{(1,K)} d\mathbf{x} \\ \vdots & & \vdots \\ \mathbf{x}^\top \mathbf{\Gamma}_{(K,1)} d\mathbf{x} & \cdots & \mathbf{x}^\top \mathbf{\Gamma}_{(K,K)} d\mathbf{x} \end{bmatrix} \mathbf{y} \\
 &= \sum_{i=1}^K \begin{bmatrix} \mathbf{x}^\top \mathbf{\Gamma}_{(1,i)} d\mathbf{x} \\ \vdots \\ \mathbf{x}^\top \mathbf{\Gamma}_{(K,i)} d\mathbf{x} \end{bmatrix} y_i \\
 &= \begin{bmatrix} \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(1,i)}^\top \right) \\ \vdots \\ \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(K,i)}^\top \right) \end{bmatrix} d\mathbf{x}. \tag{4.90}
 \end{aligned}$$

以上の式 (4.88) に対して式 (4.89) 及び式 (4.90), 式 (4.67) ($d\mathbf{h}(\mathbf{x}) = \mathcal{S}(\mathbf{c})d\mathbf{x}$) を代入することにより,

$$d\Delta \mathbf{c}_* = \left\{ \mathbf{L}_\mathbf{W}^{-\top} \begin{bmatrix} \mathbf{y}^\top \mathbf{S}_1 \\ \vdots \\ \mathbf{y}^\top \mathbf{S}_{n_c} \end{bmatrix} + \mathbf{G}^\top(\mathbf{x}) \mathcal{G}^{-1}(\mathbf{x}) \left(\mathcal{S}(\mathbf{c}) - \begin{bmatrix} \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(1,i)}^\top \right) \\ \vdots \\ \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(K,i)}^\top \right) \end{bmatrix} \right) \right\} d\mathbf{x}$$

を得, ヤコビ行列はその係数行列

$$\mathbf{L}_\mathbf{W}^{-\top} \begin{bmatrix} \mathbf{y}^\top \mathbf{S}_1 \\ \vdots \\ \mathbf{y}^\top \mathbf{S}_{n_c} \end{bmatrix} - \mathbf{G}^\top(\mathbf{x}) \mathcal{G}^{-1}(\mathbf{x}) \left(\mathcal{S}(\mathbf{c}) + \begin{bmatrix} \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(1,i)}^\top \right) \\ \vdots \\ \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(K,i)}^\top \right) \end{bmatrix} \right) \tag{4.91}$$

として計算できる.

Cholesky 分解の利用

本節では, より計算量の少ない目的関数の評価方法と, ヘッセ行列の近似を与える.

$\mathcal{G}(\mathbf{x})$ の Cholesky 分解

$$\mathcal{G}(\mathbf{x}) = \mathbf{L}_\mathcal{G}^\top \mathbf{L}_\mathcal{G} \tag{4.92}$$

Algorithm 7 コレスキー分解を介した線形方程式 $\mathcal{G}(\mathbf{x})\mathbf{u} = \mathbf{v}$ の求解
: Solution of the Linear System $\mathcal{G}(\mathbf{x})\mathbf{u} = \mathbf{v}$ via Cholesky Decomposition

Input: $\mathcal{G}, \mathbf{v}$

Output: $\mathbf{u}$

- 1: $\mathcal{G}$ の Cholesky 分解 $\mathbf{L}_{\mathcal{G}}$ を計算.
 - 2: $\mathbf{L}_{\mathcal{G}}^{\top} \mathbf{g} = \mathbf{v}$ および $\mathbf{L}_{\mathcal{G}} \mathbf{u} = \mathbf{g}$ を後退代入により求解.
-

は、次の恒等式

$$\mathcal{G}^{-1}(\mathbf{x}) = \mathbf{L}_{\mathcal{G}}^{-\top} \mathbf{L}_{\mathcal{G}}^{-1} \quad (4.93)$$

に基づき連立一次方程式 $\mathcal{G}(\mathbf{x})\mathbf{u} = \mathbf{v}$ を数値的に安定に解く方法を与える．その手順を Algorithm 7 に示す．

Cholesky 分解を利用することで、コスト関数は次のようにも表現される：

$$\mathcal{L}(\mathbf{x}) = \|\mathbf{L}_{\mathcal{G}}^{-1/2} \mathbf{h}(\mathbf{x})\|_2^2. \quad (4.94)$$

この式変形により、目的関数は要素数 n_p のベクトルに対する二乗和から要素数 L のベクトルに対する二乗和になり、計算量が大きく削減される．

目的関数のヘッセ行列

最後に、目的関数のヘッセ行列に対する近似を行う．式 (4.94) に現れるベクトルを、次のように定義する：

$$\mathbf{f}(\mathbf{x}) := \mathbf{L}_{\mathcal{G}}^{-1/2} \mathbf{h}(\mathbf{x}). \quad (4.95)$$

この記法を用いると、目的関数は $\mathcal{L}(\mathbf{x}) = \|\mathbf{f}(\mathbf{x})\|_2^2$ と書き表される．Gauss–Newton 法では、このような非線形最小二乗問題に対して、 $\nabla^2 \mathcal{L}(\mathbf{x})$ の対称正定値な近似として $2\mathbf{J}_f^{\top} \mathbf{J}_f$ (ただし $\mathbf{J}_f$ は $\mathbf{f}(\mathbf{x})$ のヤコビ行列) が用いられる．ここで、 $\mathbf{f}(\mathbf{x})$ の全微分 $d\mathbf{f}(\mathbf{x})$ は次のように与えられる：

$$d\mathbf{f}(\mathbf{x}) = d\left(\mathbf{L}_{\mathcal{G}}^{-1/2}(\mathbf{x})\mathbf{h}(\mathbf{x})\right) = \mathbf{L}_{\mathcal{G}}^{-1/2} d\mathbf{h}(\mathbf{x}) + (d\mathbf{L}_{\mathcal{G}}^{-1/2})\mathbf{h}(\mathbf{x}), \quad (4.96)$$

しかしながら、 $d\mathbf{L}_{\mathcal{G}}^{-1/2}$ を解析的に計算することが困難であるため、 $\mathbf{J}_f$ を明示的に表現することができない．

この問題に対し、文献 [21] では次のような近似を導入した:

$$\mathrm{d}\left(\mathbf{L}_{\mathcal{G}}^{-1/2}(\mathbf{x})\mathbf{h}(\mathbf{x})\right) \approx \mathbf{L}_{\mathcal{G}}^{-1/2}\mathrm{d}\mathbf{h}(\mathbf{x}) - \frac{1}{2}\mathbf{L}_{\mathcal{G}}^{-\top}(\mathrm{d}\mathcal{G}(\mathbf{x}))\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x}). \quad (4.97)$$

この近似により定義される擬似ヤコビ行列 ([98] を参照のこと) を用いることで、ヘッセ行列の近似とする手法が考案されている. 擬似ヤコビ行列 $\mathbf{J}_f^p$ は、 $\mathbf{y}$ の定義式 ($\mathcal{G}^{-1}\mathbf{h}(\mathbf{x}) = \mathbf{y}$) 及び式 (4.90) を式 (4.97) に代入することにより、直ちに

$$\mathbf{J}_f^p = \mathbf{L}_{\mathcal{G}}^{-1/2}\mathrm{d}\mathcal{S}(\mathbf{c}) - \frac{1}{2}\mathbf{L}_{\mathcal{G}}^{-\top} \begin{bmatrix} \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(1,i)}^\top \right) \\ \vdots \\ \mathbf{x}^\top \left(\sum_{i=1}^K y_i \mathbf{\Gamma}_{(K,i)}^\top \right) \end{bmatrix} \quad (4.98)$$

と計算できる.

さらに、この擬似ヤコビ行列を用いた勾配計算においては、 $\nabla \mathcal{L}(\mathbf{x}) = 2(\mathbf{J}_f^p)^\top \mathbf{f}(\mathbf{x})$ が成立し、目的関数が、式 (4.56) のように $\mathbf{h}^\top(\mathbf{x})\mathcal{G}^{-1}(\mathbf{x})\mathbf{h}(\mathbf{x})$ の形を持つ場合には、この擬似ヤコビ行列が通常のヤコビ行列と同じ停留点を与え、ヘッセ行列に対する近似も同様に与えることが [98] で示されている.

4.4.3 勾配およびヤコビアン の計算効率化に向けた構造の活用

前節で導出した勾配及びヤコビアンは、構造化行列の構造を明示的に記述することで、計算手順の具体化および計算量の評価が可能となる. 実際の問題に現れる構造化行列は多種多様であるが、それら全てに対して個別に勾配及びヤコビ行列を計算する必要があるわけではない. この主張は次の定理によって支持される.

定理 36 ([21, Lemma 3]). 行列構造 $\mathcal{S}' : \mathbb{R}^{n_c} \rightarrow \mathbb{R}^{K \times L'}$ に対して、任意の行列 $\Xi \in \mathbb{R}^{L' \times L}$ が $\text{rank } \Xi = L \leq L'$ を満たすとする. このとき、

$$\mathcal{S}(\mathbf{c}) = \mathcal{S}'(\mathbf{c})\Xi \quad (4.99)$$

と表される新たな構造 $\mathcal{S} : \mathbb{R}^{n_c} \rightarrow \mathbb{R}^{K \times L}$ に対して、

$$\mathcal{L}_{\mathcal{S}'}(\Xi \mathbf{x}) = \mathcal{L}_{\mathcal{S}}(\mathbf{x}) \quad (4.100)$$

が成り立つ. ここで $\mathcal{L}_{\mathcal{S}'}$ および $\mathcal{L}_{\mathcal{S}}$ はそれぞれ対応する構造に基づく SLRA 問題 (4.12) の目的関数である.

この定理は、ある行列構造に対して定められる SLRA 問題を、それと特定の線形関係を持つ別の構造に帰着して解くことが可能であることを意味する。特に、実用上よく現れる多くの Affine 構造化行列 $\mathcal{S}(\mathbf{c})$ は、モザイク型 Hankel 行列 $\mathcal{H}_{m,n}(\mathbf{c})$ と式 (4.100) の関係にあることが知られている。例えば、Toeplitz 行列や Sylvester 行列がその一例である [2]。

この関係性を活用し、文献 [21] ではモザイク型 Hankel 行列 $\mathcal{H}_{m,n}(\mathbf{c})$ に対する目的関数およびその勾配、最適化変数のヤコビアン の計算方法が詳細に述べられている。また、同文献では、それらの評価が $O(KL^2)$ で実行可能であることが示された。ただし、同文献では本論文とは異なり、横長の構造化行列を対象としているため、本論文と合わせて議論する際には、行列及び関連する式を転置したものを考える必要があることを付記する。

4.5 零空間表現以外の低ランク性の記述法

これまでに、SLRA 問題に対する定式化として、零空間表現に基づく方法を紹介した。しかし、低ランク性を表現する手法はこれに限らず、他にも像空間表現や、核ノルム最小化による凸緩和など、異なる特徴を持つ定式化が提案されている。本節では、これら零空間表現以外の代表的なアプローチについて概説する。

4.5.1 像空間表現法 [3]

文献 [3] では、次の目的関数を最小化する手法が提案された。

$$\min_{\mathbf{Q}, \mathbf{P}} \|\pi_{\mathcal{S}}(\mathbf{QP}) - \mathcal{S}(\mathbf{c})\|_{\mathbf{W}}^2 + \lambda \|\mathbf{QP} - \pi_{\mathcal{S}}(\mathbf{QP})\|_{\mathbf{F}}^2, \quad (4.101)$$

ここで、 $\pi_{\mathcal{S}}(\mathbf{D})$ は行列 $\mathbf{D}$ を写像 $\mathcal{S}$ の値域の要素のうち最も近いものへと移す写像であり、次の補題 37 で与えられる。また、第 2 項は構造を持つ行列との距離をペナルティとして課す項であり、 λ は小さい値から大きい値へと適応的に増加可能である。 λ が固定された場合、非構造化の場合と同様に、目的関数は交互最小二乗法により最小化される。この定式化において、目的関数の第一項は前節で扱った定式化とは異なり、行列のノルムであることに注意が必要である。

補題 37 (構造化行列空間への直交射影 [3]). 仮定 31 を満たす写像 $\mathbf{S}$ が所与であるとする. その像空間 $\text{Im}(\mathbf{S})$ への行列 $\mathbf{D}$ の直交射影 $\pi_{\mathbf{S}}(\mathbf{D})$ は,

$$\pi_{\mathbf{S}}(\mathbf{D}) := \mathbf{S}(\mathbf{S}^\dagger \text{vec}(\mathbf{D})) \quad (4.102)$$

で与えられる.

補題 37 の証明. $\text{Im}(\mathbf{S})$ に属する行列のうち, 非構造化行列 $\mathbf{D}$ に Frobenius ノルムの意味で最も近いものは, 次の最小化問題の解として得られる:

$$\min_{\hat{\mathbf{c}} \in \mathbb{R}^{np}} \|\mathbf{D} - \mathbf{S}(\hat{\mathbf{c}})\|_{\text{F}}^2. \quad (4.103)$$

これはさらに, vec 作用素を用いて以下のように表せる:

$$\min_{\hat{\mathbf{c}} \in \mathbb{R}^{np}} \|\text{vec}(\mathbf{D}) - (\text{vec}(\mathbf{S}_0) + \mathbf{S}\hat{\mathbf{c}})\|_2^2. \quad (4.104)$$

仮定 31 より, $\mathbf{S}$ の列ベクトルたちは互いに線形独立であることにより $\mathbf{S}^\top \text{vec}(\mathbf{S}_0) = \mathbf{0}$ であるから, 式 (4.104) は解が一意に定まる最小二乗問題となる. この解は,

$$\hat{\mathbf{c}}^* = (\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top (\text{vec}(\mathbf{D}) - \text{vec}(\mathbf{S}_0)) \quad (4.105)$$

$$= (\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top \text{vec}(\mathbf{D}) \quad (4.106)$$

$$= \mathbf{S}^\dagger \text{vec}(\mathbf{D}) \quad (4.107)$$

と表される. したがって,

$$\pi_{\mathbf{S}}(\mathbf{D}) = \mathbf{S}(\mathbf{S}^\dagger \text{vec}(\mathbf{D})) \quad (4.108)$$

となり, 補題の主張が示された. $\square$

直交射影 $\pi_{\mathbf{S}}$ に関して, 行列構造の定義 4.5 及び式 (4.4) より,

$$\begin{aligned} \text{vec } \pi_{\mathbf{S}}(\mathbf{D}) &= \text{vec}(\mathbf{S}(\mathbf{S}^\dagger \text{vec}(\mathbf{D}))) \\ &= \text{vec}(\mathbf{S}^{(0)}) + \mathbf{S}\mathbf{S}^\dagger \text{vec}(\mathbf{D}) \end{aligned}$$

であるため, 次の等式が成立する:

$$\text{vec } \pi_{\mathbf{S}}(\mathbf{D}) = \text{vec}(\mathbf{S}^{(0)}) + \Pi_{\mathbf{S}} \text{vec } \mathbf{D}. \quad (4.109)$$

ここで, $\Pi_S = SS^\dagger = S(S^\top S)^{-1}S^\top$ は, S の像に対する直交射影を与える行列である. 式 (4.109) はのちに補題の証明に用いる.

注 38 (直交射影の解釈). ベクトル化された $K \times L$ 行列 $\mathbf{D}$ に対して $S^\dagger$ を適用する操作は, 構造パラメータベクトルを生成するものであり, それは $\mathbf{D}$ の中で同じ $S^{(k)}$ に対応する要素の平均をとることに対応する. 実際, $S^\top \text{vec}(\mathbf{D})$ は, 各 S_k に対応する要素の総和を含むベクトルである. また, 仮定 31 より, $S^\top S$ は対角行列であり, その対角成分はそれぞれの S_k に含まれる非零要素の個数と一致する. すなわち,

$$S^\top S = \begin{bmatrix} \|S_1\|_F^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \|S_{n_p}\|_F^2 \end{bmatrix} = \begin{bmatrix} \#nz(S_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \#nz(S_{n_p}) \end{bmatrix}, \quad (4.110)$$

ただし, $\#nz$ は非零要素数を表す. したがって, $(S^\top S)^{-1}S^\top$ を掛ける操作は, 構造パラメータの各要素に対応する行列成分の平均をとることに相当する. 特に, 構造化行列をベクトル化したものに対して $S^\dagger$ を適用すると, その構造パラメータが抽出される. すなわち,

$$S^\dagger \text{vec}(\mathcal{S}(\mathbf{c})) = (S^\top S)^{-1}(S^\top S)\mathbf{c} = \mathbf{c}. \quad (4.111)$$

◇

式 (4.101) の目的関数第一項 $\|\pi_S(\mathbf{QP}) - \mathcal{S}(\mathbf{c})\|_W^2$ を行列ノルムからベクトルノルムに書き換える.

$$\begin{aligned} \text{vec}(\mathcal{S}(\mathbf{c}) - \pi_S(\mathbf{QP})) &= S\mathbf{c} + \text{vec}(S^{(0)}) - (\Pi_S \text{vec}(\mathbf{QP}) + \text{vec}(S^{(0)})) \\ &= S\mathbf{c} - SS^\dagger \text{vec}(\mathbf{QP}) \end{aligned} \quad (4.112)$$

であるため,

$$\begin{aligned} \|\mathcal{S}(\mathbf{c}) - \pi_S(\mathbf{QP})\|_W^2 &= (S\mathbf{c} - SS^\dagger \text{vec}(\mathbf{QP}))^\top W (S\mathbf{c} - SS^\dagger \text{vec}(\mathbf{QP})) \\ &= (\mathbf{c} - S^\dagger \text{vec}(\mathbf{QP}))^\top S^\top W S (\mathbf{c} - S^\dagger \text{vec}(\mathbf{QP})) \\ &= \|\mathbf{c} - S^\dagger \text{vec}(\mathbf{QP})\|_W^2 \end{aligned} \quad (4.113)$$

と変形でき, (4.101) と同値な最小化問題として

$$\min_{\mathbf{Q}, \mathbf{P}} \|\mathbf{c} - S^\dagger \text{vec}(\mathbf{QP})\|_W^2 + \lambda \|\mathbf{QP} - \pi_S(\mathbf{QP})\|_F^2, \quad (4.114)$$

を得る．ここで

$$\overline{\mathbf{W}} := \mathbf{S}^\top \mathbf{W} \mathbf{S} \quad (4.115)$$

と定めた．以下での利用を想定し， $\overline{\mathbf{W}}$ に対する最大階数分解を $\overline{\mathbf{W}} = \mathbf{L}_{\overline{\mathbf{W}}}^\top \mathbf{L}_{\overline{\mathbf{W}}}$ と表す．

アルゴリズム

本手法の基本的な考え方は，非構造化行列に対する交互射影を用いた低ランク近似手法と同様に，最小化問題 (4.114) を $\mathbf{Q}$ を固定した下での $\mathbf{P}$ に関する最小化問題

$$\min_{\mathbf{P}} \|\mathbf{c} - \mathbf{S}^\dagger \text{vec}(\mathbf{Q}\mathbf{P})\|_{\overline{\mathbf{W}}}^2 + \lambda \|\mathbf{Q}\mathbf{P} - \pi_{\mathbf{S}}(\mathbf{Q}\mathbf{P})\|_{\mathbf{F}}^2 \quad (4.116)$$

及び， $\mathbf{P}$ を固定した下での $\mathbf{Q}$ に関する最小化問題

$$\min_{\mathbf{Q}} \|\mathbf{c} - \mathbf{S}^\dagger \text{vec}(\mathbf{Q}\mathbf{P})\|_{\overline{\mathbf{W}}}^2 + \lambda \|\mathbf{Q}\mathbf{P} - \pi_{\mathbf{S}}(\mathbf{Q}\mathbf{P})\|_{\mathbf{F}}^2 \quad (4.117)$$

を交互に解くことで近似解の更新を繰り返すことに基づく．

補題 39. 問題 (4.116) および (4.117) は，以下のような最小二乗問題と等価である：

$$(4.116) \Leftrightarrow \min_{\mathbf{P}} \left\| \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{S}^\dagger \\ \sqrt{\lambda} \mathbf{\Pi}_{\mathbf{S}_\perp} \end{bmatrix} (\mathbf{I}_L \otimes \mathbf{Q}) \text{vec}(\mathbf{P}) - \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{c} \\ \sqrt{\lambda} \text{vec}(\mathbf{S}_0) \end{bmatrix} \right\|_2^2, \quad (4.118)$$

$$(4.117) \Leftrightarrow \min_{\mathbf{Q}} \left\| \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{S}^\dagger \\ \sqrt{\lambda} \mathbf{\Pi}_{\mathbf{S}_\perp} \end{bmatrix} (\mathbf{P}^\top \otimes \mathbf{I}_K) \text{vec}(\mathbf{Q}) - \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{c} \\ \sqrt{\lambda} \text{vec}(\mathbf{S}_0) \end{bmatrix} \right\|_2^2,$$

ただし， $\mathbf{\Pi}_{\mathbf{S}_\perp}$ は $\mathbf{S}^\top$ の零空間に対する直交射影子を表す．

補題 39 の証明. 恒等式 (3.19) を用いると，次の関係が得られる：

$$\text{vec}(\mathbf{Q}\mathbf{P}) = (\mathbf{I}_L \otimes \mathbf{Q}) \text{vec}(\mathbf{P}) = (\mathbf{P}^\top \otimes \mathbf{I}_K) \text{vec}(\mathbf{Q}). \quad (4.119)$$

まず，問題 (4.116) を考える．式 (4.109) および (4.119) を用いると，式 (4.116) で定

められる最小化問題は次のように書き換えられる：

$$\begin{aligned}
 & \min_{\mathbf{P}} \|\mathbf{c} - \mathbf{S}^\dagger \text{vec}(\mathbf{QP})\|_{\overline{\mathbf{W}}}^2 + \lambda \|\mathbf{QP} - \pi_{\mathbf{S}}(\mathbf{QP})\|_{\mathbf{F}}^2 \\
 \iff & \min_{\mathbf{P}} \|\mathbf{L}_{\overline{\mathbf{W}}}(\mathbf{c} - \mathbf{S}^\dagger \text{vec}(\mathbf{QP}))\|_2^2 + \lambda \|\text{vec}(\mathbf{QP}) - \text{vec}(\pi_{\mathbf{S}}(\mathbf{QP}))\|_2^2, \\
 \iff & \min_{\mathbf{P}} \|\mathbf{L}_{\overline{\mathbf{W}}}(\mathbf{c} - \mathbf{S}^\dagger \text{vec}(\mathbf{QP}))\|_2^2 + \lambda \|\text{vec}(\mathbf{QP}) - \text{vec}(\mathbf{S}^{(0)}) - \mathbf{\Pi}_{\mathbf{S}} \text{vec}(\mathbf{QP})\|_2^2, \\
 \iff & \min_{\mathbf{P}} \|\mathbf{L}_{\overline{\mathbf{W}}} \mathbf{S}^\dagger \text{vec}(\mathbf{QP}) - \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{c}\|_2^2 + \|\sqrt{\lambda} \mathbf{\Pi}_{\mathbf{S}_\perp} \text{vec}(\mathbf{QP}) - \sqrt{\lambda} \text{vec}(\mathbf{S}^{(0)})\|_2^2, \\
 \iff & \min_{\mathbf{P}} \left\| \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{S}^\dagger \\ \sqrt{\lambda} \mathbf{\Pi}_{\mathbf{S}_\perp} \end{bmatrix} \text{vec}(\mathbf{QP}) - \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{c} \\ \sqrt{\lambda} \text{vec}(\mathbf{S}^{(0)}) \end{bmatrix} \right\|_2^2, \\
 \iff & \min_{\mathbf{P}} \left\| \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{S}^\dagger \\ \sqrt{\lambda} \mathbf{\Pi}_{\mathbf{S}_\perp} \end{bmatrix} (\mathbf{I}_n \otimes \mathbf{Q}) \text{vec}(\mathbf{P}) - \begin{bmatrix} \mathbf{L}_{\overline{\mathbf{W}}} \mathbf{c} \\ \sqrt{\lambda} \text{vec}(\mathbf{S}^{(0)}) \end{bmatrix} \right\|_2^2.
 \end{aligned}$$

$\mathbf{Q}$ に関する導出も同様に行うことができる。 $\square$

補題 39 における両方の問題は、いずれも最小二乗問題であり、解析解が閉形式で表される。ここでは、固定された λ に対して、交互最小二乗法およびブロック座標降下法 (block coordinate descent) の枠組みに基づくアルゴリズムを紹介する。すなわち、式 (4.118) に示された最小二乗問題を順次解くことで、 $\mathbf{P}$ および $\mathbf{Q}$ の近似解を交互に改善する。アルゴリズムの概要を、Algorithm 8 に示す。

Algorithm 8 行列因子分解による構造化低ランク近似

: Structured low-rank approximation by factorization

Input: $\mathbf{c} \in \mathbb{R}^{n_p}$, $\mathbf{S}^{(0)} \in \mathbb{R}^{K \times L}$, $\mathbf{S} \in \mathbb{R}^{(KL) \times n_p}$, $\overline{\mathbf{W}} = \mathbf{L}_{\overline{\mathbf{W}}}^\top \mathbf{L}_{\overline{\mathbf{W}}} \in \mathbb{R}^{n_p \times n_p}$, $r \in \mathbb{N}$, $\mathbf{Q}_0 \in \mathbb{R}^{K \times r}$.

