論文 / 著書情報
Article / Book Information

題目(和文)	教師付き学習における射影学習族に関する研究
Title(English)	
著者(和文)	平林晃
Author(English)	Akira Hirabayashi
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:乙第3381号, 授与年月日:1999年12月31日, 学位の種別:論文博士, 審査員:
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:乙第3381号, Conferred date:1999/12/31, Degree Type:Thesis doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

教師付き学習における
射影学習族に関する研究

平林　晃

目 次

本論文で用いる記号

$*$: 共役作用素 (p.10)
$\dagger$: ムーア・ペンローズ一般逆 (p.10)
$\dot{+}$: 部分空間の直和 (p.14)
$\oplus$: 部分空間の直交直和 (p.24)
$\langle \cdot, \cdot \rangle$: 内積 (Hの内積：p.8, シュミットの内積：p.24)
$\| \cdot \|$: ノルム ($\mathbf{C}^M$のノルム：p.24, Hのノルム：p.13)
$(\cdot \otimes \bar{\cdot})$: ノイマン・シャッテン積 (p.9)
A : 標本化作用素 (p.9)
A_s : AP_Sで定義される，部分空間 Sの元に対する標本化作用素 (p.23)
$A^{(J)}$: J学習作用素 (p.9)
$A^{(\mathrm{M1})}$: 第 1 種記憶学習作用素 (p.10)
$A^{(\mathrm{M2})}$: 第 2 種記憶学習作用素 (p.11)
$A^{(\mathrm{P})}$: 射影学習作用素 (p.13)
$A^{(\mathrm{PTP})}$: 部分射影学習作用素 (p.14)
$A^{(\mathrm{AP})}$: 平均射影学習作用素 (p.15)
$\mathbf{C}^M$: 標本値空間 (p.9)
H : 部分空間 Sを含む再生核ヒルベルト空間 (p.8)
J : 学習の評価基準 (p.9)
$J_n[X]$: $E_{\boldsymbol{n}} \| X\boldsymbol{n} \|^2$で定義される雑音成分 $X\boldsymbol{n}$に関する汎関数 (p.13)
J_{n0} : $X = A^{(\mathrm{SL})}$で与えられる $J_n[X]$ の最小値 (p.24)
K : 第 m, n 成分が $K(\boldsymbol{x}_m, \boldsymbol{x}_n)$ で与えられる行列 (p.59)
$K(\boldsymbol{x}, \boldsymbol{x}')$: Hの再生核 (p.8)
L : Sにおける $S \cap \mathcal{N}(A)$ の補空間 (p.20)
L_N : $\mathcal{R}(A_s^\dagger P_t Q)$ で定義される $\mathcal{R}(A_s^*)$ の部分空間 (p.32)
$\mathcal{N}(B)$: 作用素 Bの零空間 (p.10)
P : Sの中でみれば $S \cap \mathcal{N}(A)$ に沿った L への射影となるような線形作用素 (p.20)
P_t : $Q\mathcal{R}(A_s)^\perp \oplus S_t^\perp$に沿った $\mathcal{R}(A_s)$ への射影行列 (p.24)
P_1 : $\mathcal{R}(P_{\mathcal{R}(As)^\perp} Q P_{\mathcal{R}(As)^\perp})$ への正射影行列 (p.79)
P_2 : $I - Q(P_{\mathcal{R}(As)^\perp} Q P_{\mathcal{R}(As)^\perp})^\dagger$で定義される $Q\mathcal{R}(A_s)^\perp$に沿った

$\mathcal{N}(QP_{\mathcal{R}(A_s)^{\perp}})$ への射影行列 (p.79)

P_{S_1} : 一般の部分空間 S_1への正射影作用素 (p.10)

P_W : $RA^*(ARA^*)^{\dagger}A$ で定義される$\overline{\mathcal{R}(R)}$の中でみれば$\overline{\mathcal{R}(R)} \cap \mathcal{N}(A)$ に沿った $\mathcal{R}(RA^*)$ への射影となる作用素(p.28)

P_c : PP_Sで定義される $\{S \cap \mathcal{N}(A)\} \oplus S^{\perp}$に沿った L への射影作用素 (p.82)

P_α : $S \cap \mathcal{N}(A)$ への正射影作用素 (p.82)

Q : $\{\boldsymbol{n}\}$ の相関行列 (p.11)

R : $\{f\}$ の相関作用素 (p.15)

$\mathcal{R}(B)$: 作用素 Bの値域 (p.10)

S : 学習対象関数 fが含まれる Hの閉部分空間 (p.20)

S_b : $\mathcal{R}(A) + \mathcal{R}(Q)$ で定義される $\mathbf{C}^M$の部分空間 (p.11)

S_t : $\mathcal{R}(A_s) + \mathcal{R}(Q)$ で定義される $\mathbf{C}^M$の部分空間 (p.24)

U_{AP} : $ARA^* + Q$ で定義される $M \times M$行列 (p.15)

U_P : $AA^* + Q$ で定義される $M \times M$行列 (p.13)

U_{PTP} : $AP_SA^* + Q$ で定義される $M \times M$行列 (p.15)

V_{AP} : $RA^*U_{AP}^{\dagger}ARA^*$で定義される $\mathbf{C}^M$から Hへの作用素 (p.15)

V_P : $A^*U_P^{\dagger}A$ で定義される Hから Hへの作用素 (p.13)

V_{PTP} : $P_SA^*U_{PTP}^{\dagger}AP_S$で定義される Hから Hへの作用素 (p.15)

X : 学習作用素 (p.9)

Y : 学習作用素の一般形を表すときに用いる $\mathbf{C}^M$から Hへの任意の作用素 (p.10)

f : 学習対象関数 (p.8)

f_0 : 学習によって得られた関数 (p.8)

f_{AP} : 平均射影学習によって得られた関数 (p.16)

f_{M} : 記憶学習によって得られた関数 (p.12)

f_{M1} : 第 1 種記憶学習によって得られた関数 (p.11)

f_{M2} : 第 2 種記憶学習によって得られた関数 (p.12)

f_{MV} : 最小分散射影学習によって得られた関数 (p.32)

f_{P} : 射影学習によって得られた関数 (p.14)

f_{PTP} : 部分射影学習によって得られた関数 (p.15)

f_{SL} : SL 射影学習によって得られた関数 (p.25)

n_m : 標本点$\boldsymbol{x}_m$における雑音 (p.8)

$\boldsymbol{n}$: m 番目の要素が n_mである M次元ベクトル (p.9)

$\boldsymbol{x}$: fの入力 (p.8)

$\boldsymbol{x}_m$: 標本点 (p.8)

y_m : 標本点$\boldsymbol{x}_m$に対応する $f(\boldsymbol{x}_m) + n_m$で与えられる教師信号 (p.8)

$\boldsymbol{y}$: m 番目の要素が y_mである M次元ベクトル (p.8)

第 1 章

序論

1.1 本研究の背景と目的

1.1.1 教師付き学習

ヒトは，例えば子供が言葉の使い方を覚えていくように，あるいは学生が授業で教えられたことのない問題を解くことができるようになるように，教えられたことだけでなく，教えられていないことにも対応できるようになっていく．このような学習を機械に行わせることは，科学および工学における重要な課題の 1 つである．

機械による学習は，大きく教師付き学習と教師なし学習に分かれる [6, 61]．前者は，訓練データとして，問題と解答，或いは入力とそれに対応する理想的な出力等を用いる学習法である．問題や入力を一般に教材といい，それらに対応する解答や理想的な出力を教師信号という．訓練データとして教材だけを用いる学習法が教師なし学習である．本論文では，教師付き学習について論じる．教師付き学習の応用範囲は広い．例えばロボットの制御では，感覚運動マップの獲得に教師付き学習が用いられている [36, 64]．文字認識における識別関数の学習にも教師付き学習が用いられている [61]．

教師付き学習の問題をより明確に述べるために，階層型ニューラルネットワークの学習を考えてみる．一つのニューラルネットワークは，その入出力関係に着目すれば，一つの関数とみなすことができる．従って，ニューラルネットワークによる学習は，もとの関数 f の標本点における値から，ある評価基準の基で最適な近似関数 f_0 を求める問題として定式化できる [48, 53]．この f_0 を，ニューラルネットワークの内部変数を逐次変更することにより求めようとする研究が多く行われている [16, 18]．しかし，よく考えてみればこの問題は，図 1.1 に示すように，評価基準を決定する問題と，その評価基準のもとで最適な関数 f_0 を求める問題，そして，その f_0 をニューラルネットワーク上で実現する問題に分けて考えることができる．ところで第 3 の問題は，f_0 をどのように構成するかという問題であるから，技術の発展状況に依存して変化する．しかし，第 1 と第 2 の問題は，訓練データと汎化能力の関係を論じるものであり，技術の発展状況には依存しない問題である．従って，第 1 と第 2 の問題は情報科学の問題であり，第 3 の問題は情報工学の問題であるということができる [53]．

このように科学の問題と工学の問題に意識的に分けて考えることは重要である．例えば，ニューラルネットワークの中間素子数を多くし過ぎると過学習が生じるといわれることがよくある [58]．しかしその真の原因は，多すぎる中間素子数にあるのではなく，そこで用いた学習アルゴリズムが，訓練誤差だけを小さくしようとするものであるからであ

評価基準の決定

⇓

最良近似関数 f_0 の決定

⇓

f_0 の実現

図 1.1　教師付き学習の基本問題

る．実際，中間素子数をいくら多くとっても，同じ汎化能力をもたせることができることがわかっている [53]．このように過学習の真の原因を解明できたのは，汎化能力という科学の問題と，中間素子数という工学の問題を明確に分けて考えたからである．本論文では，図 1.1 の最初の 2 段階の問題，即ち，ある評価基準を決定し，その評価基準のもとで最適な近似関数を求めるという情報科学に属する問題を論じる．

これらの問題の定式化にも様々なものがある．まず訓練データに関しては，表 1.1 に示すように，(1) データ数が非常に多くなった場合の様相を調べる漸近理論や，(2) 有限個の訓練データが確率的に与えられる場合に訓練データの取り方に関して平均的に汎化能力を高めようとする研究や，(3) 訓練データの与えられ方に関わらず手元にある有限個のデータから高い汎化能力を得ようとする研究がある．統計的学習理論では，第 1 の立場 [3, 4, 5] や第 2 の立場 [63] から研究が行なわれている．計算論的学習理論でも，第 1 の立場から論じる場合 [38] と，第 2 の立場から論じる場合 [10] がある．小川によって提案された作用素論的学習理論 [48, 49, 53] では，第 3 の立場から論じられている．この第 3 の立場では，汎化能力を評価する際，訓練データに関して平均をとることはしないで，あくまでも手元にあるデータだけから得られた関数の汎化能力を評価するのである．現実の問題では，必ずしも十分でない有限個の訓練データしか使えない場合があるし，また，能動学習の問題では，いかに最適な訓練データを設計するかという問題を考えなければいけない．これらの問題に答えるためには，第 3 の立場が重要になってくる．本論文では，作用素論的手法を使って第 3 の立場から学習の問題を論じることにする．

学習に用いる前提知識にも様々なものがある．例えば学習対象の関数が属している空間が未知の場合と既知の場合がある．モデル選択の問題は前者に属している [1, 6, 40, 60, 63, 66] が，モデルを選択した後の学習の問題は後者に属す [63]．後者に属す場面はこの他にも多くのものがある．例えば，ロボットの感覚運動マップの学習のように，問題の物理的モデルから関数空間を決めることができる場合もある [64]．また，support vector machine の関数近似への応用では，核関数を用いてどのように最良近似を求めるかという議論が行われている [14, 59, 63]．核関数と関数空間は 1 対 1 に対応しており，核関数を

表 1.1　訓練データからみた学習問題の分類

	(1)	(2)	(3)	
訓練データの個数	無限個	有限個		
訓練データの与えられ方	確率的	確率的		確定的
訓練データの使い方	確率的	確率的	確定的	確定的

1 つ固定した段階で，関数空間を 1 つ固定したことになる．即ち，これは関数空間が既知である場合の議論になっている．正則化の手法も関数空間が既知の場合に該当している．即ち正則化では，

$$\text{訓練誤差} + \text{正則化項} \tag{1.1}$$

という形の評価基準が使われるが，この正則化項は，例えば学習結果の関数の微分のノルムの形式がとられる [55]．ノルムを決めた段階で，微分した関数が属す空間を決めたことになるし，それは暗黙のうちに学習結果およびもとの関数が属している空間を何かに決めていることになる．文献 [66] でも正則化の議論がなされているが，そこでは，対象の関数が再生核ヒルベルト空間 [12] に属している場合が論じられている．本論文でも，学習対象の関数が属している空間が既知の場合について論じることにする．

ところで作用素論的学習理論では，学習の問題は画像復元や信号復元などと数学的に同じ構造をもつ逆問題として定式化されている．従って，それらの分野で得られている多くの深い結果が，学習の問題にも活用できるのである [53]．本論文でも，これらの結果を随時，学習の言葉に置き換えて用いていくことにする．

1.1.2 学習族の理論

作用素論的学習理論では，これまでに射影学習 [53]，部分射影学習 [50]，平均射影学習 [65]，ウィーナー学習 [51] 等が提案されてきた．そして，これらの学習方式に対して，学習作用素の一般形が求められたり，雑音抑制能力や汎化能力が個別に論じられてきた．

しかし，議論を更に深めていくためには，個々の学習方式を個別に論じるのではなく，多くの種類の学習方式を同時に統一的に論じることが必要になってくる．

例えば，関数解析の分野でも，個々の作用素の性質を論じる '作用素論' の段階から，作用素の集合の性質を論じる '作用素環の理論'，それを更に抽象化した 'ノルム環の理論' へと進むことによって，多くの成果が得られてきた．この成果をフーリエ解析の問題へ適用することにより，従来 100 ページもかかっていたウィーナーのタウバー型定理の証明が，わずか数行で済んでしまったという有名な出来事もあった．ウィーナーのタウバー型定理がなぜ成立するかという本当の理由が，ノルム環の立場からみて始めてわかったのである [42]．

このように，個々の対象の研究から，それらの対象を含む集合の性質の研究へと進むことにより，物事の本質がわかってくるのである．学習の分野においても，個々の学習方式を個別に論じる段階から，多くの種類の学習方式を含む学習族の理論へと進むことにより，学習に対する理解が一段と深まるものと思われる．

このような考え方に基づいて，射影学習，部分射影学習，平均射影学習を含む無限に多くの種類の学習方式を統一的に扱う射影学習族の概念が提案された [43, 37, 23]．今井等は，この射影学習族の概念を，パラメトリック射影学習族へと発展させている [35]．こうして学習理論は，個々の学習方式に対する理論から，無限に多くの種類の学習方式を統一的に論じる学習族の理論へと新しい一歩を踏み出していった．しかし，射影学習族の定義そのものに不自然な部分が残っている等，まだまだ最初の一歩にとどまったままの状態である．そこで本論文では，射影学習族のより自然な定義を与え，汎化能力等の諸性質を統一的に解明すると共に，最小分散射影学習の提案，許容性理論の確立，記憶学習の射影学習族における適用範囲の解明等を行ない，射影学習族の理論を構築していく．

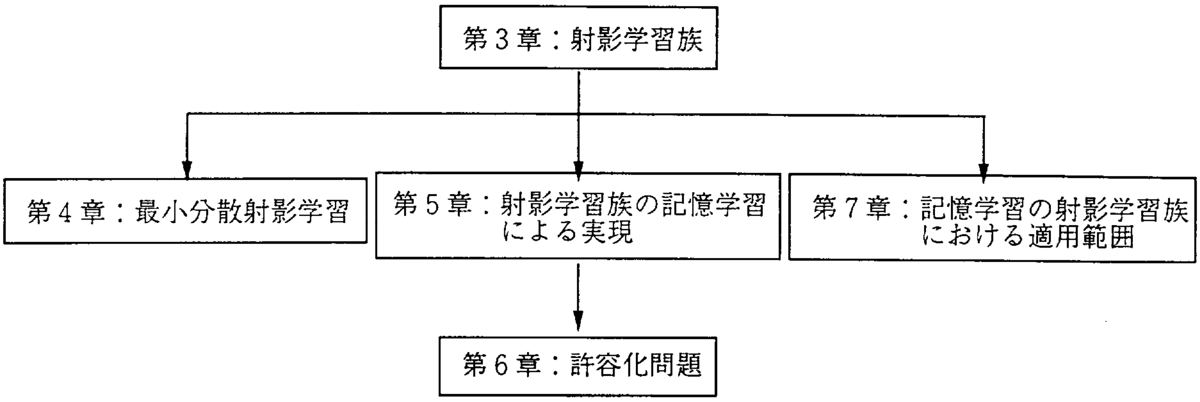


図 1.2　本論文の構成

1.2　本論文の構成

本論文は次の 8 章からなっている．本論文の構成を図 1.2 に示す．

第 1 章は序論である．

第 2 章では，作用素論的手法を用いて学習問題を定式化する．まず，記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習の 2 種類があることを示す．つづいて，射影学習，部分射影学習，平均射影学習について整理し，本論文で学習問題を論じるための準備を行なう．

第 3 章では，まず射影学習族の従来の定義の問題点を整理し，より自然な定義を与える．そして，この定義を基に，学習作用素の一般形を射影学習族に含まれるすべての学習方式に対して統一的に求めると共に，雑音抑制能力や汎化能力等の諸性質を統一的に解明し，射影学習族の基礎理論を構築する．

第 4 章では，前章で構築した射影学習族の理論の一つの応用として，最小分散射影学習の概念を提案する．学習によって得られる関数は，教師信号に含まれている雑音によって，雑音がない場合の学習結果のまわりに分布することになる．この分布の広がりが小さければ小さいほど，雑音に影響されず安定した学習結果を得ることができる．この広がりを射影学習族の中で最も小さくする学習方式が最小分散射影学習である．本章では，その特徴づけを行なうと共に，最小分散射影学習のもつ様々な性質を明らかにする．

第 5 章では，射影学習族に含まれる各学習方式を記憶学習によって実現する問題を論じる．教師付き学習では，訓練誤差を最小にする記憶学習が広く用いられている．しかし，学習の本当の目的は記憶学習を行うことではなく，高い汎化能力を獲得することである．従って，記憶学習はあくまで本来行ないたい学習方式の実現手段として用いられていることになる．本来行いたい学習方式を別の学習方式によって実現できるとき，本来の学習方式は実現手段として用いられた別の学習方式を許容するという．これまでに，教師信号に雑音が含まれていない場合に射影学習，部分射影学習，平均射影学習，およびウィーナー学習が第 2 種記憶学習を許容するための必要十分条件が求められ [51, 70, 50]，また，教師信号に雑音が含まれている場合にウィーナー学習が第 2 種記憶学習を許容するための必要十分条件が求められている [70]．その結果，許容性が成立するかどうかが，教材の選び方に密接に関係していることがわかった．そして，これらの条件を基にして，許容性を成立させる教材の選び方が与えられている [71]．

ところで，教師信号に雑音が含まれていない場合の射影学習の第 2 種記憶学習に対する許容性では，逆許容性と完全許容性の 2 種類の許容性しか現れないが，部分射影学習，平均射影学習の第 2 種記憶学習に対する許容性では，非許容性以外の 4 種類の許容性が現れる．しかし，このような違いが生じる理由はまだわかっていない．また，これらの条件は，教師信号に雑音が含まれている場合に対しては求められていない．更に，第 1 種記憶学習を実現手段として用いる場合の許容性は論じられていない．そこで本章では，射影学習族に含まれる学習方式が第 1 種記憶学習および第 2 種記憶学習をそれぞれ許容するための必要十分条件を統一的に求め，なぜある許容性がある学習方式に対して成立し，別の学習方式に対して成立しないかという背後にある法則性を解明すると共に，雑音および前提知識が許容性に与える影響等を解明していく．

第 6 章では，前章の議論に関連して，新しく許容化問題と名付けた問題を論じる．これは，与えられた教材に対して許容性が成立していない場合に，教材の一部を修正することにより許容性を成立させる問題である．本章では，射影学習の実現手段として記憶学習を用いる場合の許容化問題を解明する．

第 7 章では，‘学習方式の適用範囲’ という問題を新しく提起し，記憶学習の射影学習族における適用範囲を解明する．第 5 章では，本来行ないたい学習方式が先にあり，その実現手段として別の学習方式を用いるという問題を論じた．これとは逆に，ある学習方式が手元にあり，それを様々な学習方式の実現手段として用いる場面を考える．このとき，手元にある学習方式がどの範囲の学習方式にまで適用できるかを明らかにする問題が，適用範囲の問題である．第 5 章で論じた問題がユーザーの立場からの問題設定であるのに対して，本章で扱う問題はメーカーの立場からの問題設定である．例えば，ある学習用のソフトを開発したときに，それを初期の目的以外にもどんどん応用し，様々な分野で活用していこうという立場からの問題設定である．このような問題を論じるためには，(1) 現在手元にある学習方式と，(2) 適用可能な学習方式を探す探索範囲を決めなくてはいけない．本論文では，(1) として訓練誤差だけを最小にするという最も素朴な形の記憶学習を採用する．(2) としては，第 3 章で論じた無限に多くの種類の学習方式を含んでいる射影学習族を採用する．まず初めに，教師信号に雑音が含まれていない場合の記憶学習の適用範囲を解明する．次に，雑音が含まれている場合の記憶学習の適用範囲を解明し，両者を比較することにより，雑音が適用範囲に与える影響を明らかにする．

第 8 章では，本論文で得られた成果を総括すると共に，今後解決すべき問題を明らかにする．

第 2 章

教師付き学習

2.1 まえがき

本章では，文献 [48, 49, 53] に従って，作用素論的手法を用いて学習問題を定式化する．まず，記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習の 2 種類があることを示す．つづいて，射影学習，部分射影学習，平均射影学習について整理し，本論文で学習問題を論じるための準備を行なう．

2.2 学習問題の定式化

文献 [48, 49, 53] に従って，本論文で扱う学習問題の枠組を説明する．教師付き学習の問題とは，例題とそれに対応する教師信号の組である訓練データを使って，その背後にある規則を推定する問題である．その規則が関数として表現できる場合には，関数解析的手法を使うことにより関数近似の問題として定式化することができる．

学習対象関数と学習によって得られる関数を，それぞれ f および f_0 で表す．訓練データは，例題とそれに対応する教師信号をそれぞれ $\{\boldsymbol{x}_m\}_{m=1}^{M}$ および $\{y_m\}_{m=1}^{M}$ で表せば，$\{\boldsymbol{x}_m, y_m\}_{m=1}^{M}$ と表される．y_m は加法性雑音 n_m を含む場合を考える：

$$y_m = f(\boldsymbol{x}_m) + n_m \tag{2.1}$$

$\boldsymbol{x}_m$, y_m は，関数近似の立場から，それぞれ標本点および標本値とよばれることもある．

本論文では，関数 f が含まれる空間 H が与えられている場合を論じていく．そして，H が再生核複素ヒルベルト空間である場合を考える．再生核 [12] とは，次の 2 つの条件を満たす関数 $K(\boldsymbol{x}, \boldsymbol{x}')$ である：

1. 任意に固定された $\boldsymbol{x}'$ に対して，$K(\boldsymbol{x}, \boldsymbol{x}') \in H$ である．

2. 任意の $f \in H$ と $\boldsymbol{x}'$ に対して，

$$\langle f, K(\cdot, \boldsymbol{x}') \rangle = f(\boldsymbol{x}') \tag{2.2}$$

となる．ここで，$\langle \cdot, \cdot \rangle$ は H の内積である．

ヒルベルト空間の議論では，一つの関数を H の中の 1 点とみなしているので，ある関数 f のある点 $\boldsymbol{x}$ における値 $f(\boldsymbol{x})$ というものを一般には議論できない．しかし，H が

再生核をもてば，点 $\boldsymbol{x}$ における値 $f(\boldsymbol{x})$ を式 (2.2) のように内積というヒルベルト空間の言葉を使って論じることができるのである．学習の問題では，式 (2.1) に示すように，関数の一点における値が重要になってくる．ヒルベルト空間が再生核をもつ場合を考えたのは，このためである．

$\{\boldsymbol{x}_m\}_{m=1}^M$ を固定すれば，f に対して $\{f(\boldsymbol{x}_m)\}_{m=1}^M$ は一意に定まる．この値を縦に並べたベクトルに f を対応付ける作用素を標本化作用素とよび，A で表す：

$$Af = \begin{pmatrix} f(\boldsymbol{x}_1) \\ f(\boldsymbol{x}_2) \\ \vdots \\ f(\boldsymbol{x}_M) \end{pmatrix} \tag{2.3}$$

ここで注意すべきことは，f が非線形な関数であっても，A は自動的に線形作用素になるということである．実際 A は，

$$A = \sum_{m=1}^{M} (\boldsymbol{e}_m \otimes \overline{\psi_m}) \tag{2.4}$$

と表すことができる．ここで，$\boldsymbol{e}_m$ は $\mathbf{C}^M$ の標準基底，即ち，第 m 成分が 1 でそれ以外は 0 の M 次元ベクトルである．ψ_m は，

$$\psi_m(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}_m) \tag{2.5}$$

で定義される関数である．$(\cdot \otimes \bar{\cdot})$ はノイマン・シャッテン積とよばれ，

$$(\boldsymbol{e}_m \otimes \overline{\psi_m})f = \langle f, \psi_m \rangle \boldsymbol{e}_m \tag{2.6}$$

で定義される作用素である [57]．式 (2.6) の右辺に現れる $\langle f, \psi_m \rangle$ は，式 (2.5)，(2.2) より，

$$\langle f, \psi_m \rangle = f(\boldsymbol{x}_m) \tag{2.7}$$

となる．この性質より，ψ_m は標本化関数とよばれる．

$\{y_m\}_{m=1}^M$, $\{n_m\}_{m=1}^M$ をそれぞれ縦に並べたベクトルを $\boldsymbol{y}$, $\boldsymbol{n}$ で表す．これらのベクトルは，M 次元ユニタリ空間 $\mathbf{C}^M$ の元である．この空間 $\mathbf{C}^M$ を標本値空間とよぶことにする．f から $\boldsymbol{y}$ への対応は

$$\boldsymbol{y} = Af + \boldsymbol{n} \tag{2.8}$$

と表すことができる．$\boldsymbol{y}$ を f_0 へ対応付ける作用素を X で表す：

$$f_0 = X\boldsymbol{y} \tag{2.9}$$

X は線形な範囲で考えることにする．

式 (2.8)，(2.9) より，学習の問題は，f にできるだけ近い f_0 を与える X を求める問題と等価になる．そこで X を学習作用素とよぶことにする．

f と f_0 の近さの測り方には次節で述べるように様々な方法がある．それらの評価基準に従って，様々な学習法が提案されている．次節では，そのいくつかの例を紹介するが，そのような評価基準を一般に J で表す．そして，J を満足する X を J 学習作用素とよび，$A^{(J)}$ で表す．$A^{(J)}$ は一般に一意に定まらないので，その全体を $A\{J\}$ で表す．$A^{(J)}$ による学習を J 学習とよぶ．

2.3 様々な学習法

学習法は大きく 2 つに分類することができる．第 1 は標本点における誤差，いわゆる訓練誤差だけを小さくしようとする学習法である．第 2 は標本点だけでなく，すべての入力に対する誤差，いわゆる汎化誤差を小さくしようとする学習法である．本節では，まず第 1 の学習法である記憶学習を紹介し，その後で，射影学習など第 2 の学習法のいくつかの例を紹介する．

2.3.1 記憶学習

記憶学習は，f_0 の標本点における値 Af_0 と，式 (2.8) の教師信号 $\boldsymbol{y}$ との誤差をできるだけ小さくしようとする学習である．この学習法を実現するアルゴリズムが，誤差逆伝搬法である．記憶学習には，$\boldsymbol{y}$ に関する前提知識を用いる場合と用いない場合の 2 種類の定義がある．

まず，前提知識を用いない場合を考える．この場合，$\boldsymbol{y}$ は $\mathbf{C}^M$ のどこに現れるかわからないので，$\boldsymbol{y}$ が $\mathbf{C}^M$ のどのベクトルであっても対応できるようにすることを考える．即ち，次のように定義する．

定義 2.1 （第 1 種記憶学習 [41]）$\mathbf{C}^M$ の任意のベクトル $\boldsymbol{y}$ に対して，汎関数

$$J_M[X] = \|AX\boldsymbol{y} - \boldsymbol{y}\|^2 \tag{2.10}$$

を最小にする X を第 1 種記憶学習作用素（memorization learning operator of the first kind）とよび，$A^{(\mathrm{M1})}$ で表す．ここで $\|\cdot\|$ は $\mathbf{C}^M$ のノルムである．$A^{(\mathrm{M1})}$ による学習を第 1 種記憶学習（memorization learning of the first kind）という．

A の共役作用素と値域，零空間をそれぞれ A^* および $\mathcal{R}(A), \mathcal{N}(A)$ で表し，$\mathcal{R}(A)$ への正射影作用素を $P_{\mathcal{R}(A)}$ で表す．A に対して，

1. $AXA = A$
2. $XAX = X$
3. $(AX)^* = AX$
4. $(XA)^* = XA$

の中の i., j., $\cdots$, k. 番目の条件を満たす X を A の i, j, $\cdots$, k-逆といい，$A^{(\mathrm{i,j,\cdots,k})}$ で表す．i, j, $\cdots$, k-逆は一般に無限に存在するので，それらの全体を $A\{\mathrm{i,j,\cdots,k}\}$ で表す．但し，1. から 4. のすべての条件を満たす X は一意に定まる．それをムーア・ペンローズ一般逆といい，$A^\dagger$ で表す [11].

第 1 種記憶学習作用素の一般形は次のように与えられる．

命題 2.1 作用素 X が第 1 種記憶学習作用素になるための必要十分条件は，X が

$$AX = P_{\mathcal{R}(A)} \tag{2.11}$$

を満たすこと，即ち X が A の 1,3-逆になることである．$A^{(\mathrm{M1})}$ の一般形は，

$$A^{(\mathrm{M1})} = A^\dagger + (I - A^\dagger A)Y \tag{2.12}$$

で与えらえる．ここで，I は H 上の恒等作用素であり，Y は $\mathbf{C}^M$ から H への任意の作用素である．第 1 種記憶学習によって得られる関数 f_{M1} は

$$f_{\mathrm{M1}} = A^{\dagger}\boldsymbol{y} + (I - A^{\dagger}A)Y\boldsymbol{y} \tag{2.13}$$

で与えられる．

命題の証明はすべて付録 C に示す．式 (2.13) の右辺には任意の作用素 Y があるので，f_{M1} は一意に定まらない．しかし，それらを A で変換すれば，式 (2.13), (2.8) より，

$$Af_{\mathrm{M1}} = Af + P_{\mathcal{R}(A)}\boldsymbol{n} \tag{2.14}$$

となり，f_{M1} の標本点における値は Y によらずすべて同じものになる．しかし，この値は一般には $\boldsymbol{y}$ にならず，Af と $P_{\mathcal{R}(A)}\boldsymbol{n}$ の和になる．このことは，第 1 種記憶学習が $\boldsymbol{n}$ を $\mathcal{R}(A)$ に正射影することによって雑音抑制を行なっていると解釈することができる．

次に，$\boldsymbol{y}$ に関する前提知識を用いる記憶学習を考える．$\boldsymbol{n}$ の相関行列を Q で表す：

$$Q = E_{\boldsymbol{n}}(\boldsymbol{n} \otimes \overline{\boldsymbol{n}}) \tag{2.15}$$

ここで，$E_{\boldsymbol{n}}$ は雑音 $\{\boldsymbol{n}\}$ に関する平均を表す．部分空間 S_b を

$$S_b = \mathcal{R}(A) + \mathcal{R}(Q) \tag{2.16}$$

で定義する．$\mathcal{R}(A)$ は，H の元を A で変換して得られるベクトルの全体である．また，$\mathcal{R}(Q)$ は雑音 $\boldsymbol{n}$ が平均 2 乗的に含まれる最小の空間である [68]．そこで $\mathcal{R}(Q)$ に含まれる $\boldsymbol{n}$ を考えることにする．このとき式 (2.8) より，$\boldsymbol{y}$ は S_b に含まれることになる．このように，部分空間 S_b は教師信号 $\boldsymbol{y}$ の偏り（bias）を表す空間であるので，添字には bias の b が用いてある．この前提知識，即ち $\boldsymbol{y}$ が S_b に属しているという前提知識を用いる場合には，記憶学習を次のように定義できる．

定義 2.2 （第 2 種記憶学習）S_b の任意のベクトル $\boldsymbol{y}$ に対して，式 (2.10) を最小にする X を第 2 種記憶学習作用素（memorization learning operator of the second kind）とよび，$A^{(\mathrm{M2})}$ で表す．$A^{(\mathrm{M2})}$ による学習を第 2 種記憶学習（memorization learning of the second kind）という．

第 1 種記憶学習と第 2 種記憶学習の違いは，式 (2.10) の $\boldsymbol{y}$ の動き得る範囲が，前者が $\mathbf{C}^M$ 全体であるのに対して，後者が $\mathbf{C}^M$ の部分空間 S_b にとどまるということである．このことから，第 1 種記憶学習作用素は，同時に第 2 種記憶学習作用素にもなっている

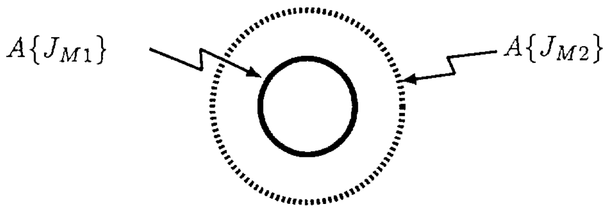


図 2.1　第 1 種記憶学習作用素と第 2 種記憶学習作用素の関係

ことがわかる．即ち，第 1 種および第 2 種記憶学習作用素の全体をそれぞれ $A\{J_{M1}\}$ および $A\{J_{M2}\}$ で表せば，図 2.1に示すように，

$$A\{J_{M1}\} \subset A\{J_{M2}\} \tag{2.17}$$

になっているのである．なお，第 6 章，第 7 章で行なう議論のように，これらの記憶学習を同時に議論できるときには，単に記憶学習とよぶこともある．

S_b への正射影行列を P_{S_b} で表す．第 2 種記憶学習作用素の一般形は次のようになる．

命題 2.2 作用素 X が第 2 種記憶学習作用素になるための必要十分条件は，X が

$$AXP_{S_b} = P_{\mathcal{R}(A)} \tag{2.18}$$

を満たすことである．$A^{(\mathrm{M2})}$ の一般形は，

$$A^{(\mathrm{M2})} = A^{\dagger} + Y - A^{\dagger}AYP_{S_b} \tag{2.19}$$

で与えられ，第 2 種記憶学習によって得られる関数 f_{M2} は，

$$f_{\mathrm{M2}} = A^{\dagger}\boldsymbol{y} + (I - A^{\dagger}A)Y\boldsymbol{y} \tag{2.20}$$