Output: 問題 (4.114) に対する局所最適解 $\mathbf{Q} \in \mathbb{R}^{K \times r}$ 及び $\mathbf{P} \in \mathbb{R}^{r \times L}$.

- 1: 次を計算: $\mathbf{S}^\dagger = (\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top$.
 - 2: 初期化: $\mathbf{Q} = \mathbf{Q}_0$, $\lambda_1 = 1$, $j = 0$.
 - 3: **repeat**
 - 4: $j = j + 1$, $k = 0$
 - 5: **repeat**
 - 6: $k = k + 1$
 - 7: 部分問題 (4.116) の求解により $\mathbf{P}$ を更新.
 - 8: 部分問題 (4.117) の求解により $\mathbf{Q}$ を更新.
 - 9: **until** (停止基準を満たすまで)
 - 10: ペナルティ係数 λ_{j+1} の値を $\lambda_{j+1} > \lambda_j$ となるように更新.
 - 11: **until** (停止基準を満たすまで)
-

派生手法

このアプローチの拡張として、Lagrange の未定乗数法に基づく手法が [99] で提案されている。また、複数のランク制約を扱うために、近接法（Proximal Methods）を用いた拡張が [100] により提案されている。一般的に、ランク r の行列の集合は多様体を形成するが、この多様体上での最適化手法は、式 (4.101) ように定式化された SLRA 問題に適用可能である。

4.5.2 核ノルム緩和法

低ランク制約は特別な場合を除いて非凸な制約であり、最適化を難しくしている要因の一つである。これを核ノルムの制約に置き換えることで、半正定値計画（Semi-Definite Program）問題へと変換することが可能である¹。文献 [101, 92] では、この考えに基づいた SLRA 手法が提案されている。半正定値計画問題は、既存のアルゴリズムによって収束が保証された形で解くことができるため、高品質なソフトウェア実装が容易である。特に CVX [102] パッケージがツールとして知られている。また、核ノルム緩和法の利点の一つとして、問題設定の柔軟性が挙げられる。たとえば、問題に Affine 不等式制約を追加しても、問題の凸性は保持される。そのため、現実のさまざまな制約条件を考慮しやすい。しかし、核ノルム緩和法を用いる際の欠点として、最適化変数の数がデータ点の数の 2 乗に比例して増加する点がある。そのため、この手法は（パラメータベクトルの次元 n_p が 1000 を超えるような）大規模な問題には適用が困難である。このように、小規模な問題に適した手法であることには注意が必要である。

正則化項つき核ノルム最小化

定義 40. 行列 $\mathbf{M}$ の核ノルムは、その特異値の総和として定義される：

$$\|\mathbf{M}\|_* = \sum_i \sigma_i(\mathbf{M}) \quad (4.120)$$

ここで、 $\sigma_i(\mathbf{M})$ は $\mathbf{M}$ の i 番目に小さい特異値である。

¹対象の行列が正定値対称であればトレース、対称行列に対しては固有値の絶対和を用いる。

核ノルムを最小化することで、解の階数が低くなるように誘導可能であり、核ノルムに関する閾値 μ を適切に与えることで、ランク制約 $\text{rank } \mathbf{S}(\hat{\mathbf{c}}) \leq r$ に近似可能な核ノルム制約 $\|\mathbf{S}(\hat{\mathbf{c}})\|_* \leq \mu$ を得る。この関係を用いることで、SLRA 問題²

$$\begin{aligned} & \underset{\hat{\mathbf{c}} \in \mathbb{R}^{nz}}{\text{minimize}} && \|\hat{\mathbf{c}} - \mathbf{c}\|_2 \\ & \text{subject to} && \text{rank } \mathbf{S}(\hat{\mathbf{c}}) \leq r \end{aligned} \quad (4.121)$$

を緩和（核ノルム緩和）した形式を得る：

$$\begin{aligned} & \underset{\hat{\mathbf{c}}}{\text{minimize}} && \|\mathbf{c} - \hat{\mathbf{c}}\|_2 \\ & \text{subject to} && \|\mathbf{S}(\hat{\mathbf{c}})\|_* \leq \mu. \end{aligned} \quad (4.122)$$

核ノルムによるランク制約の緩和は、元の制約の凸緩和のうち最も厳密な凸緩和であることが示されており [101, Theorem 1]、理論的な正当性がある。ここで新たな制約なし最適化問題

$$\underset{\hat{\mathbf{c}}}{\text{minimize}} \|\mathbf{c} - \hat{\mathbf{c}}\|_2 + \gamma \|\mathbf{S}(\hat{\mathbf{c}})\|_*. \quad (4.123)$$

を考えると、これは正則化された核ノルム最小化問題であり、適切な γ を選ぶことで (4.122) で定められる問題の解と一致する。なお、式 (4.123) は次の最適化問題と等価である [23]:

$$\underset{\hat{\mathbf{c}}}{\text{minimize}} \|\mathbf{S}(\hat{\mathbf{c}})\|_* + \gamma' \|\mathbf{c} - \hat{\mathbf{c}}\|_2. \quad (4.124)$$

これは SLRA 問題の目的関数と制約条件を入れ替えた問題である、構造化行列のランク最小化問題に対する核ノルム緩和問題 [103, 23, 90, 18, 101]

$$\begin{aligned} & \underset{\hat{\mathbf{c}}}{\text{minimize}} && \|\mathbf{S}(\hat{\mathbf{c}})\|_* \\ & \text{subject to} && \|\mathbf{c} - \hat{\mathbf{c}}\|_2 \leq \tau. \end{aligned} \quad (4.125)$$

に対し正則化を適用した最適化と解釈できる。

上述のように、核ノルム緩和を用いることの利点の一つに、Affine 不等式制約の追加に対する柔軟性があった。例えば、問題 (4.124) に対して新たな線形不等式制約 $\mathbf{A}\hat{\mathbf{c}} \leq \mathbf{b}$

²本節では先行研究の記述に倣い、目的関数は単に 2 ノルムとし、また二乗ではなく一乗を考える。これは本章の冒頭で定めたものと異なっていることに注意が必要である

を追加した，次の最適化問題を考える：

$$\begin{aligned}
 & \underset{\hat{\mathbf{c}}}{\text{minimize}} && \|\mathbf{c} - \hat{\mathbf{c}}\|_2 \\
 & \text{subject to} && \|\mathcal{S}(\hat{\mathbf{c}})\|_* \leq \mu, \\
 & && \mathbf{A}\hat{\mathbf{c}} \leq \mathbf{b}.
 \end{aligned} \tag{4.126}$$

核ノルム最小化問題と半正定値最小化問題の等価性

一般の行列 $\mathbf{D}$ に対する核ノルム制約 $\|\mathbf{D}\|_* \leq \mu$ は，次の補題が示すように，半正定値制約に書き換えられる [101, Chapter 5].

補題 41 ([101, Lemma 2]). 行列 $\mathbf{D} \in \mathbb{R}^{K \times L}$ と $\mu \in \mathbb{R}$ に対して， $\|\mathbf{D}\|_* \leq \mu$ であることと，

$$\frac{1}{2}(\text{tr}(\mathbf{E}) + \text{tr}(\mathbf{F})) \leq \mu, \tag{4.127}$$

$$\begin{bmatrix} \mathbf{E} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{F} \end{bmatrix} \succeq 0 \tag{4.128}$$

なる二つの半正定値行列 $\mathbf{E} \in \mathbb{R}^{K \times K}$ 及び $\mathbf{F} \in \mathbb{R}^{L \times L}$ が存在することは同値である．ここで $\mathbf{M} \succeq 0$ は，行列 $\mathbf{M}$ が半正定値行列であることを表す．

これにより，式 (4.126) で定められる最適化問題を次のような半正定値計画問題として解くことができる：

$$\begin{aligned}
 & \underset{\hat{\mathbf{c}}}{\text{minimize}} && \|\mathbf{c} - \hat{\mathbf{c}}\|_2 \\
 & \text{subject to} && \begin{bmatrix} \mathbf{E} & \mathcal{S}^\top(\hat{\mathbf{c}}) \\ \mathcal{S}(\hat{\mathbf{c}}) & \mathbf{F} \end{bmatrix} \succeq 0, \\
 & && \frac{1}{2}(\text{tr} \mathbf{E} + \text{tr} \mathbf{F}) \leq \mu, \\
 & && \mathbf{A}\hat{\mathbf{c}} \leq \mathbf{b}.
 \end{aligned} \tag{4.129}$$

半正定値計画問題は，凸最適化のための計算パッケージである CVX [102] を使用することで効率よく求解可能である．

一方，問題のサイズが増大すると半正定値行列のサイズが増大した場合には，計算量が極端に増大してしまうため，半正定値最適化問題として扱うことも不適当となる．な

お、双対な関係にある構造化行列のランク最小化問題に対する核ノルム緩和問題 (4.125) に対しては、内点法を用いた研究や、交互乗数法 (Alternating Direction Method of Multipliers: ADMM) を用いた研究例が報告されている [90].

さらに対象を限定した場合の手法として、特定の目的関数である場合に、その性質を考慮した凸包に基づく、より厳密な緩和手法も提案されている [104, 105, 106], [107, §VI].

補題 41 の証明 [101]. (十分性の証明) $\|\mathbf{D}\|_* \leq \mu$ を仮定する. $\mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{D}$ ($\mathbf{U} \in \mathbb{R}^{K \times (K-r)}$, $\Sigma \in \mathbb{R}^{r \times r}$, $\mathbf{V} \in \mathbb{R}^{L \times (L-r)}$) を $\mathbf{D}$ の SVD とする. $\mathbf{E} = \mathbf{U}\Sigma\mathbf{U}^\top + \gamma\mathbf{I}_K$, $\mathbf{F} = \mathbf{V}\Sigma\mathbf{V}^\top + \gamma\mathbf{I}_L$ とおくと, $\gamma > 0$ に対してこれら二つの行列は明らかに半正定値行列であり,

$$\text{tr } \mathbf{E} + \text{tr } \mathbf{F} = 2 \text{tr } \Sigma + \gamma(K + L) = 2\|\mathbf{D}\|_* + \gamma(K + L) \quad (4.130)$$

が従う. 不等式 (4.127) の成立を示すためには, このように定めた $\mathbf{E}, \mathbf{F}$ に対して

$$\frac{1}{2}(\text{tr } \mathbf{E} + \text{tr } \mathbf{F}) - \mu \leq 0 \quad (4.131)$$

が成立する $\gamma > 0$ が存在することを示す必要がある. 不等式 (4.131) の左辺に式 (4.130) を代入することで

$$\frac{1}{2}(\text{tr } \mathbf{E} + \text{tr } \mathbf{F}) - \mu = \|\mathbf{D}\|_* + \frac{\gamma}{2}(K + L) - \mu \quad (4.132)$$

を得る. 仮定より $\mu - \|\mathbf{D}\|_* \geq 0$ であるため,

$$\gamma \leq \frac{2(\mu - \|\mathbf{D}\|_*)}{K + L} \quad (4.133)$$

を満たす γ を採用すれば, 不等式 (4.131) が成立する. 続いて式 (4.128) の左辺について

$$\begin{bmatrix} \mathbf{E} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{U}\Sigma\mathbf{U}^\top & \mathbf{V}\Sigma\mathbf{U}^\top \\ \mathbf{U}\Sigma\mathbf{V}^\top & \mathbf{V}\Sigma\mathbf{V}^\top \end{bmatrix} + \gamma \begin{bmatrix} \mathbf{I}_K & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_L \end{bmatrix} \quad (4.134)$$

$$= \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \Sigma \begin{bmatrix} \mathbf{U}^\top & \mathbf{V}^\top \end{bmatrix} + \gamma\mathbf{I}_{K+L} \quad (4.135)$$

となるため, 式 (4.128) が成立する.

(必要性の証明) $\mathbf{E}$ 及び $\mathbf{F}$ が不等式 (4.127) 及び式 (4.128) の条件を満たすことを仮定する. 新たに

$$\begin{bmatrix} \mathbf{U}\mathbf{U}^\top & -\mathbf{U}\mathbf{V}^\top \\ -\mathbf{V}\mathbf{U}^\top & \mathbf{V}\mathbf{V}^\top \end{bmatrix}$$

なる行列を考えると, この行列が半正定値であることは簡単な計算により確認できる. 半正定値行列同士の積のトレースは非負であるため,

$$\text{tr} \left(\begin{bmatrix} \mathbf{U}\mathbf{U}^\top & -\mathbf{U}\mathbf{V}^\top \\ -\mathbf{V}\mathbf{U}^\top & \mathbf{V}\mathbf{V}^\top \end{bmatrix} \begin{bmatrix} \mathbf{E} & \mathbf{D}^\top \\ \mathbf{D} & \mathbf{F} \end{bmatrix} \right) \geq 0 \quad (4.136)$$

である. この不等式の左辺を展開することで

$$\text{tr} \mathbf{U}\mathbf{U}^\top \mathbf{E} - \text{tr} \mathbf{U}\mathbf{V}^\top \mathbf{D} - \text{tr} \mathbf{V}\mathbf{U}^\top \mathbf{D}^\top + \text{tr} \mathbf{V}\mathbf{V}^\top \mathbf{F} \geq 0 \quad (4.137)$$

を得る. $\mathbf{U}$ は部分直交行列であるため, これを拡張することで直交行列を得る, すなわち

$$\exists \tilde{\mathbf{U}} \text{ such that } [\mathbf{U} \tilde{\mathbf{U}}][\mathbf{U} \tilde{\mathbf{U}}]^\top = \mathbf{U}\mathbf{U}^\top + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top = \mathbf{I}_K. \quad (4.138)$$

また同様に,

$$\exists \tilde{\mathbf{V}} \text{ such that } [\mathbf{V} \tilde{\mathbf{V}}][\mathbf{V} \tilde{\mathbf{V}}]^\top = \mathbf{V}\mathbf{V}^\top + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top = \mathbf{I}_L \quad (4.139)$$

である. 仮定より $\text{tr} \mathbf{E} \geq 0$ であり, 加えて $\text{tr} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top \geq 0$ であることから $\text{tr} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top \mathbf{E} \geq 0$ が従い,

$$\text{tr} \mathbf{U}\mathbf{U}^\top \mathbf{E} \leq \text{tr} \mathbf{U}\mathbf{U}^\top \mathbf{E} + \text{tr} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top \mathbf{E} = \text{tr} \mathbf{E} \quad (4.140)$$

なる不等式を得る. 同様に $\text{tr} \mathbf{V}\mathbf{V}^\top \mathbf{F} \leq \text{tr} \mathbf{F}$ も成立する. さらに, $\text{tr} \mathbf{V}\mathbf{U}^\top \mathbf{D} = \text{tr} \mathbf{V}\Sigma\mathbf{V}^\top = \text{tr} \Sigma$, $\text{tr} \mathbf{U}\mathbf{V}^\top \mathbf{D}^\top = \text{tr} \mathbf{U}\Sigma\mathbf{U}^\top = \text{tr} \Sigma$ であることを踏まえると, 式 (4.137) より $\text{tr} \mathbf{E} + \text{tr} \mathbf{F} - 2\text{tr} \Sigma \geq 0$ が成立し, 仮定している不等式 (4.127) より

$$\text{tr} \Sigma = \|\mathbf{D}\|_* \leq \mu \quad (4.141)$$

が帰結し, 題意が成立する. □

4.5.3 常微分方程式の活用

近年, 一連の研究 [108, 6, 109] により, SLRA の解を, 固有値または特異値に関する微分方程式の解として求めることが可能であることが示された. このため, 対応す

る常微分方程式 (Ordinary Differential Equation: ODE) を積分することで SLRA を解く手法が提案されている。このアプローチは、ある種の行列構造（例えば、Sylvester 行列 [108, 6]）に対しては非常に良好に機能するが、Hankel 行列の場合の挙動は完全には解明されておらず、アルゴリズムに対する適宜の調整が必要とされる [109]。

4.5.4 IRLS を適用する手法

階数の代替指標（サロゲート）の考えに基づき、IRLS (Iteratively Reweighted Least Squares) による最適化を行う手法も提案されており、これらの手法は局所収束性や理論的保証を備えている [110, 111]。また、これらの手法は高速に実装可能である点が優れている [112]。しかしながら、著者の知る限り、この分野においてこの手法は主要な方法論とは認識されていないのが現状である。

第5章 STLN-SLRAにおける初期方程式の選択方法

第1章で述べたように、零空間表現に基づく SLRA 問題では、以下の近似式を考える:

$$\mathbf{S}(\mathbf{c})\mathbf{x} \approx \mathbf{0}. \quad (5.1)$$

この式は、修正項 $\Delta\mathbf{c}$ を含む厳密な等式 $\mathbf{S}(\mathbf{c} + \Delta\mathbf{c})\mathbf{x} = \mathbf{0}$ において、 $\Delta\mathbf{c} = \mathbf{0}$ を代入したものに由来する。この近似を用いることで、SLRA 問題を解くための数学的な枠組みを定めることが可能となる。

STLN に基づく SLRA では、第4.3節でも述べたように、低ランク性の制約式 (5.1) の代わりに次の式（本論文では、第 i 初期方程式と呼称）を用いる。

第 i 初期方程式

$$\mathbf{S}_{(i)}\mathbf{x}_{(i)} = \mathbf{s}_i. \quad (5.2)$$

ここで、 $\mathbf{S}_{(i)} := \mathbf{S}(\mathbf{c})\tilde{\mathbf{I}}_i$, $\mathbf{s}_i := -\mathbf{S}(\mathbf{c})\mathbf{e}_i$ である。

歴史的な経緯を振り返ると、STLN に基づく SLRA（以下、STLN-SLRA）が提案された当初は、主に最初または最後の初期方程式が用いられていた [63, 7]。実際、これらが適切となる状況も存在するが、限定的である。文献 [25] では、初期方程式の選択が STLN-SLRA の収束速度や近似精度に影響することが示されており、異なる初期方程式を選択することで STLN-SLRA の性能向上が期待できる。したがって、初期方程式の選択は、STLN-SLRA の実用的な応用において重要な要素である。しかし、文献 [25] で提案された初期方程式の選択基準は必要な計算量が大きく、 $\mathbf{S}(\mathbf{c})$ の列数 L に対して $O(L^4)$ を要し、理論的妥当性の検討も不十分である。本章¹では元の問題と初期方程式との関係を理論的に解析し、 $O(L^3)$ の計算量で実現可能な新たな選択基準を提案する。さらに、Sylvester 行列及び Hankel 行列に対する数値実験を通じて、提案する選

¹本章の内容は、著者らによる研究 [4] (Copyright(c)2025 Elsevir B.V.) に基づいて記述されている。

択基準が初期方程式の選択にかかる時間のみならず，STLN-SLRA 全体の計算時間の短縮にも寄与することを確認する．

5.1 STLN-SLRA における問題点

本節では，ランク制約を表す元々の式から初期方程式への変形の過程（式 (4.15)–(4.17)）に数値誤差を生み出す要因があることと，さらにその数値誤差が選択する初期方程式によって変化することを具体例を用いて確認する²．

5.1.1 $\mathcal{S}(c)$ が雑音を含まない場合の各列ベクトルの線形従属性

SLRA 問題 (4.12) において

$$\mathcal{S}(c)x = 0 \quad (5.3)$$

を厳密に満たす非零ベクトル x の存在を仮定すると，その最適解は明らかに $\Delta c = 0$ であり，式 (5.3) は問題 (4.12) の特殊な形式であると言える．式 (5.3) に対して式 (4.15) から (4.17) までと同様の変形を行い，式 (5.2) の形式で解いた際の計算誤差を比較する．

例 3 (x に零要素が含まれる場合)．はじめに，選択すべきでない初期方程式について説明する．これは，第 4.3 節で説明した通り， $x_i = 0$ であるような第 i 初期方程式である．実際

$$\begin{aligned} \mathcal{S}(c)x = 0 &\iff x_1 s_1 + \cdots x_{i-1} s_{i-1} + x_i s_i + x_{i+1} s_{i+1} + \cdots + x_L s_L = 0 \\ &\iff x_1 s_1 + \cdots x_{i-1} s_{i-1} + x_{i+1} s_{i+1} + \cdots + x_L s_L = 0 \end{aligned} \quad (5.4)$$

となるため，式 (5.3) からは s_i が $\mathcal{S}(c)$ の他の列ベクトルたちの張る線形部分空間に属するかどうかを判別できず，第 i 初期方程式 $\mathcal{S}_{(i)} x_{(i)} = s_i$ に解が存在する保証がない．

より具体的な例として，二つの多項式の組

$$\begin{aligned} p^{(1)}(t) &= (t+1)(t-1)(t-2) = t^3 - 2t^2 - 1t + 2 \\ p^{(2)}(t) &= (t+0)(t-1)(t-2) = t^3 - 3t^2 + 2t + 0 \end{aligned}$$

²本節は，著者らによる研究 [113](Copyright(c)2022 IEICE) に基づいて記述している．

表 5.1: 線形方程式 (5.6) の解に基づく誤差の比較 ($\mathbf{x}$ に零要素が存在する場合)
: Error Comparison Based on Solutions of the Linear Equation (5.6) (with zero in $\mathbf{x}$)[113].

列番号	$\ \mathcal{S}_{(i)}\mathbf{x}_{(i)} - \mathbf{s}_i\ _2$	$\ \mathbf{x} - \phi_i(\mathbf{x}_{(i)})\ $
1	1.02×10^{-15}	2.94×10^{-16}
2	1.09×10^{-15}	2.24×10^{-16}
3	1.04×10^{-15}	3.35×10^{-16}
4	2.45	1.41

を考えると、これらの GCD は $(t-1)(t-2)$ であり、その次数は 2 である。これ故に

$$\mathcal{S}(\mathbf{c}) = \text{fSyl}_2(\mathbf{p}^{(n)}) = \begin{pmatrix} 1 & 0 & -1 & 0 \\ -3 & 1 & 2 & -1 \\ 2 & -3 & 1 & 2 \\ 0 & 2 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad (5.5)$$

に対して³ $\mathcal{S}(\mathbf{c})\mathbf{x} = \mathbf{0}$ であり、さらに $i \neq 4$ に対して $x_i \neq 0$ であるため

$$\mathcal{S}_{(i)}\mathbf{x}_{(i)} = \mathbf{s}_i \quad (5.6)$$

は解を持つ。この線形方程式の解 $\mathbf{x}_{(i)}$ から $\phi_i(\mathbf{x}_{(i)})$ を構成すると、これらは理想的にはスカラ倍を除いて $\mathbf{x}$ と一致するはずである。 $\phi_i(\mathbf{x}_{(i)})$ たちをノルムが 1 になるように正規化し、同様に正規化した真値 $\mathbf{x}$ との誤差を比較した結果を表 5.1 の第 3 列に示した。また、 $i \neq 4$ に対して $\mathbf{s}_i \in \mathcal{S}_{(i)}$ であることから、ベクトル $\mathcal{S}_{(i)}\mathbf{x}_{(i)} - \mathbf{s}_i$ も $\mathbf{0}$ になることが望ましい。この誤差についても表 5.1 の第 2 列に示した。

$\text{fSyl}_2(\mathbf{p}^{(n)})$ の第 4 列を右辺に移動した場合の誤差がどちらも 0 から大きく離れている。そのためこの例において明らかに第 4 初期方程式に対する求解は意味をなさない。誤差が大きくなる原因としては、先も述べた通り $\mathbf{x}$ の第 4 成分が 0 であり、第 4 初期方程式が解を持たないことに起因する。

例 4 ($\mathbf{x}$ に零要素が存在しない場合). 次に、 $x_i \neq 0$ であるような i が複数ある場合にも、

³第 2.2 節で定めた、複数の多項式に対する記法を使用している。

表 5.2: 線形方程式 (5.6) の解に基づく誤差の比較 ($\mathbf{x}$ に零要素が存在しない場合)
: Error Comparison Based on Solutions of the Linear Equation (5.6) (w/o zero in $\mathbf{x}$)[113].

列番号	$\ \mathcal{S}_{(i)}\mathbf{x}_{(i)} - \mathbf{s}_i\ _2$	$\ \mathbf{x} - \phi_i(\mathbf{x}_{(i)})\ $	$ x_i $
1	4.55×10^{-9}	5.06×10^{-11}	1
2	4.70×10^{-9}	1.73×10^{-11}	10^2
3	1.25×10^{-10}	3.59×10^{-13}	10^3
4	1.24×10^{-7}	1.76×10^{-12}	1
5	1.34×10^{-8}	1.47×10^{-11}	20
6	2.06×10^{-9}	1.44×10^{-12}	10^2

選択する初期方程式によって数値的に良い結果を得ることを説明する．具体例として

$$p^{(1)}(t) = (t+2)(t^2 + 100t + 1000) = t^3 + 102t^2 + 1200t + 2000$$

$$p^{(2)}(t) = (t+2)(t^2 + 20t + 100) = t^3 + 22t^2 + 140t + 200$$

を考える．これら 2 つの多項式の GCD は $(t+2)$ であり，その次数は 1 であることから，

$$\mathcal{S}(\mathbf{c}) = \text{fSyl}_1(\mathbf{p}^{(n)}) = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 22 & 1 & 0 & -102 & -1 & 0 \\ 140 & 22 & 1 & -1200 & -102 & -1 \\ 200 & 140 & 22 & -2000 & -1200 & -102 \\ 0 & 200 & 140 & 0 & -2000 & -1200 \\ 0 & 0 & 200 & 0 & 0 & -2000 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 100 \\ 1000 \\ 1 \\ 20 \\ 100 \end{pmatrix}$$

に対して $\mathcal{S}(\mathbf{c})\mathbf{x} = \mathbf{0}$ を満たす．例 3 と同様に初期方程式の求解に関連する誤差を表 5.2 に示した．なお，この例では零空間の真値との誤差および初期方程式の残差を最小とする列番号は $\mathbf{x}$ の絶対値最大の要素番号に一致している．

上記の表 5.1 及び表 5.2 の第 2 列目で評価した残差ベクトル

$$\mathcal{S}_{(i)}\mathbf{x}_{(i)} - \mathbf{s}_i = (\mathcal{S}_{(i)}\mathcal{S}_{(i)}^\dagger - \mathbf{I}_K)\mathbf{s}_i \quad (5.7)$$

は $\mathbf{s}_i$ の $\text{Im}(\mathcal{S}_{(i)})$ の直交補空間への直交射影であり，厳密に $\mathbf{s}_i \in \text{Im}(\mathcal{S}_{(i)})$ であれば $\mathbf{0}$ になる．そのため

$$\min_{\mathbf{x}_{(i)}} \|\mathcal{S}_{(i)}\mathbf{x}_{(i)} - \mathbf{s}_i\|_2 = \|(\mathcal{S}_{(i)}\mathcal{S}_{(i)}^\dagger - \mathbf{I}_K)\mathbf{s}_i\|_2 \quad (5.8)$$

が大きいベクトル $\mathbf{s}_i$ を，以後線形従属度の低いベクトルと表現する．

5.1.2 $\mathcal{S}(\mathbf{c})$ が雑音を含む場合

改めて Affine 構造化行列 $\mathcal{S}(\mathbf{c})$ がパラメータベクトル $\mathbf{c}$ の摂動の影響でフルランク行列となっている場合、つまり全ての列が線形独立となっている状況を考える。 $\mathcal{S}(\mathbf{c})$ の摂動成分を STLN 法によって補償し SLRA を行うとき、フルランク行列 $\mathcal{S}(\mathbf{c})$ から第 i 列ベクトル $\mathbf{s}_i$ を抜き出し、

$$(\mathbf{s}_i + \Delta \mathbf{s}_i(\Delta \mathbf{c})) \in \text{Im}(\mathcal{S}_{(i)} + \Delta \mathcal{S}_{(i)}(\Delta \mathbf{c})) \quad (5.9)$$

を満たす最小の修正差分 $\Delta \mathbf{c}$ によって計算される $\mathcal{S}(\mathbf{c} + \Delta \mathbf{c})$ が $\mathcal{S}(\mathbf{c})$ に対する近似結果となる。ただし

$$\Delta \mathbf{s}_i(\Delta \mathbf{c}) := (\mathcal{S}(\Delta \mathbf{c}) - \mathbf{S}_0) \mathbf{e}_i \quad (5.10)$$

$$\Delta \mathcal{S}_{(i)}(\Delta \mathbf{c}) := (\mathcal{S}(\Delta \mathbf{c}) - \mathbf{S}_0) \tilde{\mathbf{I}}_i \quad (5.11)$$

である。この操作はフルランク行列 $\mathcal{S}(\mathbf{c})$ のいずれかの列ベクトルを、その他の列ベクトルたちに対して線形従属にしているとも捉えることができる。最初期の STLN-SLRA では、そのような列ベクトルとして $\mathbf{s}_1$ あるいは $\mathbf{s}_L$ を用いていたが、残りの列ベクトルたちとの線形従属度が低い場合、それらが線形従属となるように矯正する労力が余分にかかることになる。また、先に確認したように、線形従属度が低い列ベクトルを右辺に移動した場合、その初期方程式の解 $\mathbf{x}_{(i)} = \mathcal{S}_{(i)}^\dagger \mathbf{s}_i$ から構成される $\phi_i(\mathbf{x}_{(i)})$ は、真の値 $\mathbf{x}$ から大きく乖離する。STLN-SLRA では $\mathbf{c}$ のみならず $\mathbf{x}$ についても最適化を局所的に行うが、このように乖離の大きい $\phi_i(\mathbf{x}_{(i)})$ はその初期値として不適切であり、それを導く初期方程式もまた望ましくない。したがって、STLN-SLRA の実行にあたっては、適切な初期方程式を選択するための基準が必要である。

5.2 従来手法

文献 [25] では、Algorithm 5 の手順 3 において初期方程式を選択する基準として次の二つが提案された。

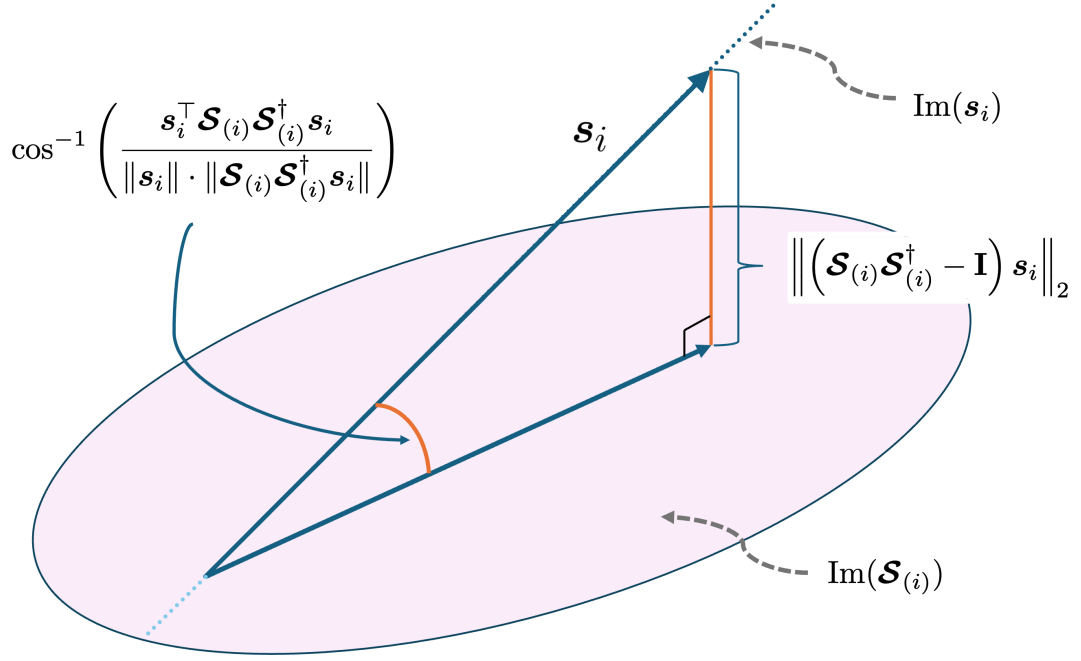


図 5.1: 選択基準 1 及び 2 の幾何的解釈

: Geometric Interpretation of Selection Criterion 1 and 2.

選択基準 1

$$i_{\text{residual}} := \arg \min_i \text{dist}_{\text{residual}}(\mathcal{S}_{(i)}, \mathbf{s}_i) \quad (5.12)$$

$$\text{dist}_{\text{residual}}(\mathcal{S}_{(i)}, \mathbf{s}_i) := \|\mathcal{S}_{(i)}(\mathcal{S}_{(i)}^\dagger \mathbf{s}_i) - \mathbf{s}_i\|^2 = \|(\mathbf{I}_K - \mathcal{S}_{(i)} \mathcal{S}_{(i)}^\dagger) \mathbf{s}_i\|^2 \quad (5.13)$$

選択基準 2

$$i_{\text{angle}} := \arg \min_i \text{dist}_{\text{angle}}(\mathcal{S}_{(i)}, \mathbf{s}_i) \quad (5.14)$$

$$\text{dist}_{\text{angle}}(\mathcal{S}_{(i)}, \mathbf{s}_i) := \cos^{-1} \left(\frac{\mathbf{s}_i^\top \mathcal{S}_{(i)} \mathcal{S}_{(i)}^\dagger \mathbf{s}_i}{\|\mathbf{s}_i\| \cdot \|\mathcal{S}_{(i)} \mathcal{S}_{(i)}^\dagger \mathbf{s}_i\|} \right) \quad (5.15)$$

選択基準 1 は, $\|\mathbf{r}(\mathbf{0}, \mathbf{x}_{(i)})\|^2$ を最小化することに基づいており, これは Algorithm 5 の手順 4 で指定されている内容と同じである. この基準は, $\mathbf{x}_{(i)}$ に対する誤差が最小になるように初期方程式を選択することを目的としている.

一方, 選択基準 2 は, 図 5.1 に表されるように, 部分空間 $\text{Im}(\mathcal{S}_{(i)})$ と $\text{Im}(\mathbf{s}_i)$ の間の主要角 (principal angle) を最小化することに基づいている. STLN の目的は, 適切な修正差分を加えることで $\mathbf{s}_i$ を $\text{Im}(\mathcal{S}_{(i)})$ に含めることであるため, この基準はその目的に整合している.

しかし、これらの基準によって選択された初期方程式 (5.2) が、元の近似式 (5.1) の妥当な近似となるかどうかは自明ではない。特に、選択基準 1 や 選択基準 2 で選ばれた初期方程式が、元の問題設定に対してどの程度適切であるかについての理論的な保証が不足している。

さらに、これらの基準には計算量が大きいという問題も存在する。具体的には、各 $i = 1, \dots, L$ について初期方程式を選択するために、擬似逆行列 $\mathbf{S}_{(i)}^\dagger$ を計算する必要があり、この計算の計算量は $O(L^3)$ である。そのため、全体としての計算量は $O(L^4)$ となり、大規模な問題に対しては計算負荷が非常に高くなる。したがって、実用的な観点からは、これらの基準をそのまま用いるのではなく、計算量を抑えつつ有効な初期方程式を選択するための新たな手法の導入が求められる。

5.3 提案法 [4]

本節では、新たに初期方程式の選択基準を提案する。そのような基準の構築にあたって、式 (5.1) における‘近似’の解釈を改めて行うことから始める。式 (5.1) の解を

$$\hat{\mathbf{x}} = \arg \min_{\|\mathbf{x}\| \geq 1} \|\mathbf{S}(\mathbf{c})\mathbf{x}\|^2, \quad (5.16)$$

によって表す。これより、式 (5.16) と同一の解を持つという意味で同等な線型方程式を探す。

σ を $\mathbf{S}(\mathbf{c})$ の最小特異値とし、 $\mathbf{u}$ および $\mathbf{v}$ をそれに対応する正規化された左および右特異ベクトルとする。このとき、式 (5.16) の解は、明らかに次式を満たす：

$$\mathbf{S}(\mathbf{c})\mathbf{v} = \sigma\mathbf{u}. \quad (5.17)$$

このため、初期方程式の選択は、式 (5.17) を元にして行うべきである。

式 (5.17) と 式 (5.2) の差異を解析するために、式 (5.17) において $\mathbf{S}(\mathbf{c})$ の第 i 列を左辺から右辺へ移項する：

$$\begin{aligned} \mathbf{S}(\mathbf{c})(\tilde{\mathbf{I}}_i \tilde{\mathbf{I}}_i^\top + \mathbf{e}_i \mathbf{e}_i^\top) \mathbf{v} = \sigma \mathbf{u} &\iff \mathbf{S}(\mathbf{c}) \tilde{\mathbf{I}}_i \tilde{\mathbf{I}}_i^\top \mathbf{v} = \sigma \mathbf{u} - \mathbf{S}(\mathbf{c}) \mathbf{e}_i \mathbf{e}_i^\top \mathbf{v} \\ &\iff -\frac{1}{v_i} \mathbf{S}_{(i)} \tilde{\mathbf{I}}_i^\top \mathbf{v} = \mathbf{s}_i - \frac{\sigma}{v_i} \mathbf{u}. \end{aligned} \quad (5.18)$$

次に $\Pi_{\mathbf{s}_i} = \frac{\mathbf{s}_i \mathbf{s}_i^\top}{\|\mathbf{s}_i\|^2}$ を $\text{Im}(\mathbf{s}_i)$ への直交射影子とし、 $\mathbf{u}$ 次のように分解する:

$$\mathbf{u} = \Pi_{\mathbf{s}_i} \mathbf{u} + (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u}.$$

この時、 $(\mathbf{I}_K - \Pi_{\mathbf{s}_i})$ も直交射影子である。ここで、直交射影子は冪等な対称行列であることを付記する。すなわち

$$\Pi_{\mathbf{s}_i}^\top = \Pi_{\mathbf{s}_i}, \quad (5.19)$$

$$\Pi_{\mathbf{s}_i}^2 = \Pi_{\mathbf{s}_i}, \quad (5.20)$$

$$(\mathbf{I}_K - \Pi_{\mathbf{s}_i})^\top = (\mathbf{I}_K - \Pi_{\mathbf{s}_i}), \quad (5.21)$$

$$(\mathbf{I}_K - \Pi_{\mathbf{s}_i})^2 = (\mathbf{I}_K - \Pi_{\mathbf{s}_i}). \quad (5.22)$$

式 (5.18) は次のように変形可能である:

$$\begin{aligned} -\frac{1}{v_i} \mathbf{S}_{(i)} \tilde{\mathbf{I}}_i^\top \mathbf{v} &= \left(\mathbf{s}_i - \frac{\sigma}{v_i} \Pi_{\mathbf{s}_i} \mathbf{u} \right) - \frac{\sigma}{v_i} (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u} \\ &= \left(1 - \frac{\sigma}{v_i} \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2} \right) \mathbf{s}_i - \frac{\sigma}{v_i} (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u}. \end{aligned} \quad (5.23)$$

上式の両辺を $\mathbf{s}_i$ の係数で除すことで、さらに次式を得る:

$$\frac{-\frac{1}{v_i}}{1 - \frac{\sigma}{v_i} \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2}} \mathbf{S}_{(i)} \tilde{\mathbf{I}}_i^\top \mathbf{v} = \mathbf{s}_i - \frac{\frac{\sigma}{v_i}}{1 - \frac{\sigma}{v_i} \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2}} (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u}. \quad (5.24)$$

続いて

$$\mathbf{x}_{(i)} = -\frac{\frac{1}{v_i}}{1 - \frac{\sigma}{v_i} \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2}} \tilde{\mathbf{I}}_i^\top \mathbf{v} \quad (5.25)$$

とおくことで、式 (5.24) を第 i 初期方程式に類似する形式に整理可能である:

$$\begin{aligned} \mathbf{S}_{(i)} \mathbf{x}_{(i)} &= \mathbf{s}_i - \frac{\frac{\sigma}{v_i}}{1 - \frac{\sigma}{v_i} \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2}} (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u} \\ &= \mathbf{s}_i - \left(\frac{v_i}{\sigma} - \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2} \right)^{-1} (\mathbf{I}_K - \Pi_{\mathbf{s}_i}) \mathbf{u}. \end{aligned} \quad (5.26)$$

式 (5.26) は、式 (5.17) と同一であることに注意されたい。ここで、式 (5.26) の右辺第 2 項を $\mathbf{h}$ とする。式 (5.2) が式 (5.26) を良好に近似するためには、 $\mathbf{h}$ のノルムが十分

小さいことが望ましい．そこで， $\|\mathbf{h}\|^2$ と $\|\mathbf{s}_i\|^2$ の比に着目し，以下のように定義する．

$$\kappa_i := \frac{1}{\|\mathbf{s}_i\|^2} \left(\frac{v_i}{\sigma} - \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2} \right)^{-2} \|(\mathbf{I}_K - \mathbf{\Pi}_{\mathbf{s}_i})\mathbf{u}\|^2. \quad (5.27)$$

続いて式 (5.27) に現れるいくつかの因子に対し変形を施す：

$$\left(\frac{v_i}{\sigma} - \frac{\mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2} \right)^{-2} = \frac{\sigma^2}{v_i^2} \frac{\|\mathbf{s}_i\|^4}{(\|\mathbf{s}_i\|^2 - \sigma^2)^2}, \quad (5.28)$$

$$\begin{aligned} \|(\mathbf{I}_K - \mathbf{\Pi}_{\mathbf{s}_i})\mathbf{u}\|^2 &= \mathbf{u}^\top (\mathbf{I}_K - \mathbf{\Pi}_{\mathbf{s}_i})\mathbf{u} \\ &= \mathbf{u}^\top \mathbf{u} - \frac{\mathbf{u}^\top \mathbf{s}_i \mathbf{s}_i^\top \mathbf{u}}{\|\mathbf{s}_i\|^2} \\ &= 1 - \frac{\sigma^2 v_i^2}{\|\mathbf{s}_i\|^2}. \end{aligned} \quad (5.29)$$

ここでは，式 (5.17) から $\mathcal{S}(\mathbf{c})^\top \mathbf{u} = \sigma \mathbf{v}$ が導かれ， $\mathbf{s}_i^\top \mathbf{u} = \sigma v_i$ が成立することを利用した．上式 (5.28) 及び式 (5.29) を受けて， κ_i は次の形式に縮約される：

$$\kappa_i = \frac{\|\mathbf{s}_i\|^2 - (\mathbf{s}_i^\top \mathbf{u})^2}{\left(\frac{v_i}{\sigma} \|\mathbf{s}_i\|^2 - \mathbf{s}_i^\top \mathbf{u} \right)^2} = \frac{\sigma^2}{v_i^2} \cdot \frac{\|\mathbf{s}_i\|^2 - \sigma^2 v_i^2}{(\|\mathbf{s}_i\|^2 - \sigma^2)^2}. \quad (5.30)$$

最終的に， κ_i より定数因子 σ^2 を取り除いたものとして，初期方程式の新たな選択基準を次に提案する．

提案選択基準

$$i_{\text{proposed}} = \arg \min_i \frac{\frac{\|\mathbf{s}_i\|^2}{v_i^2} - \sigma^2}{(\|\mathbf{s}_i\|^2 - \sigma^2)^2} \quad (5.31)$$

$\mathcal{S}(\mathbf{c})$ の全ての列ベクトルのノルムが等しい，つまり $\|\mathbf{s}_1\| = \|\mathbf{s}_2\| = \dots = \|\mathbf{s}_n\|$ となる場合は，提案選択基準はさらに次の単純な形式に帰着する：

$$i_{\text{proposed}} = \arg \max_i |v_i|. \quad (5.32)$$

5.3.1 本手法の計算量

提案した基準の計算量は容易に $O(L^3)$ であると確認できるのに対し，従来の基準の計算量は $O(L^4)$ となる．この差は，式 (5.31) の計算量が主に $\mathcal{S}(\mathbf{c})^\dagger$ で占められており，その計算量が $O(L^3)$ であることに起因する．