で与えられる．

第 1 種記憶学習作用素は，式 (2.12) より

$$A^{(\mathrm{M1})} = A^{\dagger} + Y - A^{\dagger}AY \tag{2.21}$$

となっているが，この式の第 3 項が $A^{\dagger}AY$ であるのに対して，第 2 種記憶学習作用素では式 (2.19) の第 3 項が $A^{\dagger}AYP_{S_b}$ になっている．このように，作用素の段階でみれば，第 1 種記憶学習作用素と第 2 種記憶学習作用素は異なったものになっている．しかし，式 (2.20) と式 (2.12) の右辺は同じ式であるから，関数の段階でみれば

$$f_{\mathrm{M1}} = f_{\mathrm{M2}} \tag{2.22}$$

となり，第 1 種記憶学習で学習しても第 2 種記憶学習で学習しても同じ関数を得るのである．そこで，以下の議論では，$f_{\mathrm{M1}}, f_{\mathrm{M2}}$ を単に f_{M} と表すことにする：

$$f_{\mathrm{M}} = f_{\mathrm{M1}} = f_{\mathrm{M2}} \tag{2.23}$$

このことから，f_{M1} に関する性質が f_{M2} に対してもそのまま成り立つことになる．即ち，第 2 種記憶学習によって得られる関数は一意に定まらないが，標本点における値 Af_{M2} は Y によらず一意に定まり，式 (2.14) で与えられる．第 2 種記憶学習も $\boldsymbol{n}$ を $\mathcal{R}(A)$ に正射影することによって雑音抑制を行なっているのである．

2.3.2 射影学習

射影学習は汎化誤差そのものを小さくしようとする学習法の一つであり，次のように定義される．

定義 2.3 （射影学習 [53]）条件

$$XA = P_{\mathcal{R}(A^*)} \tag{2.24}$$

を満たす作用素 X の中で，

$$J_n[X] = E_{\boldsymbol{n}}\|X\boldsymbol{n}\|^2 \tag{2.25}$$

を最小にする X を射影学習作用素（projection learning operator）といい，$A^{(\mathrm{P})}$ で表す．ここで，$\|\cdot\|$ は H におけるノルムである．$A^{(\mathrm{P})}$ による学習を射影学習（projection learning）という．

射影学習の概念は，次のような考え方から導かれたものである [53]．式（2.8），(2.9) より，

$$f_0 = XAf + X\boldsymbol{n} \tag{2.26}$$

となる．式（2.26）の第 1 項と第 2 項はそれぞれ，f_0 の信号成分，雑音成分とよばれている．射影学習では，信号成分 XAf を f にできるだけ近付け，雑音成分 $X\boldsymbol{n}$ をできるだけ小さくすることによって，f_0 を f に近付けようとする．まず信号成分に関しては，f そのものはわからないけれども H の元であることはわかっているので，f が H のどの元であっても XAf が f の最良近似になるようにする．H のすべての元に対する XAf の全体は部分空間 $\mathcal{R}(XA)$ になる．この空間における f の最良近似は，f の $\mathcal{R}(XA)$ への正射影 $P_{\mathcal{R}(XA)}f$ である．そこで，XAf が $P_{\mathcal{R}(XA)}f$ に一致するようにする：

$$XAf = P_{\mathcal{R}(XA)}f \qquad f \in H \tag{2.27}$$

この式は，

$$XA = P_{\mathcal{R}(XA)} \tag{2.28}$$

となる．従って，$\mathcal{R}(XA) = \mathcal{R}((XA)^*) = \mathcal{R}(A^*X^*) \subset \mathcal{R}(A^*)$ より，

$$\mathcal{R}(XA) \subset \mathcal{R}(A^*) \tag{2.29}$$

を得る．$P_{\mathcal{R}(XA)}f$ の f に対する近似の精度は，$\mathcal{R}(XA)$ が大きくなればなるほど高くなる．そこで，

$$\mathcal{R}(XA) = \mathcal{R}(A^*) \tag{2.30}$$

の場合を考えることにする．こうして式（2.28）より，条件（2.24）が得られるのである．

雑音成分に関しては，$\boldsymbol{n}$ はランダム雑音であるので，個々の $\boldsymbol{n}$ について論じても意味がない．そこで，$\boldsymbol{n}$ に関する平均の意味で $\|X\boldsymbol{n}\|^2$ ができるだけ小さくなるようにする．このことを表しているのが式（2.25）の $J_n[X]$ である．

命題 2.3 （射影学習作用素 [44]）作用素 X が射影学習作用素になるための必要十分条件は，X が

$$XU_P = V_P^\dagger A^* \tag{2.31}$$

を満たすことである．ここで，U_P および V_P はそれぞれ，

$$U_P = AA^* + Q \tag{2.32}$$

$$V_P = A^*U_P^\dagger A \tag{2.33}$$

で定義される行列と作用素である．$A^{(\mathrm{P})}$ の一般形は，

$$A^{(\mathrm{P})} = V_P^\dagger A^* U_P^\dagger + Y(I_M - U_P U_P^\dagger) \tag{2.34}$$

で与えられる．ここで，I_M は $\mathbf{C}^M$ 上の単位行列であり，Y は $\mathbf{C}^M$ から H への任意の作用素である．射影学習によって得られる関数 f_{P} は，

$$f_{\mathrm{P}} = V_P^\dagger A^* U_P^\dagger \boldsymbol{y} \tag{2.35}$$

で与えられる．

式 (2.34) の右辺には作用素 Y が含まれているので，射影学習作用素は一般には一意に定まらない．しかし，それらを $\boldsymbol{y}$ に作用させて得られる関数 f_{P} は，式 (2.35) からわかるように，Y によらず一意に定まるのである．

射影学習作用素は次のように特徴づけられる．まず，式 (2.16) で定義した部分空間 S_b は，

$$S_b = \mathcal{R}(A) \dot{+} Q\mathcal{R}(A)^\perp \tag{2.36}$$

と直和分解できる [44]．この直和分解に関して，次の命題が成立する．

命題 2.4 [44] 作用素 X が射影学習作用素になるための必要十分条件は，

$$Xu = \begin{cases} A^\dagger u & : u \in \mathcal{R}(A) \\ 0 & : u \in Q\mathcal{R}(A)^\perp \end{cases} \tag{2.37}$$

が成立することである．

H の元 f を A で変換して得られるベクトル Af は $\mathcal{R}(A)$ に入る．それに $\boldsymbol{n}$ が加わることにより，$\boldsymbol{y}$ は $\mathcal{R}(A)$ からはみ出してしまう．しかし，$\boldsymbol{n}$ は $\mathcal{R}(Q)$ の元であるので，S_b から出ることはない．このとき射影学習は，$\boldsymbol{y}$ を $\mathcal{R}(A)$ の S_b における補空間に沿って $\mathcal{R}(A)$ に射影することにより雑音を抑制しようとする．そのような補空間の中で雑音を最も多く含むものが $Q\mathcal{R}(A_s)^\perp$ であることが，式 (2.37) の第 2 式が 0 になっていることからわかるのである．

2.3.3 部分射影学習

部分射影学習は，f が H の閉部分空間 S に属すことがわかっている場合に，射影学習の考え方を適用できるようにしたものである．

定義 2.4 （部分射影学習 [50]）S を H の閉部分空間とする．条件

$$XAP_S = P_{\mathcal{R}(P_S A^*)} P_S \tag{2.38}$$

を満たす作用素 X の中で，式 (2.25) の $J_n[X]$ を最小にする X を部分射影学習作用素 (partial projection learning operator) といい，$A^{(\mathrm{PTP})}$ で表す．$A^{(\mathrm{PTP})}$ による学習を部分射影学習（partial projection learning）という．

命題 2.5 （部分射影学習作用素 [47]）作用素 X が部分射影学習作用素になるための必要十分条件は，X が

$$XU_{PTP} = V_{PTP}^{\dagger}A^{*} \tag{2.39}$$

を満たすことである．ここで，U_{PTP} および V_{PTP} はそれぞれ，

$$U_{PTP} = AP_SA^{*} + Q \tag{2.40}$$
$$V_{PTP} = P_SA^{*}U_{PTP}^{\dagger}AP_S \tag{2.41}$$

で定義される行列と作用素である．$A^{(\mathrm{PTP})}$ の一般形は，

$$A^{(\mathrm{PTP})} = V_{PTP}^{\dagger}A^{*}U_{PTP}^{\dagger} + Y(I_M - U_{PTP}U_{PTP}^{\dagger}) \tag{2.42}$$

で与えられる．ここで，Y は $\mathbf{C}^M$ から H への任意の作用素である．部分射影学習によって得られる関数 f_{PTP} は，$f \in S$ に対して

$$f_{\mathrm{PTP}} = V_{PTP}^{\dagger}A^{*}U_{PTP}^{\dagger}\boldsymbol{y} \tag{2.43}$$

で与えられる．

部分射影学習作用素も，式 (2.42) の右辺に作用素 Y が含まれているので，射影学習作用素と同様に一意に定まらない．しかし，それらを $\boldsymbol{y}$ に作用させて得られる関数 f_{PTP} は，式 (2.43) からわかるように，Y によらず一意に定まるのである．

2.3.4 平均射影学習

平均射影学習は，頻繁に現れる関数ほど学習精度を上げたい場合に，射影学習の考え方を適用できるようにしたものである．E_f を $\{f\}$ に関する平均とするとき，次のように定義される．

定義 2.5 （平均射影学習 [65]）汎関数

$$J_{\mathrm{AP}}[X] = E_f\|XAf - f\|^2 \tag{2.44}$$

を最小にする作用素 X の中で，式 (2.25) の $J_n[X]$ を最小にする X を平均射影学習作用素（averaged projection learning operator）といい，$A^{(\mathrm{AP})}$ で表す．$A^{(\mathrm{AP})}$ による学習を平均射影学習（averaged projection learning）という．

$\{f\}$ の相関作用素を R で表す：

$$R = E_f(f \otimes \overline{f}) \tag{2.45}$$

命題 2.6 （平均射影学習作用素 [67]）作用素 X が平均射影学習作用素になるための必要十分条件は，X が

$$XU_{AP} = RA^{*}V_{AP}^{\dagger}RA^{*} \tag{2.46}$$

を満たすことである．ここで，U_{AP} および V_{AP} はそれぞれ，

$$U_{AP} = ARA^{*} + Q \tag{2.47}$$
$$V_{AP} = RA^{*}U_{AP}^{\dagger}ARA^{*} \tag{2.48}$$

で定義される行列と作用素である．$A^{(\mathrm{AP})}$ の一般形は，

$$A^{(\mathrm{AP})} = RA^* V_{AP}^\dagger RA^* U_{AP}^\dagger + Y(I_M - U_{AP} U_{AP}^\dagger) \tag{2.49}$$

で与えられる．ここで，Y は $\mathbf{C}^M$ から H への任意の作用素である．平均射影学習によって得られる関数 f_{AP} は，$f \in \mathcal{R}(R)$ に対して

$$f_{\mathrm{AP}} = RA^* V_{AP}^\dagger RA^* U_{AP}^\dagger \boldsymbol{y} \tag{2.50}$$

で与えられる．

平均射影学習作用素も，式 (2.49) の右辺には作用素 Y が含まれているので，射影学習作用素と同様に一意に定まらない．しかし，それらを $\boldsymbol{y}$ に作用させて得られる関数 f_{AP} は，式 (2.50) からわかるように，Y によらず一意に定まるのである．

作用素 $A^{(\mathrm{AP})}A$ は，$\mathcal{R}(R)$ の閉包 $\overline{\mathcal{R}(R)}$ の中でみれば，$\overline{\mathcal{R}(R)} \cap \mathcal{N}(A)$ に沿った $\mathcal{R}(RA^*)$ への射影作用素になっている [68]．

2.4 まとめ

本章では，文献 [48, 49, 53] に従って，作用素論的手法を用いて学習問題を定式化した．まず，記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習の 2 種類があることを示した．つづいて，射影学習，部分射影学習，平均射影学習について整理し，本論文で学習問題を論じるための準備を行った．

第 3 章

射影学習族

3.1 まえがき

作用素論的学習理論ではこれまでに，前章で述べた射影学習，部分射影学習，平均射影学習などが提案されてきた．そして，これらの学習方式それぞれに対して，学習作用素が求められたり，雑音抑制能力や汎化能力が個別に論じられてきた．

しかし，議論を更に深めていくためには，個々の学習方式を個別に論じるのではなく，多くの種類の学習方式を同時に統一的に論じることが必要になってくる．

例えば，関数解析の分野でも，個々の作用素の性質を論じる‘作用素論’の段階から，作用素の集合の性質を論じる‘作用素環の理論’，それを更に抽象化した‘ノルム環の理論’へと進むことによって，多くの成果が得られてきた．この成果をフーリエ解析の問題へ適用することにより，従来 100 ページもかかっていたウィーナーのタウバー型定理の証明が，わずか数行で済んでしまったという有名な出来事もあった．ウィーナーのタウバー型定理がなぜ成立するかという本当の理由が，ノルム環の立場からみて始めてわかったのである [42]．

このように，個々の対象の研究から，それらの対象を含む集合の性質の研究へと進むことにより，物事の本質がわかってくるのである．従って，個々の学習方式を個別に論じる段階から，多くの種類の学習方式を含む学習族の理論へと進むことにより，学習に対する理解が一段と深まるものと思われる．

このような考え方に基づいて，射影学習族の概念が提案された [43, 37, 23]．射影学習族とは，射影学習，部分射影学習，平均射影学習に共通する 2 種類の性質

性質 1 作用素 XA がある部分空間でみれば射影になっている．

性質 2 性質 1 を満たす X の中で式 (2.25) を最小にする．

を満たす無限に多くの学習方式である．

これまでに射影学習族の定義が 3 種類与えられている [43, 37, 23]．しかし，これらの定義にはまだ不自然な部分が残っており，理論を更に発展させていくためには不便であった．そこで本章では，まずこれらの定義の問題点を整理し，射影学習族のより自然な定義を与えることにする．そして，この定義を基に，学習作用素の一般形を射影学習族に含まれるすべての学習方式に対して統一的に求めると共に，雑音抑制能力や汎化能力等の諸性質を統一的に解明し，射影学習族の基礎理論を構築する．

3.2 射影学習族の従来の定義とその問題点

本節では，これまでに与えられている射影学習族の 3 種類の定義 [43, 37, 23] を示し，それぞれの問題点を指摘する．

まず文献 [43] では，3 種類の射影学習作用素が連立作用素方程式で特徴づけられることに着目し，それらをまとめて

$$XA = \mathcal{B} \tag{3.1}$$

$$X\mathcal{C} = Y\mathcal{A}^* \tag{3.2}$$

と表された．式 (3.1) , (3.2) の解を X_0 で表せば，X_0 による学習は，

$$\mathcal{A} = A,\quad \mathcal{B} = P_{\mathcal{R}(A^*)},\quad \mathcal{C} = Q \tag{3.3}$$

のとき射影学習になり，

$$\mathcal{A} = AP_S,\quad \mathcal{B} = P_S P_{\mathcal{R}(P_S A^*)},\quad \mathcal{C} = Q \tag{3.4}$$

のとき部分射影学習になり，

$$\mathcal{A} = ARA^*,\quad \mathcal{B} = RA^*,\quad \mathcal{C} = Q \tag{3.5}$$

のとき平均射影学習になる．

文献 [43] では，X_0 の様々な数学的性質が論じられている．それらは 3 種類の射影学習に共通する性質をより一般化したものであるが，それらの性質をそのまま学習の立場から解釈することは難しい．また，性質 1 と性質 2 は，式 (3.1) , (3.2) だけからは導くことができない．このように考えると，すべての $\mathcal{A}, \mathcal{B}, \mathcal{C}$ に対する X_0 による学習の全体は，射影学習族とよぶには広すぎることがわかる．射影学習族の定義としては，学習の範囲を適切に制限することにより，そのままで学習の立場から様々な議論ができるものが望ましい．

文献 [37] で与えられたものを示す．

定義 3.1 （R_F射影学習 [37]）H 上の任意に固定した自己共役作用素 R_F に対して，ある作用素 C と共に，連立作用素方程式

$$XAR_F = R_F P_{\mathcal{R}(R_F A^*)} \tag{3.6}$$

$$XQ = CR_F A^* \tag{3.7}$$

を満たす作用素 X を R_F 射影学習作用素といい，$A^{(\mathrm{RF})}$ で表す．$A^{(\mathrm{RF})}$ による学習を R_F 射影学習という．すべての自己共役作用素 R_F に対する R_F 射影学習の全体を，R_F から導かれる射影学習族という．

R_F 射影学習は，

$$R_F = I \tag{3.8}$$

のとき射影学習になり，

$$R_F = P_S \tag{3.9}$$

のとき部分射影学習になり，

$$R_F = R^{1/2} \tag{3.10}$$

のとき平均射影学習になる．

R_F 射影学習作用素が性質 1，性質 2 を満たすことは，第 3.4 節で示す定理 3.2からわかる．しかし，定義 3.1 からこれらの性質を直接くみとることは容易なことではない．

文献 [23] で与えられた射影学習族の定義を示す．T を H 上の半正値自己共役作用素とする．$\overline{\mathcal{R}(T)}$ は，

$$\overline{\mathcal{R}(T)} = \mathcal{R}(TA^*) \dot{+} \{\overline{\mathcal{R}(T)} \cap \mathcal{N}(A)\} \tag{3.11}$$

と直和分解できる [68]．$\mathcal{R}(TA^*)$ は有限次元部分空間であるから閉じている．従って，$\mathcal{R}(TA^*)$ への射影を考えることができる．

定義 3.2 （T射影学習 [23]）T を任意に固定した H 上の半正値自己共役作用素とする．閉部分空間 $\overline{\mathcal{R}(T)}$ の中でみれば$\overline{\mathcal{R}(T)} \cap \mathcal{N}(A)$ に沿った $\mathcal{R}(TA^*)$ への射影となる H 上の線形作用素を P_T で表す．条件

$$XAP_{\overline{\mathcal{R}(T)}} = P_T P_{\overline{\mathcal{R}(T)}} \tag{3.12}$$

を満たす作用素 X の中で，

$$J_n[X] = E_{\boldsymbol{n}} \|X\boldsymbol{n}\|^2 \tag{3.13}$$

を最小にする X を T 射影学習作用素といい，$A^{(\mathrm{TP})}$ で表す．$A^{(\mathrm{TP})}$ による学習を T 射影学習という．すべての半正値自己共役作用素 T に対する T 射影学習の全体を，T から導かれる射影学習族という．

T 射影学習は

$$T = I \tag{3.14}$$

のとき射影学習になり，

$$T = P_S \tag{3.15}$$

のとき部分射影学習になり，

$$T = R \tag{3.16}$$

のとき平均射影学習になる．

定義 3.2 は，性質 2 を直接表現したものになっている．一方，性質 1 は「部分空間」および「射影」という 2 種類の要素からなりたっているが，定義 3.2 は，この中の「射影」の部分だけを直接表現しており,「部分空間」に関しては，$\overline{\mathcal{R}(T)}$, $\mathcal{R}(TA^*)$ のように，作用素 T を使って間接的に表現している．従って，定義 3.1 よりは素直な表現になっているが，まだ不自然さが残っている．そこで次節では，射影学習族の本質的な部分だけをとり出した，より自然な定義を与えることにする．

3.3 射影学習族の新しい定義

射影学習，部分射影学習，平均射影学習では，f がそれぞれ H，S, $\overline{\mathcal{R}(R)}$ に含まれていた．本節では，これらの空間を一般的に，H の閉部分空間 S で表すことにする．即ち，

f が S に含まれていることがわかっている場合を考える．S の中でみれば S の部分空間への射影になっている作用素を P で表せば，性質 1 は

$$XAP_S = PP_S \tag{3.17}$$

となる．式 (3.17) はいつでも解 X をもつわけではない．それは次の場合に限られるのである．

補題 3.1 式 (3.17) が解をもつための必要十分条件は，

$$\mathcal{N}(P) \supset S \cap \mathcal{N}(A) \tag{3.18}$$

となることである．

式 (3.18) より，

$$S \cap \mathcal{N}(P) \supset S \cap \mathcal{N}(A) \tag{3.19}$$

となる．一方 S は，

$$S = \mathcal{R}(PP_S) \dot{+} \{S \cap \mathcal{N}(P)\} \tag{3.20}$$

と直和分解できる．従って，式 (3.17) は，

$$XAf = \begin{cases} f : f \in \mathcal{R}(PP_S) \\ 0 : f \in S \cap \mathcal{N}(P) \end{cases} \tag{3.21}$$

と同値である．XAf は式 (2.26) で述べたように f_0 の信号成分であるから，式 (3.21) の第 1 式に現れる $\mathcal{R}(PP_S)$ はできるだけ大きく，第 2 式に現れる $S \cap \mathcal{N}(P)$ はできるだけ小さい方がよい．そこで，式 (3.19) で $S \cap \mathcal{N}(P)$ が最小になる P，即ち，

$$S \cap \mathcal{N}(P) = S \cap \mathcal{N}(A) \tag{3.22}$$

となる P を考える．このとき式 (3.20) より，$\mathcal{R}(PP_S)$ は $S \cap \mathcal{N}(A)$ の S における補空間になる．それを L で表す：

$$L = \mathcal{R}(PP_S) \tag{3.23}$$

こうして，次の射影学習族の定義を得る．

定義 3.3 （SL 射影学習）S を H の閉部分空間とし，L を S における $S \cap \mathcal{N}(A)$ の補空間とする：

$$S = L \dot{+} \{S \cap \mathcal{N}(A)\} \tag{3.24}$$

S の中でみれば $S \cap \mathcal{N}(A)$ に沿った L への射影となる線形作用素を P で表す．条件

$$XAP_S = PP_S \tag{3.25}$$

を満たす作用素 X の中で，

$$J_n[X] = E_{\boldsymbol{n}} \|X\boldsymbol{n}\|^2 \tag{3.26}$$

を最小にする X を SL 射影学習作用素（SL projection learning operator）といい，$A^{(\mathrm{SL})}$ で表す．$A^{(\mathrm{SL})}$ による学習を **SL 射影学習**（SL projection learning）という．すべての S および L に対する SL 射影学習の全体を，S と L から導かれる射影学習族という．

定義 3.3 は，第 3.1 節に示した性質 1 と性質 2 を自然に表現している．従って，射影学習族の自然な定義になっている．

ところで，この定義では，SL 射影学習が汎化能力をどのようにとらえているかということは，それほど明確ではない．そこで，汎化という立場からこの定義を見直してみることにする．まず，雑音がない場合を考える．本論文では，学習対象関数 f が部分空間 S に属しているという前提知識が与えられた場合を考えている．しかし，S の中のどの元かわからないので，f が S のどの元であっても対応できるようにしたい．即ち，できれば，f が S の中のどの元であっても，Af から学習作用素 X によって完全に f を推定できるように，

$$X(Af) = f \quad : f \in S \tag{3.27}$$

が成立するようにしたい．しかし，$S \cap \mathcal{N}(A)$ の元は A によって零元に変換されるので，X として線形作用素を考えている限り，$S \cap \mathcal{N}(A)$ の元 f に対しては，$X(Af) = 0$ になってしまう．

一方，$S \cap \mathcal{N}(A)$ の S における補空間 L の非零元 f に対しては，Af が 0 になることはない．そこで，L の元 f に対しては，Af から学習作用素 X によって完全に f を推定できるようにしたい．即ち，L の中の任意の f に対して

$$X(Af) = f \quad : f \in L \tag{3.28}$$

が成立するようにしたい．

これら 2 つの状況，即ち，$S \cap \mathcal{N}(A)$ の元 f に対しては $X(Af) = 0$ になり，L の元 f に対しては式 (3.28) が成立するという状況を，作用素 XA に着目して表現したものが式 (3.25) である．即ち，作用素 XA は $S \cap \mathcal{N}(A)$ の元を零にし，L の元を不変に保つので，空間 S の中でみれば，$S \cap \mathcal{N}(A)$ に沿った L への射影作用素になっているのである．X は線形作用素であるので，これら 2 つの状況から X は本質的には決まってしまう．従って，L の中の任意の f を Af から学習作用素 X によって完全に推定できるかどうかということが，SL 射影学習における汎化能力の評価になっているのである．

雑音がある場合には，学習結果は式 (2.26) に示すように，$X\boldsymbol{n}$ によって XAf のまわりに分布するので，式 (3.26) で表される散らばりを小さくするのである．

再び雑音がない場合に話をもどす．$f \in S$ は式 (3.24) より，L の元 f_1 と $S \cap \mathcal{N}(A)$ の元 f_2 に

$$f = f_1 + f_2 \tag{3.29}$$

と一意に分解できる．そして，前にも述べたように，XA は S の中でみれば，$S \cap \mathcal{N}(A)$ に沿った L への射影学習作用素になっている．従って，

$$X(Af) = f_1 \tag{3.30}$$

となり，X によって f の L 成分が得られるのである．

ところで，標本点 $\{\boldsymbol{x}_m\}_{m=1}^{M}$ が固定されているとき，$S \cap \mathcal{N}(A)$ は S によって一意に決まってしまうが，補空間 L の取り方には無限の自由度が残っている．従って，作用素 XA によって射影される方向は $S \cap \mathcal{N}(A)$ に固定されているが，射影された結果を受けとめる空間 L の選び方には自由度が残っている．しかも，L の取り方によって，式 (3.29) の f_1 と f_2 が共に変化する．こうして，空間 L の決め方が重要になってくる．

L は，どのような学習を行いたいかによって決まってくる．例えば，射影学習では，$S=H$ の個々の元 f に対して，$X(Af)$ が L の中でみれば f の最良近似，即ち，f の L への正射影になるようにするために，

$$L=\mathcal{R}(A^*) \tag{3.31}$$

が採用されている．同様に部分射影学習では，$S(\subset H)$ の個々の元 f に対して，$X(Af)$ が L の中でみれば f の最良近似，即ち，f の L への正射影になるようにするために，

$$L=\mathcal{R}(P_S A^*) \tag{3.32}$$

が採用されている．

また平均射影学習は，命題 2.6 のすぐ下で述べたように，$S=\overline{\mathcal{R}(R)}, L=\mathcal{R}(RA^*)$ となっているが，この L のもつ意味は次のように特徴づけることができる．$\mathbf{C}^M$ から H への有界線形作用素の全体を $\mathcal{B}$ で表し，ある L に対して式 (3.25) を満たす X の全体を $\mathcal{B}(L)$ で表す．また，式 (2.44) の $J_{\mathrm{AP}}[X]$ に対して，

$$J_0=\min_{X\in\mathcal{B}} J_{\mathrm{AP}}[X] \tag{3.33}$$

$$J[L]=\min_{X\in\mathcal{B}(L)} J_{\mathrm{AP}}[X] \tag{3.34}$$

とおく．$J_{\mathrm{AP}}[X]=J_0$ となる作用素 X が平均射影学習作用素である．式 (3.34) の右辺の値は空間 L の取り方によって変ってくる．それを $J[L]$ で表してある．式 (3.33)，(3.34) の X の探索範囲には $\mathcal{B}\supset\mathcal{B}(L)$ なる関係があるので，一般には J_0 よりも $J[L]$ の方が大きくなる：

$$J_0\leq J[L] \tag{3.35}$$

しかし，

$$J_0=J[\mathcal{R}(RA^*)] \tag{3.36}$$

となっているので，

$$L=\mathcal{R}(RA^*) \tag{3.37}$$

ととることにより，式 (3.35) で等号を成立させることができる．従って，次の定理が成立する．

定理 3.1 平均射影学習は，$S=\overline{\mathcal{R}(R)}$ としたとき，式 (3.34) の $J[L]$ を最小にする L に対する SL 射影学習である。

なお，文献 [69] の系 2 より，このような $L\subset\overline{\mathcal{R}(R)}$ は一意に定まることがわかる．

これらの学習以外にも，次のような SL 射影学習を考えることができる．平均射影学習では，頻繁に現れる関数ほど精度よく学習することを考えた．逆に，まれにしか現れない関数ほど精度よく学習したい場合もある．この学習を，逆平均射影学習とよぶことにする．即ち，スカラー k に対して $k^\dagger$ を

$$k^\dagger=\begin{cases}1/k & k\neq 0\\ 0 & k=0\end{cases} \tag{3.38}$$

で定義し，f の出現確率を $p(f)$ で表すとき，逆平均射影学習は，汎関数

$$J[X]=\int_{f\in H}\|XAf-f\|^2 p(f)^\dagger df \tag{3.39}$$

を最小にする X の中で，式 (3.26) を最小にする学習である．ただし，式 (3.39) の積分が存在する場合のみを考える．

作用素 G を

$$G = \int_{f \in H} (f \otimes \overline{f}) p(f)^{\dagger} df \tag{3.40}$$

で定義すれば，逆平均射影学習は，

$$S = \overline{\mathcal{R}(G)}, \quad L = \mathcal{R}(GA^*) \tag{3.41}$$

の場合の SL 射影学習になっている．

3.4 各種定義の相互関係

これまでに述べてきた射影学習族に対する 3 種類の定義，即ち，定義 3.1，定義 3.2，定義 3.3 が，集合としてはすべて一致することを次に示す．

定理 3.2 S と L から導かれる射影学習族と T から導かれる射影学習族，R_F から導かれる射影学習族はすべて一致する．

ところで，定義の良さは

1. 本来備えているべき性質をどれだけ素直に表現しているか．
2. その定義から多くの結果をいかに自然に導き出せるか．

という 2 つの立場から評価しなければいけない．定理 3.2 より，3 種類の定義はすべて論理的には等価であったけれども，この 2 つの基準からみれば，定義 3.3 が最も自然な定義になっている．実際，第 1 の基準については，性質 1 を直接表現しているものは定義 3.3 だけである．また第 2 の基準については，次節で示すように，SL 射影学習の学習機構を素直に解釈できる学習作用素の一般形を，定義 3.3 から自然に導くことができる．また，第 4 章で論じる最小分散射影学習の概念など，定義 3.3 にもとづき多くの結果を自然に得ることができるのである．従って今後は，定義 3.3 を射影学習族の定義として用いていくことにする．

3.5 射影学習族の諸性質

本節では定義 3.3 に従って，SL 射影学習の性質を論じていく．まず，SL 射影学習作用素の一般形を与える．作用素 A_s を，

$$A_s = AP_S \tag{3.42}$$

で定義する．A_s は，S の中でみたときの A の働きを表している．SL 射影学習では，部分空間 S の元をどのように扱うかが重要である．従って，A よりも A_s が本質的な役割を果たすことになる．

ところで，S の元全体を A で変換したものが $\mathcal{R}(A_s)$ であるから，$\mathcal{R}(A_s)$ の中のすべての元が教師信号 $\boldsymbol{y}$ の中の信号成分になり得る．即ち，$\mathcal{R}(A_s)$ の中では，信号と雑音を区別することができない．雑音を信号から区別できるのは，$\mathcal{R}(A_s)$ の外の部分だけであ

る．従って，$\mathcal{R}(A_s)$ が $\mathbf{C}^M$ の真の部分空間になっている場合にのみ，即ち，A_s が全射でない場合にのみ，雑音を抑制できるのである．本論文では，一般的に $\mathcal{R}(A_s) \subset \mathbf{C}^M$ の場合について議論を進めていくことにする．

部分空間 S_t を，

$$S_t = \mathcal{R}(A_s) + \mathcal{R}(Q) \tag{3.43}$$

で定義する．$\mathcal{R}(A_s)$ は S の元を A で変換したものの全体 AS である．また，$\mathcal{R}(Q)$ は $\boldsymbol{n}$ が含まれる部分空間であった．従って，$f \in S$ に対する教師信号 $\boldsymbol{y}$ は S_t に含まれるのである．S_t の添字 t は，この "教師信号（teacher signal）が含まれる空間" の意味で，teacher signal の t が用いてある．S_t への正射影行列を P_{S_t} で表す．

S_t は，

$$S_t = \mathcal{R}(A_s) \dot{+} Q\mathcal{R}(A_s)^{\perp} \tag{3.44}$$

と直和分解できる [37]．従って $\mathbf{C}^M$ は，

$$\begin{aligned} \mathbf{C}^M &= S_t \oplus S_t^{\perp} \\ &= \mathcal{R}(A_s) \dot{+} \{Q\mathcal{R}(A_s)^{\perp} \oplus S_t^{\perp}\} \end{aligned} \tag{3.45}$$

と直和分解できる．$Q\mathcal{R}(A_s)^{\perp} \oplus S_t^{\perp}$ に沿った $\mathcal{R}(A_s)$ への射影行列を P_t で表す．

定理 3.3 作用素 X が SL 射影学習作用素になるための必要十分条件は，X が

$$XP_{S_t} = PA_s^{\dagger}P_t \tag{3.46}$$

を満たすことである．$A^{(\mathrm{SL})}$ の一般形は

$$A^{(\mathrm{SL})} = PA_s^{\dagger}P_t + Y(I_M - P_{S_t}) \tag{3.47}$$

で与えられる．ここで，Y は $\mathbf{C}^M$ から H への任意の作用素である．$J_n[X]$ の最小値 J_{n0} は

$$J_{n0} = \langle PA_s^{\dagger}P_tQ, PA_s^{\dagger}\rangle \tag{3.48}$$

で与えられる．ここで，$\langle\cdot,\cdot\rangle$ は作用素のシュミットの内積である[1]．

式（3.46）によって，作用素 X が SL 射影学習作用素になっているかどうかを判定することができる．この判定法はまた，後で述べる定理 3.4 や系 3.3，あるいは定理 5.2 の証明で有効な働きをする．

また，P_t の定義からわかるように

$$P_t = P_tP_{S_t} \tag{3.49}$$

であるから，式（3.46）より直ちに次の系を得る．

[1] C^M から H への有界線形作用素 B, C のシュミットの内積 $\langle B,C\rangle$ は，C^M の任意の正規直交基底 $\{\varphi_m\}_{m=1}^M$ を用いて

$$\langle B,C\rangle = \sum_{m=1}^{M}\langle B\varphi_m, C\varphi_m\rangle$$

で定義される．ここで，右辺の $\langle\cdot,\cdot\rangle$ は H の内積である．この式の右辺の値は，$\{\varphi_m\}_{m=1}^M$ のとり方によらず一定の値になっている [57]．

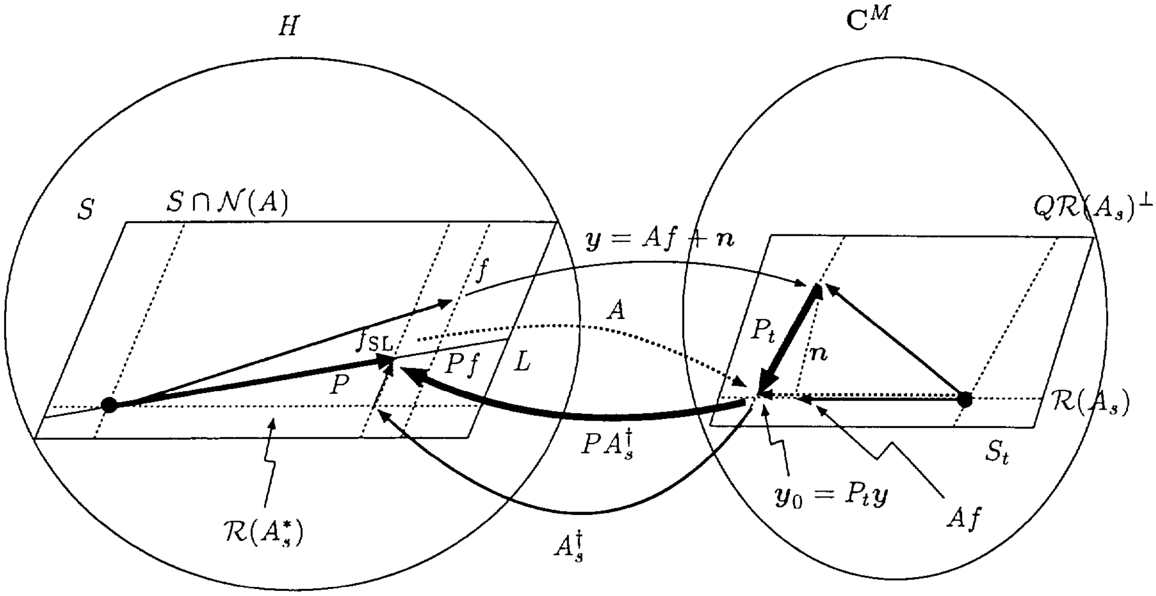


図 3.1　SL 射影学習の学習機構.