5.4 数値実験

本節では、数値例を用いて、提案する選択基準を用いた STLN-SLRA が従来法よりも計算時間を削減できることを確認する。提案基準が初期方程式の選択にかかる計算量を削減することは自明であるが、提案基準が全体の実行時間も削減するかどうかは明らかではないためである。また、提案基準を用いた STLN-SLRA が従来法と同等の精度を維持していること、並びに SLRA を介した各種推定問題における推定性能を検証する。本章で行ったすべての数値実験は 2.3 GHz CPU を搭載した計算機で行い、すべてのアルゴリズムは MATLAB R2022b [114] で実装した。

5.4.1 近似 GCD 計算への適用例

ここでは、STLN-SLRA を用いて近似 GCD を計算することにより、2つの多項式に対する GCD の推定問題 18 を解く。

真の多項式を次のように設定する：

$$\begin{aligned}\bar{p}^{(1)}(t) &= \bar{g}^{(1)}(t)\bar{w}(t) \in \mathbb{R}^{n_1}[t], \\ \bar{p}^{(2)}(t) &= \bar{g}^{(2)}(t)\bar{w}(t) \in \mathbb{R}^{n_2}[t],\end{aligned}\tag{5.33}$$

ただし、 $\bar{g}^{(1)}(t) \in \mathbb{R}^{n_1-d}[t]$ 及び $\bar{g}^{(2)}(t) \in \mathbb{R}^{n_2-d}[t]$ 、 $\bar{w}(t) \in \mathbb{R}^d[t]$ の係数は、いずれも標準正規分布に従って独立に生成し、 $\bar{g}^{(1)}(t)$ と $\bar{g}^{(2)}(t)$ が互いに疎であることを確認した。この時

$$\deg \gcd(\bar{p}^{(n)}(t)) = d \geq 1$$

である。さらに、 $\bar{p}^{(1)}$ 及び $\bar{p}^{(2)}$ はそれぞれ $\|\bar{p}^{(1)}\| = \|\bar{p}^{(2)}\| = 1$ となるよう正規化する。雑音項は係数が標準正規分布に従う多項式 $\eta^{(1)}(t)$ 及び $\eta^{(2)}$ を用いて $\alpha\eta^{(1)}(t)$ と $\alpha\eta^{(2)}(t)$ とし、 α は雑音の標準偏差を表す。観測多項式は次のように与えられる：

$$\begin{aligned}p^{(1)}(t) &= \bar{p}^{(1)}(t) + \alpha\eta^{(1)}(t) \\ p^{(2)}(t) &= \bar{p}^{(2)}(t) + \alpha\eta^{(2)}(t).\end{aligned}\tag{5.34}$$

この観測多項式に対して、次の SLRA 問題を STLN-SLRA で解き、真の多項式の係数ベクトル $\bar{p}^{(n)}$ に対する推定値 $\hat{p}_*^{(n)}$ を得る：

$$\begin{aligned}\hat{p}_*^{(n)} &:= \arg \min_{\hat{p}^{(n)}} \|\hat{c} - c\|_2^2 \\ \text{subject to } \text{rank fSyl}_d(\hat{p}^{(n)}) &\leq L - 1.\end{aligned}\tag{5.35}$$

実験では、 $\bar{\mathbf{p}}^{(n)}$ 及び $\boldsymbol{\eta}^{(n)}$ をランダムに生成し、これを異なる値の α に対して各 1000 回繰り返した。上述の設定では $\|\mathbf{p}^{(1)}\| \approx \|\mathbf{p}^{(2)}\|$ であることに着目し、提案選択基準の計算には、式 (5.32) を使用した。

計算精度の評価

まず、提案選択基準を採用した STLN-SLRA の挙動と性能を調査し、従来の選択基準を採用したものと比較した。SLRA アルゴリズムの出力の妥当性を、次の 3 つの指標により評価した。

- (a) 結果の低ランク性
- (b) 修正差分の大きさ
- (c) アルゴリズムでの反復計算回数

近似の結果得られる行列 $\text{fSyl}_d(\hat{\mathbf{p}}_*^{(n)})$ は、その最小特異値が 10^{-15} 未満である場合に、低ランク性を有するとみなす。指標 (a) は、1000 回の試行のうち、この条件を満たして低ランク性が得られた試行の割合である。指標 (b) は、 $\|\mathbf{p}^{(1)} - \hat{\mathbf{p}}_*^{(1)}\|^2 + \|\mathbf{p}^{(2)} - \hat{\mathbf{p}}_*^{(2)}\|^2$ により計算され、式 (4.2) における $\|\Delta \mathbf{c}\|^2$ に対応する。指標 (c) は、初期方程式の選択より後に実行される部分の計算コストを評価する目的で採用した。

加えて、真の GCD の推定問題における性能評価を、以下の 2 つの指標に基づいて行った。

- (d) 真の多項式の推定誤差
- (e) 零空間基底の推定誤差

指標 (d) は、 $\|\bar{\mathbf{p}}^{(1)} - \hat{\mathbf{p}}_*^{(1)}\|^2 + \|\bar{\mathbf{p}}^{(2)} - \hat{\mathbf{p}}_*^{(2)}\|^2$ により計算される。零空間基底の推定誤差を示す指標 (e) は、 $\cos^{-1}(|\bar{\mathbf{v}}^\top \hat{\mathbf{v}}_*|)$ により定義され、真の商多項式の推定精度を表す。ここで、 $\bar{\mathbf{v}}$ および $\hat{\mathbf{v}}_*$ は、それぞれ $\text{fSyl}_d(\bar{\mathbf{p}}^{(n)})$ および $\text{fSyl}_d(\hat{\mathbf{p}}_*^{(n)})$ の零特異値に対応する右特異ベクトルであり、これは商多項式の係数ベクトル $\bar{\mathbf{g}}^{(n)}$ の推定精度を評価するものである。本評価では、 $n_1 = n_2 = 28$, $d = 19$ に設定した。

表 5.3: 指標 (a) から (e) の比較 ($\alpha = 0.001$): Comparison of Measures (a) to (e) for $\alpha = 0.001$.

	選択なし ($i = 1$)	選択基準 1	選択基準 2	提案選択基準
(a)	0.99	0.99	0.99	0.98
(b)	3.1×10^{-5}	1.9×10^{-5}	1.9×10^{-5}	1.9×10^{-5}
(c)	23	23	23	24
(d)	5.1×10^{-5}	3.9×10^{-5}	3.9×10^{-5}	3.9×10^{-5}
(e)	0.19	0.17	0.17	0.17

表 5.4: 指標 (a) から (e) の比較 ($\alpha = 0.02$): Comparison of Measures (a) to (e) for $\alpha = 0.03$.

	選択なし ($i = 1$)	選択基準 1	選択基準 2	提案選択基準
(a)	0.87	0.87	0.87	0.87
(b)	9.7×10^{-3}	7.7×10^{-3}	7.7×10^{-3}	7.6×10^{-3}
(c)	166	165	164	163
(d)	2.0×10^{-2}	1.8×10^{-2}	1.8×10^{-2}	1.8×10^{-2}
(e)	0.62	0.62	0.63	0.62

表 5.3 には、各選択基準を採用した際の指標 (a)～(e) の結果を示す。この表により、いずれの基準においても、ほぼ同等の結果が得られたことが確認された。実際、多くの場合で各基準で同じ初期方程式が選択された。

表 5.4 に示すように、 $\alpha = 0.02$ の場合にも類似する結果を得た。この結果の $\alpha = 0.01$ の場合との差異は、反復計算回数の増加である。具体的には、提案基準を用いた STLN-SLRA は、ノイズの分散が大きい場合に、従来のものを用いた場合よりも多くの反復計算を必要とする可能性がある。しかし、このような状況では、STLN を低ランク近似の手段として使用することは適切ではないと言える。なぜなら、STLN-SLRA の結果の行列が低ランク性を十分に満たさないケースが多くなるためである。

実行時間の評価

図 5.2 は、 $\alpha = 0.001$ での各基準を採用した場合の STLN-SLRA の実行時間を示す箱ひげ図であり、提案選択基準が最も短い平均実行時間を達成することを示す。この例では前節と同じように、多項式の次数は $n_1 = n_2 = 28$ 、 $d = 19$ に設定した。これらの結果は、指標 (e) が比較的小さく、すべての基準間でほぼ同一であることから、全体

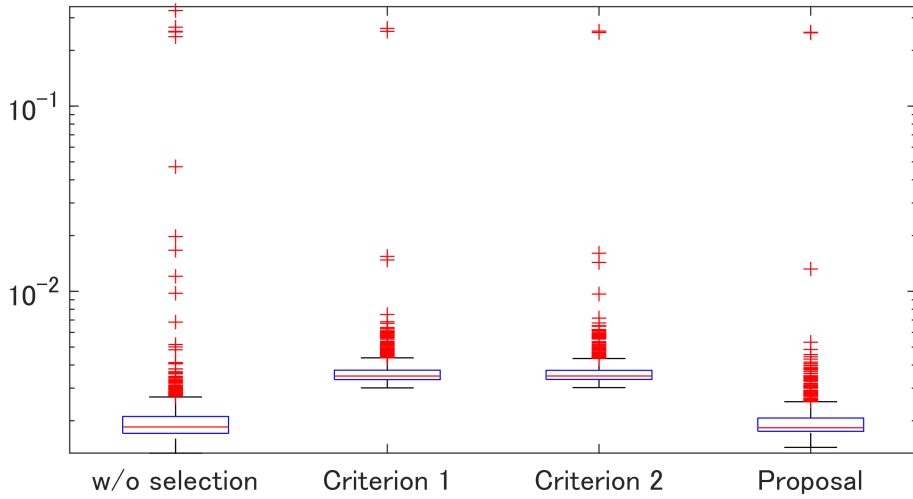


図 5.2: 実行時間の対数の箱ひげ図 ($\alpha = 0.001$)

: Box-and-Whisker Plots of the Logarithm of the Execution Times for $\alpha = 0.001$. [4]

の実行時間は列選択の計算時間に大きく依存していることを示している。

表 5.5 は、雑音強度 α に対する提案基準と従来の基準の実行時間の比較である。この結果から、提案基準は雑音強度が低い場合、実行時間の短縮に貢献したと結論づけられる。雑音強度が高いほど、STLN の反復計算回数が増加し、全体の実行時間は STLN の実行時間に大きく依存するためである。

拡張性の評価

次に、提案手法の多項式の次数に関する拡張性を確認するために、GCD の次数 d を 19 に固定したまま、 $\mathbf{p}^{(1)}(t)$ の次数 n_1 と $\mathbf{p}^{(2)}(t)$ の次数 n_2 を増加させる例を考える。言い換えれば、 $\bar{\mathbf{g}}^{(1)}(t)$ と $\bar{\mathbf{g}}^{(2)}(t)$ の次数を増加させることになる。ここで雑音強度 α は 0.001 に設定した。

図 5.3 は、提案基準と従来の基準を用いた STLN-SLRA の実行時間を比較したものである。この結果では、提案選択基準を採用した STLN-SLRA のアルゴリズム全体での実行時間は、従来の基準を使用したものと比較して短縮された。その一方ですべての選択基準で推定精度が同程度であったことは注目に値する。選択基準 1 と選択基準 2 の計算量は $O(L^4)$ であるのに対し、提案された選択基準は $O(L^3)$ であったことから、高次多項式を扱う場合は、提案された選択基準が他の 2 つの基準と比較して計算時間の点で有利であると言える。

表 5.5: Sylvester 行列に対する STLN-SLRA の実行時間 (秒) の, 雑音強度 α に関する提案基準と従来の基準との比較

: Comparison of Execution Time (sec) of STLN-SLRA for Sylvester Matrices between Proposed and Conventional Criteria with respect to Noise Level α . [4]

α	選択基準 1	選択基準 2	提案選択基準
0.6	8.8×10^{-2}	8.7×10^{-2}	8.5×10^{-2}
0.4	8.4×10^{-2}	8.5×10^{-2}	8.4×10^{-2}
0.2	8.1×10^{-2}	8.4×10^{-2}	8.0×10^{-2}
0.1	6.9×10^{-2}	6.9×10^{-2}	7.0×10^{-2}
0.01	4.3×10^{-2}	4.3×10^{-2}	4.1×10^{-2}
0.02	2.8×10^{-2}	2.7×10^{-2}	2.7×10^{-2}
10^{-3}	8.1×10^{-3}	8.0×10^{-3}	6.6×10^{-3}
10^{-4}	3.6×10^{-3}	3.6×10^{-3}	2.3×10^{-3}

その他の近似 GCD 計算手法との比較

本節では, 提案選択基準を採用する STLN-SLRA と, STLN を使用しない次の三つの近似 GCD 計算手法とを比較する:

- **UVGCD** - Numerical Algebraic Computing Toolbox (NAClab) [115] に含まれる `uvGCDfixedDegree` 関数
- **ODE** - 文献 [116] で提案された, 常微分方程式に基づく方法 (著者による実装を使用)
- **VP-SLRA** - 第 2.2.5 節に記載した手法. 実装は SLRA toolbox [57] に含まれる `gcd_nls` 関数を使用

表 5.6 は, 第 5.4.1 節と同じ実験条件を使用し, 評価項目 (a), (b), (d), (e), および実行時間を測定した際の, 複数の近似 GCD 計算手法の比較を示す. 測定項目ごとに最良の結果を太字で強調した. 他の 3 つの近似 GCD 計算手法とは異なり, UVGCD は, 近似 GCD と商多項式を直接計算し, $\hat{p}_*^{(n)}(t)$ を出力しないため, それらを GCD に商を掛けたものとして計算した. したがって, UVGCD の測定 (a) は必然的に 1 となる. ODE では, の最小特異値のしきい値を 10^{-16} に設定したにもかかわらず, ODE の評価項目

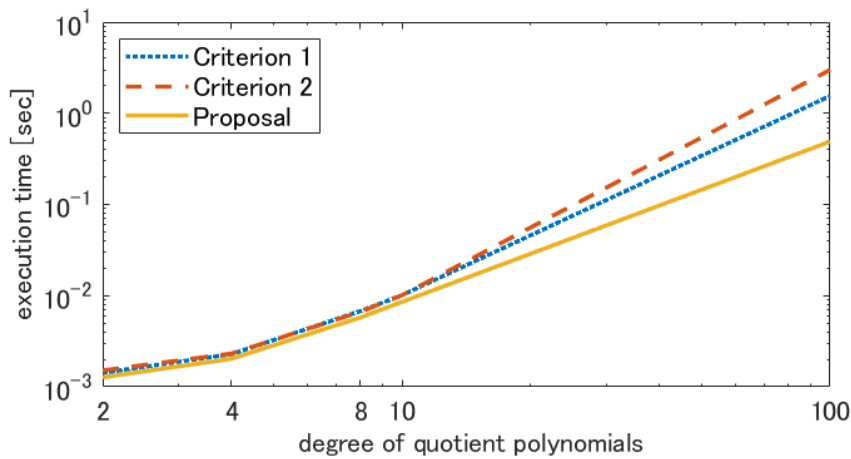


図 5.3: 商多項式の次数ごとの実行時間比較

: Comparison of Execution Time with respect to the Degree of Quotient Polynomials. [4]

表 5.6: 近似 GCD 計算方法の比較

: Comparison of Approximate GCD Computation Methods. [4]

	Proposal	UVGCD	ODE	VP-SLRA
(a)	0.98	1.0	0	1.0
(b)	2.0×10^{-5}	3.7×10^{-2}	9.3×10^{-1}	7.0×10^{-5}
time	6.7×10^{-3}	1.5×10^{-3}	1.6×10^0	1.1×10^{-2}
(d)	4.0×10^{-5}	3.7×10^{-2}	9.3×10^{-1}	9.0×10^{-5}
(e)	1.7×10^{-1}	7.6×10^{-1}	5.0×10^{-1}	7.7×10^{-1}

(a) は 0 に留まった。この結果は、今回の実験条件下で ODE が望ましくない局所最適解に収束する傾向に起因する可能性が考慮される。

近似結果の低ランク性という点では、指標 (a) が 1 であることから、VP-SLRA が最良の手法であると言える。最短の計算時間は UVGCD によって達成された。一方、STLN-SLRA は修正差分が最小であるだけでなく近似 GCD 計算精度の点でも最良の結果を示した。各手法は異なる性能指標においてそれぞれの強みを有していることが分かる。

5.4.2 雑音除去への適用例

本節では、Hankel 行列の SLRA における提案手法の性能を検証する。第 2.1.6 節で述べたように、この種の SLRA は雑音除去に応用可能である。ここでは信号を実信号

に限定し、ランク $2q$ の真の信号 $\bar{\mathbf{c}} \in \mathbb{R}^{n_p}$ を次式で生成する：

$$\bar{c}_t = \sum_{j=1}^q a_j \sin(2\pi\omega_j k + \phi_j), \text{ for } t = 0, 1, \dots, n_p - 1, \quad (5.36)$$

where $a_j \neq 0$, $0 < \omega_j < 1$ for $j = 1, \dots, q$.

これは、式 (2.2) において、 t を変数とした多項式 $p^{(j)}(t)$ を t によらない定数 a_j とし、減衰率を $d_j = 0$ とした場合に相当する。この真の信号 $\bar{\mathbf{c}}$ に雑音 $\alpha\boldsymbol{\eta}$ を重畳した信号

$$\mathbf{c} = \bar{\mathbf{c}} + \alpha\boldsymbol{\eta} \in \mathbb{R}^{n_p} \quad (5.37)$$

を観測信号とする。ここで $\boldsymbol{\eta}$ は平均 0、分散 1 の白色 Gauss 性雑音であり、 α は雑音の強度を表す正の実数である。この設定の下、次式に示す Hankel 行列の SLRA を STLN-SLRA によって計算し、真の信号 $\bar{\mathbf{c}}$ の推定を行う：

$$\begin{aligned} \hat{\mathbf{c}}_* &:= \arg \min_{\hat{\mathbf{c}}} \|\hat{\mathbf{c}} - \mathbf{c}\|_2^2 \\ \text{subject to } \text{rank } \mathcal{H}_{2q}(\hat{\mathbf{c}}) &\leq 2q - 1. \end{aligned} \quad (5.38)$$

例 5. 式 (5.38) において $q = 2$ 及び $n_p = 50$, $\alpha = 0.01$, $a_1 = 1$, $a_2 = \frac{3}{10}$, $\omega_1 = \frac{4}{50}$, $\omega_2 = \frac{3}{50}$, $\phi_1 = \phi_2 = 0$ とする。この時、真の信号 $\bar{\mathbf{c}}$ は次の形をとる：

$$\bar{c}_t = \sin\left(\frac{4}{50}2\pi t\right) + \frac{3}{10} \cos\left(\frac{3}{50}2\pi t\right) \quad \text{for } t = 0, 1, \dots, n_p - 1. \quad (5.39)$$

図 5.4 は、信号の推定誤差 $\|\bar{\mathbf{c}} - \hat{\mathbf{c}}_*\|^2 / \|\bar{\mathbf{c}}\|^2$ と 100 回の試行における全体的な実行時間の箱ひげ図を示す。参考までに、第 5.4.1 節の 近似 GCD 問題の例で良好な性能を示した SLRA 法である VP-SLRA の結果も含めた。

例 6. より多くの正弦波の和で表される信号に対しても同様の検証を行う。式 (5.38) において $q = 3$ 及び $n_p = 100$, $\alpha = 0.01$, $a_1 = \frac{4}{5}$, $a_2 = \frac{6}{10}$, $a_3 = \frac{7}{10}$, $\omega_1 = \frac{4}{100}$, $\omega_2 = \frac{7}{100}$, $\frac{1}{100}$, $\phi_1 = 0$, $\phi_2 = 1$, $\phi_3 = \frac{\pi}{2}$ とする。この時、真の信号 $\bar{\mathbf{c}}$ は次の形をとる：

$$\begin{aligned} \bar{c}_t &= \frac{4}{5} \sin\left(\frac{4}{100}2\pi t\right) + \frac{6}{10} \sin\left(\frac{7}{100}2\pi t + 1\right) + \frac{7}{10} \cos\left(\frac{1}{100}2\pi t\right) \\ &\quad \text{for } t = 0, 1, \dots, n_p - 1, \end{aligned} \quad (5.40)$$

図 5.5 は、信号の推定誤差 $\|\bar{\mathbf{c}} - \hat{\mathbf{c}}_*\|^2 / \|\bar{\mathbf{c}}\|^2$ と 100 回の試行にわたる全体的な実行時間の箱ひげ図を示す。この図には、例 5 と同様、VP-SLRA を使用した計算結果も含めた。

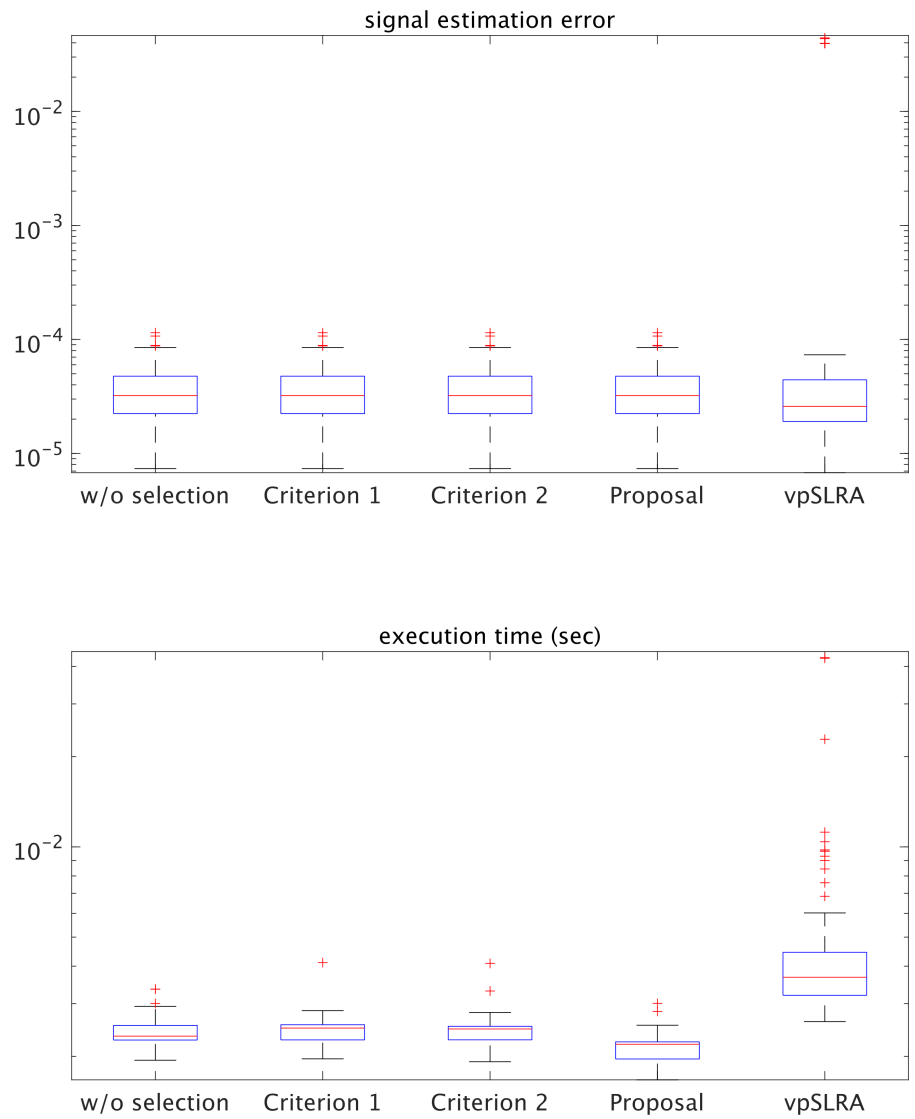


図 5.4: 例 5: 信号推定誤差（上段）と実行時間（秒）（下段）の比較

: Example 5: Comparison of Signal Estimation Error (upper row) and Execution Time in sec (lower row.) [4]