系 **3.1** 作用素 X が SL 射影学習作用素になるための必要十分条件は，X が

$$Xu = \begin{cases} PA_s^\dagger P_t u : u \in S_t \\ Yu \quad\quad : u \in S_t^\perp \end{cases} \tag{3.50}$$

を満たすことである．

即ち $A^{(\mathrm{SL})}$ は，S_t の元 u を $PA_s^\dagger P_t u$ に，$S_t^\perp$ の元を H の任意の関数 Yu に変換することにより特徴づけられるのである．このことは，式 (3.47) にも反映されている．即ち，$I - P_{S_t}$ は $S_t^\perp$ への正射影行列であるから，$A^{(\mathrm{SL})}$ は S_t の元を式 (3.47) の第 1 項の $PA_s^\dagger P_t$ で変換し，$S_t^\perp$ の元を第 2 項の Y で変換している．$\boldsymbol{y}$ は S_t の元であり，$S_t^\perp$ に含まれることはない．従って，$S_t^\perp$ の元に対して，Y を自由に選ぶことができるのである．

以上のことから，直ちに次の系を得る．

系 **3.2** SL 射影学習によって得られる関数 f_{SL} は，

$$f_{\mathrm{SL}} = PA_s^\dagger P_t \boldsymbol{y} \tag{3.51}$$

と表される．

式 (3.51) から次のことがわかる．まず行列 P_t によって，教師信号 $\boldsymbol{y}$ を $\mathcal{R}(A_s)$ へ変換し，更に作用素 $PA_s^\dagger$ でベクトル空間 $\mathcal{R}(A_s)$ から関数空間 L へ変換している．この様子を図 3.1 に示す．これら 2 種類の変換の意味を理解するために，次の系を用意する．

系 **3.3** 作用素 X が SL 射影学習作用素になるための必要十分条件は，X が次式を満たすことである．

$$Xu = \begin{cases} PA_s^\dagger u : u \in \mathcal{R}(A_s) \\ 0 \quad\quad\; : u \in Q\mathcal{R}(A_s)^\perp \end{cases} \tag{3.52}$$

式 (3.42) と文献 [47] の式 (43) より，S は

$$S = \mathcal{R}(A_s^*) \oplus \{S \cap \mathcal{N}(A)\} \tag{3.53}$$

と直交直和分解できるので，

$$P_S = A_s^\dagger A_s + P_{S\cap\mathcal{N}(A)} \tag{3.54}$$

が成立する．従って，$PP_S = P(A_s^\dagger A_s + P_{S\cap\mathcal{N}(A)}) = PA_s^\dagger A_s$ となり，

$$PP_S = PA_s^\dagger A_s \tag{3.55}$$

を得る．よって，定義 3.3 の式 (3.25) は，式 (3.52) の第 1 式と同値である．しかも，式 (3.24) からわかるように，A は L から $\mathcal{R}(A_s)$ への全単射になっている．従って，X が SL 射影学習作用素になるかどうか，即ち，式 (3.25) あるいは式 (3.52) の第 1 式のもとで X が式 (3.26) を最小にするかどうかは，$\mathcal{R}(A_s)$ の S におけるどの補空間を零に変換するかによって決まってくる．式 (3.52) の第 2 式は，補空間 $Q\mathcal{R}(A_s)^\perp$ を零に変換することによって式 (3.26) を最小にできることを示している．即ち，$\mathcal{R}(A_s)$ の外で最も多くの雑音を含んでいる空間が $Q\mathcal{R}(A_s)^\perp$ なのである．この雑音抑制を行う行列が P_t である．

そこで，

$$\boldsymbol{y}_0 = P_t \boldsymbol{y} \tag{3.56}$$

とおけば，式 (3.51) は

$$f_{\mathrm{SL}} = PA_s^\dagger \boldsymbol{y}_0 \tag{3.57}$$

となる．すぐ上でも述べたように，A は L から $\mathcal{R}(A_s)$ への全単射になっている．従って，A で変換すれば $\boldsymbol{y}_0$ になる L の元，即ち，$\boldsymbol{y}_0$ の L における A に関する原像は一意に決まる．実は，式 (3.57) の f_{SL} はこの原像になっており，作用素 $PA_s^\dagger$ は $\mathcal{R}(A_s)$ から L への A の逆変換になっているのである．それを示すために，新しく S の部分空間 $\mathcal{R}(A_s^*)$ を考える．$\mathcal{R}(A_s^*)$ と $\mathcal{R}(A_s)$ でみれば，A_s と $A_s^\dagger$ が互いの逆変換になっている．また，L と $\mathcal{R}(A_s^*)$ でみれば，$P_{\mathcal{R}(A_s^*)}$ と P が互いの逆変換になっている．更に，

$$P_{\mathcal{R}(A_s^*)} = A_s^\dagger A_s \tag{3.58}$$

であるから，L と $\mathcal{R}(A_s)$ でみれば，A と $PA_s^\dagger$ が互いの逆変換になっている．従って，f_{SL} は $\boldsymbol{y}_0$ の L における A に関する原像になっているのである．

ところで，SL 射影学習による雑音の抑制効果は，本来式 (3.26) のように関数空間 H の中で評価されていた．しかし，上述の議論からわかるように，教師信号 $\boldsymbol{y}$ に含まれる雑音をベクトル空間 $\mathbf{C}^M$ の中で射影行列 P_t によって抑制することにより，式 (3.26) の本来の意味での雑音抑制を実現できるのである．

ところで，式 (3.51) は f_{SL} の式 (2.9) に対応した表現である．式 (2.26) に対応した表現を次に示す．

系 **3.4** f_{SL} は，

$$f_{\mathrm{SL}} = Pf + PA_s^\dagger P_t \boldsymbol{n} \tag{3.59}$$

と表すことができる．

式 (3.59) より，f_{SL} の信号成分 XAf が Pf になっており，条件 (3.25) が実現されていることがわかる．雑音成分 $X\boldsymbol{n}$ は $PA_s^\dagger P_t\boldsymbol{n}$ になり，その広がりを表したものが式 (3.48) である．

次に，f_{SL} の標本点における値がどのようになっているかを示す．

系 **3.5** Af_{SL} は，

$$Af_{\mathrm{SL}} = \boldsymbol{y}_0 \tag{3.60}$$
$$= Af + P_t\boldsymbol{n} \tag{3.61}$$

で与えられる．

即ち，標本点における値は，真の値 Af でもなく，教師信号 $\boldsymbol{y}$ でもなく，$\boldsymbol{y}_0$ に一致しているのである．その値は，式 (3.61) からわかるように，$\boldsymbol{n}$ が $\mathcal{N}(P_t)\cap S_t = Q\mathcal{R}(A_s)^\perp$ に含まれているときは，真の値 Af に完全に一致するのである．

雑音の性質によっては，SL 射影学習作用素および学習結果の関数をより簡単に表現することができる．

定理 3.4 条件

$$Q\mathcal{R}(A_s)^\perp \subset \mathcal{R}(A_s)^\perp \tag{3.62}$$

が成立するとき，またそのときに限り，

$$A^{(\mathrm{SL})} = PA_s^\dagger + Y(I_M - P_{S_t}) \tag{3.63}$$
$$f_{\mathrm{SL}} = PA_s^\dagger\boldsymbol{y} = Pf + PA_s^\dagger\boldsymbol{n} \tag{3.64}$$

と表現できる．

式 (3.63) では，式 (3.47) の右辺第 1 項の $PA_s^\dagger P_t$ が単に $PA_s^\dagger$ になっている．一般に，$\mathcal{N}(A_s^\dagger) = \mathcal{R}(A_s)^\perp$ であるので，$\mathcal{R}(A_s)^\perp$ の元は $A_s^\dagger$ によって 0 に変換される．従って，式 (3.62) が成立しているとき，$Q\mathcal{R}(A_s)^\perp$ の元も $A_s^\dagger$ によって 0 に変換される．よって，射影行列 P_t を用いなくても，$A_s^\dagger$ だけで雑音を抑制できるのである．

例えば，雑音 n_m の分散が等しくσ^2 であり，互いに無相関であるとき，$Q = \sigma^2 I$となり，この条件が成立する．

A_s が全射ならば，$\mathcal{R}(A_s)^\perp = \{0\}$ となり，式 (3.62) が常に成立する．しかし式 (3.63) の右辺第 2 項は零になり，$A^{(\mathrm{SL})}$ は

$$A^{(\mathrm{SL})} = PA_s^\dagger \tag{3.65}$$

と更に簡単な形に表現できる．しかしこのとき，$A^{(\mathrm{SL})}$ によって雑音を抑制することはできないので，式 (3.65) の表現は，雑音がない場合にのみ有効になるのである．

最後に，射影学習族のもとになった 3 種類の射影学習に対する学習作用素を特徴付ける作用素方程式とその一般形の新しい表現を示す．

(i) 射影学習作用素

$$XP_{S_b} = A^\dagger P_t \tag{3.66}$$
$$A^{(\mathrm{P})} = A^\dagger P_t + Y(I - P_{S_b}) \tag{3.67}$$

(ii) 部分射影学習作用素

$$XP_{S_t} = A_s^{\dagger} P_t \tag{3.68}$$

$$A^{(\mathrm{PTP})} = A_s^{\dagger} P_t + Y(I - P_{S_t}) \tag{3.69}$$

(iii) 平均射影学習作用素

$$XP_{S_t} = P_W A_s^{\dagger} P_t \tag{3.70}$$

$$A^{(\mathrm{AP})} = P_W A_s^{\dagger} P_t + Y(I - P_{S_t}) \tag{3.71}$$

ここで，$P_W = RA^*(ARA^*)^{\dagger}A$ である．

式 (3.67) , (3.69) , (3.71) はそれぞれ，式 (2.34) , (2.42) , (2.49) に対応するものである．式 (2.34) , (2.42) , (2.49) に示してあるこれまでに求められていた学習作用素の一般形からは，雑音抑制などの学習機構を直接解釈することは難しかった．一方，式 (3.67) , (3.69) , (3.71) の一般形は，図 3.1に示した学習機構が直接表現された，わかりやすいものになっている．

3.6 簡単な例

H として，Ω 以下に帯域制限された 1 変数複素数値関数の全体であり，内積が

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx \tag{3.72}$$

で定義される空間を考える．H の再生核 $K(x, x')$ は，

$$K(x, x') = \frac{\Omega}{\pi}\mathrm{sinc}\frac{\Omega}{\pi}(x - x') \tag{3.73}$$

となる．ここで，

$$\mathrm{sinc}(x) = \begin{cases} \dfrac{\sin(\pi x)}{\pi x} & : x \neq 0 \\ 1 & : x = 0 \end{cases} \tag{3.74}$$

である．$\{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5\} = \{-\frac{2\pi}{\Omega}, -\frac{\pi}{\Omega}, 0, \frac{\pi}{\Omega}, \frac{2\pi}{\Omega}\}$ とし，関数 φ_n を

$$\varphi_n(x) = K(x, \hat{x}_n) \tag{3.75}$$

で定義するとき，$S = \mathcal{L}\{\{\varphi\}_{n=1}^{5}\}$ である場合を考える．ここで，$\mathcal{L}\{\cdot\}$ は括弧の中に示された関数によって張られる部分空間である．標本点数を 5 個とし，

$$\{x_1, x_2, x_3, x_4, x_5\} = \{-\frac{3\pi}{\Omega}, -\frac{\pi}{\Omega}, 0, \frac{\pi}{\Omega}, \frac{3\pi}{\Omega}\}$$

とする．このとき，

$$S \cap \mathcal{N}(A) = \mathcal{L}\{\varphi_1, \varphi_5\} \tag{3.76}$$

になる．L として，

$$L = \mathcal{L}\{\varphi_1 + \varphi_2, \varphi_3, \varphi_4 + \varphi_5\} \tag{3.77}$$

を考える．雑音の相関行列が，σ を正の定数として，

$$Q = \sigma^2 \begin{pmatrix} 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \end{pmatrix} \tag{3.78}$$

である場合を考える．学習したい関数 f として，

$$f = \sum_{n=1}^{5} a_n \varphi_n \tag{3.79}$$

を考える．$\boldsymbol{y}$ は，

$$\begin{aligned} &(y_1 \;\; y_2 \;\; y_3 \;\; y_4 \;\; y_5) = \\ &\quad (n_1 \;\; f(x_2)+n_2 \;\; f(x_3)+n_3 \;\; f(x_4)+n_4 \;\; n_5) \end{aligned}$$

を転置したものになる．P_t は

$$P_t = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -0.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.80}$$

となるので，$\boldsymbol{y}_0$ は

$$\boldsymbol{y}_0 = \begin{pmatrix} 0 \\ y_2 - 0.5y_1 \\ y_3 \\ y_4 - 0.5y_5 \\ 0 \end{pmatrix} \tag{3.81}$$

となる．即ち，P_t による雑音抑制が，x_1, x_5 における標本値を 0 にし，x_2, x_4 における標本値からそれぞれ $0.5y_1$, $0.5y_5$ を引くことにより，実現されていることがわかる．

次に $\boldsymbol{y}_0$ を $PA_s^\dagger$ で変換することにより f_{SL} を求めると，

$$f_{\mathrm{SL}} = \frac{\pi}{\Omega}\{(y_2 - 0.5y_1)(\varphi_1 + \varphi_2) + y_3\varphi_3 + (y_4 - 0.5y_5)(\varphi_4 + \varphi_5)\} \tag{3.82}$$

となる．この式により標本値 $\{y_m\}_{m=1}^5$ から f_{SL} を求めることができる．また，式 (3.82)，(3.77) より，f_{SL} が L の元になっていることがわかる．式 (3.82) を更に変形すれば，

$$\begin{aligned} f_{\mathrm{SL}} = &\{a_2(\varphi_1 + \varphi_2) + a_3\varphi_3 + a_4(\varphi_4 + \varphi_5)\} \\ &+ \frac{\pi}{\Omega}\{(n_2 - 0.5n_1)(\varphi_1 + \varphi_2) + n_3\varphi_3 + (n_4 - 0.5n_5)(\varphi_4 + \varphi_5)\} \end{aligned} \tag{3.83}$$

となる．この式は，f_{SL} の式 (2.26) に対応した表現であり，括弧 $\{\}$ で示した第 1 項と第 2 項がそれぞれ，信号成分と雑音成分である．まず，式 (3.77) より信号成分が L の元になっていることがわかる．また，式 (3.79) より f と $(a_2 - a_1)\varphi_1 + (a_4 - a_5)\varphi_5$ だけ

異なることがわかる．そしてこの成分は，式 (3.76) より $S \cap \mathcal{N}(A)$ の元になっているので，信号成分が f を $S \cap \mathcal{N}(A)$ に沿って L に射影したものになっていることがわかる．また，$a_1 = a_2$ かつ $a_4 = a_5$ であれば信号成分は f に一致する．このとき $f \in L$ となっていることも，式 (3.79)，(3.77) から理解できる．一方，式 (3.83) の第 2 項である雑音成分も，式 (3.77) より L の元になっている．従って，f_{SL} が信号成分のまわりに L の中で分布していることがわかるのである．

3.7 まとめ

本章では，まず射影学習族の従来の定義の問題点を整理し，SL 射影学習の概念を導入することにより，射影学習族のより自然な定義を与えた．そして，それが従来の定義と等価であることを示した．この新しい定義を基に，学習作用素の一般形を射影学習族に含まれるすべての学習方式に対して統一的に求めると共に，雑音抑制能力や汎化能力等の諸性質を統一的に解明し，射影学習族の基礎理論を構築した．

第 4 章

最小分散射影学習

4.1 まえがき

本章では，前章で構築した射影学習族の理論の一つの応用として，最小分散射影学習の概念を提案する．学習によって得られる関数は，教師信号に含まれている雑音によって，雑音がない場合の学習結果のまわりに分布することになる．この分布の広がりが小さければ小さいほど，雑音に影響されず安定した学習結果を得ることができる．この広がりを射影学習族の中で最も小さくする SL 射影学習が最小分散射影学習である．本章では，その特徴づけを行なうと共に，最小分散射影学習のもつ様々な性質を明らかにする．

4.2 最小分散射影学習

SL 射影学習では，任意に固定した S に対して，式 (3.24) を満たす限り L を自由に選ぶことができる．ところで，式 (3.51) より f_{SL} は L の元になっている．そして式 (3.59) より，f_{SL} は雑音成分 $PA_s^{\dagger}P_t\boldsymbol{n}$ によって信号成分 Pf のまわりに L の中で分布することになる．この分布の大きさを表す値が式 (3.48) の J_{n0} であった．この値が小さければ小さいほど，雑音に影響されず，安定した学習結果を得ることができる．そこで，J_{n0} を最小にするように L を決定することにする．以上の考えに基づき，次の概念を定義する．

定義 4.1 任意に固定した S と，式 (3.48) の J_{n0} を最小にする L に対する SL 射影学習を，最小分散射影学習 (minimum variance projection learning) という．

J_{n0} は，式 (3.59) より信号成分 Pf のまわりの分散であり，通常用いる平均のまわりの分散とは異なるものになっている．しかし，例えば雑音の平均が 0 である場合には，両者は一致する．

4.3 最小分散射影学習の諸性質

本節では，最小分散射影学習の特徴づけを行ない，雑音抑制機構などの諸性質を明らかにしていく．まず，最小分散射影学習は次のように特徴づけられる．

定理 4.1 SL 射影学習が最小分散射影学習になるための必要十分条件は，L が

$$L \supset L_N \tag{4.1}$$

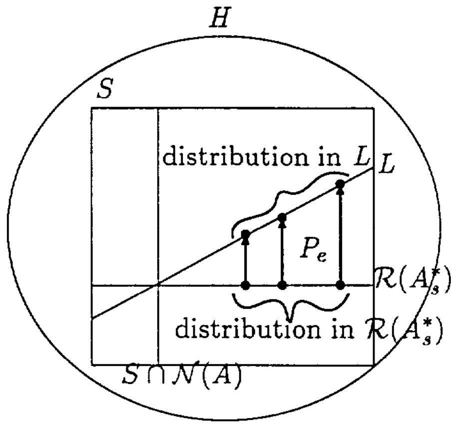


図 4.1　P による分布の拡大. $\mathcal{R}(A_s^*)$ における分布を P で L に射影して得られた分布は, もとの分布よりも大きなものになっている.

を満たすことである. ここで,

$$L_N = \mathcal{R}(A_s^\dagger P_t Q) \tag{4.2}$$

である. このとき, J_{n0} の値は,

$$J_{n0} = \langle A_s^\dagger P_t Q, A_s^\dagger \rangle \tag{4.3}$$

となる.

条件 (4.1) は次のように解釈できる. 式 (3.59) の第 2 項 $PA_s^\dagger P_t \boldsymbol{n}$ に現れる $A_s^\dagger P_t \boldsymbol{n}$ は, $\mathcal{R}(A_s^*)$ の元である. 従って $PA_s^\dagger P_t \boldsymbol{n}$ の分布は, 図 4.1 に示すように, $\mathcal{R}(A_s^*)$ における $A_s^\dagger P_t \boldsymbol{n}$ の分布を, P で $S \cap \mathcal{N}(A)$ に沿って L に射影したものになっている. ところで $\mathcal{R}(A_s^*)$ は, S における $S \cap \mathcal{N}(A)$ の直交補空間である :

$$S = \mathcal{R}(A_s^*) \oplus \{S \cap \mathcal{N}(A)\} \tag{4.4}$$

従って, $\mathcal{R}(A_s^*)$ における分布より, それを L に射影して得られる分布の方が, 一般には大きなものになってしまう. しかし, L がすべての $A_s^\dagger P_t \boldsymbol{n}$ を含んでいれば, P はその分布をそのまま $\mathcal{R}(A_s^*)$ の中に保ち, 拡大することはない. すべての $A_s^\dagger P_t \boldsymbol{n}$ を含む最小の空間は, $\boldsymbol{n}$ が $\mathcal{R}(Q)$ のベクトルであるので, $\mathcal{R}(A_s^\dagger P_t Q) = L_N$ になる. 従って, 式 (4.1) が成立しているとき, SL 射影学習は最小分散射影学習になるのである.

最小分散射影学習によって得られる関数を f_{MV} で表す.

定理 4.2 f_{MV} は,

$$f_{\mathrm{MV}} = Pf + A_s^\dagger P_t \boldsymbol{n} \tag{4.5}$$

と表される.

式 (4.5) では, 式 (3.59) の雑音成分 $PA_s^\dagger P_t \boldsymbol{n}$ の P がなくなり, 単に $A_s^\dagger P_t \boldsymbol{n}$ になっている. 即ち,

$$PA_s^\dagger P_t \boldsymbol{n} = A_s^\dagger P_t \boldsymbol{n} \tag{4.6}$$

となり，最小分散射影学習では，P が雑音成分の広がりを拡大することなく，最小のまま保っていることがわかる．

定義 2.3 ～ 定義 2.5 の 3 種類の射影学習と，最小分散射影学習の関係を述べる．まず，射影学習に関しては $S = H$ であるので，$A_s = A$ となる．従って，式 (3.31), (4.2) より

$$L = \mathcal{R}(A^*) \supset \mathcal{R}(A^\dagger P_t Q) = L_N$$

となり，定理 4.1より射影学習は最小分散射影学習になっている．

部分射影学習に関しては，式 (3.32), (3.42), (4.2) より

$$L = \mathcal{R}(P_S A^*) = \mathcal{R}(A_s^*) \supset L_N$$

となるので，やはり最小分散射影学習になっている．

平均射影学習に関しては，式 (3.37) より $L = \mathcal{R}(RA^*)$ であるので，$\{f\}$ の分布，即ち，$\{f\}$ の相関作用素 R によって事情は異なる．即ち，

$$\mathcal{R}(RA^*) \supset L_N \tag{4.7}$$

であるとき，またそのときに限って，最小分散射影学習になる．例えば，$\mathcal{R}(A^*)$ が R の固有空間であるとき $\mathcal{R}(RA^*) = \mathcal{R}(A^*)$ となり，式 (4.7) が成立する．

条件 (4.1) で注意すべきことは，S を固定しても J_{n0} を最小にする L は一意に定まらないということである．即ち，最小分散射影学習は一般に，1 個の S に対して無数に存在することになる．このことは，次のように説明できる．S は式 (4.4) のように直交直和分解でき，L_N は $\mathcal{R}(A_s^*)$ の部分空間である．従って，S は $L_N \oplus \{S \cap \mathcal{N}(A)\}$ とその S における任意の補空間 L_C に直和分解できる：

$$S = L_C \dot{+} [L_N \oplus \{S \cap \mathcal{N}(A)\}] \tag{4.8}$$

よって，L を

$$L = L_N \dot{+} L_C \tag{4.9}$$

と定義すれば，任意の L_C に対して L が式 (4.1) を満たす．即ち，L_N が $\mathcal{R}(A_s^*)$ の真部分空間になっている限り，L は無数に存在するのである．逆に，L_N が $\mathcal{R}(A_s^*)$ と一致するとき，最小分散射影学習は一意に定まるのである．

L_N が $\mathcal{R}(A_s^*)$ と一致するかどうかは，雑音の性質によって変わる．そこで，これらの空間が一致し，最小分散射影学習が一意に定まるための条件を与える．

定理 4.3 最小分散射影学習が S に対して一意に定まるための必要十分条件は，

$$\mathcal{R}(P_t Q) = \mathcal{R}(A_s) \tag{4.10}$$

が成立することである．このとき最小分散射影学習は，$S = H$ であれば射影学習に，$S \subset H$ であれば部分射影学習になる．

式 (4.10) の左辺の $\mathcal{R}(P_t Q)$ は，抑制しきれずに残った雑音 $P_t \boldsymbol{n}$ を含む最小の部分空間である．一方，式 (4.10) の右辺の $\mathcal{R}(A_s)$ は，S のすべての元 f に対する Af からなる空間である．定理 4.3 は，これら 2 つの部分空間が $\mathbf{C}^M$ の中で一致することが，関数空間 H の 2 つの部分空間 L_N と $\mathcal{R}(A_s^*)$ が一致することと同値になることを意味しているのである．

例えば，$\mathcal{R}(Q) = \mathbf{C}^M$ のとき，

$$\mathcal{R}(P_t Q) = \mathcal{R}(P_t) = \mathcal{R}(A_s)$$

となるので，式 (4.10) が成立する．

4.4 計算機実験

本節では，最小分散射影学習とそうでない SL 射影学習によって得られた関数を比較するために，計算機実験を行なう．H として，帯域が Ω 以下に制限された 1 変数複素数値関数の全体であり，内積が式 (3.72) で定義された空間を考える．H の再生核 $K(x,x')$ は，式 (3.73) となる．

部分空間 S を，

$$\varphi_m(x) = K(x, x_m) \quad (m = 1, 2, \cdots, 5) \tag{4.11}$$

で張られる 5 次元空間とする．ここで，

$$\{x_1, x_2, x_3, x_4, x_5\} = \{-\frac{3\pi}{\Omega}, -\frac{\pi}{\Omega}, \frac{3\pi}{\Omega}, 0, \frac{4\pi}{\Omega}\}$$

である．$\Omega = \pi$ のときのこれらの関数を図 4.2(a) に示す．標本点として，$\{x_1, x_2, x_3\}$ を用いることにする．これらの点に対応する関数 $\varphi_m(x)$ は実線で，残りの 2 個の関数は破線で示してある．雑音の相関行列 Q として，その成分がすべて $\sigma^2 = 0.09$ である 3×3 行列を考える．このとき L_N は，

$$L_N = \mathcal{L}\{\varphi_1 + \varphi_2 + \varphi_3\} \tag{4.12}$$

となる．従って，L_N は，図 4.2(a) に実線で示した関数の和で張られる 1 次元部分空間になる．

部分空間 L が

$$L_1 = \mathcal{L}\{\varphi_1 + \varphi_4 + \varphi_5,\ \varphi_2 - \varphi_4,\ \varphi_3 - \varphi_5\}$$

で与えられる場合の SL 射影学習を，SL_1射影学習とよぶことにする．L_1 は式 (4.12) の L_N を含むので，SL_1射影学習は最小分散射影学習になっている．また，$L_1 \neq \mathcal{R}(A_s^*)$ であるので，SL_1射影学習は部分射影学習とは異なるものである．

一方，$\hat{\varphi} = 3\varphi_4 + \varphi_5$ とおくとき，部分空間 L が

$$L_2 = \mathcal{L}\{\varphi_1 + \hat{\varphi},\ \varphi_2 + \hat{\varphi},\ \varphi_3 + \hat{\varphi}\}$$

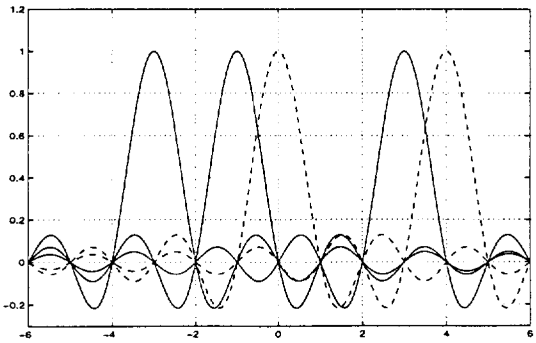
で与えられる場合の SL 射影学習を，SL_2射影学習とよぶことにする．$L_2 \not\supset L_N$ であるので，SL_2射影学習は最小分散射影学習にはなっていない．

学習対象関数 $f(x)$ を，

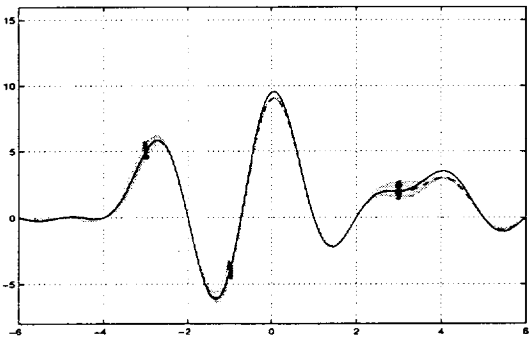
$$f(x) = 5\varphi_1(x) - 4\varphi_2(x) + 2\varphi_3(x) + 19/2\varphi_4(x) + 7/2\varphi_5(x) \tag{4.13}$$

とする．この関数は，学習結果を比較しやすいように，$S \cap \mathcal{N}(A)$ に沿った L_1 への射影と L_2 への射影が一致するように選んだものである．

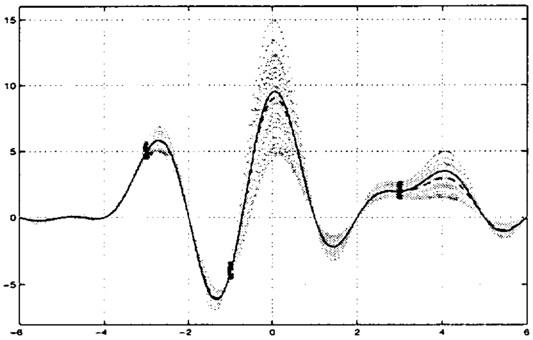
訓練データが 30 回与えられたとき，SL_1射影学習と SL_2射影学習によって得られた 30 個の関数をそれぞれ，図 4.2(b), (c) に示す．これらの図には同時に，$f(x)$ を実線で，$(Pf)(x)$ を破線で示してある．これらの図より，最小分散射影学習がそれ以外の SL 射影学習より，雑音に対して安定した学習結果を与えていることがわかる．



(a)



(b)



(c)

図 4.2　最小分散射影学習とそうでない SL 射影学習の比較実験．(a) 部分空間 S の基底関数．L_N は実線で示した関数の和で張られる 1 次元部分空間になり，$S \cap \mathcal{N}(A)$ は破線で示した関数で張られる 2 次元空間になる．(b) 30 種類の訓練データが与えられた場合の最小分散射影学習によって得られた関数．f, Pf, および f_{MV} をそれぞれ実線，破線，および点線で示す．(c) (b) と同じ訓練データから最小分散射影学習でない SL 射影学習によって得られた関数．f, Pf, および f_{SL} をそれぞれ実線，破線，および点線で示す．(b) の方が散らばりが小さいことがわかる．

4.5 まとめ

第 3 章で構築した射影学習族の理論の一つの応用として，最小分散射影学習の概念を提案した．これは，学習によって得られる関数の教師信号に含まれている雑音に関する分散を，射影学習族の中で最も小さくする学習方式である．まず，SL 射影学習が最小分散射影学習になるための L に関する必要十分条件を与え，各 S に対して最小分散射影学習になるような L が一般には無数に存在することを示した．更に，L が一意に定まるための条件を与え，この条件が成立するとき，最小分散射影学習は射影学習および部分射影学習に限ることを示した．最後に，計算機実験により最小分散射影学習の有効性を示した．

第 5 章

射影学習族の記憶学習による実現

5.1 まえがき

本章では，射影学習族に含まれる各学習方式を記憶学習によって実現する問題を論じる．これは一見奇妙に感じられる問題設定であるが，実は現在行われている多くの研究が，明確には述べられていないけれども，この範疇に属しているのである．例えば，誤差逆伝搬法に代表される訓練誤差を最小にする学習方式を考えてみよう．これは訓練誤差だけを小さくすることを要請する学習方式であるので，記憶学習とよばれている [51]．しかし，記憶学習を行った後で，新しい入力に対してニューラルネットワークが正しく応答しているかどうかを実験的に検証しているということは，本当は汎化誤差を最小にする関数 f_0 を求めたいのであるが，それを記憶学習によって実現しようとしているのである．即ち，汎化誤差最小学習の実現手段として記憶学習が用いられているのである．

この考え方がもう少しはっきりと現れているものに，早期停止法 [7, 54] の研究がある．これは過学習 [17] を防ぐために導入された学習法である．過学習とは，記憶学習を行っているときに，訓練誤差を小さくし過ぎるとかえって汎化誤差が大きくなるという現象である．そこで，訓練データを学習用と検査用の 2 つに分割し，学習用データを使って記憶学習を行う．そして，学習の途中で検査用データに対する誤差を調べる．この誤差が減少している間は記憶学習を続け，この誤差が増加し始めると，そこで学習を終了するのである．このような学習方式が早期停止法である．この学習方式では，検査用データに対する誤差が，汎化誤差の近似として用いられている．このように，この学習方式でも，本当は汎化誤差を最小にする関数 f_0 を求めたいのであるが，それを早期停止法という学習法によって実現しようとしているのである．即ち，汎化誤差最小学習の実現手段として早期停止法が用いられているのである．

正則化学習 [16, 63, 15] もこの範疇に属している．これは，訓練誤差を小さくするだけでは汎化誤差を十分に小さくできないとき，訓練誤差を ‘訓練誤差＋正則化項’ のように修正した新しい評価基準を用いる学習法である．ここでも，本当は汎化誤差を最小にする関数 f_0 を求めたいのであるが，それを正則化学習によって実現しようとしているのである．即ち，汎化誤差最小学習の実現手段として正則化学習が用いられているのである．

このように，明確には述べられていないけれども，非常に多くの研究が，本来行いたい学習を別の学習方式によって実現しようとする問題を扱っている．そして，始めてこの種の問題を明確に定式化したのが，小川等によって始められた許容性の理論である [50, 51, 70, 71]．

許容性の概念の意義を理解するために，先ずは，それ以前の研究で抱えていた問題点について述べる．ここで扱うべき問題をもう一度繰り返すと，

「本来行いたい学習を別の学習方式によって実現する」

ということになる．この中で，実現手段としての学習方式は，上の例に示したように，例えば，記憶学習，早期停止法，正則化学習というように，明確に述べられている．しかし，本来行いたい学習に関しては，単に「汎化誤差を最小にする学習」というだけで，その実態が明確に述べられていないのである．学習対象となる関数を f で表し，本来行いたい学習法によって得られた関数を f_0 で表す．汎化誤差は，例えば

$$\int |f_0(x) - f(x)|^2 dx \tag{5.1}$$

で評価できる．しかし,「式 (5.1) を最小にする f_0 を求めよ」というだけでは，まだ問題が作られていないのである．学習対象となる関数 f は 1 つしかない．しかし，だからといって，それ以外の関数は考えないということはできない．もし本当に 1 つの関数 f しか考えないのであれば，式 (5.1) の値を求めることはできないし，式 (5.1) を最小にする f_0 を求めることもできない．また，例えば上述の早期停止法では，検査用データに対する誤差を汎化誤差の近似として使っていたが，それが式 (5.1) の近似になっているということを論理的に主張することはできないのである．早期停止法では，あらかじめわかっている f に対して早期停止法を適用してみて良い結果が得られるから，おそらく現実の場面でも良い結果が得られるであろうという論理の展開がなされている．しかし，現実の場面では f はわかっていないのであり，本物の f 以外は考えていないのであるから，既知の関数に対して良い結果が得られたという事実が，今問題になっている f に対しては何の役にも立たないのである．本当に式 (5.1) を最小にする f_0 を求めようとするのであれば，f が属している空間とか，f の出現確率等のような f に関する条件も併せて指定しない限り，この問題は解けない．しかも，これだけではまだ不十分であり，式 (5.1) を最小にする f_0 を求める為には，f_0 を探す範囲，即ち，解の探索範囲も指定しなければいけない．これらの条件を何も指定しないで，ただ「式 (5.1) を最小にする f_0 を求めよ」といっただけでは，問題がまだでき上がっていないのである．これらの条件をすべて指定し，問題を論理的に議論できる形に記述して始めて，問題ができたと主張できるのである．ここで，問題を論理的に議論できるとは，例えば解の存在を論じたり，解そのものを求めたり，別の手段で求めた解が本来の学習にとってどれだけ良い近似になっているかを論じることができるということである．このようにきちんと問題を作るということは，実は図 1.1 の第 1 段階の「評価基準の決定」に対応する部分であり，従来の研究で最もおろそかにされていた部分である．小川は文献 [48, 53] でこの問題をきちんと定式化し，その結果を用いて許容性の概念を導入することにより，実現手段としての学習方式の問題を厳密に定式化したのである [50, 51, 70, 71].

本来行いたい学習方式を別の学習方式によって実現できるとき，本来の学習方式は実現手段として用いられた別の学習方式を許容するという．これまでに，教師信号に雑音が含まれていない場合に射影学習，部分射影学習，平均射影学習，およびウィーナー学習が第 2 種記憶学習を許容するための必要十分条件が求められ [51, 70, 50]，また，教師信号に雑音が含まれている場合にウィーナー学習が第 2 種記憶学習を許容するための必要十分条件が求められている [70]．その結果，許容性が成立するかどうかが，教材の選び方に密接に関係していることがわかった．そして，これらの条件を基にして，許容性を成立させる教材の選び方が与えられている [71].

ところで，教師信号に雑音が含まれていない場合の射影学習の第 2 種記憶学習に対する許容性では，逆許容性と完全許容性の 2 種類の許容性しか現れないが，部分射影学習，平均射影学習の第 2 種記憶学習に対する許容性では，非許容性以外の 4 種類の許容性が現れる．しかし，このような違いが生じる理由はまだわかっていない．また，これらの条件は，教師信号に雑音が含まれている場合に対しては求められていない．更に，第 1 種記憶学習を実現手段として用いる場合の許容性は論じられていない．

そこで本章では，SL 射影学習が第 1 種記憶学習および第 2 種記憶学習をそれぞれ許容するための必要十分条件を求め，なぜある許容性がある学習方式に対して成立し，別の学習方式に対して成立しないかという背後にある法則性を解明すると共に，雑音および前提知識が許容性に与える影響等を解明していく．

5.2 学習の許容性

許容性の概念を説明する．J 学習の実現手段として J' 学習を用いる場合を考える．ある J' 学習作用素 $A^{(J')}$ で J 学習を実現できるかどうかは，その $A^{(J')}$ が評価基準 J を満足するかどうかで決まる．このことを評価するために，次の概念が導入された．

定義 5.1 （学習作用素の許容性 [51]） $A^{(J')}$ が評価基準 J を満たすとき，J 学習は $A^{(J')}$ を許容するという．

学習作用素は，例えば式（2.19）に示すように，一般には一意に定まらない．従って，J' 学習で J 学習を実現できるかどうかは，すべての $A^{(J')}$ と評価基準 J との関係によって決まる．この考えに基づき，学習の許容性は次のように定義されている（図 5.1）．

定義 5.2 （許容性 [51]）

(a) （非許容性）J 学習がすべての $A^{(J')}$ を許容しないとき，即ち，

$$A\{J\} \cap A\{J'\} = \emptyset \tag{5.2}$$

が成立するとき，J 学習は J' 学習を許容しない（do not admit），あるいは許容性は成立しないという．

(b) （部分許容性）J 学習が少なくとも一つの $A^{(J')}$ を許容するとき，即ち，

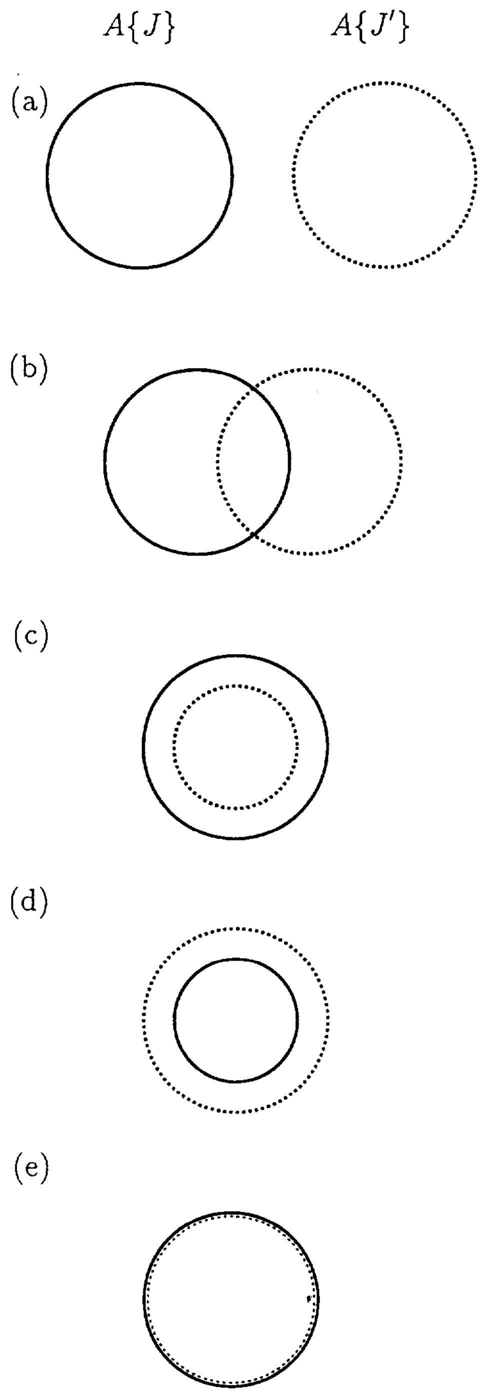
$$A\{J\} \cap A\{J'\} \neq \emptyset \tag{5.3}$$

が成立するとき，J 学習は J' 学習を部分的に許容する（partially admit），あるいは部分許容性（partial admissibility）が成立するという．

(c) （順許容性）J 学習がすべての $A^{(J')}$ を許容するとき，即ち

$$A\{J\} \supset A\{J'\} \tag{5.4}$$

が成立するとき，J 学習は J' 学習を常に許容する（always admit），あるいは順許容性（forward admissibility）が成立するという．



1. 非許容性

$$A\{J\} \cap A\{J'\} = \emptyset \qquad \text{(a)}$$

J 学習は J' 学習を許容しない.

2. 部分許容性

$$A\{J\} \cap A\{J'\} \neq \emptyset \qquad \text{(b} \sim \text{e)}$$

J 学習は J' 学習を部分的に許容する.

3. 順許容性

$$A\{J\} \supset A\{J'\} \qquad \text{(c, e)}$$

J 学習は J' 学習を常に許容する.

4. 逆許容性

$$A\{J\} \subset A\{J'\} \qquad \text{(d, e)}$$

J 学習は J' 学習に許容される.

5. 完全許容性

$$A\{J\} = A\{J'\} \qquad \text{(e)}$$

J 学習は J' 学習を完全に許容する.

図 5.1　学習の許容性

表 5.1　第 2 種記憶学習に対する許容性の研究状況

本来の学習	$n=0$	$n \neq 0$
ウィーナー学習	小川, 山崎 [51]	山崎, 小川 [70]
射影学習	小川, 山崎 [51]	
部分射影学習	小川, 山崎 [50]	
平均射影学習	小川, 山崎 [51]	

表 5.2　教師信号に雑音が含まれていない場合に各種射影学習が第 2 種記憶学習を許容するための必要十分条件

	射影学習	部分射影学習	平均射影学習 ウィーナー学習
非許容性	成立しない	成立しない	成立しない
部分許容性	常に成り立つ	常に成り立つ	常に成り立つ
順許容性	$\mathcal{N}(A)=\{0\}$ or $\mathcal{N}(A)=H$	$\mathcal{N}(A)=\{0\}$ or $\mathcal{N}(A)\supset S$	$\mathcal{N}(A)=\{0\}$ or $\mathcal{N}(A)\supset\overline{\mathcal{R}(R)}$
逆許容性	常に成り立つ	$\mathcal{N}(A)+S=H$	$\mathcal{N}(A)+\overline{\mathcal{R}(R)}=H$
完全許容性	$\mathcal{N}(A)=\{0\}$ or $\mathcal{N}(A)=H$	$\begin{cases} S=H \\ \mathcal{N}(A)=\{0\} \end{cases}$ or $\mathcal{N}(A)=H$	$\begin{cases} \overline{\mathcal{R}(R)}=H \\ \mathcal{N}(A)=\{0\} \end{cases}$ or $\mathcal{N}(A)=H$

(d) （逆許容性）J' 学習がすべての $A^{(J)}$ を許容するとき，即ち

$$A\{J\} \subset A\{J'\} \tag{5.5}$$

が成立するとき，J 学習は J' 学習に許容される，あるいは逆許容性（inverse admissibility）が成立するという．

(e) （完全許容性）J 学習が J' 学習を許容し，逆も成り立つとき，即ち，

$$A\{J\} = A\{J'\} \tag{5.6}$$

が成立するとき，J 学習は J' 学習を完全に許容する（completely admit），あるいは完全許容性（complete admissibility）が成立するという．

部分許容性が成立する場合には，J' 学習作用素を適切に選ぶことにより，J 学習と同じ結果を得ることができる．順許容性が成立する場合には，どの J' 学習作用素を使っても J 学習と同じ結果を得ることができる．即ち，J' 学習を J 学習の実現手段として常に使用することができる．逆許容性が成立する場合には，J' 学習作用素を適切に選ぶことにより，すべての J 学習作用素を実現することができる．完全許容性が成立する場合に

は，どの J' 学習作用素を使っても J 学習と同じ結果を得ることができるだけでなく，すべての J 学習作用素を J' 学習作用素で実現することができる．

これまでに，表 5.1 に示すように，教師信号に雑音が含まれていない場合にウィーナー学習，射影学習，部分射影学習，および平均射影学習が第 2 種記憶学習を許容するための必要十分条件が求められ [51, 70, 50]，また，教師信号に雑音が含まれている場合にウィーナー学習が第 2 種記憶学習を許容するための必要十分条件が求められている [70]．その結果，許容性が成立するかどうかが，教材の選び方に密接に関係していることがわかった．そして，これらの条件を基にして，許容性を成立させる教材の選び方が与えられている [71]．

ところで，表 5.2 に示すように，教師信号に雑音が含まれていない場合の射影学習の第 2 種記憶学習に対する許容性では，逆許容性と完全許容性の 2 種類の許容性しか現れないが，部分射影学習，平均射影学習の第 2 種記憶学習に対する許容性では，非許容性以外の 4 種類の許容性が現れる．しかし，このような違いが生じる理由はまだわかっていない．また，これらの条件は，教師信号に雑音が含まれている場合に対しては求められていない．更に，第 1 種記憶学習を実現手段として用いる場合の許容性は論じられていない．そこで本章では，SL 射影学習が第 1 種記憶学習および第 2 種記憶学習をそれぞれ許容するための必要十分条件を求め，なぜある許容性がある学習方式に対して成立し，別の学習方式に対して成立しないかという背後にある法則性を解明すると共に，雑音および前提知識が許容性に与える影響等を解明していく．

5.3 SL 射影学習の第 1 種記憶学習による実現

本節では，SL 射影学習の実現手段として第 1 種記憶学習を用いる場合を論じる．まず，部分許容性からはじめる．

定理 5.1 （部分許容性）SL 射影学習が第 1 種記憶学習を部分的に許容するための必要十分条件は，

$$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.7}$$

が成立することである．

条件（5.7）の意味を説明する．式（3.42）より，

$$\mathcal{R}(A_s) \subset \mathcal{R}(A) \tag{5.8}$$

となるので，式（5.7）が成立しているとき，

$$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A_s)^{\perp} \tag{5.9}$$

となる．即ち，$\mathcal{R}(A_s)^{\perp}$ は Q の不変部分空間になっている．しかし，それだけでは部分許容性が成立するためには十分でなく，Q は $\mathcal{R}(A_s)^{\perp}$ を更に小さな空間 $\mathcal{R}(A)^{\perp}$ に変換しなければいけないのである．

系 3.3 で示したように，SL 射影学習は $Q\mathcal{R}(A_s)^{\perp}$ の元を 0 に変換することにより雑音抑制を行なっている．一方，式（2.13）からわかるように，第 1 種記憶学習は $\mathcal{R}(A)^{\perp}$ の元を 0 に変換することにより雑音抑制を行なっている．従って，式（5.7）が成立していれば，第 1 種記憶学習によって SL 射影学習の雑音抑制を完全に達成できるのである．

このことと，学習作用素を適切に選ぶことにより第 1 種記憶学習で SL 射影学習を実現できることが同値になっているのである．

ところで，式 (5.7) は，それぞれの学習方式が何を 0 に変換することによって雑音抑制を行なっているかを表している．これを，抑制しきれなかった部分に着目して表現したものが次の系である．

系 5.1 式 (5.7) が成立するための必要十分条件は，任意の $\boldsymbol{n} \in \mathcal{R}(Q)$ に対して，

$$P_t \boldsymbol{n} = P_{\mathcal{R}(A)} \boldsymbol{n} \tag{5.10}$$

が成立することである．

式 (3.61)，(2.14) からわかるように，式 (5.10) の左辺と右辺がそれぞれ，SL 射影学習作用素と第 1 種記憶学習作用素によって抑制しきれなかった雑音成分である．それらが一致しているとき，またそのときに限って部分許容性が成立するのである．

このとき，$A^{(\mathrm{SL})}$, f_{SL}, および Af_{SL} がどのようになっているかを示す．

系 5.2 部分許容性が成立しているとき，$f \in S$ に対して次の関係が成立する．

$$A^{(\mathrm{SL})} = PA_s^\dagger + Y(I_M - P_{S_t}) \tag{5.11}$$

$$f_{\mathrm{SL}} = Pf + PA_s^\dagger \boldsymbol{n} \tag{5.12}$$

$$Af_{\mathrm{SL}} = Af + P_{\mathcal{R}(A)} \boldsymbol{n} = Af_{\mathrm{M}} \tag{5.13}$$

即ち，部分許容性が成立しているとき，式 (5.11) の一部を式 (2.12) の $A^{(\mathrm{M1})}$ を使って，また，式 (5.12) の f_{SL} を式 (2.13) を使って実現できるのである．そして，式 (5.13) に示すように，一般には異なる Af_{SL} と Af_{M} が，S の元 f に対しては一致するようになるのである．

ところで，たとえ S が式 (5.7) を満たしていても，学習作用素を適切に選ばない限り，第 1 種記憶学習によって SL 射影学習を実現することはできない．即ち，式 (5.11) の SL 射影学習作用素を実現できない第 1 種記憶学習作用素が存在するのであり，この記憶学習で SL 射影学習を常に実現できるわけではない．それが可能になるのは，次の順許容性が成立する場合である．

定理 5.2 （順許容性）SL 射影学習が第 1 種記憶学習を常に許容するための必要十分条件は，

$$\mathcal{N}(A) = \{0\} \tag{5.14}$$

$$Q\mathcal{R}(A_s)^\perp \subset \mathcal{R}(A)^\perp \tag{5.15}$$

が同時に成立するか，または，

$$\mathcal{N}(A) \supset S \tag{5.16}$$

$$Q = 0 \tag{5.17}$$

が同時に成立することである．

式 (5.16), (5.17) が同時に成立するということは，S の元 f に対するすべての教師信号 $\boldsymbol{y}$ が 0 になることを意味する．従って，学習の立場からはあまり意味がない．よって，順許容性が成立するためには式 (5.14), (5.15) が同時に成立しなければいけないのである．

このとき，式 (3.24) より

$$L = S \tag{5.18}$$

となる．従って，式 (5.14) が成立している場合は，一つの S に対する SL 射影学習は $L = S$ の一種類しか存在しないのである．$L = S$ の場合の SL 射影学習を S 射影学習とよぶことにする．

式 (5.14) は

$$\mathcal{R}(A^*) = H \tag{5.19}$$

と同値である．この条件が成立するためには，訓練データの数 M は H の次元以上でなければいけない．従って H が無限次元の場合，有限個の教材 $\{\boldsymbol{x}_m\}_{m=1}^M$ をどのように選んでも式 (5.19)，即ち，式 (5.14) を満足させることはできない．一方，H が有限次元の場合には，教材 $\{\boldsymbol{x}_m\}_{m=1}^M$ をうまく選ぶことによって，順許容性を成立させることができるのである．そのような場合の例を，本節の最後に示す．

式 (5.14) は，

$$A^\dagger A = I \tag{5.20}$$

と同値であり，式 (2.13) より第 1 種記憶学習によって得られる関数が

$$f_{\mathrm{M}} = A^\dagger \boldsymbol{y} \tag{5.21}$$

と一意に定まることを意味する．しかし，この条件だけでは第 1 種記憶学習を SL 射影学習に適用できない．更に，部分許容性のための条件である式 (5.15) が成立しなければいけないのである．このとき，$A^{(\mathrm{SL})}$, f_{SL} および $A^{(\mathrm{M1})}$, f_{M} は次のようになる．

系 5.3 式 (5.14), (5.15) が同時に成立しているとき，SL 射影学習作用素および第 1 種記憶学習作用素は，

$$A^{(\mathrm{SL})} = A^\dagger P_{S_t} + Y(I_M - P_{S_t}) \tag{5.22}$$

$$A^{(\mathrm{M1})} = A^\dagger \tag{5.23}$$

と表され，SL 射影学習によって得られる関数 f_{SL} は，

$$f_{\mathrm{SL}} = f + A^\dagger \boldsymbol{n} = f_{\mathrm{M}} \tag{5.24}$$

となる．

系 5.3 より，第 1 種記憶学習でなぜ SL 射影学習を実現できるかがわかる．即ち式 (5.22) では，式 (3.47) の右辺の第 1 項 $PA_s^\dagger P_t$ が $A^\dagger P_{S_t}$ になっている．$I_M - P_{S_t}$ は $S_t^\perp$ への正射影行列であるから，この SL 射影学習は S_t および $S_t^\perp$ のベクトルをそれぞれ $A^\dagger$ および Y で変換している．教師信号 $\boldsymbol{y}$ は S_t のベクトルであり，$S_t^\perp$ に現れることはない．従って，$S_t^\perp$ のベクトルに対しては Y を自由に選ぶことができる．逆に S_t のベクトルをどのように変換するかが SL 射影学習を特徴付けるのであり，式 (5.22) の SL 射影学習は $A^\dagger$ で変換する．一方 $A^{(\mathrm{M1})}$ は，式 (5.14) が成立しているときには式 (5.23) で

与えられる．従って，第 1 種記憶学習は $\mathbf{C}^M$ のすべてのベクトルを $A^\dagger$ で変換することになり，S_t の元も $A^\dagger$ で変換する．即ち，S_t の元に対して，第 1 種記憶学習は式 (5.22) の SL 射影学習と同じ働きをするのである．

ところで，式 (5.22)，(5.23) からわかるように，式 (5.14)，(5.15) が同時に成立していても，式 (5.23) の $A^{(\mathrm{M1})}$ で実現できない $A^{(\mathrm{SL})}$ は無数に存在する．例えば，S_t が $\mathbf{C}^M$ の真部分空間になっている限り，式 (5.22) で $Y=0$ とおいた $A^{(\mathrm{SL})}=A^\dagger P_{S_t}$ を実現することはできない．このような第 1 種記憶学習作用素で実現できない SL 射影学習作用素の中に，例えば非常に簡単に計算できるといった良い性質をもつものが存在するかもしれない．こうした SL 射影学習作用素も第 1 種記憶学習作用素で実現できた方がよい．そこで次に，この記憶学習作用素ですべての SL 射影学習作用素を実現できるための条件を求める．

定理 5.3 （逆許容性）SL 射影学習が第 1 種記憶学習に許容されるための必要十分条件は，

$$\mathcal{N}(A)+S=H \tag{5.25}$$

$$Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp \tag{5.26}$$

が同時に成立するか，または，

$$\mathcal{N}(A)=H \tag{5.27}$$

が成立することである．

式 (5.27) は，H のすべての元が A で標本化すれば 0 になることを意味しており，学習の立場からはあまり意味がない．従って，逆許容性が成立するためには式 (5.25)，(5.26) が同時に成立しなければいけないのである．

定理 5.3 では逆許容性のための条件が式 (5.25) のように S を用いて表現されているが，次のように L を使って表現することもできる．

系 5.4 SL 射影学習が第 1 種記憶学習に許容されるための必要十分条件は，

$$\mathcal{N}(A)\dotplus L=H \tag{5.28}$$

$$Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp \tag{5.29}$$

が同時に成立するか，または，

$$\mathcal{N}(A)=H \tag{5.30}$$

が成立することである．

即ち，式 (5.25) と式 (5.28) が同値になっているのである．この式は，H が A で変換すれば 0 になる関数の空間 $\mathcal{N}(A)$ と，SL 射影学習が信号成分を完全に復元できる関数の空間 L の直和になっていることを意味している．

逆許容性が成立しているとき，学習によって得られる関数の標本点における値に関して，次の系が成立する．

系 5.5 式 (5.25)，(5.26) が同時に成立するとき，H の任意の f に対して式 (5.13) が成立する．

表 5.3 SL 射影学習の第 1 種記憶学習に対する各種許容性が成立するための必要十分条件

許容性	雑音がある場合	雑音がない場合
非許容性	$Q\mathcal{R}(A_s)^{\perp} \not\subset \mathcal{R}(A)^{\perp}$	成立しない
部分許容性	$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A)^{\perp}$	常に成立する
順許容性	$\begin{cases}\mathcal{N}(A) = \{0\} \\ Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A)^{\perp}\end{cases}$ または $\begin{cases}\mathcal{N}(A) \supset S \\ Q = 0\end{cases}$	$\mathcal{N}(A) = \{0\}$ または $\mathcal{N}(A) \supset S$
逆許容性	$\begin{cases}\mathcal{N}(A) + S = H \\ Q\mathcal{R}(A)^{\perp} = \mathcal{R}(A)^{\perp}\end{cases}$ または $\mathcal{N}(A) = H$	$\begin{cases}\mathcal{N}(A) + S = H \\ \mathcal{R}(A) = \mathbf{C}^M\end{cases}$ または $\mathcal{N}(A) = H$
完全許容性	$\begin{cases}S = H,\ \mathcal{N}(A) = \{0\} \\ Q\mathcal{R}(A)^{\perp} = \mathcal{R}(A)^{\perp}\end{cases}$ または $\begin{cases}\mathcal{N}(A) = H \\ Q = 0\end{cases}$	$\begin{cases}S = H,\ \mathcal{N}(A) = \{0\} \\ \mathcal{R}(A) = \mathbf{C}^M\end{cases}$ または $\mathcal{N}(A) = H$

即ち，部分許容性が成立しているときには，S の元 f に対してのみ Af_{SL} と Af_{M} が一致していたのに対して，逆許容性が成立しているときには，H のすべての元 f に対して一致するのである．

定理 5.2 と定理 5.3 より，次の定理を得る．

定理 5.4 （完全許容性）SL 射影学習が第 1 種記憶学習を完全に許容するための必要十分条件は，

$$S = H \tag{5.31}$$

$$\mathcal{N}(A) = \{0\} \tag{5.32}$$

$$Q\mathcal{R}(A)^{\perp} = \mathcal{R}(A)^{\perp} \tag{5.33}$$

が同時に成立するか，または，

$$\mathcal{N}(A) = H \tag{5.34}$$

$$Q = 0 \tag{5.35}$$

が同時に成立することである．

式 (5.34)，(5.35) が同時に成立するということは，H の元 f に対するすべての教師信号 $\boldsymbol{y}$ が 0 になることを意味する．従って，学習の立場からはあまり意味がない．完全許容性を成立させるためには式 (5.31)，(5.32)，(5.33) を同時に成立させなければいけないのである．このとき式 (3.24) より，$L = S = H$ となる．即ち，完全許容性が成立するような SL 射影学習は，$\mathcal{R}(A^*) = H$ という特別な射影学習に限られるのである．

SL 射影学習の第 1 種記憶学習に対する各種許容性の条件をまとめて表 5.3 に示す．この表には，教師信号に雑音が含まれていない場合も併せて示してある．これは，表 5.3 の‘雑音がある場合’の欄の各条件で $Q=0$ とおくことによって得られたものである．この表より，教師信号に雑音が含まれている場合にはどの SL 射影学習に対しても 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかる．

表 5.3 には，SL 射影学習の S は現れるが，L は現れない．しかしこのことは，L を任意に選べるということではない．例えば，順許容性に関していえば，$\mathcal{N}(A)=\{0\}$ が成立するときには $L=S$ に，また $\mathcal{N}(A)\supset S$ が成立するときには $L=\{0\}$ に，自動的に定まってしまうのである．また逆許容性では，式 (5.28) を満たすように L を決めなければいけないのである．

表 5.3 には，$\mathcal{N}(A)$, $\mathcal{R}(A)$, $\mathcal{R}(A_s)$ 等，A に関係した条件が現れている．A および A_s は式 (2.3), (2.4), (3.42) からわかるように，教材 $\{\boldsymbol{x}_m\}_{m=1}^{M}$ から決まってくるものである．このことは，教材の取り方によって許容性を制御できることを意味している．この問題については，次章で詳しく論じることにする．

表 5.3 より更に次のことがわかる．教師信号に雑音が含まれている場合の条件は，$\mathcal{N}(A)=\{0\}$ のように Q に関係しないものと，$Q\mathcal{R}(A_s)^{\perp}\subset\mathcal{R}(A)^{\perp}$ のように Q に関係するものに分かれている．そして，前者はそのまま雑音が存在しない場合の条件になっている．一方，後者の事情は異なり，更に 2 種類に分かれる．例えば，順許容性のための条件 $Q\mathcal{R}(A_s)^{\perp}\subset\mathcal{R}(A)^{\perp}$ は，$Q=0$ の場合に自動的に成立するので，表 5.3 の雑音がない場合の欄ではこの条件は消えている．しかし，逆許容性および完全許容性のための条件 $Q\mathcal{R}(A)^{\perp}=\mathcal{R}(A)^{\perp}$ は，$Q=0$ の場合に自動的に成立はしない．左辺が $Q\mathcal{R}(A)^{\perp}=\{0\}$ となるので，右辺も $\mathcal{R}(A)^{\perp}=\{0\}$ とならなければいけない．即ち，$\mathcal{R}(A)=\mathbf{C}^M$ となるのである．表 5.3 の雑音がない場合の欄にはこの条件が現れている．このように Q を含む条件には，(a) $Q=0$ の場合に自動的に成立する条件と，(b) $Q=0$ の場合に $\{\boldsymbol{x}_m\}_{m=1}^{M}$ に関する制約になる条件の 2 種類があるのである．

定理 5.1 ~ 定理 5.4 より，直ちに次の定理を得る．

定理 5.5 射影学習，部分射影学習，および平均射影学習の第 1 種記憶学習に対する各種許容性が成立するための必要十分条件は表 5.4 に示すとおりである．

この表では，

$$\mathcal{R}(AP_{\overline{\mathcal{R}(R)}})=\mathcal{R}(AR) \tag{5.36}$$

の関係を用いてある．表 5.4 から次のことがわかる．まず，どの射影学習についても一般の SL 射影学習の場合と同様に，教師信号に雑音が含まれている場合には 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には 4 種類の許容性が現れる．順許容性が成立するための条件を比較してみると，括弧‘{’によってまとめられた 2 組の条件の論理和になっているが，本質的なものは上に示された条件である．これらの条件には $\mathcal{N}(A)=\{0\}$ が共通して含まれており，第 1 種記憶学習でこれらの各種射影学習を実現するためには H は有限次元でなければいけないことがわかる．もう一方の条件を比較してみると，例えば射影学習と部分射影学習では，$\mathcal{R}(A)\supset\mathcal{R}(AP_S)$ であるので，

$$\mathcal{R}(A)^{\perp}\subset\mathcal{R}(AP_S)^{\perp} \tag{5.37}$$

となり，射影学習より部分射影学習の方が，Q はより大きな空間を $\mathcal{R}(A)^{\perp}$ に変換しなければいけないのである．従って，部分射影学習が第 1 種記憶学習を常に許容するための

表 5.4　射影学習，部分射影学習，および平均射影学習の第 1 種記憶学習に対する各種許容性が成立するための必要十分条件

（a）教師信号に雑音が含まれている場合

許容性	射影学習	部分射影学習	平均射影学習
S と L	$S=H,\ L=\mathcal{R}(A^*)$	$S\subset H,\ L=\mathcal{R}(P_SA^*)$	$S=\overline{\mathcal{R}(R)},\ L=\mathcal{R}(RA^*)$
非許容性	$Q\mathcal{R}(A)^\perp\not\subset\mathcal{R}(A)^\perp$	$Q\mathcal{R}(AP_S)^\perp\not\subset\mathcal{R}(A)^\perp$	$Q\mathcal{R}(AR)^\perp\not\subset\mathcal{R}(A)^\perp$
部分許容性	$Q\mathcal{R}(A)^\perp\subset\mathcal{R}(A)^\perp$	$Q\mathcal{R}(AP_S)^\perp\subset\mathcal{R}(A)^\perp$	$Q\mathcal{R}(AR)^\perp\subset\mathcal{R}(A)^\perp$
順許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp\subset\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(AP_S)^\perp\subset\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)\supset S\\ Q=0\end{cases}$	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(AR)^\perp\subset\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)\supset\overline{\mathcal{R}(R)}\\ Q=0\end{cases}$
逆許容性	$Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp$ または $\mathcal{N}(A)=H$	$\begin{cases}\mathcal{N}(A)+S=H\\ Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp\end{cases}$ または $\mathcal{N}(A)=H$	$\begin{cases}\mathcal{N}(A)+\overline{\mathcal{R}(R)}=H\\ Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp\end{cases}$ または $\mathcal{N}(A)=H$
完全許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}S=H,\ \ \mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}\overline{\mathcal{R}(R)}=H,\ \ \mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp=\mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$

（b）教師信号に雑音が含まれていない場合

許容性	射影学習	部分射影学習	平均射影学習
非許容性	成立しない	成立しない	成立しない
部分許容性	常に成立する	常に成立する	常に成立する
順許容性	$\mathcal{N}(A)=\{0\}$ または $\mathcal{N}(A)=H$	$\mathcal{N}(A)=\{0\}$ または $\mathcal{N}(A)\supset S$	$\mathcal{N}(A)=\{0\}$ または $\mathcal{N}(A)\supset\overline{\mathcal{R}(R)}$
逆許容性	$\mathcal{R}(A)=\mathbf{C}^M$ または $\mathcal{N}(A)=H$	$\begin{cases}\mathcal{N}(A)+S=H\\ \mathcal{R}(A)=\mathbf{C}^M\end{cases}$ または $\mathcal{N}(A)=H$	$\begin{cases}\mathcal{N}(A)+\overline{\mathcal{R}(R)}=H\\ \mathcal{R}(A)=\mathbf{C}^M\end{cases}$ または $\mathcal{N}(A)=H$
完全許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ \mathcal{R}(A)=\mathbf{C}^M\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}S=H,\ \ \mathcal{N}(A)=\{0\}\\ \mathcal{R}(A)=\mathbf{C}^M\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}\overline{\mathcal{R}(R)}=H,\ \ \mathcal{N}(A)=\{0\}\\ \mathcal{R}(A)=\mathbf{C}^M\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$

条件は，射影学習が第 1 種記憶学習を常に許容するための条件より厳しいものになっている．同様に，平均射影学習が第 1 種記憶学習を常に許容するための条件は，射影学習が第 1 種記憶学習を常に許容するための条件より厳しいものになっている．部分射影学習や平均射影学習では，Q の他に S および R という前提知識を用いている．これに対して射影学習で用いる前提知識は，Q の他には H だけである．従って，部分射影学習や平均射影学習を実現するための条件の方が，射影学習を実現するための条件より厳しくなったのである．

表 5.4 の条件をより深く理解するために，例えば射影学習の第 1 種記憶学習に対する順許容性の条件を説明する．この中で本質的な条件は，

$$\mathcal{N}(A) = \{0\} \tag{5.38}$$

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.39}$$

が同時に成立することである．式 (5.38) は，H が無限次元である場合には有限個の訓練データをどのように選んでも成立させることはできないが，H が有限次元である場合には適切な訓練データを選ぶことにより成立させることができるのである．式 (5.39) は $\mathcal{R}(A)^{\perp}$ が Q の不変部分空間であることを意味している．例えば $Q = \sigma^2 I_M$ であるとき，σ そのものの値はわからなくても式 (5.39) が成立することはわかるのである．

以下では計算機実験により，標本点のとり方によって順許容性が成立する場合としない場合があることを示す．x をスカラーとして，$\{\varphi_n\}_{n=1}^{9}$ を区間 $[-\pi, \pi]$ で定義された関数

$$\varphi_n(x) = \begin{cases} 1 & : n = 1 \\ \sqrt{2}\cos\dfrac{n}{2}x & : n = 2, 4, 6, 8 \\ \sqrt{2}\sin\dfrac{n-1}{2}x & : n = 3, 5, 7, 9 \end{cases} \tag{5.40}$$

とする．H を，$\{\varphi_n\}_{n=1}^{9}$ の線形結合の全体に対して，内積が

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx \tag{5.41}$$

で定義された 9 次元関数空間を考える．この空間の再生核は，

$$K(x, x') = \begin{cases} \dfrac{\sin\frac{9(x-x')}{2}}{\sin\frac{(x-x')}{2}} & : x \neq x' \\ 9 & : x = x' \end{cases} \tag{5.42}$$

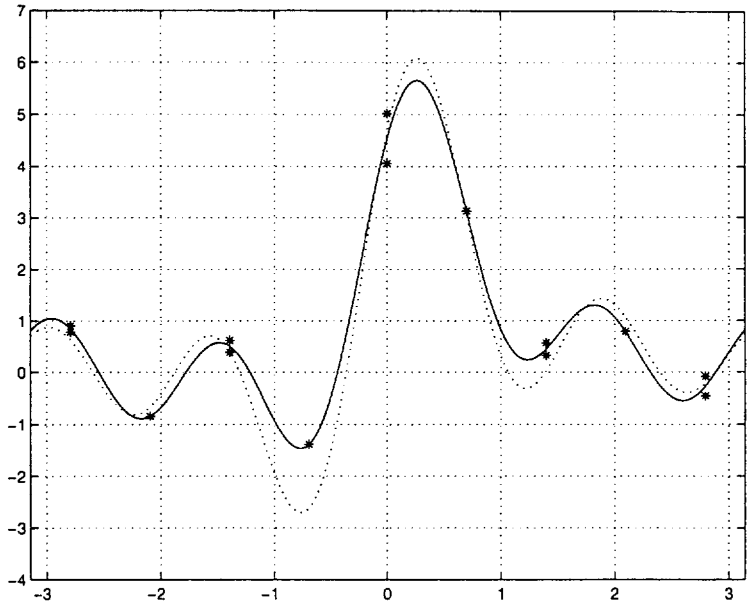
となる．学習したい関数 $f(x)$ が

$$f(x) = \sum_{n=1}^{9} \varphi_n(x) \tag{5.43}$$

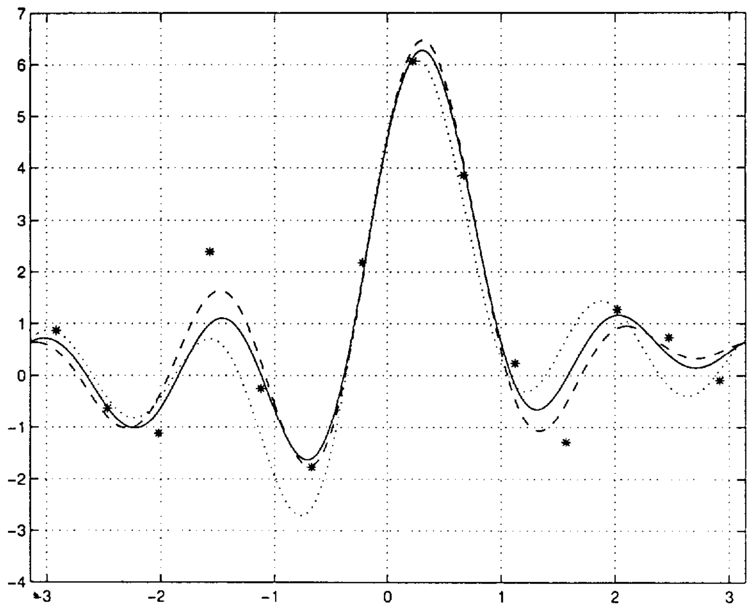
である場合を考える．この関数を，14 個の標本点からなる 2 組の訓練データを使って学習する．

まず，標本点

$$x_m = \begin{cases} \dfrac{2m-10}{9}\pi & : m = 1, 2, \cdots, 9 \\ x_{2m-19} & : m = 10, 11, \cdots, 14 \end{cases} \tag{5.44}$$



(a)



(b)

図 5.2　射影学習と第 1 種記憶学習によって得られた関数. (a) 射影学習が第 1 種記憶学習を許容する場合. (b) 射影学習が第 1 種記憶学習を許容しない場合.