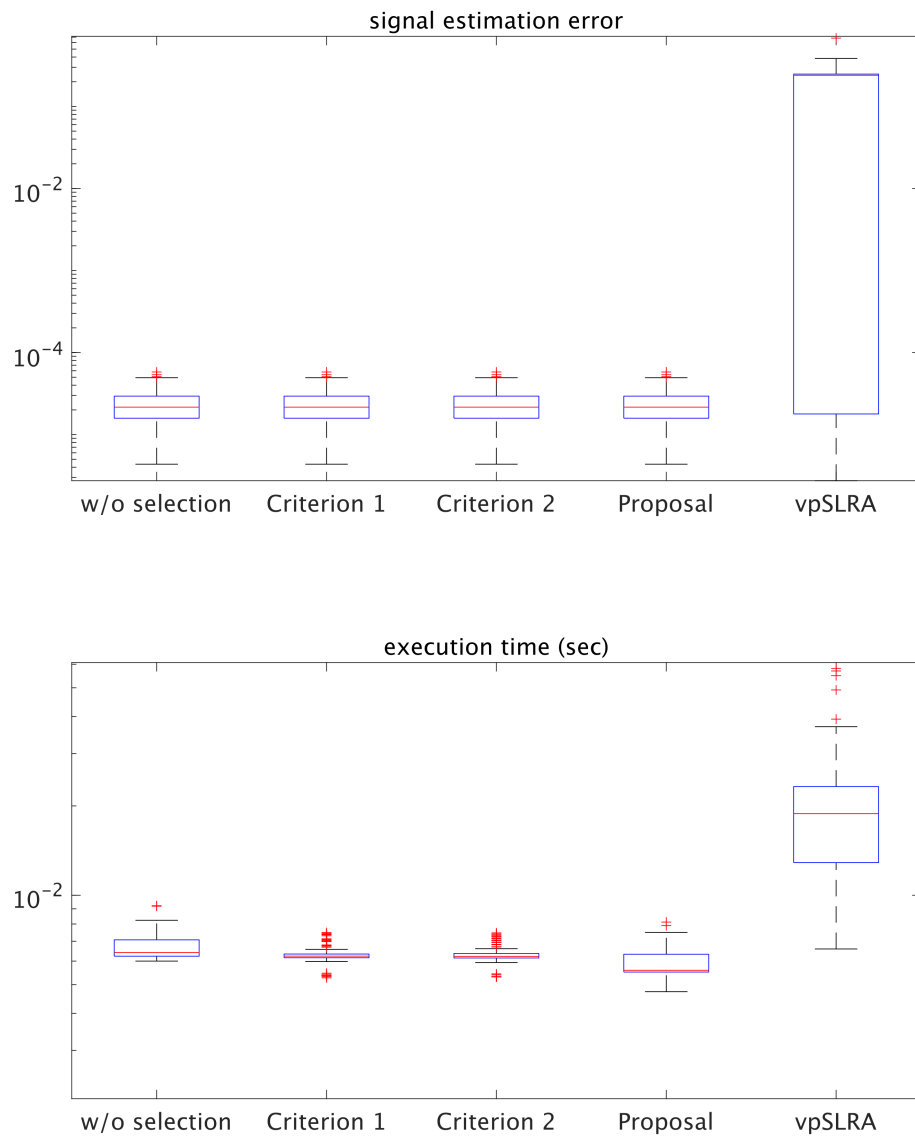


図 5.5: 例 6: 信号推定誤差（上段）と実行時間（秒）（下段）の比較

: Example 6: Comparison of Signal Estimation Error (upper row) and Execution Time in sec (lower row.) [4]

考察

上記の2つの例で、提案された選択基準を持つ STLN-SLRA は、Hankel 行列の SLRA においても、他の初期方程式選択基準を使用した STLN-SLRA と比較して、精度を損なうことなく計算時間を短縮した。さらに、上記の STLN-SLRA の信号推定精度は、VP-SLRA と同等程度であった。実際、提案法が VP-SLRA よりも優れている状況も存在した。しかし、ここで検討されている応用例に現れる Hankel 行列は列数が少なく、選択基準を計算するために必要な時間がアルゴリズム全体の計算時間に占める割合がごく僅かであるため、計算時間を短縮する効果は限定的である。提案法は、列数が多い Hankel 行列を扱う状況、つまり次数が高い LRR を扱う場面でより効果的であると期待される。

5.5 本章のまとめ

この章では、STLN-SLRA の初期方程式を詳細に分析し、その方程式の新しい選択基準を提案した。従来の基準の計算量が $O(L^4)$ であるのに対し、提案選択基準の計算量は $O(L^3)$ に改善された。さらに数値例を用いて、提案された基準が STLN-SLRA に必要な計算量を削減することを確認した。

第6章 低ランク性の零空間表現を使用するSLRA手法の初期化戦略

6.1 はじめに

行列の低ランク性を零空間表現するアルゴリズム群においては、零空間の基底ベクトルも最適化変数として扱われている。SLRA問題はランク制約が非凸であるが、実用上凸最適化手法が多く用いられている。非凸な問題に対して凸最適化手法を用いるため、そのようなアルゴリズムは初期値依存性が高いことに注意が必要である。多くの先行研究では、零空間の基底の初期値として、所与の行列の最小特異値に対応する右特異ベクトルが使用される。この初期化戦略のもとでは、所望のSLRAの零空間の基底が観測された行列の最小特異値に対応する右特異ベクトルと高い類似性を示すと仮定できる場合に所望の近似結果へのアルゴリズムの収束が期待される。

本章¹では、後述する「特異値の交代」という現象により、上述の二つのベクトルが整合しない、つまり従来の初期化戦略が想定通りに機能しないことがあり得ることを指摘し、当該問題への対処が必要であることを明確にする。続いて、行列の低ランク性を零空間表現するアルゴリズムとしてVP-SLRAを取り上げ、VP-SLRAにて特異値の交代現象に対処するための初期化戦略を提案するとともに、数値実験により当該初期化戦略が実用上有益であることを示す。

6.2 特異値の交代

行列に摂動が加わると、その特異値にも変化が生じる。このとき、最小特異値は相対的に摂動の影響を受けやすいことが知られている。このため、摂動後の行列では、元

¹本章は、著者らによる研究成果[32] (Copyright(C)2025 IEICE), 及び[31] (Copyright(C)2025 IEEE)に基づいている。

の最小特異値に対応していた特異ベクトルに対して類似度の高いベクトルが、最小ではない特異値に対応し、別のベクトルが新たな最小特異値に対応する可能性がある。

この現象を数式で整理する。ある行列構造 $\mathbf{S}$ およびそのパラメータベクトル $\bar{\mathbf{c}}$ と、それに対して何らかの摂動が加算されたベクトル $\mathbf{c} = \bar{\mathbf{c}} + \boldsymbol{\eta}$ を考える。 $\mathbf{S}(\bar{\mathbf{c}}) \in \mathbb{R}^{K \times L}$ ($K \geq L$) の特異値を、小さい順に $0 \leq \bar{\sigma}_1 \leq \bar{\sigma}_2 \leq \dots \leq \bar{\sigma}_L$ とする。また、 $\mathbf{S}(\mathbf{c}) \in \mathbb{R}^{K \times L}$ の特異値を $\sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_L$ とし、 $\bar{\sigma}_i$ 及び σ_i に対応する右特異ベクトルをそれぞれ $\bar{\mathbf{v}}_i, \mathbf{v}_i$ とする。構造化パラメータに加えられた摂動 $\boldsymbol{\eta}$ によって引き起こされる $\bar{\sigma}_i$ の摂動を $\delta\bar{\sigma}_i$ とすると、 $i \neq j$ に対して一般に次のような大小関係

$$\sigma_i \simeq \delta\bar{\sigma}_i + \bar{\sigma}_i \leq \delta\bar{\sigma}_{i+1} + \bar{\sigma}_{i+1} \simeq \sigma_{i+1}$$

が成立する保証はない。このように、摂動によって特異値の大小関係が入れ替わる現象を、本研究では**特異値の交代 (singular value switching)** と呼称する。図 6.1 は特異値の交代現象の模式図である。特に、 $\bar{\sigma}_1 = 0$ の場合に

$$\sigma_1 > \sigma_j \simeq \bar{\sigma}_1 + \delta\bar{\sigma}_1$$

なる j , ($2 \leq j \leq L$) が存在する（零特異値の交代が発生している）場合、真の右特異ベクトル $\bar{\mathbf{v}}_1$ に対してより高い類似度をもつのは $\mathbf{v}_1$ ではなく $\mathbf{v}_j$ である可能性が生じる²。

特異値の交代に着目する理由は、実用上重要な問題に関係している。第 2 章に述べたように、多くの工学的応用において、ある行列が一次元の零空間を持ち、その零空間の非零元に信号やシステムのパラメータが格納されているという構造を仮定したモデルが用いられる。このような場面では、元の行列がランク欠落していること、すなわち零特異値を持っていることが前提となっている。しかし、実際の観測では雑音の影響を完全に排除することはほぼ不可能であり、雑音の影響により特異値の交代が生じる可能性がある。このとき、本来零空間に含まれるはずのパラメータベクトルが、直感的に対応する最小特異値にはもはや対応しなくなるという現象が起こる。

このように、特異値の交代は、零空間からモデルパラメータを抽出する手法の信頼性や頑健性に直接関わる現象である。本研究が対象とする、低ランク性の零空間表現

²行列の摂動は特異値だけでなく特異ベクトルにも影響を及ぼす。このため摂動が加わった結果、真の右特異ベクトル $\bar{\mathbf{v}}_1$ と高い類似度をもつベクトル $\mathbf{v}_j$, ($j = 1, 2, \dots, L$) が存在しない場合も十分考えられる。

に基づくを行う SLRA 手法においては、零空間に属するベクトルを暗に推定する構造を持つため、同様にこの問題の影響を受ける。したがって、特異値の交代に対する適切な理解と対応が必要不可欠である。

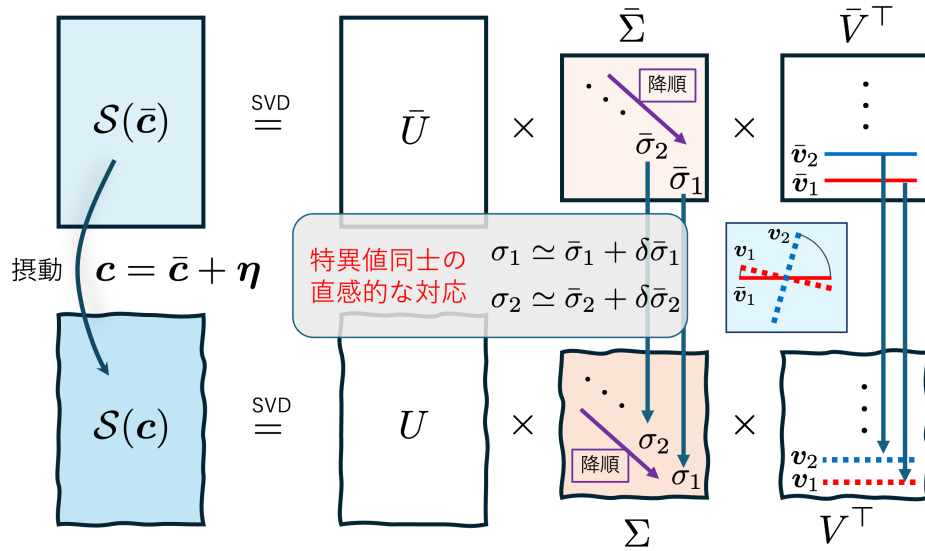
特異値の交代が及ぼす具体的な影響として、変数射影法による SLRA 計算における初期値選定の問題が挙げられる。 $\mathbf{x}$ の初期値は、 $\mathbf{S}(\mathbf{c})$ の（重みを考慮しない）非構造化低ランク近似の結果、つまり $\mathbf{S}(\mathbf{c})$ の最小特異値に対応する右特異ベクトルが用いられる場合が多い。しかしながら、 $\mathbf{S}(\mathbf{c})$ で零特異値の交代が発生した場合、 $\mathbf{v}_1$ よりも $\bar{\mathbf{v}}_1$ と高い類似度をもつ $\mathbf{v}_j$ が存在する可能性があり、 $\mathbf{v}_1$ を $\mathbf{x}$ の初期値とすることは不適当となる。その場合、 $\mathbf{x}$ の初期値としてより適切な、最小ではない特異値に対応する右特異ベクトル $\mathbf{v}_j$ の存在が期待される。

より具体的に、SLRA による近似 GCD 計算において特異値の交代現象を持つ意味を示すために、真の多項式 $\bar{p}^{(n)}(x)$ に摂動が加えられた多項式の組 $p^{(n)}(x)$ に対する近似 GCD 問題を考える。ここでは、変数射影法により得られる近似 GCD が、（未知の）真の GCD に一致、あるいは非常に近いと仮定する³。前述のように、文献 [1] では、 $\text{fSyl}_d^N(\mathbf{p}^{(n)})$ の最小特異値に対応する右特異ベクトル $\mathbf{v}_1$ を商多項式の初期推定値 $\hat{\mathbf{g}}_{\text{ini}}^{(n)}$ として用いている。しかし、真の多項式から構成される行列 $\text{fSyl}_d^N(\bar{\mathbf{p}}^{(n)})$ と、摂動後の行列 $\text{fSyl}_d^N(\mathbf{p}^{(n)})$ の間で零特異値の交代が生じた場合、 $\mathbf{v}_1$ よりも $\bar{\mathbf{g}}^{(n)}$ に近い特異ベクトル $\mathbf{v}_j$ ($j \neq 1$) が存在する可能性がある。このような場合、 $\mathbf{v} = \mathbf{v}_1$ として Algorithm 1 を実行することは、大域的最適解への収束を妨げ、意図しない多項式を出力する原因となり得る。実際に、摂動の前後でフル・Sylvester 行列における零特異値の交代が発生することを、次の予備実験にて確認した。

6.2.1 予備実験

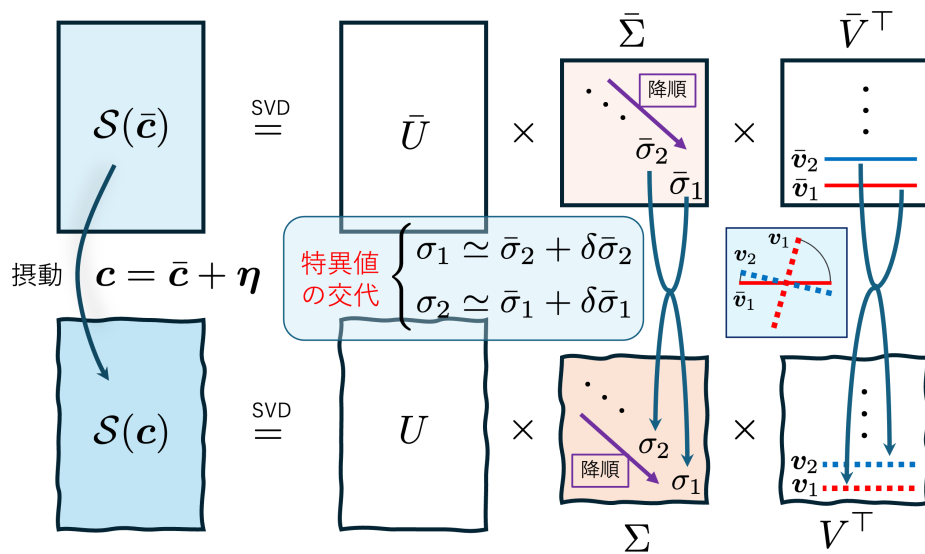
第 5.4.1 節と同様の多項式の組を考える。次数 $d = 10$ の GCD を持つように真の多項式 $\bar{p}^{(n)}(x)$ を生成し、さらにそれに雑音を加えた多項式 $p^{(n)}(x)$ を生成する。それぞれのフル・Sylvester 行列 $\text{fSyl}_d^N(\mathbf{p}^{(n)})$ および $\text{fSyl}_d^N(\bar{\mathbf{p}}^{(n)})$ の特異値分解を行い、右特異ベクトルの類似度を比較することで特異値の交代が実際に発生することを確認する。

³実際、次章の数値実験結果は、雑音強度が比較的低い場合にこの仮定が成立することを支持している。



(a) 特異値が大小関係を保存したまま対応する場合

: When Singular Values Correspond while Preserving Their Order of Magnitude



(b) 特異値の交代が発生した場合: When the Singular Value Switching Happens

図 6.1: 行列に対して摂動が加わる前後での特異値特異ベクトルの対応関係の模式図
: Correspondence between Singular Values and Singular Vectors before and after a Perturbation [31]

$n_1 = n_2 = 30$ とし, $\bar{\mathbf{g}}^{(1)} \in \mathbb{R}^{21}$, $\bar{\mathbf{g}}^{(2)} \in \mathbb{R}^{21}$, $\bar{\mathbf{h}} \in \mathbb{R}^{11}$ の各要素を, 標準正規分布より独立にサンプリングし設定する. この時, $\bar{\mathbf{g}}^{(1)}(x)$ と $\bar{\mathbf{g}}^{(2)}(x)$ が互いに素であることを確認する. 次に $\bar{\mathbf{p}}^{(i)}(x) = \bar{\mathbf{h}}(x) \cdot \bar{\mathbf{g}}^{(i)}(x)$ とし, 係数に対する正規化: $\bar{\mathbf{p}}^{(i)} \leftarrow \bar{\mathbf{p}}^{(i)} / \|\bar{\mathbf{p}}^{(i)}\|$ を行う. これに $\eta^{(i)}(x)$ を加えた $\mathbf{p}^{(n)}(x) = \bar{\mathbf{p}}^{(n)}(x) + \sigma \eta^{(n)}(x)$ を生成する. ただし, $\eta^{(i)}$ の個々の要素は平均 0 の Gauss 性雑音であり, ベクトルのノルムが 1 となるように正規化を行なっている. $\text{fSyl}_d^N(\mathbf{p}^{(n)})$ に対しては, 小さい順に 5 つの特異値に対応する右特異ベクトル $\mathbf{v}_1, \dots, \mathbf{v}_5$ を抽出し, $\text{fSyl}_d^N(\bar{\mathbf{p}})$ に対しては, 唯一の零特異値に対応する右特異ベクトル $\bar{\mathbf{v}}_1$ を取得した. 各試行において, $\bar{\mathbf{v}}_1$ と $\mathbf{v}_j$ ($j = 1, \dots, 5$) の類似度を以下の類似度指標

$$\psi(\mathbf{a}, \mathbf{b}) := \frac{|\mathbf{a}^\top \mathbf{b}|}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \quad (6.1)$$

により評価し, 最も類似度が高いベクトルを特定した. この類似度が 0.75 以上である場合を有意な一致とみなし, その回数をカウントした. 上記回数のうち特に, $\mathbf{v}_1$ 以外のベクトルが最も高い類似度を示した場合には, 特異値の交代が発生したと判断される.

この操作を雑音強度 $\sigma = 0.01, 0.05, 0.1$ に対してそれぞれ 100 回の試行で実施し, 得た結果を表 6.1 に示す. 比較的雑音の影響が小さい $\sigma = 0.01$ においても, 100 回中 4 回の交代が確認された. $\sigma = 0.05$ では交代の発生回数が 18 回に増加し, 雑音強度の上昇に伴い交代の頻度が高まる傾向が見られた. 一方, さらに雑音を強めた $\sigma = 0.1$ では交代の回数が 16 回とやや減少したが, この条件下では, $\bar{\mathbf{v}}_1$ と有意な一致を示すベクトル $\mathbf{v}_j$ ($j = 1, \dots, 5$) 自体が少なくなっており, そもそも一致が成立しにくい状況であることに留意すべきである. 有意な一致を示すベクトルが $\{\mathbf{v}_1, \dots, \mathbf{v}_5\}$ 内に存在しない要因としては, 真の GCD である $\bar{\mathbf{h}}(x)$ とは大きく異なる近似 GCD が存在した結果, 最小特異値に対応する特異ベクトルが真の商多項式の係数ベクトルとの類似性を持たなくなった可能性や, 特異ベクトル自体が大きく摂動され, いずれの $\mathbf{v}_j$ とも整合しない状況が生じた可能性が考えられる.

以上の予備実験から, フル・Sylvester 行列における零特異値の交代が実際に発生し, さらに雑音強度が高い状況においては特異ベクトル自体も比較的大きな摂動を受けることが確認された. このような不安定性は, 計算結果に重大な影響を及ぼす可能性があるため, 計算量が定数倍程度増加する場合であっても, この問題を回避する価値が

表 6.1: 特異値の交代の発生回数: Count of the number of Singular Value Switching[32]

σ	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$
0.01	94	4	0	0	0
0.05	55	16	2	0	0
0.1	29	10	6	0	0

あると考える．次節 6.3 及び第 6.4 節ではこの現象に着目し，雑音に対して頑健な初期値構成法を与える．なお，実験結果より，零特異値との交代が起こり得るものとして，少なくとも 3 番目に小さい特異値までを考慮することが妥当であると判断した．より次数の大きい多項式を扱う場合など，本実験では検討していない条件下における交代現象については，今後の詳細な調査が必要である．

以降では，行列の低ランク性を零空間で表現する SLRA アルゴリズムの具体例として，変数射影法に基づく手法（VP-SLRA）を用いる．これは議論の対象を明確化することを目的としたものであり，同様の議論は他のアルゴリズムにも適用可能である．

6.3 提案法 1：特異ベクトルの選択

一般に，大域的最適解に近い初期値を用いて最適化を行うことで，大域的最適解への収束確率が高まると期待される．前節の実験結果に加え，この知見を踏まえると，アルゴリズムに与える初期値として， $\mathbf{v}_1$ のみを考えるのではなく，大域的最適化において広く用いられるマルチスタート法 [117] に着想を得て， α 個の複数の右特異ベクトル $\mathbf{v}_i$, $i = 1, \dots, \alpha$ を初期値の候補とし，その中からより適切な初期ベクトル $\mathbf{x}$ を何らかの基準に従って選定するという戦略が考えられる．

選定のための基準としては，各候補 $\mathbf{v}_i$ ($i = 1, \dots, \alpha$) を初期値として用いた場合に得られる SLRA 問題の局所最適解に対して，目的関数 (4.56) を評価し，その値が最小となるものを選択する：

$$i_1 := \arg \min_{i \in \{1, \dots, \alpha\}} \mathcal{L}^{(\text{VP})}(\mathbf{v}_i).$$

ただし， $\mathcal{L}^{(\text{VP})}(\mathbf{v}) := \mathcal{L}(\mathbf{c}_*^{(\text{VP})}(\mathbf{v}))$ は， $\mathbf{v}$ を $\mathbf{x}$ の初期値として VP 法により計算した SLRA 問題の最適関数値であり， $\mathbf{c}_*^{(\text{VP})}(\mathbf{v})$ は，対応する局所最適解である．

この基準は 近似 GCD の定義に直接基づくものであり、より小さな修正差分によって所与のパラメータベクトルに対する近似が行われることが期待される点で、計算結果の妥当性を保証する指標となる。なお、応用的な立場においては、SLRA の目的関数値がある程度小さければ十分とみなし、それよりも重要視する値を最小化の基準として設けることも有用な指針となりうる。それらの基準と区別するため、以後、このように SLRA 問題本来の目的関数の大きさを（基準 A）と呼称する。

6.4 提案法2：特異ベクトルの線型結合

実際には、摂動は特異値のみならず特異ベクトルにも影響を及ぼす。これらの摂動の大きさや方向によっては、 $\mathcal{S}(\mathbf{c})$ の右特異ベクトル $\mathbf{v}_i$ ($i = 1, \dots, L$) のいずれも、真の特異ベクトル $\bar{\mathbf{v}}_1$ と十分に整合しない可能性がある。このような場合に、単一の特異ベクトルを $\mathbf{x}$ の初期推定値として用いることは、性能の劣化を招くおそれがある。このような摂動による影響を緩和するために、提案法1の初期値の候補となるベクトルの集合を拡張する方法を提案する。具体的には、追加する候補ベクトルの本数を β とし、以下の形式

$$\mathbf{x}^{(i)} = \begin{cases} \mathbf{v}_i & \text{for } i = 1, \dots, \alpha \\ \mathbf{B}\mathbf{c}^{(i-\alpha)} & \text{for } i = \alpha + 1, \dots, \gamma, \end{cases} \quad (6.2)$$

で構成される $\gamma := \alpha + \beta$ 個の初期ベクトルを用いる。ここで、 $\mathbf{B} := [\mathbf{v}_1, \dots, \mathbf{v}_\alpha] \in \mathbb{R}^{L \times \alpha}$ は候補特異ベクトルの集合を表し、 $\mathbf{c}^{(i)} \in \mathbb{R}^\alpha$ は区間 $[-1, 1]$ 上の一様分布から独立にサンプリングされた係数ベクトルである。各候補 $\mathbf{x}^{(i)}$ ($i = 1, \dots, \gamma$) に対して VP-SLRA の目的関数 $\mathcal{L}^{(\text{VP})}(\mathbf{x}^{(i)})$ を評価し、その値が最小となるものを選択する：

$$i_2 := \arg \min_{i \in \{1, \dots, \gamma\}} \mathcal{L}^{(\text{VP})}(\mathbf{x}^{(i)}). \quad (6.3)$$

6.5 数値実験

6.5.1 数値例：雑音除去性能の評価

提案する初期化戦略の、第2.1.6に示した Hankel 行列の SLRA を介した雑音除去での有効性を評価する。ここでは信号を実数（式(2.2)で生成される低ランク信号）に限定する。本実験では、 $p_j(k) = 1$ と設定することで、モデルは指数減衰型サイン波の和

に簡略化される．パラメータは $\alpha = 3$, $\beta = 15$ に設定し，重み行列 $\mathbf{W}$ には単位行列を用いた．数値実験には MATLAB [118] を使用し，最適化問題 (4.56) の解法としては，計算パッケージ [57] にて提供されている実装を既定のパラメータ設定のまま用いた．

信号モデル (2.2) に従って， $R = 5$, $n_p = 40$ の設定のもと，クリーンな信号 $\bar{\mathbf{p}} \in \mathbb{R}^{n_p}$ を生成した．この際，減衰率 d_j は区間 $[0.9, 1.1]$ 上の一様分布から，周波数 ω_j は $[10^{-6}, 1 - 10^{-6}]$ から，位相 ϕ_j は $[0, \pi]$ からそれぞれランダムにサンプリングした．観測信号 $\mathbf{p}$ は，以下のように Gauss 性雑音を加えることで構成した：

$$\mathbf{p} = \bar{\mathbf{p}} + \boldsymbol{\xi}.$$

ここで，雑音項 $\boldsymbol{\xi}$ は標準偏差 η をもつ白色 Gauss 分布に従うものとする⁴．本実験では，基準手法である VP-SLRA (C) と，提案手法 1 及び 2 (以降，M1・M2 と呼称) の性能を以下の 2 つの指標に基づいて比較した：

1. 推定誤差： $\|\mathbf{p}^{(\text{VP})}(\mathbf{x}^{(i)}) - \bar{\mathbf{p}}\|$,
2. 正規化目的関数値： $(\mathcal{L}^{(\text{VP})}(\mathbf{x}^{(i)}))^{\frac{1}{2}} / \|\mathbf{p}\|$.

これらの指標はそれぞれ，信号の推定精度および VP-SLRA の目的関数に基づく解の品質を定量的に評価するものである．図 6.2 は各手法に対する推定誤差および正規化目的関数値の箱ひげ図を，100 回の試行に基づいて示したものである．提案手法により，目的関数値をより小さくする (図 6.2 右) とともに，極端な推定誤差の発生頻度が大幅に低減されていることが確認できる (図 6.2 左)．特に，誤差のばらつきは M1 において縮小し，M2 ではさらに顕著に抑制されており，初期化の不適切さに対する頑健性の向上を示している．この改善をさらに定量的に評価するために，推定誤差が加えた雑音の標準偏差を超える試行回数を数えた．すなわち，復元された信号が真の信号よりも雑音付き観測信号から遠ざかっている場合である．このような試行は，C で 15 回，M1 で 6 回，M2 では 2 回であった．これらの結果は，提案する初期化戦略が，雑音環境下における VP-SLRA の信頼性向上に有効であることを明確に示している．

提案手法の頑健性をさらに評価するために，雑音強度 $\eta = 0.02, 0.04, \dots, 0.2$ にわたって性能を検証した．評価には，前節で導入した指標のうち二つ (推定誤差が雑音の

⁴雑音が Gauss 性のものであるため，SLRA が真の信号に対する最尤推定になっていることを付記する．

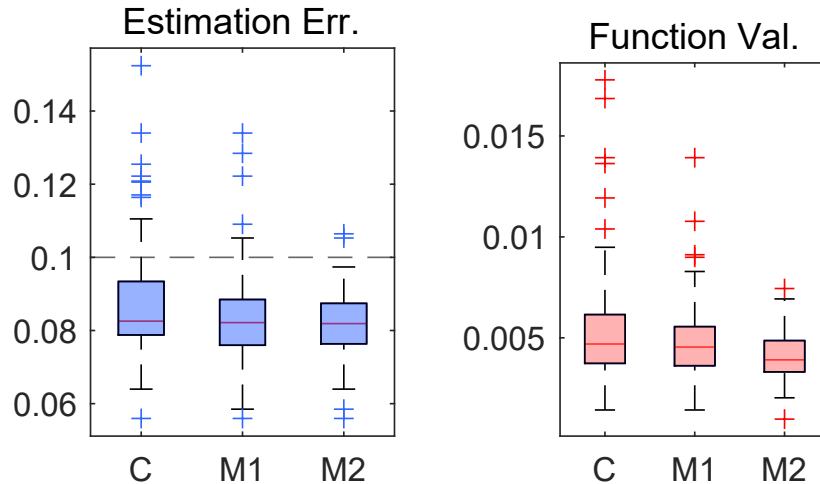


図 6.2: 雑音強度 $\eta = 0.1$ における C, M1, M2 の性能比較. 左側のグラフでは破線が雑音の標準偏差を表している.

Performance Comparison of C, M1, and M2 at Noise Level $\eta = 0.1$. In the left pannel, the ashed horizontal line indicates the noise standard deviation. [31]

標準偏差を超える試行回数及び正規化目的関数値) を用いた. 図 6.3 は, これら 2 つの指標に基づく評価結果をまとめたものである. 上段のグラフは, 推定誤差が, 雑音強度を上回った試行回数を示しており, 推定された信号が元の観測信号よりも不正確である事例を可視化している. 図 6.3 に示された結果では, 復元誤差が雑音強度を超えるような, 明らかに復元に失敗した試行回数は, C から M1, さらに M2 へと大幅に減少している. これは, 提案手法が大きな推定誤差の抑制に有効であることを示している⁵. 下段のグラフは, VP-SLRA の目的関数の正規化値を示しており, ここでも M1 および M2 は基準手法 C を上回る性能を示している. 特に M2 は, 全ての雑音強度において最も低い目的関数値を達成している. これらの結果は, 提案する初期化戦略が, 雑音環境下における VP-SLRA の頑健性を向上させることを明確に示している.

信号長の変化に伴う推定性能を検証するため, 雑音強度を $\eta = 0.02$ に固定し, 指数減衰による数値的不安定性を避けるために, 5 つの正弦波から構成される合成信号の復元を考える. 各信号長 $N = 50, 100, 200, 400, 800$ に対して, 信号および雑音をランダムに 100 回生成し, 前節と同様の 2 つの指標 (平均推定誤差および推定誤差が雑音の標準偏差を超えた試行回数) を評価した. 結果を表 6.2 に示す. M1 および M2 はいずれ

⁵大きな誤差の発生頻度を低減するだけでなく, M1 および M2 は平均推定誤差においても, C と同等またはそれ以下の値を安定して達成していることを確認した.

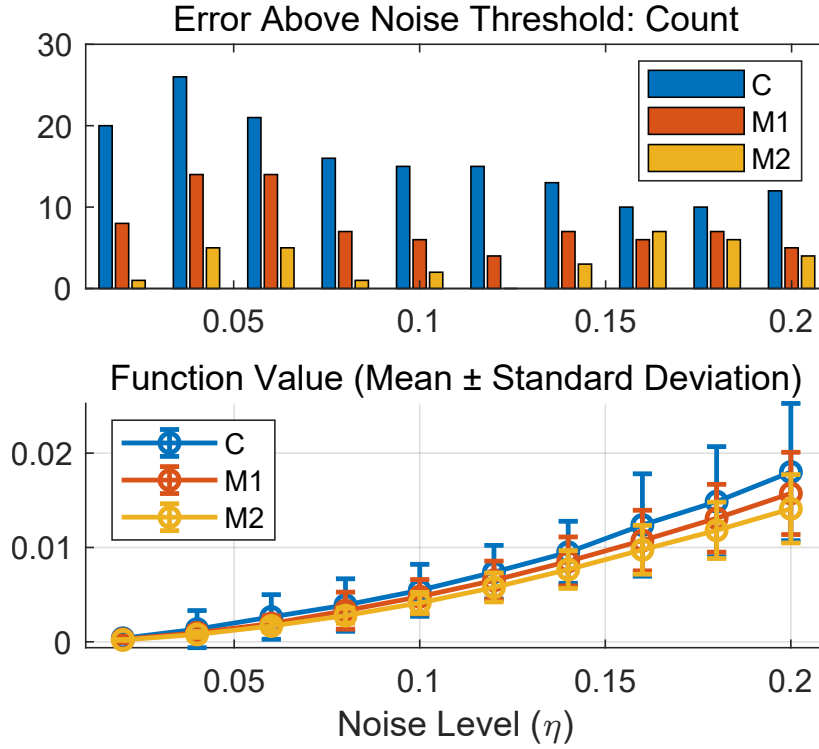


図 6.3: 雑音強度 $\eta = 0.02, 0.04, \dots, 0.2$ における C, M1, M2 の性能比較. 上段: 復元誤差が雑音標準偏差を超えた試行回数. 下段: 正規化した目的関数値

Performance comparison of C, M1, and M2 under varying noise levels $\eta = 0.02, 0.04, \dots, 0.2$. Upper: Number of Trials where the Reconstruction Error Exceeds the Noise Standard Deviation. Lower: Normalized VP Objective Function Value. [31]

も基準手法 C より優れた性能を示し、特に M2 はすべての信号長において復元誤差を最も小さく抑えるだけでなく、明らかに復元に失敗した回数もほぼ 0 に押さえ込むことに成功した。

6.5.2 VP-SLRA による近似 GCD 計算での数値例

VP-SLRA を介して近似 GCD 計算を行う場合、先に述べた (基準 A) だけではなく、この計算方法の構造的性質に着目することで、もう一つの選択/最適化基準を考えることができる。VP-SLRA を介した近似 GCD 計算では、命題 20 にて示されている関係性を利用することで、近似 GCD 問題の目的関数を直接最小化するのではなく、モザイク型 Hankel 行列に対する SLRA 問題へと帰着させて解を求めている。このようなアプローチでは、近似 GCD 問題において本来常に満たされるべき制約である式 (2.57) が、

表 6.2: まざまな信号長にわたる推定性能: Reconstruction Performance across Varying Signal Lengths ($\eta = 0.02$). [31]

N	50	100	200	400	800
Mean Estimation Error					
C	0.031	0.043	0.039	0.049	0.091
M1	0.025	0.016	0.011	0.014	0.036
M2	0.013	0.009	0.006	0.005	0.003
Error Above Noise Threshold: Count					
C	14	16	10	10	18
M1	9	8	1	3	9
M2	0	0	0	1	0

数値的には厳密に成立しない可能性がある．すなわち，式 (2.57) が満たされない場合，得られた解が本来の近似 GCD 問題とは異なる最適化問題の解となっている可能性を否定できない．この観点から，得られた解が本来の問題設定に対してどの程度整合しているかを評価する指標として，次に示す式 (2.57) の残差ノルムを用いることは有効であると考えられる．

(基準 B) 式 (2.57) に対する残差ノルム:

$$\|(\Pi_*(\mathbf{v}_j) - \mathbf{I}_{n_p}) \hat{\mathbf{p}}_*^{(n)}(\mathbf{v}_j)\|. \quad (6.4)$$

(基準 A) を採用する場合，より小さい修正差分により所与の多項式に対する近似を行うことが期待できるという点で計算結果に妥当性がある一方で，(基準 B) を採用する場合は，結果的に得られる $\hat{h}_*(x)$ と $\hat{g}_*^{(n)}(x)$ の積が $\hat{p}_*^{(n)}(x)$ と数値的に整合するという点で，計算結果に正当性があると言える．なお，(基準 B) は変数射影法の目的関数である修正差分の大きさと直接の関係を持たないが，応用的な立場においては，修正差分がある程度小さければ十分とみなし，修正差分の最小性よりも式 (2.57) の数値的整合性を重視する選択が合理的で有用な指針となりうる．

本節では，提案法 2 を拡張し，特異ベクトルの線型結合係数を粒子群最適化 (Particle Swarm Optimization: PSO[119]) により最適化するものを，提案法 2 の代わりに用いる．この手法は，次の最小化問題

$$\mathbf{a}_* := \arg \min_{\mathbf{a}} \mathcal{L}^{(\text{VP})}(\mathbf{B}\mathbf{a}). \quad (6.5)$$

を PSO により解くものであり、ランダムな線型結合からの選択方式と区別するため、提案法 2(PSO) と呼ぶ。

提案法 1 は同一の近似 GCD 問題に対して異なる α 個の商多項式の初期推定値を与え、それぞれを個別に解くため、計算量は Algorithm 1 の α 倍程度である。一方提案法 2(PSO) では、PSO による最適化の過程で、初期集団生成時及び集団の更新の度に、全ての粒子の位置に対し Algorithm 1 を実行する必要がある。このため、関数値が収束せず、集団の更新の回数が事前に設定した上限に達するような場合、1 回の提案法 2(PSO) の実行中に評価回数の上限である (集団の更新回数の上限 + 1) $\times$ (集団内の粒子数) 回 Algorithm 1 が実行される。

(基準 A/B) に関する選択、及び最適化により得る計算結果が、それらに期待される (第 6.3 節で述べた) 正当性を有することを数値実験により検証する (検証項目 1)。さらに、本実験では、近似 GCD 計算を真の GCD の推定問題 18 に応用した場合の提案手法の性能 (検証項目 2) も併せて評価する。実験に使用した計算機は、プロセッサが Intel Xeon CPU E5-2697 v4@2.30GHz (18 コア)、メモリが 128GB であり、OS が Windows10 という構成のものである。本実験には、科学技術計算ソフトウェア MATLAB[118] を使用し、PSO の実装は、MATLAB の `particleswarm` 関数 [120] を使用する。PSO のパラメータは、`SwarmSize` を $5 \times \alpha$ 、`InitialSwarmSpan` を 2、`FunctionTolerance` を 10^{-2} 、`MaxIterations` を 5 と設定し、言及していないものはデフォルトのものを使用する。商多項式の初期推定値の候補となるベクトルの本数 α は 3 とする。以降 (基準 A/B) を採用した際の提案法 1 を SEL(A/B)、提案法 2(PSO) を PSO(A/B) と呼称する。

6.5.3 異なる次数の多項式に対する提案法の有効性

問題 (18) において、真の多項式として [71, §4.6] に類似する多項式を考え、

$$\begin{aligned}\bar{g}^{(1)}(t) &= (t^{kl_1} - 1)(t^{kl_2} - 2)(t^{kl_3} - 3), \\ \bar{g}^{(2)}(t) &= (t^{kl_1} + 1)(t^{kl_2} + 5)(t^{kl_3} + 2)\end{aligned}$$

とし、 $\bar{h}(t) = t^4 + 10t^2 + t - 1$ とする。

$$\bar{p}^{(1)}(t) = \bar{g}^{(1)}(t) \cdot \bar{h}(t), \quad \bar{p}^{(2)}(t) = \bar{g}^{(2)}(t) \cdot \bar{h}(t)$$

とし、係数のノルムによって正規化を行う．これ対し、雑音 $\eta^{(n)}(x)$ を加えた $p^{(n)}(x) = \bar{p}^{(n)}(x) + 10^{-2}\eta^{(n)}(x)$ の近似 GCD を、 $\xi = \hat{h}$ とした提案法 1, 2 でそれぞれ計算したものを $\bar{h}(x)$ に対する推定値とし、解の性質を以下の項目

$$J_a := \frac{\mathcal{L}_1(\hat{h}_*(v_*))}{\|p^{(n)}\|^2} \quad (6.6)$$

$$J_b := \left\| (\Pi_*(v_*) - \mathbf{I}_{n_p}) \frac{\hat{p}_*^{(n)}(v_*)}{\|p^{(n)}\|} \right\| \quad (6.7)$$

$$J_c := \psi(\bar{h}, \hat{h}_*(v_*)) \quad (6.8)$$

$$J_d := \psi(\bar{g}^{(n)}, \hat{g}_*^{(n)}(v_*)) \quad (6.9)$$

$$J_e := \text{cond}(\mathcal{M}_{k,l}(\xi_*(v_*))) \quad (6.10)$$

により評価する．ここで $\text{cond}(\cdot)$ は所与の行列の条件数である．なお、 $\eta^{(i)}$ ($i = 1, \dots, \alpha$) は標準正規分布に従う各要素を持つベクトルとしてサンプリングされ、そのノルムが 1 になるように正規化されている．

評価項目 J_a は正規化した修正差分の 2 乗ノルムであり、上記評価項目のうち、近似 GCD の定義本来の基準を評価する唯一の項目である． J_b は正規化した残差ノルム、 J_e はアルゴリズムの出力 $\hat{p}_*^{(n)}(v_*)$, $\xi_*(v_*)$ に対し定まる y_* についての方程式 (2.57) の悪条件度合を表し、これら二つは得られた解が本来満たすべき式 (2.57) にどの程度整合しているかを評価する．以上 J_a, J_b, J_e は検証項目 1 のための指標である．

評価項目 J_c と J_d は、近似 GCD 問題を通じて推定される真の GCD 及び商多項式に対する推定精度⁶を、係数ベクトルのコサイン類似度により評価するものであり、検証項目 2 のための指標である．多項式同士の類似度は零点集合の近接性により比較することも可能であるが、零点の計算は数値的に不安定であり、本実験のように雑音の影響が大きい場合には適さないと考えられる．また、画像復元 [10] や残響除去 [121] への近似 GCD の応用では、推定対象は多項式の係数そのものである．係数ベクトルの類似度を測る方法としては、正規化後の差分ノルムなども考えられるが、向きやスケールの任意性に影響されない指標が好ましい．これらの理由によりコサイン類似度を用いている．

⁶注 17 で述べたように、この条件では、近似 GCD 問題は真の GCD に対する最尤推定として機能する．

比較のため、 v_1 を商多項式の初期値とする Algorithm 1（従来法）でも計算を行う。 v_* は従来法に対しては v_1 を表す。

$k = 1, 2, \dots, 5$ に対して以上の計算を 50 の異なる雑音に対して行った際の、 J_a から J_e の平均値と標準偏差を表 6.3 に示した。なお、上記の計算において、PSO を用いる手法はいずれも 90 回（関数評価回数の上限）の Alogorithm 1 の計算を必要とした。

（基準 A）を用いる手法は、他手法と比較して多項式の次数に依存せず、より微少な修正差分を与えた。特に、提案法 2(PSO) により（基準 A）を最小化することで、他の手法と比べて顕著に小さな修正差分を得た。また、 J_b や J_e が従来法と比べて著しく悪化することはなく、数値的整合性や安定性の観点でも一定の信頼性が保たれている。

（基準 B）を用いる手法は、修正差分をある程度小さく保ちつつ、数値的整合性を重視する基準に基づいて初期値を選定した。実際、 J_a は従来法より平均的に大きくなったものの、加えられた雑音ノルムと比較して十分に小さい値であり、（基準 B）の設計意図に沿った結果と評価できる。また、PSO(B) は GCD 及び商多項式の推定精度が高く、特に $k \geq 3$ においては他の手法と比較して顕著に良好な結果を示した。

6.5.4 異なる雑音強度に対する提案法の有効性

3つの多項式 ($N = 3$) に対する真の GCD の推定を、近似 GCD 計算により行う。真の多項式 $\bar{g}^{(1)}(t), \bar{g}^{(2)}(t), \bar{g}^{(3)}(t) \in \mathcal{P}_{20}$ 及び $\bar{h}(t) \in \mathcal{P}_{10}$ の各要素を、範囲が -1 から 1 の連続一様乱数により決定する。 $\bar{p}^{(n)}(t) = (\bar{g}^{(n)} \cdot \bar{h})(t)$ とし、係数のノルムによる正規化を行う。これに対し、 $\eta^{(i)}$ が平均が 0 で標準偏差が α の Gauss 性雑音 $\eta^{(i)}(t)$ を加えた $p^{(n)}(t) = \bar{p}^{(n)}(t) + \eta^{(n)}(t)$ の近似 GCD 問題を、 $\xi = \hat{g}^{(n)}$ とした提案法 1, 2(PSO) でそれぞれ計算し、最適性及び GCD の推定精度を第 6.5.3 節と同様の指標により評価する。

以上の計算を異なる雑音の標準偏差 $\alpha = 0.4, 0.2, 0.1, 0.05$ に対しそれぞれ 50 回計算した際の、 J_a から J_e の平均及び標準偏差を表 6.4 に示した。この実験においても、PSO を用いる手法はいずれも 90 回の Alogorithm 1 の計算を必要とした。