からなる訓練データを考える．これらの標本点は，x_1 から x_9 までを $[-\pi,\pi]$ で等間隔に，残りの 5 個を x_1,x_3,x_5,x_7,x_9 ととったものである．なお，x_1 と x_9 の‘外側の間隔’，即ち $\{x_1-(-\pi)\}+(\pi-x_9)$ も，それらの内側にある隣合う標本点の間隔と同じになっている．雑音の相関行列は，互いに異なる正の実数 $\sigma_m(m=1,2,\cdots,9)$ に対して，

$$Q=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_9^2,\sigma_1^2,\sigma_3^2,\cdots,\sigma_9^2) \tag{5.45}$$

である場合を考える．ここで，diag(·) は括弧 () の中に示される値を対角成分にもつ対角行列である．実際に用いた $\sigma_m(m=1,2,\cdots,9)$ の値はそれぞれ，

$$\begin{aligned}\sigma_1=0.20,\quad \sigma_2=0.24,\quad \sigma_3=0.52,\quad \sigma_4=1.36,\quad \sigma_5=0.60,\\ \sigma_6=0.44,\quad \sigma_7=0.40,\quad \sigma_8=0.22,\quad \sigma_9=0.26\end{aligned} \tag{5.46}$$

である．このとき，式 (5.38)，(5.39) が共に成立するので，順許容性が成立する．これらの標本点からなる訓練データを使って学習した関数を図 5.2(a) に示す．f_{P}, f_{M}, f をそれぞれ，実線，破線，点線で，訓練データを‘*’で示してある．この図には破線が現れていないが，これは $f_{\mathrm{P}}=f_{\mathrm{M}}$ となっているからである．即ち，順許容性が成立していることがわかる．

次に，標本点

$$x_m=\frac{2m-15}{14}\pi \quad (m=1,2,\cdots,14) \tag{5.47}$$

からなる訓練データを考える．これらの標本点は，$[-\pi,\pi]$ で等間隔になるようにとったものである．式 (5.44) の x_1,x_9 と同様に，x_1 と x_{14} の‘外側の間隔’も内側にある隣合う標本点の間隔と同じになっている．雑音の相関行列は，式 (5.46) の $\sigma_m(m=1,2,\cdots,9)$ と，5 個の正の実数 $\sigma_m(m=10,11,\cdots,14)$ を用いて，

$$Q=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_{14}^2) \tag{5.48}$$

である場合を考える．ここで，$\sigma_1,\sigma_2,\cdots,\sigma_{14}$ はすべて互いに異なるものであり，実際に用いた $\sigma_m(m=10,11,\cdots,14)$ の値はそれぞれ，

$$\sigma_{10}=0.34,\quad \sigma_{11}=1.14,\quad \sigma_{12}=0.54,\quad \sigma_{13}=0.38,\quad \sigma_{14}=0.30 \tag{5.49}$$

である．このとき，式 (5.38) は成立するが，式 (5.39) は成立しないので，順許容性は成立しない．これらの標本点からなる訓練データを使って学習した関数を図 5.2(b) に示す．$f_{\mathrm{P}}, f_{\mathrm{M}}, f$ をそれぞれ，実線，破線，点線で，訓練データを‘*’で示してある．この図では実線と破線は異なっており，$f_{\mathrm{P}}\neq f_{\mathrm{M}}$ となっていることがわかる．即ち，順許容性は成立していない．このように，順許容性が成立するかどうかは，標本点によって変わるのである．

5.4 SL 射影学習の第 2 種記憶学習による実現

本節では，SL 射影学習の実現手段として第 2 種記憶学習を用いる場合を論じる．第 1 種記憶学習では $\boldsymbol{y}$ の動き得る範囲が $\mathbf{C}^M$ 全体であったのに対して，本節で扱う第 2 種記憶学習では，$\mathbf{C}^M$ の部分空間 S_b にとどまっている．式 (2.16) に示したように $S_b=\mathcal{R}(A)+\mathcal{R}(Q)$ であり，教師信号 $\boldsymbol{y}$ はすべて S_b に属している．即ち，第 2 種記憶学習は，$\boldsymbol{y}$ が S_b に属しているという前提知識を用いた記憶学習である．

以下では，まず各種許容性が成立するための必要十分条件を求め，前節で求めた条件と比較することにより，この前提知識が許容性に与える影響を解明する．まず，部分許容性と順許容性はそれぞれ，次のようになる．

定理 5.6 （部分許容性）SL 射影学習が第 2 種記憶学習を部分的に許容するための必要十分条件は，

$$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.50}$$

が成立することである．

定理 5.7 （順許容性）SL 射影学習が第 2 種記憶学習を常に許容するための必要十分条件は

$$\mathcal{N}(A) = \{0\} \tag{5.51}$$

$$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.52}$$

が同時に成立するか，または，

$$\mathcal{N}(A) \supset S \tag{5.53}$$

$$Q = 0 \tag{5.54}$$

が同時に成立することである．

このように，これら 2 種類の許容性が成立するための必要十分条件は，第 1 種記憶学習を用いた場合の条件と全く同じものになる．この理由は次のとおりである．式 (3.43)，(2.16)，(5.8) より，S_t, S_b, および $\mathbf{C}^M$ には

$$S_t \subset S_b \subset \mathbf{C}^M \tag{5.55}$$

という関係がある．定理 5.1 と定理 5.2 で重要なことは，式 (3.43) で定義される S_t が，第 1 種記憶学習で $\boldsymbol{y}$ が動き得る範囲 $\mathbf{C}^M$ の部分空間であるということである：

$$S_t \subset \mathbf{C}^M \tag{5.56}$$

この性質は，式 (5.55) より，第 2 種記憶学習でも保存されている．即ち S_t は，第 2 種記憶学習で $\boldsymbol{y}$ が動き得る範囲 S_b の部分空間にもなっている：

$$S_t \subset S_b \tag{5.57}$$

従って，部分許容性，順許容性が成立するための必要十分条件は，第 1 種記憶学習を用いた場合の条件と全く同じものになるのである．

系 5.3 に対応して，次の系が成立する．

系 5.6 式 (5.51)，(5.52) が同時に成立しているとき，SL 射影学習作用素は式 (5.22) で表され，SL 射影学習によって得られる関数 f_{SL} は式 (5.24) で表される．第 2 種記憶学習作用素は

$$A^{(\mathrm{M2})} = A^{\dagger} + Y(I_M - P_{S_t}) \tag{5.58}$$

と表される．ここで，Y は $\mathbf{C}^M$ から H への任意の作用素である．

証明は略す．式 (5.51)，(5.52) が同時に成立しているとき，式 (5.23) に示したように第 1 種記憶学習作用素は一意に定まった．一方，第 2 種記憶学習作用素には，式 (5.58) に Y が含まれるので，無限に多くのものが存在する．従って，第 2 種記憶学習作用素は，SL 射影学習作用素の全体 $A\{J_{SL}\}$ のより広い部分を実現できているのである．

次に逆許容性が成立するための条件，即ち，すべての SL 射影学習作用素を第 2 種記憶学習作用素を用いて実現できるための条件を与える．

定理 5.8 (逆許容性) SL 射影学習が第 2 種記憶学習に許容されるための必要十分条件は，

$$\mathcal{N}(A) + S = H \tag{5.59}$$

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.60}$$

が同時に成立することである．

このように，第 2 種記憶学習を用いた場合には，第 1 種記憶学習を用いた場合の条件 (5.26) が式 (5.60) にゆるめられ，条件 (5.27) がなくなるのである．図 2.1 に示したように，第 2 種記憶学習作用素の全体 $A\{J_{M2}\}$ は，第 1 種記憶学習作用素の全体 $A\{J_{M1}\}$ よりも広い．このために，すべての SL 射影学習作用素を実現できるための条件はゆるくなったのである．

系 5.5 に対応して，次の系が成立する．

系 5.7 式 (5.59)，(5.60) が同時に成立するとき，H の任意の f に対して式 (5.13) が成立する．

証明は略す．このように，第 2 種記憶学習を用いた場合には，第 1 種記憶学習を用いた場合よりゆるい条件で式 (5.13) が成立するのである．

定理 5.7 と 定理 5.8 より，次の定理を得る．

定理 5.9 (完全許容性) SL 射影学習が第 2 種記憶学習を完全に許容するための必要十分条件は，

$$S = H \tag{5.61}$$

$$\mathcal{N}(A) = \{0\} \tag{5.62}$$

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.63}$$

が同時に成立するか，または，

$$\mathcal{N}(A) = H \tag{5.64}$$

$$Q = 0 \tag{5.65}$$

が同時に成立することである．

完全許容性のための条件も，第 1 種記憶学習を用いた場合と比較して，条件 (5.33) が式 (5.63) にゆるめられている．逆許容性の条件がゆるくなったので，完全許容性の条件もゆるくなったのである．

SL 射影学習の第 2 種記憶学習に対する各種許容性の条件をまとめて表 5.5 に示す．この表には，教師信号に雑音が含まれていない場合も併せて示してある．これは，表 5.5 の

表 5.5　SL 射影学習の第 2 種記憶学習に対する各種許容性が成立するための必要十分条件

許容性	雑音がある場合	雑音がない場合
非許容性	$Q\mathcal{R}(A_s)^\perp \not\subset \mathcal{R}(A)^\perp$	成立しない
部分許容性	$Q\mathcal{R}(A_s)^\perp \subset \mathcal{R}(A)^\perp$	常に成立する
順許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A_s)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)\supset S\\ Q=0\end{cases}$	$\mathcal{N}(A)=\{0\}$ または $\mathcal{N}(A)\supset S$
逆許容性（定理 5.8）	$\begin{cases}\mathcal{N}(A)+S=H\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$	$\mathcal{N}(A)+S=H$
完全許容性	$\begin{cases}S=H,\ \mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}S=H\\ \mathcal{N}(A)=\{0\}\end{cases}$ または $\mathcal{N}(A)=H$

表 5.6　射影学習，部分射影学習，および平均射影学習の第 2 種記憶学習に対する各種許容性が成立するための必要十分条件

許容性	射影学習	部分射影学習	平均射影学習
S と L	$S=H,\ L=\mathcal{R}(A^*)$	$S\subset H,\ L=\mathcal{R}(P_SA^*)$	$S=\overline{\mathcal{R}(R)},\ L=\mathcal{R}(RA^*)$
非許容性	$Q\mathcal{R}(A)^\perp \not\subset \mathcal{R}(A)^\perp$	$Q\mathcal{R}(AP_S)^\perp \not\subset \mathcal{R}(A)^\perp$	$Q\mathcal{R}(AR)^\perp \not\subset \mathcal{R}(A)^\perp$
部分許容性	$Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp$	$Q\mathcal{R}(AP_S)^\perp \subset \mathcal{R}(A)^\perp$	$Q\mathcal{R}(AR)^\perp \subset \mathcal{R}(A)^\perp$
順許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(AP_S)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)\supset S\\ Q=0\end{cases}$	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(AR)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)\supset \overline{\mathcal{R}(R)}\\ Q=0\end{cases}$
逆許容性	$Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp$	$\begin{cases}\mathcal{N}(A)+S=H\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$	$\begin{cases}\mathcal{N}(A)+\overline{\mathcal{R}(R)}=H\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$
完全許容性	$\begin{cases}\mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}S=H,\ \mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$	$\begin{cases}\overline{\mathcal{R}(R)}=H\\ \mathcal{N}(A)=\{0\}\\ Q\mathcal{R}(A)^\perp \subset \mathcal{R}(A)^\perp\end{cases}$ または $\begin{cases}\mathcal{N}(A)=H\\ Q=0\end{cases}$

‘雑音がある場合’の欄の各条件で $Q=0$ とおくことによって得られたものである．この表より，教師信号に雑音が含まれている場合にはどのSL射影学習に対しても 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかる．

表 5.5 には，SL 射影学習の S は現れるが，L は現れない．しかしこのことは，表 5.3 と同様に，L を任意に選べるということではない．例えば，順許容性に関していえば，$\mathcal{N}(A)=\{0\}$ が成立するときには $L=S$ に，また $\mathcal{N}(A)\supset S$ が成立するときには $L=\{0\}$ に，自動的に定まってしまうのである．また逆許容性では，$\mathcal{N}(A)\dot{+}L=H$ を満たすように L を決めなければいけないのである．

表 5.5 には，表 5.3 と同様に，$\mathcal{N}(A)$, $\mathcal{R}(A)$, $\mathcal{R}(A_s)$ 等，A に関係した条件が現れている．従って，この表からも許容性を教材の取り方によって制御できることがわかる．

表 5.5 より更に次のことがわかる．教師信号に雑音が含まれている場合の条件は，Q に関係しないものとするものに分かれている．そして前者は，第 1 種記憶学習の場合と同様に，そのまま雑音が存在しない場合の条件になっている．第 1 種記憶学習の場合と異なることは，後者がすべて $Q=0$ の場合に自動的に成立するものであるということである．即ち，第 1 種記憶学習に関して述べた Q に関する条件の分類の (a) に属すものだけになっているのである．

表 5.5 と表 5.3 の条件を各許容性ごとに比較すれば，非許容性，部分許容性，順許容性の条件は一致しているが，逆許容性と完全許容性の条件は第 2 種記憶学習を用いる場合の方がゆるくなっていることがわかる．逆許容性，完全許容性はすべての SL 射影学習作用素を実現しようとするものであり，図 2.1 に示したように第 2 種記憶学習作用素の集合の方が第 1 種記憶学習作用素の集合より大きいので，これらの許容性の条件がそれだけゆるくなったのである．従って，SL 射影学習と同じ結果を得ようとするだけであればどちらの記憶学習を用いても同じであるが，すべての SL 射影学習作用素を実現しようとするときには第 2 種記憶学習を用いた方がよいということになるのである．

定理 5.6 ～ 定理 5.9 より，直ちに次の定理を得る．

定理 5.10 射影学習，部分射影学習，および平均射影学習の第 2 種記憶学習に対する各種許容性が成立するための必要十分条件は表 5.6 に示すとおりである．

表 5.4 と比較すれば次のことがわかる．まず，どの射影学習についても，逆許容性のための条件と完全許容性のための条件がゆるくなっている．また，この表の条件で $Q=0$ とおけば表 5.2 に示した条件になり，定理 5.10 の結果がこれまでに個別に得られていた結果と符合していることがわかる．表 5.2 では，射影学習と部分射影学習，平均射影学習で許容性の現れ方が異なっていたが，その原因が，SL 射影学習という一般の場合から個別の各種射影学習を見ることにより明らかになる．即ち，射影学習では $S=H$ であるので，$\mathcal{N}(A)+S=H$ は常に成立する．従って，この式は逆許容性が成立するための条件から消えている．この結果，部分許容性が成立するための条件と逆許容性が成立するための条件が同じものになる．同様に，$S=H$ が完全許容性が成立するための条件から消えている．この結果，順許容性が成立するための条件と完全許容性が成立するための条件が同じものになる．よって，第 2 種記憶学習に対する許容性では，非許容性，逆許容性，完全許容性の 3 種類の許容性しか現れないのである．一方，部分射影学習では $S\subset H$ であり，平均射影学習では $S=\overline{\mathcal{R}(R)}$ であるので，一般に S は全空間にならない．この結果，これらの射影学習に対しては，教師信号に雑音が含まれている場合には 5 種類の許容性がすべて現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の

許容性が現れたのである．このように，許容性の現れ方が異なる原因が，実は $S=H$ かどうかにあったのである．

この結果をより一般的にしたものが次の定理である．

定理 5.11 $S=H$ のとき，SL 射影学習の第 2 種記憶学習に対する部分許容性が成立するための必要十分条件は

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.66}$$

が成立することであり，このとき逆許容性が自動的に成立する．完全許容性が成立するための必要十分条件は，

$$\mathcal{N}(A) = \{0\} \tag{5.67}$$

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{5.68}$$

が同時に成立するか，または，

$$\mathcal{N}(A) = H \tag{5.69}$$

$$Q = 0 \tag{5.70}$$

が同時に成立することである．教師信号に雑音が含まれていない場合には，逆許容性は常に成立し，完全許容性が成立するための必要十分条件は，

$$\mathcal{N}(A) = \{0\} \tag{5.71}$$

が成立するか，または

$$\mathcal{N}(A) = H \tag{5.72}$$

が成立することである．

なお，このような性質は第 2 種記憶学習に対して成立するものであり，第 1 種記憶学習に対しては $S=H$ であっても一般の場合と同じように，教師信号に雑音が含まれている場合には 5 種類すべての許容性が現れ，雑音が含まれていない場合は非許容性以外の 4 種類の許容性が現れる．2 種類の記憶学習の前提知識の違いが，許容性の構造にこのような違いをもたらすのである．

5.5 まとめ

本章では，SL 射影学習を記憶学習で実現できるための条件，即ち SL 射影学習の記憶学習に対する各種許容性が成立するための条件を求めた．記憶学習には，雑音に関する前提知識を用いない第 1 種記憶学習と前提知識を用いる第 2 種記憶学習があるが，それぞれの記憶学習に対する各種許容性のための条件から，以下のことが明らかになった．

まず，どちらの記憶学習を用いた場合にも，どの許容性が成立するかということは，教材と雑音の性質によって決まることがわかった．

SL 射影学習の第 1 種記憶学習に対する許容性では，教師信号に雑音が含まれている場合にはどの SL 射影学習に対しても 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかった．個々の条件を見てみると，雑音の相関行列 Q に関係しないものとするものに分かれており，前

者はそのまま雑音が含まれていない場合の条件になる．一方，後者は更に，(a) $Q=0$ の場合に自動的に成立する条件と，(b) $Q=0$ の場合に教材 $\{\boldsymbol{x}_m\}_{m=1}^{M}$ に関する制約になる条件の 2 種類に分かれることがわかった．

SL 射影学習の第 2 種記憶学習に対する許容性でも，教師信号に雑音が含まれている場合には一般に 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかった．しかし，射影学習のように $S=H$ であれば状況は異なっている．即ち，教師信号に雑音が含まれている場合には非許容性，逆許容性，完全許容性の 3 種類の許容性だけが現れ，雑音が含まれていない場合には逆許容性と完全許容性の 2 種類の許容性だけが現れる．従来の個別の射影学習に対する議論では，射影学習と部分射影学習，平均射影学習で，第 2 種記憶学習に対する許容性の現れ方が表 5.2 に示したように異なっていたが，その原因はわかっていなかった．しかし，許容性を射影学習族に対して統一的に論じることにより，その原因が実は $S=H$ かどうかにあることが解明できたのである．なお，このような性質は第 2 種記憶学習に対して成立するものであり，第 1 種記憶学習に対しては $S=H$ であっても一般の場合と同じように，教師信号に雑音が含まれている場合には 5 種類すべての許容性が現れ，雑音が含まれていない場合は非許容性以外の 4 種類の許容性が現れるのである．個々の条件を見てみると，雑音の相関行列 Q に関係しないものとするものに分かれており，前者はそのまま雑音が含まれていない場合の条件になる．一方，この場合の後者は $Q=0$ の場合に自動的に成立するもの，即ち上記分類の (a) に属すものだけであることを解明した．

各種許容性ごとに条件を比較してみると，非許容性，部分許容性，順許容性が成立するための条件は同じであるが，逆許容性，完全許容性が成立するための条件は，第 2 種記憶学習を用いる場合の方がゆるくなっていることがわかった．逆許容性，完全許容性はすべての SL 射影学習作用素を実現しようとするものであり，第 2 種記憶学習作用素の集合の方が第 1 種記憶学習作用素の集合より大きいので，これらの許容性の条件がそれだけゆるくなったのである．即ち，SL 射影学習と同じ結果を得ようとするだけであればどちらの記憶学習を用いても同じであるが，すべての SL 射影学習作用素を実現しようとするときには第 2 種記憶学習を用いる方がよいことが明らかになった．

第 6 章

許容化問題

6.1 まえがき

本章では，前章の議論に関連して，新しく許容化問題と名付けた問題を論じる．前章の議論より，許容性が成立するかどうかは教材に密接に関係することがわかる．従って，与えられた教材によっては許容性が成立していないこともある．そのような場合には，文献 [71] に述べられている方法によって，初めから教材を取り直すことも考えられる．しかし，折角手元にある教材を捨ててしまうのは無駄なことであるし，新しく教材を採集することが困難な場合や高価な場合もある．そのような場合，現在手元にある教材の一部を修正することにより許容性を成立させることができれば便利である．

実はこのような立場からの議論はこれまでなされていなかった．そこで本論文では，この問題を許容化問題と名付け，許容性が成立するように手元にある教材を修正する方法を与えることにする．具体的には，現在手元にある教材の一部を削除し，新しくいくつかの教材を追加することにより，許容性を成立させるのである．次節では，まず許容化問題を定義し，つづいて，射影学習の実現手段として記憶学習を用いる場合の許容化問題を論じることにする．

6.2 許容化問題

まず，許容化問題を次のように定義する．

定義 6.1 （許容化問題）手元にある教材に対して許容性が成立していないとき，それを修正することにより許容性を成立させることを許容化（realization of admissibility）という．許容化のための方法を導く問題を許容化問題（realization problem of admissibility）という．

許容化には，第 5.2 節で述べた 5 種類の許容性の（b）～（e）に対応した部分許容化，順許容化，逆許容化，完全許容化がある．この中でも特に重要なものが順許容化と完全許容化である．本章では，射影学習の実現手段として記憶学習を用いた場合に順許容化を行なうような教材の修正法を与えることにする．

6.3 射影学習の記憶学習に対する許容化問題

表 5.3，表 5.5 に示したように，射影学習の記憶学習に対する順許容性が成立するための必要十分条件は，第 1 種記憶学習を用いた場合も第 2 種記憶学習を用いた場合も同じものになっている．そこで本章では，これら 2 種類の記憶学習を単に記憶学習とよび，まとめて議論することにする．

順許容性が成立するための本質的な条件は，

$$\mathcal{N}(A) = \{0\} \tag{6.1}$$

$$Q\mathcal{R}(A)^{\perp} \subset \mathcal{R}(A)^{\perp} \tag{6.2}$$

が同時に成立することであった．式 (5.38) が成立するためには H は有限次元でなければいけない．そこで，H が N 次元空間である場合を考える．

以下の議論では，訓練データの数が変化する．それを明示するために，A および Q を A_M および Q_M と表記する．

まず，Q が対角行列である場合の手続きを示す．

手続き 1 以下の手続きによって順許容性を成立させる教材を得ることができる．
<u>第 *1* 段階</u>　現在手元にある教材 $\{\boldsymbol{x}_m\}_{m=1}^{M}$ から，式 (2.5) によって関数系 $\{\psi_m\}_{m=1}^{M}$ を作る．

Step 1. $\{\psi_m\}_{m=1}^{M}$ が 1 次従属である場合には，いくつかの $\boldsymbol{x}_m$ を削除して，残りの ψ_m が 1 次独立になるようにする．削除した後の新しい標本点を，番号をつけ替えて $\{\boldsymbol{x}_m\}_{m=1}^{M_1}$ と記すことにする．

Step 2. $M_1 < N$ であれば，$(N - M_1)$ 個の新しい標本点 $\{\boldsymbol{x}_m\}_{m=M_1+1}^{N}$ を $\{\psi_m\}_{m=1}^{N}$ が 1 次独立になるように追加する．

<u>第 *2* 段階</u>

Step 3. $(k-1)N$ 個の標本点 $\{\boldsymbol{x}_m\}_{m=N+1}^{kN}$ を，$i = 2, 3, \cdots, k$ に対して

$$\boldsymbol{x}_{(i-1)N+m} = \boldsymbol{x}_m \quad (m = 1, 2, \cdots, N). \tag{6.3}$$

となるように追加する．

第 1 段階における $\{\psi_m\}_{m=1}^{M}$ の 1 次独立性は，空間 $\mathcal{L}\{\{\psi_m\}_{m=1}^{M}\}$ の次元で評価することができる．その次元は，第 i, j 成分が $\langle\psi_j, \psi_i\rangle$ で与えられる $\{\psi_m\}_{m=1}^{M}$ のグラム行列で求めることができる．更に，式 (2.7) , (2.5) より，

$$\langle\psi_j, \psi_i\rangle = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{6.4}$$

となるので，第 1 段階では行列 $(K)_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の階数によって $\{\psi_m\}_{m=1}^{M}$ の一次独立性を判定することができる．

第 1 段階の手続きが修了した後では，$\{\psi_m\}_{m=1}^{N}$ は 1 次独立になっているので，

$$\dim\mathcal{R}(A_N^*) = \mathrm{rank}(K) = N$$

となっている．従って，$\mathcal{N}(A_N)=\{0\}$ と $\mathcal{R}(A_N)=\mathbf{C}^N$ が成立し，条件 (6.1)，(6.2) が成立する．よって，この段階で順許容性は成立しているのである．

一方，$\mathcal{R}(A_N)^\perp=\{0\}$ はまた，命題 2.4より，雑音抑制が全く機能しないことも意味する．この意味で，許容化はまだ不十分である．そこで，第 2 段階が必要になってくる．式 (6.3) は，$\boldsymbol{x}_m$ $(m=1,2,\cdots,N)$ を標本点として k 回選ぶことを意味する．この段階を理解するために，次の補題を用意する．

補題 6.1 関数 $\{\psi_m\}_{m=1}^N$ が 1 次独立であると仮定する．また，雑音の相関行列が

$$Q_N=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_N^2) \tag{6.5}$$

で与えられる場合を考える．ここで，σ_m $(m=1,2,\cdots,N)$ は正の定数である．kN 個の標本点 $\{\boldsymbol{x}_m\}_{m=1}^{kN}$ が $i=2,3,\cdots,k$ に対して式 (6.3) を満たせば，式 (6.2) が成立する．即ち，

$$Q_{kN}\mathcal{R}(A_{kN})^\perp\subset\mathcal{R}(A_{kN})^\perp \tag{6.6}$$

となる．

補題 6.1 では，

$$\dim\mathcal{R}(A_{kN})^\perp=(k-1)N \tag{6.7}$$

が成立しているので，$\mathcal{R}(A_{kN})^\perp\neq\{0\}$ となっている．従って第 2 段階の手続きにより，許容性を成立させたまま雑音抑制を行なうことができるようになるのである．

ところで，雑音の相関行列 Q が正定数 σ と共に $Q=\sigma^2I$ で与えられる場合，式 (6.2) は $\mathcal{R}(A)^\perp$ に関わらず成立する．従って，次の手続きを得る．

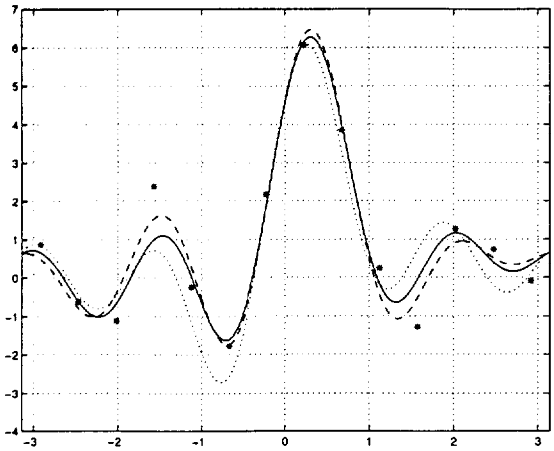
手続き 2 $\mathcal{N}(A_M)\neq\{0\}$ であり，$Q_M=\sigma^2I_M$ である場合には，次の手続きによって順許容性を成立させる教材を得ることができる．

$\mathcal{N}(A_{M+r})=\{0\}$ が成立するように r 個の教材 $\{\boldsymbol{x}_m\}_{m=M+1}^{M+r}$ を追加する．

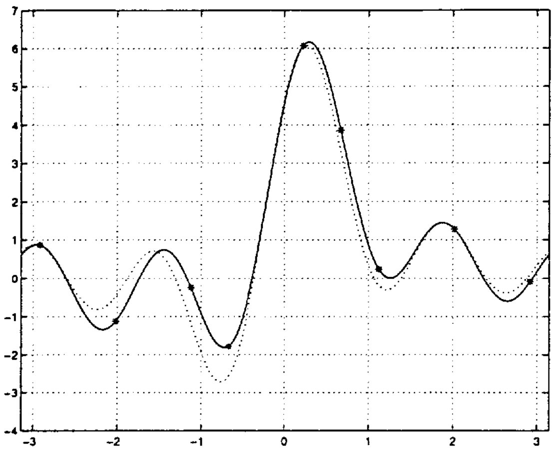
即ち，$Q_M=\sigma^2I_M$ である場合には，$\{\psi_m\}_{m=1}^{M+r}$ は必ずしも 1 次独立である必要はなく，その場合にも雑音抑制を行なうことができるのである．

6.4 計算機実験

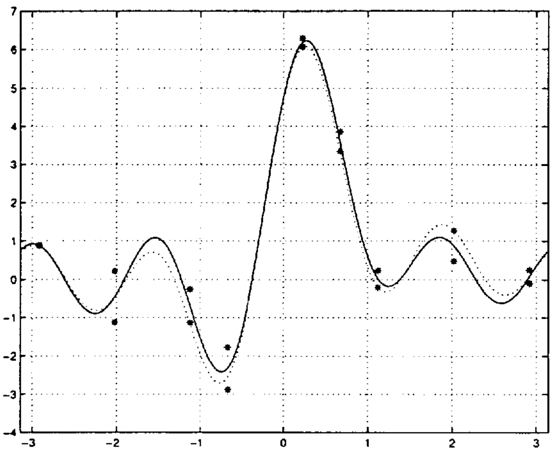
本節では，前節で得られた許容化法の有効性を計算機実験によって示す．第 5.3 節では，順許容性が成立していない場合の例を図 5.2(b) に示した．この場合に手続き 1 を適用することにより許容化を行なうことにする．この図を再び，図 6.1(a) に示す．H は 9 次元空間であり，標本点は 14 個であるので，$\{\psi_m\}_{m=1}^{14}$ は 1 次従属になっている．そこで，第 1 段階の Step.1 に従い，$x_2,x_4,x_7,x_{11},x_{13}$ を削除し，残る ψ_m $(m=1,3,5,6,8,9,10,12,14)$ を 1 次独立にする．このとき，訓練データは 9 個になっており，H の次元と一致しているので，Step 2. は必要なくなり，順許容性が成立することになる．こうして得られた訓練データを改めて，x_m $(m=1,2,\cdots,9)$ と記す．この訓練データを使って学習した関数を図 6.1(b) に示す．f_{P}, f_{M}, f をそれぞれ，実線，破線，点線で，訓練データを '*' で示してある．(a) では異なっていた f_{P} と f_{M} が，この場合には一致していることがわかる．しかし，f_{M} は雑音を含んだ教師信号を通っており，雑音は抑制されていない．



(a)



(b)



(c)

図 6.1　手続き 1 による許容化．実線，破線，点線はそれぞれ f_{P}, f_{M}, および f を表す．'$*$' は訓練データを表す．(a) 順許容性が成立していない場合の関数．図 5.2(b) と同じ図である．(b) 手続き 1 の第 1 段階に従い，(a) の $x_2, x_4, x_7, x_{11}, x_{13}$ を削除して得られた 9 個の標本点を使って学習した関数．f_{M} が f_{P} と一致しているので，順許容性が成立していることがわかる．しかし，雑音は抑制されていない．(c) 手続き 1 の第 2 段階に従い，$x_{m+9} = x_m (m = 1, 2, \cdots, 9)$ を追加して得られた 18 個の標本点を使って学習した関数．f_{M} が f_{P} に一致しているだけでなく，雑音が抑制され，f_{M} が f に近くなっていることがわかる．

次に，手続き 1 の第 2 段階に従い，

$$x_{m+9} = x_m \quad (m = 1, 2, \cdots, 9) \tag{6.8}$$

を追加して得られた 18 個の標本点を使って学習した関数を図 6.1(c) に示す．f_{M} は f_{P} と一致しているので順許容性が成立していることがわかる．更に，f_{M} は標本点において真の値に近い点を通っている．即ち，第 2 段階によって，順許容性を成立させたまま，雑音抑制が行なわれていることがわかるのである．

6.5 まとめ

本章では，第 5 章の議論に関連して，与えられた教材に対して許容性が成立していない場合に，教材の一部を修正することにより許容性を成立させる許容化問題と名付けた問題を論じた．そして，雑音の相関行列が対角行列である場合に，射影学習の記憶学習に対する順許容性を成立させるような教材の修正法を与えた．更に，計算機実験により提案した手法の有効性を確認した．

第 7 章

記憶学習の射影学習族における適用範囲

7.1 まえがき

本章では，'学習方式の適用範囲' という問題を新しく提起し，記憶学習の射影学習族における適用範囲を解明する．第 5 章では，本来行ないたい学習方式が先にあり，その実現手段として別の学習方式を用いるという問題を論じた．これとは逆に，ある学習方式が手元にあり，それを様々な学習方式の実現手段として用いる場面を考える．このとき，手元にある学習方式がどの範囲の学習方式にまで適用できるかを明らかにする問題が，適用範囲の問題である．第 5 章で論じた問題がユーザーの立場からの問題設定であるのに対して，本章で扱う問題はメーカーの立場からの問題設定である．例えば，ある学習用のソフトを開発したときに，それを初期の目的以外にもどんどん応用し，様々な分野で活用していこうという立場からの問題設定である．このような問題を論じるためには，(1) 現在手元にある学習方式と，(2) 適用可能な学習方式を探す探索範囲を決めなくてはいけない．本論文では，(1) として訓練誤差だけを最小にするという最も素朴な形の記憶学習を採用する．(2) としては，第 3 章で論じた無限に多くの種類の学習方式を含んでいる射影学習族を採用する．まず初めに，教師信号に雑音が含まれていない場合の記憶学習の適用範囲を解明する．次に，雑音が含まれている場合の記憶学習の適用範囲を解明し，両者を比較することにより，雑音が適用範囲に与える影響を明らかにする．

なお本章では，適用できるということを順許容性が成立するということで定義することにする．従って，第 5 章で得られた結果が利用できることになる．

7.2 教師信号に雑音が含まれていない場合の適用範囲

本節では，教師信号に雑音が含まれていない場合の記憶学習の適用範囲を解明する．この場合の SL 射影学習の順許容性が成立するための本質的な条件は，第 1 種記憶学習を用いた場合も第 2 種記憶学習を用いた場合も，表 5.3，表 5.5 より，

$$\mathcal{N}(A) = \{0\} \tag{7.1}$$

が成立することであった．そこで本節では，これらの 2 種類の記憶学習を単に記憶学習とよび，まとめて議論することにする．

式 (7.1) は，H が無限次元である場合，有限個の訓練データをどのように選んでも成立させることはできない．一方，H が有限次元である場合，適切な訓練データを選べば，

式 (7.1) を成立させることができる．そこで，以下では H が有限次元の場合を論じることにする．

式 (7.1) には S も L も現れない．よって，記憶学習の射影学習族に対する適用範囲は次のようになる．

定理 7.1 (教師信号に雑音が含まれていない場合の適用範囲)

1. 与えられた教材に対して $\mathcal{N}(A) = \{0\}$ が成立しているとき，記憶学習を射影学習族すべてに適用できる．

2. $\mathcal{N}(A) = \{0\}$ が成立していなければ，記憶学習を射影学習族に適用することはできない．

即ち，記憶学習の適用範囲は，$\mathcal{N}(A) = \{0\}$ が成立していれば，射影学習族全体になっているのである．

7.3 教師信号に雑音が含まれている場合の適用範囲

本節では，教師信号に雑音が含まれている場合の適用範囲を解明する．SL 射影学習の順許容性が成立するための本質的な条件は，第 1 種記憶学習を用いた場合も第 2 種記憶学習を用いた場合も，定理 5.2，定理 5.7 より，

$$\mathcal{N}(A) = \{0\} \tag{7.2}$$

$$Q\mathcal{R}(A_s)^\perp \subset \mathcal{R}(A)^\perp \tag{7.3}$$

が同時に成立することであった．そこで本節でも，これら 2 種類の記憶学習を単に記憶学習とよび，まとめて議論することにする．雑音が含まれない場合と同様に，順許容性が成立するためには手元にある教材が式 (7.2) を満たすかどうかが重要になる．一方，式 (7.3) には S が現れる．よって，記憶学習の射影学習族に対する適用範囲は次のようになる．

定理 7.2 (教師信号に雑音が含まれている場合の適用範囲)

1. 与えられた教材に対して $\mathcal{N}(A) = \{0\}$ が成立しているとき，記憶学習は式 (7.3) を満たすすべての S に対する SL 射影学習に適用でき，それ以外の SL 射影学習には適用できない．

2. $\mathcal{N}(A) = \{0\}$ が成立していなければ，記憶学習を射影学習族に適用することはできない．

即ち，記憶学習を射影学習族に適用できるためには，少なくとも $\mathcal{N}(A) = \{0\}$ でなければいけない．一方，式 (7.3) は，SL 射影学習の雑音抑制を記憶学習で達成できるための条件であった．即ち，$\mathcal{N}(A) = \{0\}$ が成立しているとき，記憶学習で達成できる雑音抑制を行なう SL 射影学習の全体に対して，記憶学習を実現手段として適用できるのである．典型的な例を以下に示す．σ を正の定数とする．

系 **7.1** $\mathcal{N}(A) = \{0\}$ が成立し，雑音の相関行列 Q が

$$Q = \sigma^2 I \tag{7.4}$$

で与えられるとき，記憶学習の射影学習族に対する適用範囲は $L = S = H$ の SL 射影学習だけになる．

例えば，m 番目の標本点に加わる雑音 n_m の平均が 0 で，分散が m によらず σ^2 であり，互いに無相関な場合，Q は式（7.4）で与えられる．

式（7.4）が成立するとき $Q\mathcal{R}(A_s)^{\perp} = \mathcal{R}(A_s)^{\perp}$ となる．従って式（3.52）より，SL 射影学習は $\mathcal{R}(A_s)^{\perp}$ の元を 0 に変換することによって雑音抑制を行なう．一方，式（2.14）より，記憶学習は $\mathcal{R}(A)^{\perp}$ の元を 0 に変換することによって雑音抑制を行なう．ところで，一般に $\mathcal{R}(A_s)^{\perp} \supset \mathcal{R}(A)^{\perp}$ であるので，$\mathcal{R}(A_s)^{\perp} = \mathcal{R}(A)^{\perp}$，即ち

$$\mathcal{R}(A_s) = \mathcal{R}(A) \tag{7.5}$$

を満たす S に対する SL 射影学習の雑音抑制のみ，記憶学習によって達成できるのである．$\mathcal{N}(A) = \{0\}$ が成立しているので，A は H から $\mathcal{R}(A)$ への全単射である．従って，式（7.5）は $S = H$ と同値である．また，$\mathcal{N}(A) = \{0\}$ が成立しているとき，式（5.18）に示したように $L = S$ である．よって，$L = S = H$ の SL 射影学習にのみ記憶学習を適用できるのである．

ところで，式（7.3）の左辺 $Q\mathcal{R}(A_s)^{\perp}$ は $\mathcal{R}(Q)$ の部分空間である．従って，次の系が成立する．

系 **7.2** $\mathcal{N}(A) = \{0\}$ が成立し，Q が

$$\mathcal{R}(Q) \subset \mathcal{R}(A)^{\perp} \tag{7.6}$$

を満たすとき，記憶学習は射影学習族すべてに適用できる．

式（7.6）が成立しているとき，任意の S に対して

$$Q\mathcal{R}(A_s)^{\perp} = \mathcal{R}(Q) \tag{7.7}$$

となる．従って，すべての SL 射影学習が雑音を完全に抑制できる．一方，式（7.6）が成立しているとき，明らかに記憶学習も雑音を完全に抑制できる．従って，記憶学習を射影学習族全体に適用できるのである．

次の系 7.3 は，ある意味で系 7.2 の逆になっている．

系 **7.3** $\mathcal{N}(A) = \{0\}$ が成立し，Q が

$$\mathcal{R}(Q) \subset \mathcal{R}(A) \tag{7.8}$$

を満たすとき，記憶学習は

$$\mathcal{R}(Q) \subset \mathcal{R}(A_s) \tag{7.9}$$

を満たすすべての S に対する SL 射影学習に適用でき，それ以外の SL 射影学習には適用できない．

S が式 (7.9) を満たしているとき，

$$Q\mathcal{R}(A_s)^{\perp} = \{0\} \tag{7.10}$$

となるので，この S に対する SL 射影学習は雑音を全く抑制できない．一方，式 (7.8) が成立しているとき，雑音はすべて $\mathcal{R}(A)$ に含まれるので，$\mathcal{R}(A)^{\perp}$ にその成分は現れない．従って，記憶学習も雑音を全く抑制することができない．よって，記憶学習を式 (7.9) を満たす S に対する SL 射影学習にのみ適用できるのである．

系 7.1 ～ 系 7.3 の 3 種類の例が示しているように，記憶学習の適用範囲は射影学習族全体になったり，1 種類だけになったりするなど，雑音の性質，即ち相関行列 Q の性質によって大きく変化する．例えば，同じ S であっても，Q によって適用できる場合とできない場合があるのである．

また，教師信号に雑音が含まれていない場合と比較すると，一般には，雑音が含まれている場合の適用範囲の方が，含まれていない場合よりも狭くなる．しかし，系 7.2 に示したように，雑音の相関行列 Q の性質によっては適用範囲が射影学習族全体になり，雑音が含まれていない場合と一致することもあるのである．

7.4 計算機実験

本節では，記憶学習を適用できる SL 射影学習と適用できない SL 射影学習の例を示す．記憶学習には，前提知識のより少ない第 1 種記憶学習を用いることにする．

x をスカラーとし，$\{\varphi_n\}_{n=1}^{8}$ を区間 $[-\pi, \pi]$ で定義された関数

$$\varphi_n(x) = \begin{cases} \cos \frac{n+1}{2}x & : n \text{ は奇数} \\ \sin \frac{n}{2}x & : n \text{ は偶数} \end{cases} \tag{7.11}$$

とする．H を，$\{\varphi_n\}_{n=1}^{8}$ の線形結合の全体に対して，内積を

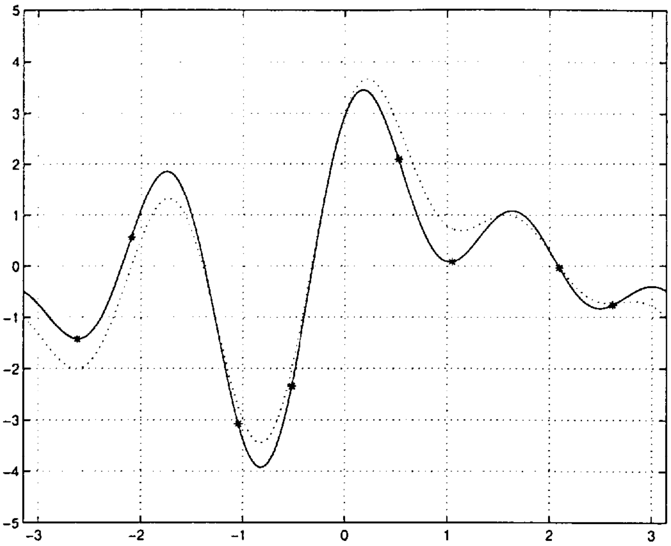
$$\langle f, g \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx \tag{7.12}$$

で定義した空間とする．標本点数を 8 とし，

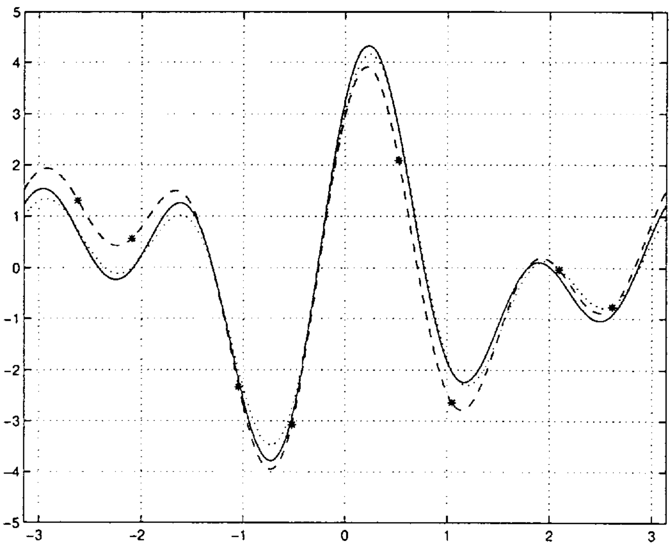
$$\begin{aligned} &\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\} \\ &= \{-\frac{5\pi}{6}, -\frac{2\pi}{3}, -\frac{\pi}{3}, -\frac{\pi}{6}, \frac{\pi}{6}, \frac{\pi}{3}, \frac{2\pi}{3}, \frac{5\pi}{6}\} \end{aligned}$$

とする．このとき $\mathcal{N}(A) = \{0\}$ が成立し，$L = S$ となる．また，$\mathcal{R}(A) = \mathbf{C}^M$ であるので，系 7.3 より，第 1 種記憶学習を SL 射影学習に適用できるかどうかを式 (7.9) で判定することができる．Q として，

$$Q = \sigma^2 \begin{pmatrix} 1&1&0&0&0&0&0&0 \\ 1&1&0&0&0&0&0&0 \\ 0&0&1&1&0&0&0&0 \\ 0&0&1&1&0&0&0&0 \\ 0&0&0&0&1&1&0&0 \\ 0&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&1&1 \end{pmatrix} \tag{7.13}$$



(a)



(b)

図 7.1　第 1 種記憶学習の SL 射影学習への適用可能性．f_{M}, f_{SL}, f をそれぞれ実線，破線，点線で示す．訓練データを '$*$' で示す．(a) 適用可能な例（$f_{\mathrm{M}} = f_{\mathrm{SL}}$）．(b) 適用不可能な例（$f_{\mathrm{M}} \neq f_{\mathrm{SL}}$）．

を考える．ここで，$\sigma = 0.5$ である．

まず，S が

$$S_1 = \mathcal{L}\{\varphi_1, \varphi_2, \varphi_4, \varphi_5, \varphi_6, \varphi_7\}$$

である S 射影学習を考え，$\mathrm{S_1}$射影学習とよぶことにする．S_1 は式 (7.9) を満たすので，$\mathrm{S_1}$射影学習に第 1 種記憶学習を適用することができる．f を

$$f = \varphi_1 + \varphi_2 + \varphi_4 + \varphi_5 + \varphi_6 + \varphi_7 \tag{7.14}$$

とする．このときの f_{M} と f_{SL} を図 7.1(a) に，それぞれ実線と破線で示す．この図には，f も点線で示してある．$f_{\mathrm{M}} = f_{\mathrm{SL}}$ になっており，第 1 種記憶学習を適用できることがわかる．

次に，S が

$$S_2 = \mathcal{L}\{\{\varphi_n\}_{n=3}^{8}\}$$

である S 射影学習を考え，これを $\mathrm{S_2}$射影学習とよぶことにする．S_2 は式 (7.9) を満たさないので，$\mathrm{S_2}$射影学習には第 1 種記憶学習を適用することができない．f が

$$f = \sum_{n=3}^{8} \varphi_n \tag{7.15}$$

である場合を考える．このときの f_{M} と f_{SL} を図 7.1(b) に，それぞれ実線と破線で示す．f も点線で示してある．$f_{\mathrm{M}} \neq f_{\mathrm{SL}}$ になっており，$\mathrm{S_2}$射影学習には第 1 種記憶学習を適用できないことがわかる．

7.5　まとめ

本章では，ある学習方式をどこまで広い範囲の学習方式の実現手段として適用していくことができるかという‘学習方式の適用範囲’の問題を提起し，記憶学習の射影学習族における適用範囲を解明した．記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習があるが，SL 射影学習が常に許容するための条件は同じであるので，これらの記憶学習をまとめて議論することができた．まず，教師信号に雑音が含まれていないとき，H が無限次元である場合には，記憶学習を射影学習族に適用することはできないが，H が有限次元である場合には，適切な訓練データが与えられれば，記憶学習は射影学習族全体に適用できることがわかった．教師信号に雑音が含まれているときには，適切な訓練データが与えられれば，記憶学習で達成できる雑音抑制を行なう SL 射影学習に対して記憶学習を適用できることがわかった．その適用範囲は，射影学習族に属するすべての SL 射影学習になったり，$L = S = H$ という 1 種類の SL 射影学習だけになったりするなど，雑音の性質によって大きく変化することがわかった．

第 8 章

結論

8.1 本研究の成果

本論文では，射影学習，部分射影学習，平均射影学習を含む無限に多くの種類の学習方式である射影学習族の自然な定義を与え，その基礎理論を構築すると共に，最小分散射影学習の提案，許容性理論の確立，記憶学習の射影学習族における適用範囲の解明等を行ない，射影学習族の理論を構築した．以下，本研究の成果を総括する．

第 2 章では，作用素論的手法を用いて学習問題を定式化した．まず，記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習の 2 種類があることを示した．つづいて，射影学習，部分射影学習，平均射影学習について整理し，本論文で学習問題を論じるための準備を行なった．

第 3 章では，まず射影学習族の従来の定義の問題点を整理し，SL 射影学習の概念を導入することにより，射影学習族のより自然な定義を与えた．そして，それが従来の定義と等価であることを示した．この新しい定義を基に，学習作用素の一般形を射影学習族に含まれるすべての学習方式に対して統一的に求めると共に，雑音抑制能力や汎化能力等の諸性質を統一的に解明し，射影学習族の基礎理論を構築した．

第 4 章では，第 3 章で構築した射影学習族の理論の一つの応用として，最小分散射影学習の概念を提案した．これは，学習によって得られる関数の教師信号に含まれている雑音に関ずる分散を，射影学習族の中で最も小さくする学習方式である．まず，SL 射影学習が最小分散射影学習になるための L に関する必要十分条件を与え，各 S に対して最小分散射影学習になるような L が一般には無数に存在することを示した．更に，L が一意に定まるための条件を与え，この条件が成立するとき，最小分散射影学習は射影学習および部分射影学習に限ることを示した．最後に，計算機実験により最小分散射影学習の有効性を示した．

第 5 章では，SL 射影学習を記憶学習で実現できるための条件，即ち SL 射影学習の記憶学習に対する各種許容性が成立するための条件を求めた．記憶学習には，雑音に関する前提知識を用いない第 1 種記憶学習と前提知識を用いる第 2 種記憶学習があるが，それぞれの記憶学習に対する各種許容性のための条件から，以下のことが明らかになった．

まず，どちらの記憶学習を用いた場合にも，どの許容性が成立するかということは，教材と雑音の性質によって決まることがわかった．

SL 射影学習の第 1 種記憶学習に対する許容性では，教師信号に雑音が含まれている場合にはどの SL 射影学習に対しても 5 種類すべての許容性が現れ，教師信号に雑音が含

まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかった．個々の条件を見てみると，雑音の相関行列 Q に関係しないものとするものに分かれており，前者はそのまま雑音が含まれていない場合の条件になる．一方，後者は更に，(a) $Q=0$ の場合に自動的に成立する条件と，(b) $Q=0$ の場合に教材 $\{\boldsymbol{x}_m\}_{m=1}^{M}$ に関する制約になる条件の 2 種類に分かれることがわかった．

SL 射影学習の第 2 種記憶学習に対する許容性でも，教師信号に雑音が含まれている場合には一般に 5 種類すべての許容性が現れ，教師信号に雑音が含まれていない場合には非許容性以外の 4 種類の許容性が現れることがわかった．しかし，射影学習のように $S=H$ であれば状況は異なっている．即ち，教師信号に雑音が含まれている場合には非許容性，逆許容性，完全許容性の 3 種類の許容性だけが現れ，雑音が含まれていない場合には逆許容性と完全許容性の 2 種類の許容性だけが現れる．従来の個別の射影学習に対する議論では，射影学習と部分射影学習，平均射影学習で，第 2 種記憶学習に対する許容性の現れ方が異なっていたが，その原因はわかっていなかった．しかし，許容性を射影学習族に対して統一的に論じることにより，その原因が実は $S=H$ かどうかにあることが解明できたのである．なお，このような性質は第 2 種記憶学習に対して成立するものであり，第 1 種記憶学習に対しては $S=H$ であっても一般の場合と同じように，教師信号に雑音が含まれている場合には 5 種類すべての許容性が現れ，雑音が含まれていない場合は非許容性以外の 4 種類の許容性が現れるのである．個々の条件を見てみると，雑音の相関行列 Q に関係しないものとするものに分かれており，前者はそのまま雑音が含まれていない場合の条件になる．一方，この場合の後者は $Q=0$ の場合に自動的に成立するもの，即ち上記分類の (a) に属すものだけであることを解明した．

各種許容性ごとに条件を比較してみると，非許容性，部分許容性，順許容性が成立するための条件は同じであるが，逆許容性，完全許容性が成立するための条件は，第 2 種記憶学習を用いる場合の方がゆるくなっていることがわかった．逆許容性，完全許容性はすべての SL 射影学習作用素を実現しようとするものであり，第 2 種記憶学習作用素の集合の方が第 1 種記憶学習作用素の集合より大きいので，これらの許容性の条件がそれだけゆるくなったのである．即ち，SL 射影学習と同じ結果を得ようとするだけであればどちらの記憶学習を用いても同じであるが，すべての SL 射影学習作用素を実現しようとするときには第 2 種記憶学習を用いる方がよいことが明らかになった．

第 6 章では，第 5 章の議論に関連して，与えられた教材に対して許容性が成立していない場合に，教材の一部を修正することにより許容性を成立させる許容化問題と名付けた問題を論じた．そして，雑音の相関行列が対角行列である場合に，射影学習の記憶学習に対する順許容性を成立させるような教材の修正法を与えた．更に，計算機実験により提案した手法の有効性を確認した．

第 7 章では，ある学習方式をどこまで広い範囲の学習方式の実現手段として適用していくことができるかという‘学習方式の適用範囲’の問題を提起し，記憶学習の射影学習族における適用範囲を解明した．記憶学習には雑音に関する前提知識を用いない第 1 種記憶学習と，前提知識を用いる第 2 種記憶学習があるが，SL 射影学習が常に許容するための条件は同じであるので，これらの記憶学習をまとめて議論することができた．まず，教師信号に雑音が含まれていないとき，H が無限次元である場合には，記憶学習を射影学習族に適用することはできないが，H が有限次元である場合には，適切な訓練データが与えられれば，記憶学習は射影学習族全体に適用できることがわかった．教師信号に雑音が含まれているときには，適切な訓練データが与えられれば，記憶学習で達成できる雑音抑制を行なう SL 射影学習に対して記憶学習を適用できることがわかった．その適用

範囲は，射影学習族に属するすべての SL 射影学習になったり，$L = S = H$ という 1 種類の SL 射影学習だけになったりするなど，雑音の性質によって大きく変化することがわかった．

8.2 今後の課題

今後，次のような問題を解決していく必要がある．

1. 本論文では，射影学習族に属す無限に多くの学習方式に対して，その一般形を統一的に求めたり，汎化能力や許容性の問題等を統一的に議論してきた．しかし，まだ本来の目的である射影学習族の構造を解明するところまでは至っていない．この構造解明が，学習族の理論における最も基本的で重要な課題である．

2. 学習方式の適用範囲の問題に関して，本論文では訓練誤差だけを最小にするという最も素朴な形の記憶学習の適用範囲を論じたが，学習方式としては他にも誤り修正型記憶学習 [41] や正則化学習 [55] 等が考えられる．このような場合の適用範囲を解明すると共に，本論文で与えた最も素朴な場合の適用範囲と比較することにより，前提知識が適用範囲に与える影響を解明していくことが第 2 の課題である．

3. 許容化問題に関して，本論文では本来行ないたい学習が射影学習である場合を論じたが，これを射影学習族へ拡張していくことが第 3 の課題である．

4. 本論文では許容化問題を，ある学習方式を実現するために教材を修正するという立場で論じたが，適用範囲を拡大するために教材を修正するという立場で論じることもできる．後者の立場から許容化問題を論じることが第 4 の課題である．

5. 本論文では，学習したい関数が含まれている空間がわかっている場合を論じた．また SL 射影学習では，部分空間 S がわかっている場合を論じた．これらの空間の決定法を与えることが第 5 の課題である．

謝辞

本研究を進めるにあたり，始終暖かく御指導頂きました東京工業大学大学院情報理工学研究科の小川英光教授に，心から感謝申し上げます．有益なご意見を頂きました，東京工業大学大学院情報理工学研究科の佐藤泰介教授，沼尾正行助教授，熊沢逸夫助教授，東京工業大学精密工学研究所の渡邊澄夫助教授に深く感謝申し上げます．

山口大学工学部の浜本義彦教授には，様々なご助言をいただき，心より感謝申し上げます．東京工業大学大学院理工学研究科の山下幸彦助教授には，私の幾多の質問に対して親切にお答え頂き，感謝申し上げます．また，理化学研究所 Sethu Vijayakumar 氏，東京工業大学中島朗子氏，杉山将氏には日頃から熱心に議論をしていただき，感謝申し上げます．佐藤京子事務官には，日頃からお世話になり，心より感謝申し上げます．最後に，いろいろとお世話になりました小川・熊沢・山下研究室の皆様に感謝申し上げます．

参考文献

[1] H. Akaike, "A new look at the statistical model identification," IEEE Trans. Appl. Comp., vol.AC-19, no.6, pp.716–723, Dec. 1974.

[2] A. Albert, "Regression and the Moore-Penrose Pseudoinverse," Academic Press, New York, 1972.

[3] S. Amari, N. Fujita, and S. Shinomoto, "Four types of learning curves," Neural Computation, vol.4, no.4, pp.605–618, July 1992.

[4] S. Amari and N. Murata,"Statistical theory of learning curves under entropic loss criterion," Neural Computation, vol.5, no.1, pp.140-153, Jan. 1993.

[5] S. Amari, "A universal theorem on learning curves," Neural Networks, vol.6, no.2, pp.161–166, 1993.

[6] 甘利俊一（編），"ニューラルネットの新展開 — 研究の最前線を探る —," サイエンス社，東京，1993.

[7] S. Amari, N. Murata, K.-R. Muller, M. Finke, and H.H. Yang, "Asymptotic statistical theory of overtraining and cross-validation," IEEE Trans. Neural Networks, vol.8, no.5, pp.985–996, Sep. 1997.

[8] 有川節夫，西野哲朗，"学習における計算論的アプローチ — 機械学習に向けて —," 情報処理，vol.32, no.3, pp.217–225, Mar. 1991.

[9] 麻生英樹，"ニューラルネットワーク情報処理— コネクショニズム入門，あるいは柔らかな記号に向けて —," 産業図書，東京，1988.

[10] E.B. Baum and D. Haussler, "What size net gives valid generalization?" Neural Computation, vol.1, no.1, pp.151–160, Jan. 1989.

[11] A. Ben-Israel and T.N.E. Greville, "Generalized Inverses : Theory and Applications," John Wiley & Sons, New York, 1974.

[12] S. Bergman, "The Kernel Function and Conformal Mapping," American Mathematical Society, Rhode Island, 1970.

[13] C. Cortes and V. Vapnik, "Support-vector networks," Machinge Learning, vol.20, no.3, pp.273–297, Sep. 1995.

[14] H. Drucker, C.J.C. Burges, L. Kaufman, A, Smola, and V. Vapnik, "Support Vector regression machines," In M. Mozer, M. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems 9, pp.155–161, MIT Press, Cambridge, 1997.

[15] F. Girosi, M. Jones, and Tomaso Poggio, "Regularization theory and neural networks architectures," Neural Computation, vol.7, no.2, pp.219–269, March 1995.

[16] A. Gupta and S.M. Lam, "Weight decay backpropagation for noisy data," Neural Networks, vol.11, no.6, pp.1127–1137, Aug. 1998.

[17] R. Hecht-Nielsen, "Neurocomputing," Addison-Wesley, Reading, Massachusetts, 1989.

[18] G.E. Hinton, D.E. Rumelhart, J.L. McClelland, and The PDP Reserch group, "Distributed Representations, In Parallel Distributed Processing Volume 1," MIT Press, Cambridge, 1986.

[19] 平林晃, 小川英光, "射影学習の記憶学習に対する許容性," 電子情報通信学会ニューロコンピューティング研究会技術研究報告, no.NC95-22, pp.25–32, June 1995.

[20] 平林晃, 小川英光, "雑音が存在する場合の射影学習の記憶学習に対する許容性," 電子情報通信学会 1996 年総合大会講演論文集, vol.6, no.D-17, p.17, March 1996.

[21] 平林晃, 坂本正臣, 小川英光, "教材の追加による過学習の抑制," 電子情報通信学会 1996 年総合大会講演論文集, vol.6, no.D-19, p.19, March 1996.

[22] A. Hirabayashi and H. Ogawa, "Admissibility of memorization learning with respect to projection learning in the presence of noise," Proceedings of ICNN'96, The 1996 IEEE International Conference on Neural Networks, Washington, D.C., vol.1, pp.335-340, June 1996.

[23] 平林晃, 小山隆明, 小川英光, "射影学習族の記憶学習に対する許容性," 日本神経回路学会第 8 回全国大会講演論文集, pp.192-193, Nov. 1997.

[24] 平林晃, 小川英光, "射影学習族," 電子情報通信学会ニューロコンピューティング研究会技術研究報告, no.NC98-49, pp.47-54, Oct. 1998.

[25] A. Hirabayashi and H. Ogawa, "Admissibility of memorization learning with respect to projection learning in the presence of noise," IEICE Trans. Inf. & Syst., vol.E82-D, no.2, pp.488–496, Feb. 1999.

[26] 平林晃, 小川英光, "射影学習族と最小分散型射影学習," 電子情報通信学会 1999 年総合大会講演論文集, vol.6, no.D-2-25, p.19, March 1999.

[27] 平林晃, 小川英光, "射影学習族," 電子情報通信学会論文誌（D-II）, vol.J83–D–II, no.2, pp.754–767, Feb. 2000.

[28] 平林晃, 小川英光, "記憶学習の射影学習族に対する適用可能性," 電子情報通信学会論文誌（D-II）, vol.J83–D–II, no.2, pp.768–775, Feb. 2000.

[29] A. Hirabayashi, H. Ogawa, and A. Nakashima, "Realization of admissibility for supervised learning," IEICE Trans. Inf. & Syst., (Submitted)

[30] A. Hirabayashi and H. Ogawa, "A class of learning for optimal generalization," Proceedings of IJCNN'99, The 1999 International Joint Conference on Neural Networks, Washington, D.C., no.246, July 1999 (CD-ROM).

[31] A. Hirabayashi and H. Ogawa, "What can memorization learning do?" Proceedings of IJCNN'99, The 1999 International Joint Conference on Neural Networks, Washington, D.C., no.247, July 1999 (CD-ROM).

[32] A. Hirabayashi and H. Ogawa, "What can memorization learning do from noisy training data?" Proceedings of ICONIP'99, The 6th International Conference on Neural Information Processing, Perth, Australia, Nov. 1999 (Accepted)

[33] A. Hirabayashi and H. Ogawa, "Projection learning of the minimum variance type," Proceedings of ICONIP'99, The 6th International Conference on Neural Information Processing, Perth, Australia, Nov. 1999 (Accepted)

[34] A. Hirabayashi, H. Ogawa, and A. Nakashima, "Realization of admissibility of memorization learning with respect to projection learning," LWA'99, Lernen, Wissensentdeckung und Adaptivität (The GI-Workshop-Days Learning, Knowledge Discovery, and Adaptivity), Magdeburg, Separate volume, pp.20–31, Sep. 1999.

[35] H. Imai, A. Tanaka, and M. Miyakoshi, "The family of parametric projection filters and its properties for perturbation," IEICE Trans. Inf. & Syst., vol.E80-D, no.8, pp.788–794, Aug. 1997.

[36] 川人光男，"脳の計算理論，" 産業図書，東京，1996.

[37] 小出裕司，山下幸彦，小川英光，"信号・画像推定のための射影フィルタ族の統一理論，" 電子情報通信学会論文誌（D-II）, vol.J77-D-II, no.7, pp.1293-1301, July 1994.

[38] P.D. Laird, "Learning from Good and Bad Data," Kluwer Academic Publishers, 1998. 邦訳，横森貴訳，"例からの学習 — 計算論的学習理論 —，" オーム社，東京，1992.

[39] D.J.C. MacKay, "Comparison of approximate methods for handling hyperparameters," Neural Computation, vol.11, no.5, pp.1035–1068, July 1999.

[40] N. Murata, S. Yoshizawa, and S. Amari, "Network information criterion – Determining the number of hidden units for an artificial neural network model," IEEE Trans. Neural Networks, vol.5, no.6, pp.865–872, Nov. 1994.

[41] A. Nakashima, A. Hirabayashi, and H. Ogawa, "Noise suppression in training data for improving generalization," Proc. 1998 Int. Joint Conf. on Neural Networks, Anchorage, pp.2236–2241, May 1998.

[42] M.A. ナイマルク，邦訳，功力金二郎他訳，"関数解析入門 I，" 共立全書 530，共立出版，東京，1968.

[43] 小川英光, “画像復元問題に現れる作用素方程式について,” 電子情報通信学会パターン認識理解研究会技術研究報告, no.PRU86-60, pp.9-16, Nov. 1986.

[44] H. Ogawa, “Projection filter regularization of ill-conditioned problem,” Proc. of SPIE, vol.808, Inverse Problems in Optics, pp.189–196, March 1987.

[45] 小川英光, “一般標本化定理,” 電子情報通信学会論文誌 (A), vol.J71-A, no.2, pp.163–170, Feb. 1988.

[46] 小川英光, 原昌司, “部分射影フィルタによる画像復元,” 電子情報通信学会論文誌 (A), vol.J71-A, no.2, pp.519–526, Feb. 1988.

[47] 小川英光, 原昌司, “部分射影フィルタの諸性質,” 電子情報通信学会論文誌 (A), vol.J71-A, no.2, pp.527–533, Feb. 1988.

[48] 小川英光, “逆問題とニューラルネットワーク,” 電子情報通信学会第 2 回回路とシステム軽井沢ワークショップ予稿集, pp.262–268, May 1989.

[49] 小川英光, “逆問題としてのニューラルネット理論,” 電子情報通信学会誌, vol.73, no.7, pp.690–695, July 1990.

[50] 小川英光, 山崎一孝, “ニューラルネットワークの汎化能力と過学習,” 電子情報通信学会ニューロコンピューティング研究会技術研究報告, no.NC91-75, pp.77–84, Jan. 1992.

[51] 小川英光, 山崎一孝, “過学習の理論,” 電子情報通信学会論文誌 (D-II), vol.J76-D-II, no.7, pp.1280–1288, June 1993.

[52] 小川英光 (編著), “パターン認識・理解の新たな展開 —挑戦すべき課題—,” コロナ社, 1994.

[53] 小川英光, “ニューラルネットワークと汎化能力,” 電子情報通信学会ニュロコンピューティング研究会技術研究報告, no.NC95-8, pp.57-64, May 1995.

[54] L. Prechelt, “Automatic early stopping using cross validation: Quantifying the criteria,” Neural Networks, vol.11, no.4, pp.761–767, June 1998.

[55] T. Poggio and F. Girosi, “Networks for approximation and learning,” Proc. IEEE, vol.78, no.9, pp.1481–1497, Sep. 1990.

[56] 坂本正臣, 平林晃, 小川英光, “教材の追加・削除による記憶学習の射影学習に対する最適許容化,” 電子情報通信学会 1998 年総合大会講演論文集, vol.6, no.D-2-15, p.21, March 1998.

[57] R. Schatten, “Norm Ideals of Completely Continuous Operators,” 2nd printing, Springer-Verlag, Heidelberg, 1970.

[58] R. Setiono, “A penalty-function approach for pruning feedforward neural networks,” Neural Computation, vol.9, no.1, pp.185-204, Jan. 1997.

[59] M. Stitson, A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, and J. Weston, "Support vector regression with ANOVA decomposition kernels," In B. Scholkopl, C.J.C. Burges, and A.J. Smola, editors, Advances in Kernel Methods – Support Vector Learning, pp.285–292, MIT Press, Cambridge, 1999.

[60] M. Sugiyama and H. Ogawa, "Functional Analytic Approach to Model Selection – Subspace Information Criterion–," IBIS'99, 1999 Workshop on Information-Based Induction Sciences, 1999.

[61] 鳥脇純一郎, "認識工学 —パターン認識とその応用—," コロナ社, 東京, 1996.

[62] L.G. Valiant, "A theory of the learnable," Communications of the ACM, vol.27, no.11, pp.1134–1142, Nov. 1984.

[63] V.N. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, New York, 1995.

[64] S. Vijayakumar, "Computational theory of incremental and active learning for optimal generalization," 東京工業大学博士論文, 1998.

[65] S. Vijayakumar, M. Sugiyama, and H. Ogawa, "Training data selection for optimal generalization with noise variance reduction in neural networks," Proceedings of WIRN Vietri-98, The 10-th Italian Workshop on Neural Nets, Salerno, Italy, pp.153–166, May 1998.

[66] G. Wahba, "Spline Models for Observational Data," Series in Applied Mathematics, vol.59, SIAM, Philadelphia, 1990.

[67] 山下幸彦, 小川英光, "平均射影フィルタによる画像復元," 電子情報通信学会論文誌（D-II）, vol.J74-D-II, no.2, pp.159–157, Feb. 1991.

[68] 山下幸彦, 小川英光, "平均射影フィルタの諸性質," 電子情報通信学会論文誌（D-II）, vol.J74-D-II, no.2, pp.142–149, Feb. 1991.

[69] 山下幸彦, 小川英光, "最適画像復元フィルタと一般逆," 電子情報通信学会論文誌（D-II）, vol.J75-D-II, no.5, pp.883–889, May 1992.