表 6.3: 異なる次数の多項式に対する計算結果. 各評価項目を (平均) $\pm$ (標準偏差) の形式で表示 (太字は各評価項目の平均値の最良). Results obtained for different degrees of polynomials. Each evaluation measure is displayed as (mean) $\pm$ (standard deviation). The best mean value for each measure is highlighted. [32]

k	従来法	SEL(A)	PSO(A)	SEL(B)	PSO(B)
1	J_a	$2.42 \times 10^{-4} \pm 4.14 \times 10^{-4}$	$1.70 \times 10^{-4} \pm 1.23 \times 10^{-4}$	$1.65 \times 10^{-4} \pm 1.23 \times 10^{-4}$	$1.10 \times 10^{-3} \pm 1.77 \times 10^{-3}$
	J_b	$2.91 \times 10^{-16} \pm 7.91 \times 10^{-17}$	$2.96 \times 10^{-16} \pm 8.16 \times 10^{-17}$	$3.17 \times 10^{-16} \pm 9.15 \times 10^{-17}$	$2.38 \times 10^{-16} \pm 4.88 \times 10^{-17}$
	J_c	0.78 ± 0.21	0.78 ± 0.22	0.77 ± 0.23	0.74 ± 0.33
	J_d	0.62 ± 0.32	0.64 ± 0.32	0.62 ± 0.34	0.66 ± 0.36
	J_e	$2.10 \times 10^0 \pm 1.28 \times 10^0$	$2.49 \times 10^0 \pm 1.68 \times 10^0$	$2.79 \times 10^0 \pm 1.94 \times 10^0$	$1.73 \times 10^0 \pm 8.59 \times 10^{-1}$
2	J_a	$6.26 \times 10^{-4} \pm 1.25 \times 10^{-3}$	$1.55 \times 10^{-4} \pm 1.88 \times 10^{-4}$	$1.22 \times 10^{-4} \pm 9.28 \times 10^{-5}$	$8.58 \times 10^{-4} \pm 1.51 \times 10^{-3}$
	J_b	$3.88 \times 10^{-16} \pm 1.03 \times 10^{-16}$	$4.17 \times 10^{-16} \pm 1.29 \times 10^{-16}$	$4.08 \times 10^{-16} \pm 9.71 \times 10^{-17}$	$3.08 \times 10^{-16} \pm 6.92 \times 10^{-17}$
	J_c	0.66 ± 0.40	0.69 ± 0.34	0.73 ± 0.29	0.87 ± 0.28
	J_d	0.63 ± 0.38	0.56 ± 0.42	0.58 ± 0.40	0.83 ± 0.29
	J_e	$2.77 \times 10^0 \pm 1.54 \times 10^0$	$2.86 \times 10^0 \pm 1.53 \times 10^0$	$2.99 \times 10^0 \pm 1.33 \times 10^0$	$1.88 \times 10^0 \pm 1.12 \times 10^0$
3	J_a	$4.00 \times 10^{-4} \pm 9.18 \times 10^{-4}$	$1.24 \times 10^{-4} \pm 9.85 \times 10^{-5}$	$1.02 \times 10^{-4} \pm 8.88 \times 10^{-5}$	$5.12 \times 10^{-4} \pm 1.24 \times 10^{-3}$
	J_b	$5.33 \times 10^{-16} \pm 1.66 \times 10^{-16}$	$4.96 \times 10^{-16} \pm 1.71 \times 10^{-16}$	$5.76 \times 10^{-16} \pm 1.82 \times 10^{-16}$	$4.01 \times 10^{-16} \pm 1.04 \times 10^{-16}$
	J_c	0.67 ± 0.29	0.70 ± 0.30	0.58 ± 0.30	0.73 ± 0.35
	J_d	0.43 ± 0.34	0.50 ± 0.39	0.36 ± 0.33	0.62 ± 0.37
	J_e	$5.92 \times 10^0 \pm 5.32 \times 10^0$	$5.53 \times 10^0 \pm 5.29 \times 10^0$	$8.30 \times 10^0 \pm 6.16 \times 10^0$	$3.04 \times 10^0 \pm 2.03 \times 10^0$
4	J_a	$4.73 \times 10^{-4} \pm 1.00 \times 10^{-3}$	$1.08 \times 10^{-4} \pm 7.15 \times 10^{-5}$	$8.19 \times 10^{-5} \pm 5.90 \times 10^{-5}$	$8.32 \times 10^{-4} \pm 1.62 \times 10^{-3}$
	J_b	$7.12 \times 10^{-16} \pm 3.37 \times 10^{-16}$	$7.19 \times 10^{-16} \pm 3.04 \times 10^{-16}$	$8.55 \times 10^{-16} \pm 2.67 \times 10^{-16}$	$4.18 \times 10^{-16} \pm 1.53 \times 10^{-16}$
	J_c	0.70 ± 0.25	0.55 ± 0.32	0.49 ± 0.24	0.79 ± 0.36
	J_d	0.39 ± 0.40	0.32 ± 0.38	0.26 ± 0.25	0.71 ± 0.39
	J_e	$1.01 \times 10^1 \pm 8.44 \times 10^0$	$1.11 \times 10^1 \pm 8.56 \times 10^0$	$1.36 \times 10^1 \pm 7.04 \times 10^0$	$3.08 \times 10^0 \pm 4.24 \times 10^0$
5	J_a	$5.85 \times 10^{-4} \pm 1.13 \times 10^{-3}$	$1.85 \times 10^{-4} \pm 5.07 \times 10^{-4}$	$7.10 \times 10^{-5} \pm 5.05 \times 10^{-5}$	$1.31 \times 10^{-3} \pm 1.87 \times 10^{-3}$
	J_b	$7.62 \times 10^{-16} \pm 3.29 \times 10^{-16}$	$7.16 \times 10^{-16} \pm 3.27 \times 10^{-16}$	$8.55 \times 10^{-16} \pm 3.37 \times 10^{-16}$	$4.61 \times 10^{-16} \pm 2.21 \times 10^{-16}$
	J_c	0.64 ± 0.24	0.69 ± 0.21	0.56 ± 0.19	0.81 ± 0.23
	J_d	0.29 ± 0.33	0.33 ± 0.36	0.21 ± 0.22	0.56 ± 0.39
	J_e	$1.28 \times 10^1 \pm 8.89 \times 10^0$	$1.28 \times 10^1 \pm 9.87 \times 10^0$	$1.65 \times 10^1 \pm 1.15 \times 10^1$	$5.72 \times 10^0 \pm 4.61 \times 10^0$
					$3.62 \times 10^0 \pm 3.76 \times 10^0$

表 6.4: 異なる雑音強度に対する計算結果. 各評価項目を (平均) $\pm$ (標準偏差) の形式で表示 (太字は各評価項目の平均値の最良). Results obtained for different noise level. Each evaluation measure is displayed as (mean) $\pm$ (standard deviation). The best mean value for each measure is highlighted. [32]

σ	従来法		SEL(A)	PSO(A)	SEL(B)	PSO(B)
0.4	J_a	$2.52 \times 10^{-2} \pm 9.96 \times 10^{-3}$	$2.20 \times 10^{-2} \pm 5.53 \times 10^{-3}$	$1.86 \times 10^{-2} \pm 4.90 \times 10^{-3}$	$2.71 \times 10^{-2} \pm 7.74 \times 10^{-3}$	$2.82 \times 10^{-2} \pm 9.92 \times 10^{-3}$
	J_b	$7.35 \times 10^{-16} \pm 4.52 \times 10^{-16}$	$1.42 \times 10^{-15} \pm 4.21 \times 10^{-15}$	$1.71 \times 10^{-15} \pm 3.80 \times 10^{-15}$	$4.83 \times 10^{-16} \pm 1.24 \times 10^{-16}$	$4.54 \times 10^{-16} \pm 8.53 \times 10^{-17}$
	J_c	0.56 ± 0.31	0.55 ± 0.30	0.43 ± 0.30	0.64 ± 0.26	0.66 ± 0.28
	J_d	0.46 ± 0.26	0.42 ± 0.27	0.33 ± 0.24	0.55 ± 0.21	0.54 ± 0.26
	J_e	$2.15 \times 10^1 \pm 2.31 \times 10^1$	$7.30 \times 10^1 \pm 3.35 \times 10^2$	$9.14 \times 10^1 \pm 3.36 \times 10^2$	$9.73 \times 10^0 \pm 9.06 \times 10^0$	$6.73 \times 10^0 \pm 4.73 \times 10^0$
0.2	J_a	$8.91 \times 10^{-3} \pm 3.82 \times 10^{-3}$	$8.01 \times 10^{-3} \pm 2.80 \times 10^{-3}$	$7.21 \times 10^{-3} \pm 1.90 \times 10^{-3}$	$1.02 \times 10^{-2} \pm 6.36 \times 10^{-3}$	$9.52 \times 10^{-3} \pm 4.78 \times 10^{-3}$
	J_b	$5.27 \times 10^{-16} \pm 3.18 \times 10^{-16}$	$5.31 \times 10^{-16} \pm 2.92 \times 10^{-16}$	$7.12 \times 10^{-16} \pm 5.09 \times 10^{-16}$	$4.20 \times 10^{-16} \pm 7.58 \times 10^{-17}$	$4.00 \times 10^{-16} \pm 5.83 \times 10^{-17}$
	J_c	0.79 ± 0.26	0.78 ± 0.25	0.76 ± 0.25	0.80 ± 0.26	0.89 ± 0.17
	J_d	0.72 ± 0.28	0.71 ± 0.29	0.61 ± 0.33	0.75 ± 0.25	0.83 ± 0.18
	J_e	$1.18 \times 10^1 \pm 1.91 \times 10^1$	$1.51 \times 10^1 \pm 2.47 \times 10^1$	$2.43 \times 10^1 \pm 4.27 \times 10^1$	$7.23 \times 10^0 \pm 4.57 \times 10^0$	$5.99 \times 10^0 \pm 2.73 \times 10^0$
0.1	J_a	$2.44 \times 10^{-3} \pm 1.61 \times 10^{-3}$	$2.37 \times 10^{-3} \pm 1.58 \times 10^{-3}$	$2.11 \times 10^{-3} \pm 5.46 \times 10^{-4}$	$5.16 \times 10^{-3} \pm 9.89 \times 10^{-3}$	$3.97 \times 10^{-3} \pm 4.34 \times 10^{-3}$
	J_b	$4.50 \times 10^{-16} \pm 2.67 \times 10^{-16}$	$5.14 \times 10^{-16} \pm 4.14 \times 10^{-16}$	$4.92 \times 10^{-16} \pm 3.27 \times 10^{-16}$	$3.91 \times 10^{-16} \pm 4.81 \times 10^{-17}$	$3.89 \times 10^{-16} \pm 5.12 \times 10^{-17}$
	J_c	0.93 ± 0.19	0.92 ± 0.20	0.92 ± 0.22	0.91 ± 0.22	0.89 ± 0.23
	J_d	0.90 ± 0.20	0.88 ± 0.25	0.89 ± 0.25	0.87 ± 0.25	0.89 ± 0.17
	J_e	$9.93 \times 10^0 \pm 2.01 \times 10^1$	$1.14 \times 10^1 \pm 2.21 \times 10^1$	$1.10 \times 10^1 \pm 2.14 \times 10^1$	$6.06 \times 10^0 \pm 4.06 \times 10^0$	$5.89 \times 10^0 \pm 2.51 \times 10^0$
0.05	J_a	$5.81 \times 10^{-4} \pm 2.54 \times 10^{-4}$	$5.81 \times 10^{-4} \pm 2.54 \times 10^{-4}$	$5.45 \times 10^{-4} \pm 1.41 \times 10^{-4}$	$1.52 \times 10^{-3} \pm 5.27 \times 10^{-3}$	$3.47 \times 10^{-3} \pm 8.30 \times 10^{-3}$
	J_b	$4.62 \times 10^{-16} \pm 3.59 \times 10^{-16}$	$4.48 \times 10^{-16} \pm 3.51 \times 10^{-16}$	$4.35 \times 10^{-16} \pm 2.95 \times 10^{-16}$	$3.93 \times 10^{-16} \pm 3.68 \times 10^{-17}$	$3.95 \times 10^{-16} \pm 4.59 \times 10^{-17}$
	J_c	0.96 ± 0.16	0.97 ± 0.13	0.98 ± 0.13	0.94 ± 0.19	0.91 ± 0.21
	J_d	0.95 ± 0.17	0.97 ± 0.12	0.98 ± 0.12	0.94 ± 0.17	0.90 ± 0.19
	J_e	$8.15 \times 10^0 \pm 1.78 \times 10^1$	$7.88 \times 10^0 \pm 1.78 \times 10^1$	$7.93 \times 10^0 \pm 1.78 \times 10^1$	$5.37 \times 10^0 \pm 2.32 \times 10^0$	$5.36 \times 10^0 \pm 2.32 \times 10^0$

(基準 A) を用いる手法は雑音の強度によらず、他手法と比べて微小な修正差分を計算した。特に雑音強度が低い場合には、PSO(A) は GCD 及び商多項式の推定精度において従来法を上回る性能を示し、SEL(A) も PSO(A) には及ばないものの、従来法より良好な結果を示した。しかしながら、雑音強度が高い場合 ($\sigma = 0.4, 0.2$) にアルゴリズムの出力 $\hat{p}_*^{(n)}(v_*)$, $\xi_*(v_*)$ に対し定まる式 (2.57) が悪条件であったことに注意が必要である。

(基準 B) を用いる手法は、アルゴリズムの出力により定まる式 (2.57) が悪条件となることを回避した。特に、PSO(B) は雑音強度が高い場合 ($\sigma = 0.4, 0.2$) に、他手法と比べて高い精度で GCD 及び商多項式を推定した。しかしながら雑音強度が低い場合に、従来法と比較してそれらの推定精度がむしろ悪化した。

6.5.5 考察

検証項目 1 については、上記の二つの数値実験で (基準 A/B) 双方が、それぞれに期待される正当性を有することを確認した。

検証項目 2 に関しては、(基準 B) を用いた手法が、第 6.5.3 節及び第 6.5.4 節におけるいくつかの設定において、推定精度の向上に寄与する結果を示した。このことは、「修正差分がある程度小さければ十分とみなし、修正差分の最小性よりも式 (2.57) の数値的整合性を重視する」という (基準 B) に基づく選択が、雑音に対し平均 0 の白色 Gauss 性が仮定される真の GCD の推定精度に何らかの影響を与えている可能性を示唆している。一方、第 6.5.4 節における雑音強度の低い設定では、(基準 B) を用いた手法が従来法と比較して推定精度を悪化させる傾向が見られた。しかしながら、これをもって (基準 B) が低雑音環境に不適であると結論づけることはできない。実際、第 6.5.3 節における $k = 3$ のケースは、第 6.5.3 節で扱う多項式の次数に近く、また設定された雑音強度は 0.01 と第 6.5.4 節よりも小さいにもかかわらず、(基準 B) を用いた手法が従来法よりも推定精度を悪化させる結果にはなっていない。このような傾向の違いを生む要因として、例えば第 6.5.3 節では重根を多く含む設定であるのに対し、第 6.5.4 節ではそのような構造を持たないなど、多項式の構造的な違いが影響している可能性が考えられる。しかしながら、これらは現時点では著者の推測の域を出るものではなく、

今後の詳細な検討が必要である。加えて、式 (2.57) の整合性と推定精度との間に理論的な因果関係が存在するかどうかについても、現時点では明らかになっておらず、今後の理論的及び実証的な検証が求められる。

6.6 本章のまとめ

本章では、行列のランクを零空間で表現する SLRA アルゴリズムにおける初期化戦略の重要性を明確化し、従来法の脆弱性を引き起こす要因として「特異値の交代」現象を指摘した。従来の初期化戦略は、観測行列の最小特異値に対応する右特異ベクトルを初期値として用いることを前提としていたが、雑音や摂動の影響によりこの仮定が成立しない場合があることを示した。この問題に対処するため、本章では VP-SLRA を対象として、次の二つの戦略を提案した。

1. 複数の候補ベクトルを評価し、最良のものを選択する方法。
2. 特異ベクトルの最良な線形結合係数を探索する方法。これはさらに、ランダムに生成した係数から最良を選ぶ方法と、PSO (Particle Swarm Optimization) により係数を最適化する方法の二つに分かれる。

さらに、修正差分のノルムおよび数値的整合性に基づく選択・最適化基準を導入し、SLRA 計算を介した真値推定問題に関する数値実験により、提案手法が従来法に比べて推定精度を改善する場合があることを確認した。特に、SLRA により計算される近似最大公約多項式を用いて真の最大公約多項式を推定する問題において、雑音が平均 0 の白色 Gauss 性を持つ場合に、提案戦略が性能向上に寄与することを示した。一方で、提案手法は計算コストが高くなる傾向があり、今後は効率的な探索アルゴリズムの導入や計算量削減が課題として残されている。また、初期値候補ベクトルの選択基準と推定精度との理論的關係については未解明であり、さらなる解析が望まれる。

第7章 結論

多くの工学の問題において、データベクトルに線形モデルを仮定した場合、その要素間の関係が、モデルに対応する行列構造を持つデータ行列と係数行列による線形方程式で表現される。言い換えると、データ行列の構造はモデルのクラスに対応し、そのランクがモデルの複雑さを決定する。一般論として、モデルの複雑さと精度のバランスは重要であり、工学的応用では単純な近似モデルが有用である。このため、所与のデータを所望のモデルに適合させる近似が必要となり、問題は構造化低ランク近似 (SLRA) として定式化される。

SLRA 問題を解くために、これまで多様なアルゴリズムが提案されてきた。しかし、現実の問題を SLRA に帰着させる場合、行列構造、計算時間の制約、要求精度、計算資源などの条件に応じて適切なアルゴリズムは異なる。本論文では、以下の二点に焦点を当て、理論的妥当性の検討と改善手法を提案した。

1. 構造化全最小ノルム法に基づくアルゴリズム (STLN-SLRA) の定式化
2. 低ランク性の零空間表現を使用する SLRA アルゴリズムにおける初期化戦略

STLN-SLRA の定式化に関する改善 STLN-SLRA では、SLRA 問題を構造化された初期方程式の近似求解問題へと変形する。初期方程式はいくつかの選択の余地があるが、従来は、初期方程式の選択に関する明確な基準がなく、全ての候補を構成した上で残差ノルム最小のものを選択する方法が採用されていた。この方法は計算量が近似対象の行列の列数 L に対し $O(L^4)$ と大きく、妥当性の検証も不十分であった。本論文では、従来の定式化を解析し、計算誤差を最小化する初期方程式の選択基準を導出した。この基準に基づく計算量は $O(L^3)$ であり、従来法に比べて大幅な効率化を達成した。さらに、数値実験により、提案基準が計算時間を短縮しつつ精度を維持できることを確認した。特に、変数の数が多い場合に高い効率性を示すことが明らかとなった。

零空間基底の初期化戦略に関する改善 零空間表現を用いる SLRA アルゴリズムでは、零空間基底ベクトルも最適化変数として扱われる。ランク制約は非凸であるため、実用上は凸最適化手法が多く用いられるが、この場合、初期値依存性が高いことが問題となる。従来は、観測行列の最小特異値に対応する右特異ベクトルを初期値として使用する戦略が一般的であった。しかし、本論文では、雑音による特異値および特異ベクトルの摂動により、零特異値に対応すべきベクトルが最小特異値に対応しなくなる「特異値の交代」現象を指摘し、この戦略の脆弱性を明確化した。

この問題に対処するため、(1) 複数の右特異ベクトルから初期値を選択する方法と、(2) それらの最良な線形結合係数を探索する方法の二つを提案した。二つ目の方法は、ランダムに生成した係数から最良を選択する手法と、PSO (Particle Swarm Optimization) により係数を最適化する手法に分かれる。さらに、両戦略に対して、修正差分のノルムと数値的整合性に基づく選択・最適化基準を導入した。数値実験では、雑音が平均 0 の白色 Gauss 性を持つと仮定される近似最大公約多項式計算において、提案手法が従来法に比べて推定精度を向上させる場面があることを確認した。一方で、最適化手法は計算コストが高くなる傾向があり、今後は計算量の削減や効率的な探索アルゴリズムの導入が課題として残されている。また、近似最大公約多項式計算に適用した際の式 (2.57) に基づく整合性と推定精度との理論的因果関係については未解明であり、さらなる検討が望まれる。

総括と今後の課題 本論文の成果は、SLRA アルゴリズムの理論的妥当性を向上させるとともに、実用上の性能改善に寄与するものである。今後は、提案手法の計算効率化、より広範な行列構造への適用、ならびに理論的解析の深化が重要な課題となる。

謝辞

本研究を行うにあたって、多忙な中、指導教員として多くの御指導、御助言を頂きました。情報数理学研究室教授 田中章先生、副指導教員として貴重なご助言をいただきました。知能情報学研究室教授 杉本雅則先生、同准教授 中村将成先生、情報解析学研究室教授 今井英幸先生、構造化知能研究室教授 松原崇先生 に深く感謝いたします。また、研究、論文作成を行うにあたり御協力頂いた情報数理学研究室内の皆様に感謝いたします。

中学からの同期の三田村拓磨様、徳山芳樹様と共に実施した、「Tu 多様体」をはじめとする、数学書の輪読ゼミを通して得られた数学的な知見や刺激が本研究に与えた影響はとて大きいものであり、ここに感謝します。北大オーケストラのヴィオラパートにおいて、共に苦楽を分かち合った同期の榎聡美様、杉山優様、高橋秀彦様、谷口諒様、森本珠美様、中田ももこ様に深く感謝申し上げます。また、札幌西高校オーケストラ部から北大オーケストラまで長きにわたり共に活動し、「優斗を追い出す曲」の演奏に関わってくださった田中優斗様、野呂菜々子様、樋口彼方様、南佐京様、米田菜乃様に心より御礼申し上げます。さらに、コロナ禍という困難な時期を『Monster Hunter World: Iceborne』と共に乗り越える中で、サークル「まろおの囲い」の皆様に多大なる支えをいただきましたことに深く感謝いたします。特に、長い時間を共に過ごしたキエ氏、シロン氏、ダイオウイカ氏、Dione 氏、マイケル氏、夜々氏に謝意を表し、今後のご活躍に幸多からんことを心よりお祈り申し上げます。また、サークル「嘘と虚無の木漏れ日」において旗下かける氏との音楽制作活動を通じて、研究に新たなインスピレーションを得ることができました。ここに深甚なる感謝の意を表します。これらの方々と過ごした時間は、大学院での研究生活において欠くことのできないものであり、本論文の完成に大きく寄与したと確信しております。

最後に、大学私生活を長年にわたりサポートしてくれていた、父 吉野巖様、母 吉野

聖子様，妹 吉野美桜様，祖母 吉野昌子様・櫻井宏子様，祖父 吉野幸男様・櫻井善介様，叔父 荒川靖夫様・加賀屋安章様，叔母 荒川真美様・加賀屋道子様にも感謝します。

参考文献

- [1] K. Usevich and I. Markovsky: “Variable projection methods for approximate (greatest) common divisor computations”, *Theoretical Computer Science*, **681**, pp. 176–198 (2017).
- [2] I. Markovsky: “Low Rank Approximation: Algorithms, Implementation, Applications”, Springer, London, 2nd edition (2018).
- [3] M. Ishteva, K. Usevich and I. Markovsky: “Factorization approach to structured low-rank approximation with applications”, *SIAM Journal on Matrix Analysis and Applications*, **35**, pp. 1180–1204 (2014).
- [4] N. Yoshino and A. Tanaka: “Computationally efficient selection criterion of initial equation in STLN-based structured low-rank approximation”, *Journal of Computational and Applied Mathematics*, **470**, p. 116703 (2025).
- [5] I. Markovsky, J. Goos, K. Usevich and R. Pintelon: “Realization and identification of autonomous linear periodically time-varying systems”, *Automatica*, **50**, pp. 1632–1640 (2014).
- [6] A. Fazzi, N. Guglielmi and I. Markovsky: “An ODE-based method for computing the approximate greatest common divisor of polynomials”, *Numerical Algorithms*, **81**, pp. 719–740 (2019).
- [7] E. Kaltofen, Z. Yang and L. Zhi: “Structured low rank approximation of a Sylvester matrix”, *Trends in Mathematics*, **41**, pp. 69–83 (2007).