[70] 山崎一孝, 小川英光, "雑音を考慮した過学習の理論," 電子情報通信学会論文誌（D-II）, vol.J76-D-II, no.11, pp.2411–2418, Nov. 1993.

[71] 山崎一孝, 小川英光, "過学習を起こさない教材の選び方," 電子情報通信学会論文誌（D-II）, vol.J76-D-II, no.11, pp.2419–2426, Nov. 1993.

付録 A

定理の証明

定理 3.2 の証明

まず，以下の補題を用意する．部分空間 $\overline{\mathcal{R}(R_F^2)}$ は，式 (3.11) と同様に，

$$\overline{\mathcal{R}(R_F^2)} = \mathcal{R}(R_F^2 A^*) \dot{+} \{\overline{\mathcal{R}(R_F^2)} \cap \mathcal{N}(A)\} \tag{A.1}$$

と直和分解できる．

補題 A.1 閉部分空間 $\overline{\mathcal{R}(R_F^2)}$ の中でみれば $\overline{\mathcal{R}(R_F^2)} \cap \mathcal{N}(A)$ に沿った $\mathcal{R}(R_F^2 A^*)$ への射影となる H 上の線形作用素を P_{RF} で表す．作用素 X が式 (3.6) を満たすための必要十分条件は，X が次式を満たすことである．

$$XAP_{\overline{\mathcal{R}(R_F^2)}} = P_{RF} P_{\overline{\mathcal{R}(R_F^2)}} \tag{A.2}$$

補題 A.2 作用素 X が R_F 射影学習作用素になるための必要十分条件は，X が条件 (A.2) のもとで式 (3.26) を最小にすることである．

補題 A.3 閉部分空間 S_1, S_2 に対して，次の 2 式は同値である．

$$S_1 \cap S_2^{\perp} = \{0\} \tag{A.3}$$

$$P_{S_1} S_2 = S_1 \tag{A.4}$$

(定理 3.2 の証明)

R_F から導かれる射影学習族は，$T = R_F^2$ とおけば，補題 A.2より T から導かれる射影学習族になる．逆も，$R_F = T^{1/2}$ とおくことにより成立する．従って，これら 2 種類の射影学習族は一致する．

次に，S と L から導かれる射影学習族と T から導かれる射影学習族が一致することを示す．そのために，条件 (3.25) , (3.12) が同値であることを示す．まず，SL 射影学習を考える．作用素 T を

$$T = P_L + P_{S \cap \mathcal{N}(A)} \tag{A.5}$$

とおく．P_L, $P_{S \cap \mathcal{N}(A)}$ は共に半正値自己共役作用素であるから，$\mathcal{R}(T)$ は右辺の 2 つの作用素の値域の和になる．また，L も $S \cap \mathcal{N}(A)$ も閉じているので，$\mathcal{R}(T)$ も閉部分空間になる．従って，式 (A.5) , (3.24) より，

$$\begin{aligned}\overline{\mathcal{R}(T)} &= \mathcal{R}(T) = \mathcal{R}(P_L) + \mathcal{R}(P_{S \cap \mathcal{N}(A)}) \\ &= L + \{S \cap \mathcal{N}(A)\} = S\end{aligned}$$

となり，
$$\overline{\mathcal{R}(T)} = S \tag{A.6}$$
となる．また，式 (3.24) より $L \cap \{S \cap \mathcal{N}(A)\} = \{0\}$ であり，$L \subset S$ であるから $L \cap S = L$ となるので，
$$L \cap \mathcal{N}(A) = \{0\} \tag{A.7}$$
となる．$\mathcal{N}(A) = \mathcal{R}(A^*)^{\perp}$ であり，$\mathcal{R}(A^*)$ も L も閉部分空間であるから，式 (A.7) と補題 A.3より
$$P_L \mathcal{R}(A^*) = L \tag{A.8}$$
となる．従って，式 (A.5)，(A.8) より，
$$\mathcal{R}(TA^*) = (P_L + P_{S \cap \mathcal{N}(A)})\mathcal{R}(A^*) = L$$
となり，
$$\mathcal{R}(TA^*) = L \tag{A.9}$$
となる．式 (A.6)，(A.9) より，SL 射影学習は T 射影学習になる．

逆に T 射影学習は，$S = \overline{\mathcal{R}(T)}$, $L = \mathcal{R}(TA^*)$ とおけばSL 射影学習になる．従って，これら 2 種類の射影学習族は一致する． ■

定理 3.3 の証明

行列 P_1, P_2 を，
$$P_1 = (P_{\mathcal{R}(As)^{\perp}} Q P_{\mathcal{R}(As)^{\perp}})(P_{\mathcal{R}(As)^{\perp}} Q P_{\mathcal{R}(As)^{\perp}})^{\dagger} \tag{A.10}$$
$$P_2 = I - Q(P_{\mathcal{R}(As)^{\perp}} Q P_{\mathcal{R}(As)^{\perp}})^{\dagger} \tag{A.11}$$
で定義する．以下の補題を用意する．

補題 A.4 次式が成立する．
$$P_{\mathcal{R}(A_s)} + P_1 = P_{S_t} \tag{A.12}$$
$$P_{S_t} P_1 = P_1 \tag{A.13}$$

補題 A.5 P_2 は $Q\mathcal{R}(A_s)^{\perp}$ に沿った $\mathcal{N}(QP_{\mathcal{R}(As)^{\perp}})$ への射影行列であり，次式が成立する．
$$P_2 P_{S_t} = P_t \tag{A.14}$$

補題 A.6 作用素 X がSL 射影学習作用素になるための必要十分条件は，X がある作用素 C と共に，次の連立作用素方程式を満たすことである．
$$XA_s = PP_S \tag{A.15}$$
$$XQ = CA_s^* \tag{A.16}$$

補題 A.7 式 (A.16) を満たす C が存在するための必要十分条件は，X が次式を満たすことである．
$$XQP_{\mathcal{R}(As)^{\perp}} = 0 \tag{A.17}$$

補題 A.8 作用素 X が SL 射影学習作用素になるための必要十分条件は，X が次の連立作用素方程式を満たすことである．

$$XP_{\mathcal{R}(A_s)} = PA_s^\dagger \tag{A.18}$$
$$XP_1 = -PA_s^\dagger Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger \tag{A.19}$$

以上の補題を使って，定理 3.3 を証明する．

（定理 3.3 の証明）

作用素 X を SL 射影学習作用素とする．補題 A.8 より式 (A.18)，(A.19) が成立するから，両辺の和をとれば，式 (A.12)，(A.11) より

$$XP_{S_t} = PA_s^\dagger P_2 \tag{A.20}$$

となる．式 (A.20)，(A.14) より，

$$XP_{S_t} = XP_{S_t}P_{S_t} = PA_s^\dagger P_2P_{S_t} = PA_s^\dagger P_t$$

となり，式 (3.46) を得る．

逆に，式 (3.46) が成立しているとき，式 (3.43) より，

$$XP_{\mathcal{R}(A_s)} = XP_{S_t}P_{\mathcal{R}(A_s)} = PA_s^\dagger P_tP_{\mathcal{R}(A_s)} = PA_s^\dagger P_{\mathcal{R}(A_s)} = PA_s^\dagger$$

となり，式 (A.18) を得る．また，式 (A.14)，(A.13)，(A.11)，(A.10) より，

$$\begin{aligned}XP_1 &= XP_{S_t}P_1 = PA_s^\dagger P_tP_1 = PA_s^\dagger P_2P_{S_t}P_1 = PA_s^\dagger P_2P_1\\ &= PA_s^\dagger\{I - Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger\}(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger\\ &= -PA_s^\dagger Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger\end{aligned}$$

となり，式 (A.19) を得る．

式 (3.47) を示す．式 (A.14) より，$X = PA_s^\dagger P_2$ は式 (3.46) の解になっている．即ち，式 (3.46) は解をもつ．その一般形は，文献 [37] の補題 2 より，

$$A^{(\mathrm{SL})} = PA_s^\dagger P_tP_{S_t}^\dagger + Y(I - P_{S_t}P_{S_t}^\dagger) \tag{A.21}$$

となる．P_{S_t} は正射影行列であるから $P_{S_t}^\dagger = P_{S_t} = P_{S_t}P_{S_t}^\dagger$ となり，$P_tP_{S_t} = P_t$ であるので，式 (3.47) を得る．

最後に式 (3.48) を示す．簡単のために $F = PA_s^\dagger P_t$ とおけば，式 (3.47)，(3.43) より

$$A^{(\mathrm{SL})}Q = FQ \tag{A.22}$$

となる．また，P_t の定義より $P_tQP_{\mathcal{R}(As)^\perp} = 0$ となるので，

$$FQP_{\mathcal{R}(As)^\perp} = 0 \tag{A.23}$$

となる．更に，$\mathcal{R}(P_t) = \mathcal{R}(A_s)$ より，

$$FP_{\mathcal{R}(A_s)} = PA_s^\dagger P_tP_{\mathcal{R}(A_s)} = PA_s^\dagger P_{\mathcal{R}(A_s)} = PA_s^\dagger$$

となり，

$$FP_{\mathcal{R}(A_s)} = PA_s^\dagger \tag{A.24}$$

となる．式 (C.12)，(A.22)，(A.23)，(A.24) より，

$$\begin{aligned} J_{n0} &= J_n[A^{(\mathrm{SL})}] = \langle A^{(\mathrm{SL})}Q, A^{(\mathrm{SL})}\rangle \\ &= \langle FQ, A^{(\mathrm{SL})}\rangle = \langle F, A^{(\mathrm{SL})}Q\rangle = \langle F, FQ\rangle \\ &= \langle FQ, F\rangle = \langle FQ(P_{\mathcal{R}(A_s)} + P_{\mathcal{R}(As)^\perp}), F\rangle \\ &= \langle FQP_{\mathcal{R}(A_s)}, F\rangle = \langle FQ, FP_{\mathcal{R}(A_s)}\rangle = \langle FQ, PA_s^\dagger\rangle \end{aligned} \tag{A.25}$$

となり，式 (3.48) を得る． ■

定理 3.4 の証明

式 (3.62) が成立しているとき，P_t は $\mathcal{R}(A_s)$ への正射影行列になり，$P_t = P_{\mathcal{R}(A_s)}$ となる．従って，

$$PA_s^\dagger P_t = PA_s^\dagger P_{\mathcal{R}(A_s)} = PA_s^\dagger$$

となり，

$$PA_s^\dagger P_t = PA_s^\dagger \tag{A.26}$$

となる．よって，式 (3.47) より式 (3.63) を得る．

逆に，$A^{(\mathrm{SL})}$ が式 (3.63) で表されているとする．$P_S - P$ は S の中でみれば L に沿った $S \cap \mathcal{N}(A)$ への射影作用素であるから，$A(P_S - P)P_S = 0$ となり，

$$APP_S = AP_S \tag{A.27}$$

となる．従って，式 (3.42) より，

$$APA_s^\dagger = APP_SA_s^\dagger = AP_SA_s^\dagger = A_sA_s^\dagger = P_{\mathcal{R}(A_s)}$$

となり，

$$APA_s^\dagger = P_{\mathcal{R}(A_s)} \tag{A.28}$$

を得る．よって，式 (3.46)，(3.63)，(A.12)，(A.10) より，

$$\begin{aligned} P_t &= P_{\mathcal{R}(A_s)}P_t = APA_s^\dagger P_t = AA^{(\mathrm{SL})}P_{S_t} \\ &= APA_s^\dagger P_{S_t} = APA_s^\dagger(P_{\mathcal{R}(A_s)} + P_1) \\ &= APA_s^\dagger + APA_s^\dagger(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger = P_{\mathcal{R}(A_s)} \end{aligned}$$

となり，$P_t = P_{\mathcal{R}(A_s)}$ となる．従って，

$$Q\mathcal{R}(A_s)^\perp = \mathcal{N}(P_t) \cap S_t \subset \mathcal{N}(P_t) = \mathcal{N}(P_{\mathcal{R}(A_s)}) = \mathcal{R}(A_s)^\perp$$

となり，式 (3.62) を得る． ■

定理 4.1 の証明

式 (3.25) の右辺に現れる作用素 PP_S をまとめて P_e で表す：

$$P_e = PP_S \tag{A.29}$$

P_e は，P が S の外の元に対してどのような変換を行なっていても，それに関係なく次に示す作用素になっている．即ち，式 (3.24) より，H は

$$H = L\dot{+}[\{S \cap \mathcal{N}(A)\} \oplus S^{\perp}] \tag{A.30}$$

と直和分解でき，次の補題が成立する．

補題 A.9 作用素 P_e は，$\{S \cap \mathcal{N}(A)\} \oplus S^{\perp}$ に沿った L への射影作用素になる．

P が S の中だけで射影作用素になっているのに対して，補題 A.9 からわかるように，P_e は全空間 H で射影作用素になっている．そして L は，P の値域 $\mathcal{R}(P)$ ではなく，P_e の値域 $\mathcal{R}(P_e)$ になっている．従って，L に関する議論を行っていくためには，PP_S をまとめて表した P_e の方が都合がよい．そこで以下の証明では，P_e を用いて議論していくことにする．

作用素 P_α を

$$P_\alpha = P_S - P_{\mathcal{R}(A_s^*)} \tag{A.31}$$

で定義する．$S \supset \mathcal{R}(A_s^*)$ であるので，これら 2 つの部分空間への正射影作用素の差である P_α は，$\mathcal{R}(A_s^*)$ における S の直交補空間 $S \cap \mathcal{N}(A)$ への正射影作用素になる．以下の補題を用意する．

補題 A.10 作用素 P_e は，

$$P_e = (I + P_\alpha P_e)P_{\mathcal{R}(A_s^*)} \tag{A.32}$$

と表現できる．ここで，I は H 上の恒等作用素である．

補題 A.11 SL 射影学習が最小分散射影学習になるための必要十分条件は，P_e が

$$P_\alpha P_e P_{L_N} = 0 \tag{A.33}$$

を満たすことである．このとき J_{n0} の値は式 (4.3) になる．

以上の補題を使って，定理 4.1 を証明する．
(定理 4.1 の証明)
SL 射影学習が最小分散射影学習であるとき，補題 A.11 より，式 (A.33) が成立する．式 (4.2) より $\mathcal{R}(A_s^*) \supset L_N$ であるから，

$$P_{\mathcal{R}(A_s^*)} P_{L_N} = P_{L_N} \tag{A.34}$$

となる．従って，式 (A.32)，(A.34)，(A.33) より，

$$P_e P_{L_N} = (I + P_\alpha P_e)P_{\mathcal{R}(A_s^*)} P_{L_N} = (I + P_\alpha P_e)P_{L_N} = P_{L_N}$$

となり，

$$P_e P_{L_N} = P_{L_N} \tag{A.35}$$

となる．よって，

$$L = \mathcal{R}(P_e) \supset \mathcal{R}(P_e P_{L_N}) = \mathcal{R}(P_{L_N}) = L_N$$

となり，式 (4.1) を得る．

逆を示す．SL 射影学習の L が式 (4.1) を満たしているとき，$\mathcal{R}(P_e) \supset \mathcal{R}(P_{L_N})$ となり，式 (A.35) が成立する．更に，$L_N \subset \mathcal{R}(A_s^*)$ であるので $L_N \perp S \cap \mathcal{N}(A)$ となり，$P_\alpha P_{L_N} = 0$ となる．従って，

$$P_\alpha P_e P_{L_N} = P_\alpha P_{L_N} = 0$$

となり，式 (A.33) を得る．よって補題 A.11 より，この SL 射影学習は最小分散射影学習になる．

式 (4.3) は補題 A.11 に示されているとおりである． ■

定理 4.2 の証明

雑音 $\boldsymbol{n}$ は $\mathcal{R}(Q)$ に含まれるので，あるベクトル $\boldsymbol{u}$ を用いて

$$\boldsymbol{n} = Q\boldsymbol{u} \tag{A.36}$$

と表すことができる．また，補題 A.11 より，P_e は式 (A.33) を満たす．従って，式 (A.36)，(A.32)，(4.2)，(A.34)，(A.33) より，

$$\begin{aligned} PA_s^\dagger P_t \boldsymbol{n} &= P_e A_s^\dagger P_t \boldsymbol{n} \\ &= P_e A_s^\dagger P_t Q\boldsymbol{u} \\ &= (I + P_\alpha P_e) P_{\mathcal{R}(A_s^*)} P_{L_N} A_s^\dagger P_t Q\boldsymbol{u} \\ &= (I + P_\alpha P_e) P_{L_N} A_s^\dagger P_t Q\boldsymbol{u} \\ &= P_{L_N} A_s^\dagger P_t Q\boldsymbol{u} = A_s^\dagger P_t \boldsymbol{n} \end{aligned}$$

となり，式 (4.6) が成立する．よって，式 (3.59) より式 (4.5) を得る． ■

定理 4.3 の証明

L_N が $\mathcal{R}(A_s^*)$ と一致しているとき，即ち，

$$L_N = \mathcal{R}(A_s^*) \tag{A.37}$$

が成立しているとき，$\mathcal{R}(P_t) = A_s A_s^\dagger \mathcal{R}(P_t)$ と式 (4.2) より，

$$\begin{aligned} \mathcal{R}(P_t Q) &= A_s A_s^\dagger \mathcal{R}(P_t Q) = A_s \mathcal{R}(A_s^\dagger P_t Q) \\ &= A_s L_N = A_s \mathcal{R}(A_s^*) = \mathcal{R}(A_s) \end{aligned}$$

となり，式 (4.10) を得る．

逆に，式 (4.10) が成立しているとき，式 (4.2) より

$$L_N = A_s^\dagger \mathcal{R}(P_t Q) = A_s^\dagger \mathcal{R}(A_s) = \mathcal{R}(A_s^*)$$

となり，式 (A.37) を得る．

定理の後半は，式 (4.9),(A.37) と $L_C=\{0\}$ より $L=\mathcal{R}(A_s^*)$ となるので，式 (3.31), (3.32) より明らかである． ■

定理 5.1 の証明

次の補題を用意する．

補題 A.12 式 (5.7) は次式と同値である．

$$P_t = P_{\mathcal{R}(A)}P_{S_t} \tag{A.38}$$

（定理 5.1 の証明）

SL 射影学習が第 1 種記憶学習を部分的に許容しているとき，式 (3.46) を満たす $A^{(\mathrm{M1})}$ が存在する．この $A^{(\mathrm{M1})}$ に対して，式 (A.28), (2.11) より，

$$\begin{aligned} P_t &= P_{\mathcal{R}(A_s)}P_t = APA_s^{\dagger}P_t \\ &= AA^{(\mathrm{M1})}P_{S_t} = P_{\mathcal{R}(A)}P_{S_t} \end{aligned}$$

となり，式 (A.38) を得る．よって，補題 A.12より，式 (5.7) を得る．

逆に，式 (5.7) が成立しているとき，補題 A.12 より式 (A.38) を得る．このとき，式 (2.12) において $Y=PA_s^{\dagger}P_t$ とおけば，式 (A.28), (A.38) より，

$$\begin{aligned} A^{(\mathrm{M1})}P_{S_t} &= \{A^{\dagger}+(I-A^{\dagger}A)PA_s^{\dagger}P_t\}P_{S_t} \\ &= A^{\dagger}P_{S_t}+PA_s^{\dagger}P_tP_{S_t}-A^{\dagger}P_{\mathcal{R}(A_s)}P_tP_{S_t} \\ &= A^{\dagger}P_{S_t}+PA_s^{\dagger}P_t-A^{\dagger}P_t \\ &= A^{\dagger}P_{S_t}+PA_s^{\dagger}P_t-A^{\dagger}P_{\mathcal{R}(A)}P_{S_t} = PA_s^{\dagger}P_t \end{aligned}$$

となるので，式 (3.46) よりこの $A^{(\mathrm{M1})}$ は SL 射影学習作用素になる． ■

定理 5.2 の証明

次の補題を用意する．

補題 A.13 任意の線形作用素を B, C, D で表す．任意の作用素 Y に対して

$$BYC = D \tag{A.39}$$

が成立すれば，

$$B=0, \quad D=0 \tag{A.40}$$

が同時に成立するか，あるいは

$$C=0, \quad D=0 \tag{A.41}$$

が同時に成立する．

（定理 5.2 の証明）

SL 射影学習が第 1 種記憶学習を常に許容するとき，すべての $A^{(\mathrm{M1})}$ が式（3.46）を満たす．従って，式（2.12）より，

$$\begin{aligned} PA_s^\dagger P_t &= A^{(\mathrm{M1})}P_{S_t} \\ &= A^\dagger P_{S_t} + (I - A^\dagger A)YP_{S_t} \end{aligned}$$

となり，任意の Y に対して

$$(I - A^\dagger A)YP_{S_t} = PA_s^\dagger P_t - A^\dagger P_{S_t} \tag{A.42}$$

となる．よって，補題 A.13 より，

$$A^\dagger A = I, \quad PA_s^\dagger P_t = A^\dagger P_{S_t} \tag{A.43}$$

が同時に成立するか，または

$$P_{S_t} = 0, \quad PA_s^\dagger P_t = A^\dagger P_{S_t} \tag{A.44}$$

が同時に成立する．式（A.43）の第 1 式より式（5.14）を得る．また，式（A.28）と式（A.43）の第 2 式より，

$$\begin{aligned} P_t &= P_{\mathcal{R}(A_s)}P_t = APA_s^\dagger P_t \\ &= AA^\dagger P_{S_t} = P_{\mathcal{R}(A)}P_{S_t} \end{aligned}$$

となり式（A.38）を得る．よって，補題 A.12より式（5.15）を得る．一方，式（A.44）より $S_t = \{0\}$ となる．よって，式（3.43）から式（5.16），(5.17）を得る．

逆に，式 (5.14),(5.15) が同時に成立しているとき，式 (2.12),(5.14) より $A^{(\mathrm{M1})} = A^\dagger$ となる．また，式（5.15）と補題 A.12 より式（A.38）が成立する．従って，式（A.28），(5.14）より，

$$\begin{aligned} A^{(\mathrm{M1})}P_{S_t} &= A^\dagger P_{S_t} = A^\dagger P_{\mathcal{R}(A)}P_{S_t} \\ &= A^\dagger P_t = A^\dagger P_{\mathcal{R}(A_s)}P_t \\ &= A^\dagger APA_s^\dagger P_t = PA_s^\dagger P_t \end{aligned} \tag{A.45}$$

となるので，式（3.46）より任意の第 1 種記憶学習作用素が SL 射影学習作用素になる．

一方，式（5.16），(5.17）が同時に成立しているとき，第 1 式より $A_s = 0$ となるので，式（3.43）より $S_t = \{0\}$ となり，$P_{S_t} = P_t = 0$ となる．よって，式（3.46）より任意の X が SL 射影学習作用素になり，任意の第 1 種記憶学習作用素も SL 射影学習作用素になる． ■

定理 5.3 の証明

次の補題を用意する．

補題 A.14 式（5.25）は次式と同値である．

$$\mathcal{R}(A_s) = \mathcal{R}(A) \tag{A.46}$$

(定理 5.3 の証明)

SL 射影学習が第 1 種記憶学習に許容されるとき，すべての $A^{(\mathrm{SL})}$ が式 (2.11) を満たす．従って，式 (3.47), (A.28) より，

$$
\begin{aligned}
P_{\mathcal{R}(A)} &= AA^{(\mathrm{SL})} \\
&= APA_s^{\dagger}P_t + AY(I - P_{S_t}) \\
&= P_{\mathcal{R}(A_s)}P_t + AY(I - P_{S_t}) \\
&= P_t + AY(I - P_{S_t})
\end{aligned}
$$

となり，任意の Y に対して

$$
AY(I - P_{S_t}) = P_{\mathcal{R}(A)} - P_t \tag{A.47}
$$

となる．従って，補題 A.13 より，

$$
A = 0, \quad P_{\mathcal{R}(A)} = P_t \tag{A.48}
$$

が同時に成立するか，または

$$
P_{S_t} = I, \quad P_{\mathcal{R}(A)} = P_t \tag{A.49}
$$

が同時に成立する．式 (A.48) の第 1 式が成立すれば，式 (3.42) より $A_s = 0$ となるので $P_{\mathcal{R}(A)} = P_t = 0$ となり，第 2 式は自動的に成立する．従って，式 (A.48) より $A = 0$ となり，式 (5.27) を得る．一方，式 (A.49) の第 1 式より，

$$
S_t = \mathbf{C}^M \tag{A.50}
$$

となる．また，式 (A.49) より式 (A.38) を得る．従って，補題 A.12 より式 (5.7) が成立する．よって，式 (A.50), (3.44), (5.7), (5.8) より，

$$
\begin{aligned}
\mathcal{R}(A) &= \mathbf{C}^M \cap \mathcal{R}(A) = S_t \cap \mathcal{R}(A) \\
&= \{\mathcal{R}(A_s) \dot{+} Q\mathcal{R}(A_s)^{\perp}\} \cap \mathcal{R}(A) \\
&= \mathcal{R}(A_s) \cap \mathcal{R}(A) = \mathcal{R}(A_s)
\end{aligned}
$$

となり，式 (A.46) を得る．従って，補題 A.14 より式 (5.25) を得る．また，式 (5.7) より式 (3.44) の直和分解は直交直和分解になるので，式 (A.46), (A.50) より，

$$
\begin{aligned}
Q\mathcal{R}(A)^{\perp} &= Q\mathcal{R}(A_s)^{\perp} = \mathcal{R}(A_s)^{\perp} \cap S_t \\
&= \mathcal{R}(A_s)^{\perp} \cap \mathbf{C}^M = \mathcal{R}(A_s)^{\perp} = \mathcal{R}(A)^{\perp}
\end{aligned}
$$

となり，式 (5.26) を得る．

逆を示す．式 (5.25), (5.26) が同時に成立しているとき，補題 A.14 より式 (A.46) を得る．従って，式 (3.44), (5.26) より，

$$
S_t = \mathcal{R}(A_s) \dot{+} Q\mathcal{R}(A_s)^{\perp} = \mathcal{R}(A) \dot{+} \mathcal{R}(A)^{\perp} = \mathbf{C}^M
$$

となり，

$$
S_t = \mathbf{C}^M \tag{A.51}
$$

となる．よって，$P_{S_t} = I_M$ となり，式 (3.47) より

$$A^{(\mathrm{SL})} = PA_s^\dagger P_t \tag{A.52}$$

となる．また，式 (5.26), (A.46) と補題 A.12 より式 (A.38) を得る．従って，式 (A.52), (A.28), (A.38), (A.46), (A.51) より，

$$\begin{aligned} AA^{(\mathrm{SL})} &= APA_s^\dagger P_t = P_{\mathcal{R}(A_s)}P_t = P_{\mathcal{R}(A_s)}P_{\mathcal{R}(A)}P_{S_t} \\ &= P_{\mathcal{R}(A)}P_{S_t} = P_{\mathcal{R}(A)} \end{aligned}$$

となるので，式 (2.11) より任意の SL 射影学習作用素が第 1 種記憶学習作用素になる．

一方，式 (5.27) が成立しているとき $A = 0$ となる．従って，式 (2.11) より任意の X が第 1 種記憶学習作用素になり，任意の SL 射影学習作用素も第 1 種記憶学習作用素になる．■

定理 5.6 の証明

SL 射影学習が第 2 種記憶学習を部分的に許容しているとき，式 (3.46) を満たす $A^{(\mathrm{M2})}$ が存在する．この $A^{(\mathrm{M2})}$ に対して，$S_b \supset S_t$ より

$$P_{S_b}P_{S_t} = P_{S_t} \tag{A.53}$$

となるので，式 (A.28), (2.18) より，

$$\begin{aligned} P_t &= P_{\mathcal{R}(A_s)}P_t = APA_s^\dagger P_t = AA^{(\mathrm{M2})}P_{S_t} \\ &= AA^{(\mathrm{M2})}P_{S_b}P_{S_t} = P_{\mathcal{R}(A)}P_{S_t} \end{aligned}$$

となり，式 (A.38) を得る．よって，補題 A.12 より，式 (5.7) を得る．

逆に，式 (5.7) が成立しているとき，補題 A.12 より式 (A.38) を得る．このとき，式 (2.19) において $Y = PA_s^\dagger P_t$ とおけば，式 (A.28), (A.53), (A.38) より，

$$\begin{aligned} A^{(\mathrm{M2})}P_{S_t} &= (A^\dagger + PA_s^\dagger P_t - A^\dagger APA_s^\dagger P_t P_{S_b})P_{S_t} \\ &= A^\dagger P_{S_t} + PA_s^\dagger P_t P_{S_t} - A^\dagger P_{\mathcal{R}(A_s)}P_t P_{S_t} \\ &= A^\dagger P_{S_t} + PA_s^\dagger P_t - A^\dagger P_t \\ &= A^\dagger P_{S_t} + PA_s^\dagger P_t - A^\dagger P_{\mathcal{R}(A)}P_{S_t} = PA_s^\dagger P_t \end{aligned}$$

となるので，式 (3.46) よりこの $A^{(\mathrm{M2})}$ は SL 射影学習作用素になる．■

定理 5.7 の証明

SL 射影学習が第 2 種記憶学習を常に許容するとき，すべての $A^{(\mathrm{M2})}$ が式 (3.46) を満たす．従って，式 (2.19), (A.53) より，

$$\begin{aligned} PA_s^\dagger P_t &= A^{(\mathrm{M2})}P_{S_t} \\ &= (A^\dagger + Y - A^\dagger AYP_{S_b})P_{S_t} \\ &= A^\dagger P_{S_t} + (I - A^\dagger A)YP_{S_t} \end{aligned}$$

となり，任意の Y に対して式 (A.42) が成立する．従って，定理 5.2 の証明と同様に，式 (5.51), (5.50) が同時に成立するか，または式 (5.53), (5.54) が成立する．

逆に，式 (5.51), (5.52) が同時に成立しているとき，式 (2.19), (5.51) より

$$A^{(\mathrm{M2})} = A^{\dagger} + Y(I_M - P_{S_b}) \tag{A.54}$$

となるので，式 (A.53) より $A^{(\mathrm{M2})}P_{S_t} = A^{\dagger}P_{S_t}$ となる．また，式 (5.52) と補題 A.12 より式 (A.38) が成立する．従って，式 (A.45) の導出と同様にして，

$$A^{(\mathrm{M2})}P_{S_t} = A^{\dagger}P_{S_t} = PA_s^{\dagger}P_t$$

となるので，式 (3.46) より任意の第 2 種記憶学習作用素が SL 射影学習作用素になる．

一方，式 (5.53), (5.54) が同時に成立しているとき，定理 5.2 の証明と同様にして，任意の第 2 種記憶学習作用素が SL 射影学習作用素になる． ■

定理 5.8 の証明

SL 射影学習が第 2 種記憶学習に許容されるとき，すべての $A^{(\mathrm{SL})}$ が式 (2.18) を満たす．従って，式 (3.47), (A.28), (A.53) より，

$$\begin{aligned}
P_{\mathcal{R}(A)} &= AA^{(\mathrm{SL})}P_{S_b} \\
&= A\{PA_s^{\dagger}P_t + Y(I_M - P_{S_t})\}P_{S_b} \\
&= APA_s^{\dagger}P_tP_{S_b} + AY(I_M - P_{S_t})P_{S_b} \\
&= P_{\mathcal{R}(A_s)}P_tP_{S_b} + AY(P_{S_b} - P_{S_t}) \\
&= P_tP_{S_b} + AY(P_{S_b} - P_{S_t})
\end{aligned}$$

となり，任意の Y に対して

$$AY(P_{S_b} - P_{S_t}) = P_{\mathcal{R}(A)} - P_tP_{S_b} \tag{A.55}$$

となる．従って，補題 A.13より，

$$A = 0, \quad P_{\mathcal{R}(A)} = P_tP_{S_b} \tag{A.56}$$

が同時に成立するか，または

$$P_{S_b} = P_{S_t}, \quad P_{\mathcal{R}(A)} = P_tP_{S_b} \tag{A.57}$$

が同時に成立する．式 (A.56) の第 1 式が成立すれば，式 (3.42) より $A_s = 0$ となるので，$P_{\mathcal{R}(A)} = P_t = 0$ となり，第 2 式は自動的に成立する．従って，式 (A.56) より $A = 0$ を得る．このとき $\mathcal{N}(A) = H$ となるので，式 (5.59) を得る．更に，$\mathcal{R}(A) = \{0\}$ となるので式 (5.60) を得る．一方，式 (A.57) の第 1 式より，

$$S_t = S_b \tag{A.58}$$

となる．また，式 (A.57) より，

$$P_{\mathcal{R}(A)}P_{S_t} = P_tP_{S_b}P_{S_t} = P_tP_{S_t} = P_t$$

となり，式 (A.38) を得る．従って，補題 A.12 より，式 (5.50) が成立する．よって，式 (2.16)，(A.58)，(3.44)，(5.50)，(5.8) より，

$$\begin{aligned}\mathcal{R}(A) &= S_b \cap \mathcal{R}(A) = S_t \cap \mathcal{R}(A) \\ &= \{\mathcal{R}(A_s) \dot{+} Q\mathcal{R}(A_s)^{\perp}\} \cap \mathcal{R}(A) \\ &= \mathcal{R}(A_s) \cap \mathcal{R}(A) = \mathcal{R}(A_s)\end{aligned}$$

となり，式 (A.46) を得る．従って，補題 A.14より式 (5.59) を得る．また，式 (5.50)，(A.46) より式 (5.60) を得る．

逆を示す．式 (5.59)，(5.60) が同時に成立しているとき，補題 A.14 より式 (A.46) を得る．従って，式 (2.16)，(3.43) より

$$S_b = S_t \tag{A.59}$$

となる．また，式 (5.60)，(A.46) と補題A.12より式 (A.38) を得る．従って，式 (3.47)，(A.59)，(A.28)，(A.38) より，

$$\begin{aligned}AA^{(\mathrm{SL})}P_{S_b} &= A\{PA_s^{\dagger}P_t + Y(I_M - P_{S_t})\}P_{S_b} \\ &= APA_s^{\dagger}P_tP_{S_b} + AY(I_M - P_{S_t})P_{S_t} \\ &= P_tP_{S_b} = P_{\mathcal{R}(A)}P_{S_t}P_{S_b} = P_{\mathcal{R}(A)}\end{aligned}$$

となり，

$$AA^{(\mathrm{SL})}P_{S_b} = P_{\mathcal{R}(A)} \tag{A.60}$$

となるので，式 (2.18) より任意のSL射影学習作用素が第 2 種記憶学習作用素になる．■

付録 B

系の証明

系 3.3 の証明

X が SL 射影学習作用素であるとき，式 (3.46) と P_t の定義より，式 (3.52) は明らかである．逆を示す．S_t の任意の元 u は，式 (3.44) より $\mathcal{R}(A_s)$ の元 $\hat{u}$ と $Q\mathcal{R}(A_s)^{\perp}$ の元 $\tilde{u}$ を用いて，

$$u = \hat{u} + \tilde{u} \tag{B.1}$$

と一意に表すことができる．また，$\hat{u} = P_t u$ である．従って，式 (3.52) が成立しているとき，

$$Xu = X\hat{u} + X\tilde{u} = PA_s^{\dagger}\hat{u} = PA_s^{\dagger}P_t u$$

となる．u は S_t の任意の元であるので，

$$XP_{S_t} = PA_s^{\dagger}P_t P_{S_t} = PA_s^{\dagger}P_t \tag{B.2}$$

となり，X は式 (3.46) を満たす． ■

系 3.4 の証明

$f \in S$ より，$Af = A_s f$ となる．従って，式 (3.55) より，

$$\begin{aligned} PA_s^{\dagger}P_t Af &= PA_s^{\dagger}P_t A_s f = PA_s^{\dagger}A_s f \\ &= PP_S f = Pf \end{aligned}$$