- [8] J. R. Winkler and J. D. Allan: “Structured total least norm and approximate GCDs of inexact polynomials”, *Journal of Computational and Applied Mathematics*, **215**, pp. 1–13 (2008).
- [9] E. J. Candès and C. Fernandez-Granda: “Super-resolution from noisy data”, *Journal of Fourier Analysis and Applications*, **19**, pp. 1229–1254 (2013).
- [10] J. R. Winkler: “Polynomial computations for blind image deconvolution”, *Linear Algebra and its Applications*, **502**, pp. 77–103 (2016).
- [11] I. Markovsky and P. L. Dragotti: “Using Hankel structured low-rank approximation for sparse signal recovery”, *Latent Variable Analysis and Signal Separation* (Eds. by Y. Deville, S. Gannot, R. Mason, M. D. Plumbley and D. Ward), Springer International Publishing, pp. 479–487 (2018).
- [12] J. Gillard and K. Usevich: “Structured low-rank matrix completion for forecasting in time series analysis”, *International Journal of Forecasting*, **34**, pp. 582–597 (2018).
- [13] X. Zhang, H. Lu, D. Guo, Z. Lai, H. Ye, X. Peng, B. Zhao and X. Qu: “Accelerated MRI reconstruction with separable and enhanced low-rank hankel regularization”, *IEEE Transactions on Medical Imaging*, **41**, pp. 2486–2498 (2022).
- [14] J. He, Y. Wang and Y. Zhou: “Three dimensional seismic data reconstruction based on truncated nuclear norm”, *Journal of Applied Geophysics*, **214**, p. 105049 (2023).
- [15] E. J. Candès and Y. Plan: “Matrix completion with noise”, *Proceedings of the IEEE*, **98**, 6, pp. 925–936 (2010).
- [16] E. J. Candès, Y. C. Eldar, T. Strohmer and V. Voroninski: “Phase retrieval via matrix completion”, *SIAM Journal on Imaging Sciences*, **6**, pp. 199–225 (2013).

-
- [17] I. Markovsky: “Structured low-rank approximation and its applications”, *Automatica*, **44**, 4, pp. 891–909 (2008).
 - [18] Z. Liu and L. Vandenberghe: “Interior-point method for nuclear norm approximation with application to system identification”, *SIAM Journal on Matrix Analysis and Applications*, **31**, pp. 1235–1256 (2010).
 - [19] É. Schost and P. J. Spaenlehauer: “A quadratically convergent algorithm for structured low-rank approximation”, *Foundations of Computational Mathematics*, **16**, pp. 457–492 (2016).
 - [20] N. Zvonarev and N. Golyandina: “Fast and stable modification of the Gauss-Newton method for low-rank signal estimation”, *Numerical Linear Algebra with Applications*, **29**, (2022).
 - [21] K. Usevich and I. Markovsky: “Variable projection for affinely structured low-rank approximation in weighted 2-norms”, *Journal of Computational and Applied Mathematics*, **272**, pp. 430–448 (2014).
 - [22] T. Saha and S. Khare: “Structured low-rank approximation: Optimization on matrix manifold approach”, *International Journal of Applied and Computational Mathematics*, **7**, p. 242 (2021).
 - [23] J. Gillard and K. Usevich: “Hankel low-rank approximation and completion in time series analysis and forecasting: a brief review”, *Statistics and Its Interface*, **16**, pp. 287–303 (2023).
 - [24] J. R. Winkler and M. Hasan: “A non-linear structure preserving matrix method for the low rank approximation of the Sylvester resultant matrix”, *Journal of Computational and Applied Mathematics*, **234**, pp. 3226–3242 (2010).
 - [25] J. R. Winkler, M. Hasan and X. Lao: “Two methods for the calculation of the degree of an approximate greatest common divisor of two inexact polynomials”, *Calcolo*, **49**, pp. 241–267 (2012).

- [26] J. R. Winkler and M. Hasan: “An improved non-linear method for the computation of a structured low rank approximation of the Sylvester resultant matrix”, *Journal of Computational and Applied Mathematics*, **237**, pp. 253–268 (2013).
- [27] J. R. Winkler: “Structured matrix methods for the computation of multiple roots of a polynomial”, *Journal of Computational and Applied Mathematics*, **272**, pp. 449–467 (2014).
- [28] K. Nagasaka: “Toward the best algorithm for approximate GCD of univariate polynomials”, *Journal of Symbolic Computation*, **105**, pp. 4–27 (2021).
- [29] Z. Zeng: “The numerical greatest common divisor of univariate polynomials”, *Randomization, Relaxation, and Complexity in Polynomial Equation Solving in Contemporary Mathematics*, **556**, pp. 187–217 (2011).
- [30] J. Gillard and K. Usevich: “Structured low-rank matrix completion for forecasting in time series analysis”, *International Journal of Forecasting*, **34**, 4, pp. 582–597 (2018).
- [31] N. Yoshino and A. Tanaka: “Robust initialization strategies for Hankel structured low-rank approximation via variable projection”, 2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1584–1589 (2025).
- [32] 吉野夏樹, 田中章: “変数射影法を用いた近似最大公約多項式計算における商多項式の初期値構成法”, *電子情報通信学会論文誌 A*, **J109-A**, 7 (2026).
- [33] C. Rao and S. K. Mitra: “Generalized inverse of a matrix and its applications”, *Berkeley Symp. on Math. Statist. and Prob.*, pp. 601–620 (1972).
- [34] N. Golyandina and A. Zhigljavsky: “Singular Spectrum Analysis for Time Series”, Springer Berlin, Heidelberg (2013).

-
- [35] G. Heinig and K. Rost: “Algebraic methods for Toeplitz-like matrices and operators”, Akademie-Verlag (1984).
 - [36] J. Razavilar, Y. Li and K. J. R. Liu: “Spectral estimation based on structured low rank matrix pencil”, 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Vol. 5, pp. 2503–2506 vol. 5 (1996).
 - [37] P. Stoica and R. Moses: “Spectral Analysis of Signals”, Prentice Hall (2005).
 - [38] L. Condat and A. Hirabayashi: “Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements”, Sampling Theory in Signal and Image Processing, **14**, pp. 17–47 (2015).
 - [39] M. Elad, P. Milanfar and G. H. Golub: “Shape from moments - an estimation theory perspective”, IEEE Transactions on Signal Processing, **52**, pp. 1814–1829 (2004).
 - [40] R. Zhang and G. Plonka: “Optimal approximation with exponential sums by a maximum likelihood modification of Prony’s method”, Advances in Computational Mathematics, **45**, pp. 1657–1687 (2019).
 - [41] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron and M. Lustig: “Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion”, Magnetic Resonance in Medicine, **72**, pp. 959–970 (2014).
 - [42] V. Oropenza and M. Sacchi: “Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis”, GEOPHYSICS, **76**, pp. V25–V32 (2011).
 - [43] J. J. Sylvester: “Sur une extension d’un théorème de clebsch relatif aux courbes du quatrième degré”, Comptes Rendus Mathématique, Académie des sciences, pp. 1532–1534 (1886).

- [44] G. R. B. Prony: “Essai experimental et analytique sur les lois de la dilatabilite de fluides elastiques et sur celles da la force expansion de la vapeur de l’alcool, a differentes temperatures”, Journal de l’Ecole Polytechnique, **1**, (1795).
- [45] K. Usevich: “On signal and extraneous roots in singular spectrum analysis”, Statistics and Its Interface, **3**, pp. 281–295 (2010).
- [46] N. Golyandina: “Analysis of time series structure SSA and related techniques” (2001).
- [47] M. Hall: “Combinatorial Theory”, Blaisdell Publishing Company (1967).
- [48] 雪江明彦: “代数学 2 環と体とガロア理論”, 日本評論社, 第 1 版 (2010).
- [49] P. L. Dragotti, M. Vetterli and T. Blu: “Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix”, IEEE Transactions on Signal Processing, **55**, pp. 1741–1757 (2007).
- [50] P. Lemmerling and S. V. Huffel: “Analysis of the structured total least squares problem for Hankel/Toeplitz matrices”, Numerical Algorithms, **27**, pp. 89–114 (2001).
- [51] N. Zvonarev and N. Golyandina: “Low-rank signal subspace: parameterization, projection and signal estimation”, Statistics and Its Interface, **16**, pp. 117–132 (2023).
- [52] M. Kuijper and J. C. Willems: “On constructing a shortest linear recurrence relation”, IEEE Transactions on Automatic Control, **42**, pp. 1554–1558 (1997).
- [53] N. Golyandina, A. Korobeynikov, A. Shlemov and K. Usevich: “Multivariate and 2d extensions of singular spectrum analysis with the Rssa package”, Journal of Statistical Software, **67**, (2015).
- [54] I. Markovsky, J. Willems, S. Van Huffel and B. De Moor: “Exact and Approximate Modeling of Linear Systems: A Behavioral Approach”, SIAM (2006).

-
- [55] I. Markovsky, A. Fazzi and N. Guglielmi: “Applications of polynomial common factor computation in signal processing”, *Latent Variable Analysis and Signal Separation* (Eds. by Y. Deville, S. Gannot, R. Mason, M. D. Plumbley and D. Ward), Cham, Springer International Publishing, pp. 99–106 (2018).
- [56] J. C. Willems: “From time series to linear system—part i. finite dimensional linear time invariant systems”, *Automatica*, **22**, pp. 561–580 (1986).
- [57] I. Markovsky and K. Usevich: “Software for weighted structured low-rank approximation”, *Journal of Computational and Applied Mathematics*, **256**, pp. 278–292 (2014).
- [58] G. Heinig: “Generalized inverses of Hankel and Toeplitz mosaic matrices”, *Linear Algebra and its Applications*, **216**, pp. 43–59 (1995).
- [59] S. Y. Kung: “A new identification and model reduction algorithm via singular value decomposition”, *12th Asilomar Conference on Circuits, Systems, and Computers*, pp. 705–714 (1978).
- [60] B. Ho and R. Kalman: “Editorial: Effective construction of linear state-variable models from input/output functions”, *Die Konstruktion von linearen Modeilen in der Darstellung durch Zustandsvariable aus den Beziehungen für Ein- und Ausgangsgrößen*, **14**, pp. 545–548 (1966).
- [61] K. Usevich, V. Emiya, D. Brie and C. Chaux: “Characterization of finite signals with low-rank STFT”, *IEEE Statistical Signal Processing Workshop, SSP 2018* (2018).
- [62] G. Heinig and P. Jankowski: “Kernel structure of block Hankel and Toeplitz matrices and partial realization”, *Linear Algebra and its Applications*, **175**, pp. 1–30 (1992).

- [63] E. Kaltofen, Z. Yang and L. Zhi: “Approximate greatest common divisors of several polynomials with linearly constrained coefficients and singular polynomials”, Proceedings of the 2006 International Symposium on Symbolic and Algebraic Computation, ISSAC ’06, New York, NY, USA, Association for Computing Machinery, pp. 169–176 (2006).
- [64] K. Nagasaka: “Approximate GCD by bernstein basis, and its applications”, Proceedings of the 45th International Symposium on Symbolic and Algebraic Computation, ISSAC ’20, New York, NY, USA, Association for Computing Machinery, pp. 372–379 (2020).
- [65] S. R. Khare, H. K. Pillai and M. N. Belur: “Real radius of controllability for systems described by polynomial matrices: SIMO case”, Proc. 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS, Vol. 5, pp. 517–522 (2010).
- [66] M. Agrawal, P. Stoica and P. Åhgren: “Common factor estimation and two applications in signal processing”, Signal Processing, **84**, pp. 421–429 (2004).
- [67] P. Stoica and T. Söderström: “Common factor detection and estimation”, Automatica, **33**, 5, pp. 985–989 (1997).
- [68] I. Markovsky, A. Fazzi and N. Guglielmi: “Applications of polynomial common factor computation in signal processing”, Latent Variable Analysis and Signal Separation, Vol. 10891, pp. 99–106 (2018).
- [69] P. Boito: “Structured matrix based methods for approximate polynomial GCD”, Edizioni della Normale Pisa, Pisa (2011).
- [70] K. Nagasaka: “Toward the best algorithm for approximate GCD of univariate polynomials”, Journal of Symbolic Computation, **105**, pp. 4–27 (2021).
- [71] D. A. Bini and P. Boito: “A fast algorithm for approximate polynomial GCD based on structured matrix computations”, Numerical Methods for Structured

- Matrices and Applications: The Georg Heinig Memorial Volume (Eds. by D. A. Bini, V. Mehrmann, V. Olshevsky, E. E. Tyrtyshnikov and M. van Barel), Basel, Birkhäuser Basel, pp. 155–173 (2010).
- [72] G. H. Golub and V. Pereyra: “The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate”, *SIAM Journal on numerical analysis*, **10**, pp. 413–432 (1973).
- [73] R. M. Corless, P. M. Gianni, B. M. Trager and S. M. Watt: “The singular value decomposition for polynomial systems”, *Proceedings of the 1995 International Symposium on Symbolic and Algebraic Computation*, Association for Computing Machinery, pp. 195–207 (1995).
- [74] W. Li, L. Zhao, Z. Lin, D. Xu and D. Lu: “Non - Local Image Inpainting Using Low - Rank Matrix Completion”, *Computer Graphics Forum* (2015).
- [75] J.-F. Cai, J. K. Choi, J. Li and G. Yin: “Restoration guarantee of image inpainting via low rank patch matrix completion”, *SIAM Journal on Imaging Sciences*, **17**, 3, pp. 1879–1908 (2024).
- [76] E. J. Candès and B. Recht: “Exact matrix completion via convex optimization”, *Foundations of Computational Mathematics*, **9**, 6, pp. 717–772 (2009).
- [77] E. J. Candès, X. Li, Y. Ma and J. Wright: “Robust principal component analysis?”, *Journal of the ACM*, **58**, 3, pp. 1–37 (2011).
- [78] J. A. Tropp, A. Yurtsever, M. Udell and V. Cevher: “Streaming low-rank matrix approximation with an application to scientific simulation”, *SIAM Journal on Scientific Computing*, **41**, 4, pp. A2430–A2463 (2019).
- [79] J. A. Tropp and R. J. Webber: “Randomized algorithms for low-rank matrix approximation: Design, analysis, and applications”, *arXiv preprint arXiv:2306.12418* (2023).

- [80] V. Klema and A. Laub: “The singular value decomposition: Its computation and some applications”, IEEE Transactions on Automatic Control, **25**, 2, pp. 164–176 (1980).
- [81] W.-S. Lu, S.-C. Pei and P.-H. Wang: “Weighted low-rank approximation of general complex matrices and its application in the design of 2-d digital filters”, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, **44**, 7, pp. 650–655 (1997).
- [82] R. Twogood and S. Mitra: “Computer-aided design of separable two-dimensional digital filters”, IEEE Transactions on Acoustics, Speech, and Signal Processing, **25**, 2, pp. 165–169 (1977).
- [83] A. Antoniou and W.-S. Lu: “Design of two-dimensional digital filters by using the singular value decomposition”, IEEE Transactions on Circuits and Systems, **34**, 10, pp. 1191–1198 (1987).
- [84] T.-B. Deng and M. Kawamata: “Frequency-domain design of 2-d digital filters using the iterative singular value decomposition”, IEEE Transactions on Circuits and Systems, **38**, 10, pp. 1225–1228 (1991).
- [85] C. Eckart and G. Young: “The approximation of one matrix by another of lower rank”, Psychometrika, **1**, pp. 211–218 (1936).
- [86] L. MIRSKY: “Symmetric gauge functions and unitarily invariant norms”, The Quarterly Journal of Mathematics, **11**, pp. 50–59 (1960).
- [87] G. H. Golub and C. F. V. Loan: “Matrix Computations”, The Johns Hopkins University Press, fourth edition (1996).
- [88] J. Manton, R. Mahony and Y. Hua: “The geometry of weighted low-rank approximations”, IEEE Transactions on Signal Processing, **51**, 2, pp. 500–514 (2003).

-
- [89] W. Zhang, L. Fu, M. Zhang and W. Cheng: “2-D seismic data reconstruction via truncated nuclear norm regularization”, *IEEE Transactions on Geoscience and Remote Sensing*, **58**, pp. 6336–6343 (2020).
- [90] Z. Liu, A. Hansson and L. Vandenberghe: “Nuclear norm system identification with missing inputs and outputs”, *Systems & Control Letters*, **62**, pp. 605–612 (2013).
- [91] I. Markovsky: “How effective is the nuclear norm heuristic in solving data approximation problems?”, *IFAC Proceedings Volumes*, **45**, pp. 316–321 (2012).
- [92] M. Fazel, T. K. Pong, D. Sun and P. Tseng: “Hankel matrix rank minimization with applications to system identification and realization”, *SIAM Journal on Matrix Analysis and Applications*, **34**, pp. 946–977 (2013).
- [93] J. B. Rosen, H. Park and J. Glick: “Total least norm formulation and solution for structured problems”, *SIAM Journal on Matrix Analysis and Applications*, **17**, pp. 110–126 (1996).
- [94] J. B. Rosen, H. Park and J. Glick: “Structured total least norm for nonlinear problems”, *Society for Industrial and Applied Mathematics*, **20**, pp. 14–30 (1998).
- [95] H. Park, L. Zhang and J. B. Rosen: “Low rank approximation of a hankel matrix by structured total least norm”, *BIT Numerical Mathematics*, **39**, pp. 757–779 (1999).
- [96] S. Boyd and L. Vandenberghe: “Convex Optimization”, Cambridge University Press (2004).
- [97] K. Usevich and I. Markovsky: “Structured low-rank approximation as a rational function minimization”, *IFAC Proceedings Volumes*, **45**, pp. 722–727 (2012).

- [98] P. Guillaume and R. Pintelon: “A Gauss-Newton-like optimization algorithm for “weighted” nonlinear least-squares problems”, *IEEE Transactions on Signal Processing*, **44**, 9, pp. 2222–2228 (1996).
- [99] C. Hage and M. Kleinstaubert: “Robust structured low-rank approximation on the grassmannian”, *Latent Variable Analysis and Signal Separation* (Eds. by E. Vincent, A. Yeredor, Z. Koldovský and P. Tichavský), Springer International Publishing, pp. 295–303 (2015).
- [100] T. Liu, I. Markovsky, T. K. Pong and A. Takeda: “A hybrid penalty method for a class of optimization problems with multiple rank constraints”, *SIAM Journal on Matrix Analysis and Applications*, **41**, 3, pp. 1260–1283 (2020).
- [101] M. Fazel: “Matrix Rank Minimization with Applications”, PhD thesis, Stanford University (2002).
- [102] I. CVX Research: “CVX: Matlab software for disciplined convex programming, version 2.0”, <https://cvxr.com/cvx> (2012).
- [103] K. Usevich and P. Comon: “Hankel low-rank matrix completion: Performance of the nuclear norm relaxation”, *IEEE Journal of Selected Topics in Signal Processing*, **10**, pp. 637–646 (2016).
- [104] J. Caprani and M. Carlsson: “Quadratic envelope regularization for structured low rank approximation”, *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8217–8221 (2019).
- [105] F. Andersson, M. Carlsson and C. Olsson: “Convex envelopes for fixed rank approximation”, *Optimization Letters*, **11**, pp. 1783–1795 (2017).
- [106] A. Al-Shabli and I. Selesnick: “Sharpening sparse regularizers”, *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4908–4912 (2019).

-
- [107] C. Grussler, A. Rantzer and P. Giselsson: “Low-rank optimization with convex constraints”, *IEEE Transactions on Automatic Control*, **63**, pp. 4000–4007 (2018).
- [108] N. Guglielmi and I. Markovsky: “An ODE-based method for computing the distance of coprime polynomials to common divisibility”, *SIAM Journal on Numerical Analysis*, **55**, pp. 1456–1482 (2017).
- [109] A. Fazzi, N. Guglielmi and I. Markovsky: “A gradient system approach for hankel structured low-rank approximation”, *Linear Algebra and its Applications*, **623**, pp. 236–257 (2021).
- [110] C. Kümmerle and C. M. Verdun: “Completion of structured low-rank matrices via iteratively reweighted least squares”, *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pp. 1–5 (2019).
- [111] C. E. Kümmerle: “Understanding and Enhancing Data Recovery Algorithms - From Noise-Blind Sparse Recovery to Reweighted Methods for Low-Rank Matrix Optimization-”, PhD thesis, Technische Universität München (2019).
- [112] G. Ongie and M. Jacob: “A fast algorithm for convolutional structured low-rank matrix recovery”, *IEEE Transactions on Computational Imaging*, **3**, pp. 535–550 (2017).
- [113] 吉野夏樹, 田中章: “誤差を含む構造化行列の一次元零空間推定における計算量削減”, *電子情報通信学会技術研究報告, 信号処理*, 第 SIP2022-51 巻, pp. 7–12 (2022).
- [114] T. M. Inc.: “MATLAB version: 9.13.0 (R2022b)” (2022).
- [115] Z. Zeng and T.-Y. Li: “NAClab: a Matlab toolbox for numerical algebraic computation”, *ACM Commun. Comput. Algebra*, **47**, 3/4, pp. 170–173 (2014).

- [116] A. Fazzi, N. Guglielmi and I. Markovsky: “Computing common factors of matrix polynomials with applications in system and control theory”, 2019 IEEE 58th Conference on Decision and Control (CDC), pp. 7721–7726 (2019).
- [117] B. Peng, F. Yang, C. Yan, X. Zeng and D. Zhou: “Efficient multiple starting point optimization for automated analog circuit optimization via recycling simulation data”, 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1417–1422 (2016).
- [118] T. M. Inc.: “MATLAB version: 24.1.0 (R2024a)” (2024).
- [119] R. Poli, J. Kennedy and T. Blackwell: “Particle swarm optimization: An overview”, *Swarm Intelligence*, **1**, pp. 33–57 (2007).
- [120] T. M. Inc.: “Global optimization toolbox version: 24.1 (R2024a)” (2024).
- [121] T. Hikichi, M. Delcroix and M. Miyoshi: “Speech dereverberation algorithm using transfer function estimates with overestimated order”, *Acoustical Science and Technology*, **27**, pp. 28–35 (2006).