となり，式 (3.59) を得る． ■

系 3.5 の証明

式 (3.60) は，f_{SL} が $\boldsymbol{y}_0$ の A に関する逆元であることから明らかである．次に，$f \in S$ と式 (3.56)，(2.8) より，

$$\boldsymbol{y}_0 = P_t\boldsymbol{y} = P_t(Af + \boldsymbol{n}) = Af + P_t\boldsymbol{n}$$

となり，式 (3.61) を得る． ■

系 5.1 の証明

$I-P_t$ は S_t の中でみれば $\mathcal{R}(A_s)$ に沿った $Q\mathcal{R}(A_s)^\perp$ への射影であるので，式 (3.43) より

$$Q\mathcal{R}(A_s)^\perp=(I-P_t)S_t=(I-P_t)\mathcal{R}(Q)$$

となり，

$$Q\mathcal{R}(A_s)^\perp=(I-P_t)\mathcal{R}(Q) \tag{B.3}$$

となる．従って，式 (5.7) が成立しているとき，

$$(I-P_t)\mathcal{R}(Q)\subset\mathcal{R}(A)^\perp \tag{B.4}$$

となり，$P_{\mathcal{R}(A)}(I-P_t)Q=0$ となる．よって，$\mathcal{R}(A)\supset\mathcal{R}(A_s)$ より，

$$P_tQ=P_{\mathcal{R}(A)}P_tQ=P_{\mathcal{R}(A)}Q$$

となり，式 (5.10) を得る．

逆に，式 (5.10) が成立しているとき，

$$(I-P_t)Q=(I-P_{\mathcal{R}(A)})Q \tag{B.5}$$

となる．従って，式 (B.3) より，

$$\begin{aligned}Q\mathcal{R}(A_s)^\perp&=\mathcal{R}((I-P_t)Q)=\mathcal{R}((I-P_{\mathcal{R}(A)})Q)\\&\subset\mathcal{R}(I-P_{\mathcal{R}(A)})=\mathcal{R}(A)^\perp\end{aligned}$$

となり，式 (5.7) を得る． ■

系 5.2 の証明

式 (5.11) は，式 (5.9) と定理 3.4 より明らかである．式 (5.12) は，式 (3.47) の替わりに式 (5.11) を用いれば，式 (3.59) と同様にして導くことができる．最後に，式 (3.61)，(A.38) より，

$$\begin{aligned}Af_{\mathrm{SL}}&=Af+P_t\boldsymbol{n}\\&=Af+P_{\mathcal{R}(A)}P_{S_t}\boldsymbol{n}\\&=Af+P_{\mathcal{R}(A)}\boldsymbol{n}\end{aligned}$$

となり，式 (5.13) を得る． ■

系 5.3 の証明

式 (5.14)，(5.15) が同時に成立しているとき，式 (A.45) より $PA_s^\dagger P_t=A^\dagger P_{S_t}$ となり，式 (3.47) より式 (5.22) を得る．式 (5.23) は定理 5.2 の証明に示したとおりである．式 (5.24) を示す．$y\in S_t$ であるから，式 (5.22)，(2.8)，(5.14) より，

$$f_{\mathrm{SL}}=A^\dagger\boldsymbol{y}=A^\dagger(Af+\boldsymbol{n})=f+A^\dagger\boldsymbol{n}$$

となり，式 (5.24) を得る． ■

系 5.4 の証明

式 (3.24) より,

$$\begin{aligned}\mathcal{N}(A)\dot{+}L &= [\mathcal{N}(A)+\{S\cap\mathcal{N}(A)\}]\dot{+}L \\ &= \mathcal{N}(A)+S\end{aligned}$$

となり, 式 (5.28) と式 (5.25) は同値になる. 従って, 定理 5.3 より系 5.4 を得る. ■

系 5.5 の証明

式 (5.25), (5.26) が同時に成立しているとき, 系 5.4 より式 (5.28) が成立する. このとき, H の任意の元 f が, L の元 f_1 と $\mathcal{N}(A)$ の元 f_2 に一意に分解できる:

$$f = f_1 + f_2 \tag{B.6}$$

$Af = Af_1$ となるので, 式 (2.8) より,

$$\boldsymbol{y} = Af_1 + \boldsymbol{n} \tag{B.7}$$

となり, $\boldsymbol{y} \in S_t$ となる. 更に, 式 (5.25) と補題 A.14 より式 (A.46) となる. よって, 式 (5.26) より式 (5.9) が成立するので, $A^{(\mathrm{SL})}$ は式 (5.11) となる. 従って, 式 (A.28), (A.46), (2.8), (2.14) より,

$$\begin{aligned}Af_{\mathrm{SL}} &= AA^{(\mathrm{SL})}\boldsymbol{y} = APA_s^{\dagger}\boldsymbol{y} = P_{\mathcal{R}(A_s)}\boldsymbol{y} \\ &= P_{\mathcal{R}(A)}(Af+\boldsymbol{n}) = Af + P_{\mathcal{R}(A)}\boldsymbol{n} = Af_{\mathrm{M}}\end{aligned}$$

となり, 式 (5.13) を得る. ■

系 7.3の証明

式 (7.8) が成立しているとき, 式 (7.3) と式 (7.9) が同値であることを示す. 式 (7.3) が成立しているとき, 式 (7.8) より, $Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(Q) \subset \mathcal{R}(A)$ となるので,

$$Q\mathcal{R}(A_s)^{\perp} \subset \mathcal{R}(A) \cap \mathcal{R}(A)^{\perp} = \{0\}$$

となり, 式 (7.10) を得る. 従って, $\mathcal{N}(Q) \supset \mathcal{R}(A_s)^{\perp}$ となり, 直交補空間をとれば式 (7.9) を得る.

逆に, 式 (7.9) が成立しているとき, その直交補空間をとれば $\mathcal{N}(Q) \supset \mathcal{R}(A_s)^{\perp}$ となる. 従って, 式 (7.10) が成立するので, 式 (7.3) を得る. ■

付録 C

命題と補題の証明

命題 2.1 の証明

式 (2.10) より，

$$J_M[X] = \|AX\boldsymbol{y} - \boldsymbol{y}\|^2 = \|AX\boldsymbol{y} - P_{\mathcal{R}(A)}\boldsymbol{y}\|^2 + \|P_{\mathcal{R}(A)^\perp}\boldsymbol{y}\|^2 \geq \|P_{\mathcal{R}(A)^\perp}\boldsymbol{y}\|^2$$

となるので，$J_M[X]$ が最小のとき，またそのときに限って，

$$AX\boldsymbol{y} = P_{\mathcal{R}(A)}\boldsymbol{y} \tag{C.1}$$

となる．式 (C.1) が $\mathbf{C}^M$ の任意のベクトル $\boldsymbol{y}$ に対して成立することと式 (2.11) は同値になる．.

次に，式 (2.12) を示す．任意の Y に対して式 (2.12) が式 (2.11) を満たすことは明らかである．逆に，式 (2.11) の任意の解を X_0 とすれば，$Y = X_0$ とおくことにより，X_0 を式 (2.12) の形に表現できる．式 (2.13) は，式 (2.9), (2.12) より明らかである．

■

命題 2.2 の証明

式 (C.1) が S_b の任意のベクトル $\boldsymbol{y}$ に対して成立することと，

$$AXP_{S_b} = P_{\mathcal{R}(A)}P_{S_b} \tag{C.2}$$

は同値である．$S_b \supset \mathcal{R}(A)$ より $P_{S_b}P_{\mathcal{R}(A)} = P_{\mathcal{R}(A)}$ となるので，両辺の共役をとれば $P_{\mathcal{R}(A)}P_{S_b} = P_{\mathcal{R}(A)}$ となる．従って，式 (C.2) と式 (2.18) は同値になる．

次に式 (2.19) を示す．任意の Y に対して式 (2.19) が式 (2.18) を満たすことは明らかである．逆に，式 (2.18) の任意の解を X_0 とすれば，$Y = X_0$ とおくことにより，X_0 を式 (2.19) の形に表現できる． ■

補題 3.1 の証明

式 (3.17) が解をもつための必要十分条件は，文献 [46] の補題 1 より，

$$P_{\mathcal{R}(P_S A^*)^\perp} P_S P^* = 0 \tag{C.3}$$

となることである．式（C.3）の両辺の共役をとれば，

$$PP_S P_{\mathcal{R}(P_S A^*)^\perp} = 0 \tag{C.4}$$

となる．式（C.4）は，

$$\mathcal{N}(P) \supset \mathcal{R}(P_S P_{\mathcal{R}(P_S A^*)^\perp}) \tag{C.5}$$

と同値である．一方，文献 [47] の定理 4 より，

$$\mathcal{R}(P_S P_{\mathcal{R}(P_S A^*)^\perp}) = S \cap \mathcal{N}(A) \tag{C.6}$$

となるから，式（C.5）は式（3.18）と同値になる． ■

補題 6.1 の証明

式（6.3）より，$m = 1, 2, \cdots, N$ および $i = 2, 3, \cdots, k$ に対して$\psi_{(i-1)N+m} = \psi_m$ となる．従って，式（2.4）より，

$$A_{kN} = \sum_{m=1}^{N} \left\{ (\sum_{i=1}^{k} e_{(i-1)N+m}^{(kN)}) \otimes \overline{\psi_m} \right\} \tag{C.7}$$

となる．ここで，$\{e_m^{(kN)}\}_{m=1}^{kN}$ は $\mathbf{C}^{kN}$ の標準基底である．また，$\{\psi_m\}_{m=1}^{N}$ が 1 次独立であるので，式（C.7）より，

$$\mathcal{R}(A_{kN}) = \mathcal{L}\{\{\sum_{i=1}^{k} e_{(i-1)N+m}^{(kN)}\}_{m=1}^{N}\} \tag{C.8}$$

となる．従って，式（C.8），(6.5) より，

$$\begin{aligned} Q_{kN}\mathcal{R}(A_{kN}) &= Q_{kN}\mathcal{L}\{\{\sum_{i=1}^{k} e_{(i-1)N+m}^{(kN)}\}_{m=1}^{N}\} \\ &= \mathcal{L}\{\{\sigma_m \sum_{i=1}^{k} e_{(i-1)N+m}^{(kN)}\}_{m=1}^{N}\} \\ &= \mathcal{L}\{\{\sum_{i=1}^{k} e_{(i-1)N+m}^{(kN)}\}_{m=1}^{N}\} \\ &= \mathcal{R}(A_{kN}) \end{aligned}$$

となり，

$$Q_{kN}\mathcal{R}(A_{kN}) = \mathcal{R}(A_{kN}) \tag{C.9}$$

となる．よって，$\mathcal{R}(A_{kN})$ への正射影行列を $P_{\mathcal{R}(A_{kN})}$ で表せば，

$$P_{\mathcal{R}(A_{kN})} Q P_{\mathcal{R}(A_{kN})} = Q P_{\mathcal{R}(A_{kN})} \tag{C.10}$$

となる．Q が自己共役であるので，式（C.10）より

$$(I - P_{\mathcal{R}(A_{kN})}) Q (I - P_{\mathcal{R}(A_{kN})}) = Q(I - P_{\mathcal{R}(A_{kN})}) \tag{C.11}$$

となり，式（6.6）を得る． ■

補題 A.1 の証明

X が式 (3.6) を満たしているとする．f が $\mathcal{R}(R_F^2A^*)$ の元であるとき，$f = R_F^2A^*u$ を満たす u が存在する．従って，式 (3.6) より，

$$
\begin{aligned}
XAf &= XAR_F^2A^*u = (XAR_F)R_FA^*u \\
&= R_FP_{\mathcal{R}(R_FA^*)}R_FA^*u = R_F^2A^*u = f
\end{aligned}
$$

となり，

$$XAf = f$$

となる．また，f が $\overline{\mathcal{R}(R_F^2)} \cap \mathcal{N}(A)$ の元であるとき，$Af = 0$ であるから，

$$XAf = 0$$

となる．従って，式 (A.2) を得る．

逆に，X が式 (A.2) を満たしているとき，$\overline{\mathcal{R}(R_F^2)} \supset \mathcal{R}(R_F^2A^*)$ であるから，

$$
\begin{aligned}
XAR_F &= XA(R_F^2A^*)(R_FA^*)^\dagger \\
&= XA(P_{\overline{\mathcal{R}(R_F^2)}}R_F^2A^*)(R_FA^*)^\dagger \\
&= (P_{RF}P_{\overline{\mathcal{R}(R_F^2)}})R_F^2A^*(R_FA^*)^\dagger \\
&= (R_F^2A^*)(R_FA^*)^\dagger = R_FP_{\mathcal{R}(R_FA^*)}
\end{aligned}
$$

となり，式 (3.6) を得る． ■

補題 A.2 の証明

補題 A.1 より，式 (A.2) と式 (3.6) は同値であるから，ある作用素が R_F射影学習作用素であることと，条件 (3.6) のもとで式 (3.26) を最小にすることが同値であることを示す．まず，式 (3.26) は

$$J_n[X] = \langle XQ, X\rangle \tag{C.12}$$

となる [46]．作用素 X_0 が R_F射影学習作用素であるとき，式 (3.6) を満たす X に対して，

$$
\begin{aligned}
\langle X_0Q, X_0\rangle &= \langle CR_FA^*, X_0\rangle = \langle C, X_0AR_F\rangle \\
&= \langle C, XAR_F\rangle = \langle CR_FA^*, X\rangle = \langle X_0Q, X\rangle
\end{aligned}
$$

となり，

$$\langle X_0Q, X_0\rangle = \langle X_0Q, X\rangle$$

となる．Q は自己共役であるから $\langle X_0Q, X_0\rangle$ は実数である．従って $\langle X_0Q, X\rangle$ も実数になり，

$$\langle X_0Q, X\rangle = \langle XQ, X_0\rangle$$

となる．よって，

$$
\begin{aligned}
J_n[X] - J_n[X_0] &= \langle XQ, X\rangle - \langle X_0Q, X_0\rangle \\
&= \langle (X - X_0)Q, X - X_0\rangle \geq 0
\end{aligned} \tag{C.13}
$$

となり，

$$J_n[X] \geq J_n[X_0]$$

となる．従って，X_0 は条件（3.6）のもとで式（3.26）を最小にする．

逆を示す．作用素 X が条件（3.6）のもとで式（3.26）を最小にしているとき，ある R_F射影学習作用素 X_0 に対して式（C.13）で等号が成立し，

$$(X - X_0)Q = 0$$

となる．従って，式（3.7）より，

$$XQ = X_0 Q = CR_F A^*$$

となり，X は C と共に式（3.7）を満たす． ■

補題 A.3の証明

S_1, S_2 は閉部分空間であるから，式（A.3）と

$$S_1^\perp + S_2 = H \tag{C.14}$$

は同値である．そこで，式（C.14）と式（A.4）が同値であることを示す．式（C.14）が成立しているとき，

$$S_1 = P_{S_1} H = P_{S_1}(S_1^\perp + S_2) = P_{S_1} S_2$$

となり，式（A.4）を得る．逆に，式（A.4）が成立しているとき，

$$\begin{aligned} H &= S_1^\perp + S_1 = S_1^\perp + P_{S_1} S_2 \\ &= S_1^\perp + (I - P_{S_1^\perp}) S_2 \\ &\subset S_1^\perp + S_2 + P_{S_1^\perp} S_2 = S_1^\perp + S_2 \subset \mathrm{H} \end{aligned}$$

となり，式（C.14）を得る． ■

補題 A.4 の証明

式（3.43）より，

$$\begin{aligned} S_t &= \mathcal{R}(A_s) + \mathcal{R}(Q) = \mathcal{R}(P_{\mathcal{R}(A_s)}) + (P_{\mathcal{R}(A_s)} + P_{\mathcal{R}(As)^\perp})\mathcal{R}(Q) \\ &\subset \mathcal{R}(P_{\mathcal{R}(A_s)}) \oplus \mathcal{R}(P_{\mathcal{R}(As)^\perp} Q) = \mathcal{R}(P_{\mathcal{R}(A_s)}) \oplus (I - P_{\mathcal{R}(A_s)})\mathcal{R}(Q) \\ &\subset \mathcal{R}(P_{\mathcal{R}(A_s)}) + \mathcal{R}(Q) = S_t \end{aligned}$$

となり，

$$S_t = \mathcal{R}(P_{\mathcal{R}(A_s)}) \oplus \mathcal{R}(P_{\mathcal{R}(As)^\perp} Q) \tag{C.15}$$

となる．また，

$$\begin{aligned} &\mathcal{R}(P_{\mathcal{R}(As)^\perp} Q P_{\mathcal{R}(As)^\perp}) = \mathcal{R}((P_{\mathcal{R}(As)^\perp} Q^{1/2})(P_{\mathcal{R}(As)^\perp} Q^{1/2})^*) \\ &= \mathcal{R}(P_{\mathcal{R}(As)^\perp} Q^{1/2}) \supset \mathcal{R}(P_{\mathcal{R}(As)^\perp} Q) \supset \mathcal{R}(P_{\mathcal{R}(As)^\perp} Q P_{\mathcal{R}(As)^\perp}) \end{aligned}$$

となるので，

$$\mathcal{R}(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp}) = \mathcal{R}(P_{\mathcal{R}(As)^\perp}Q) \tag{C.16}$$

となる．よって，式（C.15），(A.10）より，

$$S_t = \mathcal{R}(P_{\mathcal{R}(A_s)}) \oplus \mathcal{R}(P_1) \tag{C.17}$$

を得る．P_{S_t} は S_t への正射影行列であるから，式（C.17）より式（A.12）を得る．また，式（C.17）より $S_t \supset \mathcal{R}(P_1)$ であるから式（A.13）を得る． ■

補題 A.5 の証明

簡単のために $F = P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp}$ とおく．まず，$QF^\dagger$ が射影行列であることを示す．文献 [2] より，一般の作用素 B と正射影作用素 C に対して

$$C(BC)^\dagger = (BC)^\dagger, \quad (CB)^\dagger C = (CB)^\dagger \tag{C.18}$$

が成立するので，

$$P_{\mathcal{R}(As)^\perp}F^\dagger = F^\dagger = F^\dagger P_{\mathcal{R}(As)^\perp} \tag{C.19}$$

を得る．従って，

$$\begin{aligned}(QF^\dagger)^2 &= QF^\dagger QF^\dagger = QF^\dagger P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp}F^\dagger \\ &= QF^\dagger FF^\dagger = QF^\dagger\end{aligned}$$

となり，

$$(QF^\dagger)^2 = QF^\dagger \tag{C.20}$$

となる．また，式（C.16）より，

$$\begin{aligned}\mathcal{R}(QF^\dagger) &= Q\mathcal{R}(F^\dagger) = Q\mathcal{R}(F^*) = Q\mathcal{R}(F) \\ &= Q\mathcal{R}(P_{\mathcal{R}(As)^\perp}Q) = \mathcal{R}(QP_{\mathcal{R}(As)^\perp}(QP_{\mathcal{R}(As)^\perp})^*) \\ &= \mathcal{R}(QP_{\mathcal{R}(As)^\perp}) = Q\mathcal{R}(A_s)^\perp\end{aligned}$$

となり，

$$\mathcal{R}(QF^\dagger) = Q\mathcal{R}(A_s)^\perp \tag{C.21}$$

となる．更に，式（C.19），(C.16）より，

$$\begin{aligned}\mathcal{R}(F^\dagger Q) &= \mathcal{R}(F^\dagger P_{\mathcal{R}(As)^\perp}Q) = F^\dagger\mathcal{R}(P_{\mathcal{R}(As)^\perp}Q) \\ &= F^\dagger\mathcal{R}(F) = \mathcal{R}(F^\dagger F) = \mathcal{R}(F^*) \\ &= \mathcal{R}(F) = \mathcal{R}(P_{\mathcal{R}(As)^\perp}Q) = \mathcal{N}(QP_{\mathcal{R}(As)^\perp})^\perp\end{aligned}$$

となるから，$\mathcal{N}(QF^\dagger) = \mathcal{R}(F^\dagger Q)^\perp = \mathcal{N}(QP_{\mathcal{R}(As)^\perp})$ となり，

$$\mathcal{N}(QF^\dagger) = \mathcal{N}(QP_{\mathcal{R}(As)^\perp}) \tag{C.22}$$

となる．従って，式 (C.20),(C.21),(C.22) より，$QF^\dagger$ は $\mathcal{N}(QP_{\mathcal{R}(As)^\perp})$ に沿った $Q\mathcal{R}(A_s)^\perp$ への射影行列になる．よって，補題 A.5 の前半が成立する．

式 (A.14) を示す．まず，前半の証明より，

$$\mathcal{R}(P_2) = \mathcal{N}(QP_{\mathcal{R}(As)^\perp}) \supset \mathcal{N}(P_{\mathcal{R}(As)^\perp}) = \mathcal{R}(P_{\mathcal{R}(A_s)})$$

となり，

$$\mathcal{R}(P_2) \supset \mathcal{R}(P_{\mathcal{R}(A_s)})$$

となる．また，式 (3.43) より $\mathcal{R}(A_s) \subset S_t$ となる．よって，ベクトル u が $\mathcal{R}(A_s)$ の元であるとき

$$P_2 P_{S_t} u = P_2 u = u$$

となり，

$$P_2 P_{S_t} u = u \tag{C.23}$$

となる．次に，u が $Q\mathcal{R}(A_s)^\perp$ の元であるとき，u は S_t の元でもあるので，

$$P_2 P_{S_t} u = P_2 u = 0$$

となり，

$$P_2 P_{S_t} u = 0 \tag{C.24}$$

となる．u が $S_t^\perp$ の元であるとき，$P_{S_t} u = 0$ となるので，式 (C.24) が成立する．従って，式 (3.45)，(C.23)，(C.24) より式 (A.14) を得る．■

補題 A.6 の証明

式 (3.42) より，式 (3.25) と式 (A.15) は同値である．従って，作用素 X が SL 射影学習作用素であることと，X が条件 (A.15) のもとで式 (3.26) を最小にすることとは同値である．このことが，式 (A.15)，(A.16) を満たすことと同値であることを示す．

式 (A.15)，(A.16) を満たす X を X_0 で表す．式 (A.15) を満たす X に対して，

$$\begin{aligned}\langle X_0 Q, X_0 \rangle &= \langle CA_s^*, X_0 \rangle = \langle C, X_0 A_s \rangle \\ &= \langle C, XA_s \rangle = \langle CA_s^*, X \rangle = \langle X_0 Q, X \rangle\end{aligned}$$

となり

$$\langle X_0 Q, X_0 \rangle = \langle X_0 Q, X \rangle$$

となる．Q は半正値自己共役であるから，$\langle X_0 Q, X_0 \rangle$ は実数である．従って，$\langle X_0 Q, X \rangle$ も実数になり，

$$\langle X_0 Q, X \rangle = \langle XQ, X_0 \rangle$$

となる．従って，式 (C.13) と同様にして，

$$J_n[X] \geq J_n[X_0]$$

となり，X_0 は条件 (A.15) のもとで式 (3.26) を最小にする．

逆を示す．X が条件 (A.15) のもとで式 (3.26) を最小にしているとき，式 (A.15)，(A.16) を満たすある X_0 に対して，式 (C.13) で等号が成立し，

$$(X - X_0)Q = 0$$

となる．従って，式 (A.16) より

$$XQ = X_0 Q = CA_s^*$$

となり，X は C と共に式 (A.16) を満たす．■

補題 A.7の証明

式 (A.16) の C が存在するための必要十分条件は,

$$XQA_sA_s^\dagger = XQ \tag{C.25}$$

が成立することである [47]. $P_{\mathcal{R}(A_s)} = A_sA_s^\dagger$ より

$$XQP_{\mathcal{R}(As)^\perp} = XQ(I - P_{\mathcal{R}(A_s)}) = XQ(I - A_sA_s^\dagger)$$

であるから, 式 (C.25) と式 (A.17) は同値になる. ■

補題 A.8の証明

X を SL 射影学習作用素とすると, 補題 A.6 より式 (A.15), (A.16) が成立する. 式 (C.18) の第 1 式より

$$P_SA_s^\dagger = A_s^\dagger \tag{C.26}$$

が成立するので, 式 (A.15) より

$$XP_{\mathcal{R}(A_s)} = XA_sA_s^\dagger = PP_SA_s^\dagger = PA_s^\dagger$$

となり, 式 (A.18) を得る. 一方, 式 (A.16) と補題 A.7 より式 (A.17) が成立するので, 式 (A.10), (C.19), (A.17), (A.18) より,

$$\begin{aligned}
XP_1 &= X(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger \\
&= X(I - P_{\mathcal{R}(A_s)})Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger \\
&= -XP_{\mathcal{R}(A_s)}Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger \\
&= -PA_s^\dagger Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger
\end{aligned}$$

となり, 式 (A.19) を得る.

逆に, 式 (A.18), (A.19) が成立しているとき, 式 (A.18), (3.55) より,

$$XA_s = XP_{\mathcal{R}(A_s)}A_s = PA_s^\dagger A_s = PP_S$$

となり, 式 (A.15) を得る. また, 式 (A.10), (C.16) より $P_1P_{\mathcal{R}(As)^\perp}Q = P_{\mathcal{R}(As)^\perp}Q$ となるので, 両辺の共役をとれば

$$QP_{\mathcal{R}(As)^\perp}P_1 = QP_{\mathcal{R}(As)^\perp} \tag{C.27}$$

となる. 従って, 式 (A.18), (A.10), (A.19) より,

$$\begin{aligned}
XQP_{\mathcal{R}(As)^\perp} &= XP_{\mathcal{R}(A_s)}QP_{\mathcal{R}(As)^\perp} + XP_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp} \\
&= PA_s^\dagger QP_{\mathcal{R}(As)^\perp} + XP_1(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp}) \\
&= PA_s^\dagger QP_{\mathcal{R}(As)^\perp} - PA_s^\dagger Q(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp}) \\
&= PA_s^\dagger QP_{\mathcal{R}(As)^\perp} - PA_s^\dagger QP_{\mathcal{R}(As)^\perp}(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})(P_{\mathcal{R}(As)^\perp}QP_{\mathcal{R}(As)^\perp})^\dagger \\
&= PA_s^\dagger QP_{\mathcal{R}(As)^\perp} - PA_s^\dagger QP_{\mathcal{R}(As)^\perp}P_1 = 0
\end{aligned}$$

となり, 式 (A.17) を得る. よって, 補題 A.7 より式 (A.16) を得る. ■

補題 A.9 の証明

f が L の元であるとき，f は S の元でもあるので，

$$P_e f = PP_S f = Pf = f$$

となり，$P_e f = f$ を得る．また，f が $\{S \cap \mathcal{N}(A)\} \oplus S^{\perp}$ の元であるとき，$S \cap \mathcal{N}(A)$ の元 f_1 と $S^{\perp}$ の元 f_2 を使って，$f = f_1 + f_2$ と表すことができる．従って，

$$P_e f = PP_S f = PP_S(f_1 + f_2) = Pf_1 = 0$$

となり，$P_e f = 0$ となる．よって，式（A.30）より補題 A.9 を得る． ■

補題 A.10の証明

式（A.32）の証明のために，いくつかの準備をする．まず $S \supset L$ より $\mathcal{R}(P_S) \supset \mathcal{R}(P_e)$ となるから，$P_S P_e = P_e$ となる．また，$\mathcal{N}(P_S) \subset \mathcal{N}(P_e)$ と文献 [46] の補題 1 より $P_e P_S^{\dagger} P_S = P_e$ となるので，$P_e P_S = P_e$ となる．従って，

$$P_e = P_S P_e P_S \tag{C.28}$$

を得る．式（A.31）より，

$$P_S = P_{\mathcal{R}(A_s^*)} + P_\alpha \tag{C.29}$$

となる．補題 A.9 より $\mathcal{N}(P_e) \supset S \cap \mathcal{N}(A) = \mathcal{R}(P_\alpha)$ であるので，

$$P_e P_\alpha = 0 \tag{C.30}$$

が成立する．再び補題 A.9 より

$$\mathcal{N}(P_{\mathcal{R}(A_s^*)}) = \{S \cap \mathcal{N}(A)\} \oplus S^{\perp} = \mathcal{N}(P_e) = \mathcal{R}(I - P_e)$$

となり $P_{\mathcal{R}(A_s^*)}(I - P_e) = 0$ となるので，

$$P_{\mathcal{R}(A_s^*)} P_e = P_{\mathcal{R}(A_s^*)} \tag{C.31}$$

となる．式（C.28）～ 式（C.31）より，

$$\begin{aligned} P_e &= P_S P_e P_S = (P_{\mathcal{R}(A_s^*)} + P_\alpha) P_e (P_{\mathcal{R}(A_s^*)} + P_\alpha) \\ &= P_{\mathcal{R}(A_s^*)} P_e P_{\mathcal{R}(A_s^*)} + P_\alpha P_e P_{\mathcal{R}(A_s^*)} = (I + P_\alpha P_e) P_{\mathcal{R}(A_s^*)} \end{aligned}$$

となり，式（A.32）を得る． ■

補題 A.11の証明

まず，J_{n0} の最小値を求める．$P_{\mathcal{R}(A_s^*)} = A_s^{\dagger} A_s$ より $P_{\mathcal{R}(A_s^*)} A_s^{\dagger} = A_s^{\dagger}$ となる．P_α は $S \cap \mathcal{N}(A)$ への正射影作用素であるから，式（4.4）より $P_\alpha A_s^{\dagger} = 0$ となる．よって，式

(A.25), (A.32) より，

$$\begin{aligned} J_{n0} &= \langle P_e A_s^\dagger P_t Q, P_e A_s^\dagger P_t \rangle \\ &= \langle (I + P_\alpha P_e) P_{\mathcal{R}(A_s^*)} A_s^\dagger P_t Q, (I + P_\alpha P_e) P_{\mathcal{R}(A_s^*)} A_s^\dagger P_t \rangle \\ &= \langle (I + P_\alpha P_e) A_s^\dagger P_t Q, (I + P_\alpha P_e) A_s^\dagger P_t \rangle \\ &= \langle A_s^\dagger P_t Q, A_s^\dagger P_t \rangle + \langle P_\alpha P_e A_s^\dagger P_t Q, P_\alpha P_e A_s^\dagger P_t \rangle \end{aligned} \tag{C.32}$$

となる．Q は半正値であるから，

$$\langle P_\alpha P_e A_s^\dagger P_t Q, P_\alpha P_e A_s^\dagger P_t \rangle \geq 0 \tag{C.33}$$

となる．等号は，

$$P_\alpha P_e A_s^\dagger P_t Q = 0 \tag{C.34}$$

のとき，またそのときに限って成立する．例えば，$L = \mathcal{R}(A_s^*)$ とおけば $P_e = P_{\mathcal{R}(A_s^*)}$ となり，式 (C.34) は確かに成立する．こうして，SL 射影学習が最小分散射影学習になるための必要十分条件が式 (C.34) で与えられることがわかった．$P_\gamma = (A_s^\dagger P_t Q)(A_s^\dagger P_t Q)^\dagger$ より式 (C.34) と式 (A.33) は同値であるから，補題 A.11 の前半が成立する．

式 (4.3) を示す．ここまでの証明より，J_{n0} の最小値は

$$\min J_{n0} = \langle A_s^\dagger P_t Q, A_s^\dagger P_t \rangle \tag{C.35}$$

で与えられる．$I - A_s A_s^\dagger$ は $\mathcal{R}(A_s)^\perp$ への正射影行列であるので，

$$\mathcal{N}(P_t) \supset Q\mathcal{R}(A_s)^\perp = \mathcal{R}(Q(I - A_s A_s^\dagger))$$

となり，

$$P_t Q(I - A_s A_s^\dagger) = 0 \tag{C.36}$$

となる．従って，式 (C.35), (C.36) と $P_t A_s = A_s$ より

$$\begin{aligned} \min J_{n0} &= \langle A_s^\dagger P_t Q, A_s^\dagger P_t \rangle \\ &= \langle A_s^\dagger P_t Q A_s A_s^\dagger, A_s^\dagger P_t \rangle \\ &= \langle A_s^\dagger P_t Q, A_s^\dagger P_t A_s A_s^\dagger \rangle \\ &= \langle A_s^\dagger P_t Q, A_s^\dagger A_s A_s^\dagger \rangle \\ &= \langle A_s^\dagger P_t Q, A_s^\dagger \rangle \end{aligned}$$

となり，式 (4.3) を得る． ■

補題 A.12 の証明

式 (5.7) が成立しているとき，$P_{S_t} - P_t$ は $\mathcal{R}(A_s) \oplus S_t^\perp$ に沿った $Q\mathcal{R}(A_s)^\perp$ への射影作用素であるので，

$$\mathcal{N}(P_{\mathcal{R}(A)}) = \mathcal{R}(A)^\perp \supset Q\mathcal{R}(A_s)^\perp = \mathcal{R}(P_{S_t} - P_t)$$

となり，$P_{\mathcal{R}(A)}(P_{S_t} - P_t) = 0$ となる．従って，$\mathcal{R}(A) \supset \mathcal{R}(A_s)$ より，

$$P_{\mathcal{R}(A)} P_{S_t} = P_{\mathcal{R}(A)} P_t = P_t$$

となり，式 (A.38) を得る．

逆を示す．まず一般に，

$$\mathcal{N}(P_{\mathcal{R}(A)}P_{S_t}) \cap S_t \subset \mathcal{R}(A)^{\perp} \tag{C.37}$$

が成立することを示す．$u \in \mathcal{N}(P_{\mathcal{R}(A)}P_{S_t}) \cap S_t$ に対して $P_{\mathcal{R}(A)}P_{S_t}u = 0$, $P_{S_t}u = u$ が成立する．従って，

$$P_{\mathcal{R}(A)}u = P_{\mathcal{R}(A)}P_{S_t}u = 0$$

となり，$u \in \mathcal{N}(P_{\mathcal{R}(A)}) = \mathcal{R}(A)^{\perp}$ となるので，式 (C.37) を得る．よって，式 (A.38) が成立しているとき，

$$\begin{aligned} Q\mathcal{R}(A_s)^{\perp} &= \mathcal{N}(P_t) \cap S_t \\ &= \mathcal{N}(P_{\mathcal{R}(A)}P_{S_t}) \cap S_t \subset \mathcal{R}(A)^{\perp} \end{aligned}$$

となり，式 (5.7) を得る． ■

補題 A.13の証明

式 (A.39) が任意の Y に対して成立しているとき，$Y = 0$ を代入すれば $D = 0$ となる．従って，式 (A.39) は

$$BYC = 0 \tag{C.38}$$

となる．よって，任意の Y に対して

$$\mathcal{N}(B) \supset \mathcal{R}(YC) \tag{C.39}$$

となる．$C \neq 0$ であれば $\mathcal{R}(YC)$ はどのような部分空間にもなる．このとき $\mathcal{N}(B)$ は全空間となり，$B = 0$ となる．従って，式 (A.40) を得る．$C = 0$ であれば，式 (A.41) を得る． ■

補題 A.14 の証明

式 (5.25) の直交補空間をとれば，

$$\mathcal{R}(A^*) \cap S^{\perp} = \{0\} \tag{C.40}$$

となる．補題 A.3 より，式 (C.40) は

$$P_{\mathcal{R}(A^*)}S = \mathcal{R}(A^*) \tag{C.41}$$

と同値である．そこで，式 (A.46) と式 (C.41) が同値であることを示す．まず，式 (C.41) に A を作用させれば式 (A.46) を得る．逆に，式 (A.46) に $A^{\dagger}$ を作用させれば，$A^{\dagger}A = P_{\mathcal{R}(A^*)}$ であるので，式 (C.41) を得る． ■