

非線形適応信号処理技術の新潮流 ——再生核の応用——

A New Stream of Nonlinear Adaptive Signal Processing Technique :
An Application of Reproducing Kernel

湯川正裕

Abstract

最小平均二乗法 (Widrow-Hoff '60) の提案から半世紀を経て、適応アルゴリズムの研究は様々な方向へ進展している。本稿では、非線形適応信号処理技術の新潮流として注目される再生核適応フィルタに焦点を当てる。幾何学的視点から、再生核適応アルゴリズムを「出力誤差を零にするフィルタ集合 (超平面) への逐次射影法」として統一的に説明する。筆者が提唱する「多核適応フィルタ」と「辞書の適応的精練技術」のアイデアを解説し、応用事例を含めた最新動向に触れる。

キーワード：非線形適応フィルタ、再生核、反復解法、直交射影

1. はじめに

サポートベクトルマシン (再生核を用いた非線形識別器) の成功を火種に、パターン認識・機械学習・信号処理分野を中心として、再生核の理論⁽¹⁾に基づく非線形手法が盛んに研究され、非線形識別問題のみならず、非線形回帰問題・非線形関数推定などにおける有用性が広く実証されてきた^{(2)~(5)}。IEEE Signal Processing Magazine 2013年9月号に再生核を利用した信号処理研究の特集が掲載されている。本稿では、再生核に基づく「非線形適応信号処理技術」に焦点を当てる^(注1)。ここで、まずは三つの代表的な質問に対する簡単な回答を与えたい。

(1) 「非線形にすると何がうれしいの？」

一言で言うと、線形フィルタの性能限界 (例：定常入力信号に対するウィナーフィルタの性能) を超える性能が期待できる。例えば、携帯電話などでは伝搬経路特

性の非線形性が強いので、線形フィルタでは効果的な音響エコー消去が実現できない。このような場合、非線形フィルタを使用することで通話品質の大幅な改善が期待できる⁽⁶⁾。

(2) 「再生核とは？ それを使うと何がうれしいの？」

非線形フィルタを実現する上での困難性の根源は数理モデルの選択にある。どのような数理モデルで非線形フィルタを設計すればよいかは、一般に自明でない。しかし、幸いにも、再生核の一種であるガウス核を用いると、あらゆる連続関数を高精度で表現できることが知られている⁽⁷⁾ (ガウス核の普遍性^(用語)という)。

(3) 「非線形適応信号処理にはどんな応用があるの？」

非線形音響エコーキャンセラのほか、時系列予測 (気象データ、金融データほか)、生体信号処理 (ブレイン・マシンインタフェースほか) など、無限の可能性がある。応用分野の開拓は、これから楽しみな研究課題である。

再生核に基づく非線形適応信号処理 (非線形オンライン学習) の研究は、今世紀初頭に Kivinen らと Engel らによって (独立に) 発表された論文^{(8),(9)} をきっかけに開始された。それから10年が経過し、アルゴリズムの高性能化から性能解析まで、様々な進展が見られ

(注1) 再生核適応フィルタは、ニューラルネットワークのような局所最適性の問題 (非凸性に起因) を持たず、ボルテラフィルタと比べて低計算量で実現できる点が主たる利点とされている。

る^{(10),(11)}。Kivinenらの手法⁽⁸⁾とRichardらの手法⁽¹²⁾は再生核適応アルゴリズム^(用語)の典型例であるが、これら二つの手法の関係性は国際コミュニティでも広く認識されていない。本稿では、初めに、この二つの手法を紹介し、その関係性を幾何学的視点から明らかにする。次に、「どうやって良い再生核を事前に設計できるか?」という素朴な疑問から始まった「多核適応フィルタ」と「辞書の適応的精練技術」の研究⁽¹³⁾を解説し、最新動向に触れる。

2. 準備

再生核の実体をイメージして頂くために、その輪郭に少しだけ触れる。 U をユークリッド空間 \mathbb{R}^L の部分集合とし、 U 上の実数値関数 $f:U\rightarrow\mathbb{R}$ を考えよう。3.以降の話を理解して頂くためには、関数 f をある空間の点(ベクトル)と見る、関数解析^(用語)^{(14),(15)}の視点を理解しておくことが望ましい。このために、各点 $\mathbf{x}\in U$ にお

用語解説

ガウス核の普遍性 任意のコンパクトな距離空間 U に対して、ガウス核と U で規定される再生核ヒルベルト空間は、 U 上の連続関数全体で構成される(上限ノルムを具備した)ノルム空間で稠密である⁽⁷⁾。大まかに言えば、どんな連続関数も、 κ をガウス核としたときの関数空間 \mathcal{H} の元により「幾らでも精度良く」近似できることを意味する。

適応アルゴリズム 未知のシステム(ブラックボックス)の入出力が時間の経過とともに順次、観測される状況において、各時刻における観測データを基に未知システム(典型例は線形システム)を逐次的に近似(推定)するアルゴリズムを適応アルゴリズムという。オンライン学習アルゴリズムと同種のものともできるが、適応アルゴリズムは、一般に、未知システムの時間変化に対して瞬時に適応できるものが望まれる。

関数解析 線形代数で扱うベクトル空間や内積、点列の収束や連続性などの概念を関数に拡張して議論するための数理体系を与えるのが関数解析であり、線形関数解析と非線形関数解析に大きく分類される。非線形関数解析の中心的テーマである凸解析は、信号処理や最適化を含む工学的応用において特に重要な役割を演じる⁽¹⁴⁾。

一次独立性 ベクトル空間 X の部分集合 $S:=\{x_1, x_2, \dots, x_n\}\subset X$ があったとき、各々のベクトルが他のベクトルの線形結合で表現できないとき、集合 S は線形独立であるという。 $a_1x_1+a_2x_2+\dots+a_nx_n=\theta\Rightarrow a_1=a_2=\dots=a_n=0$ が成り立つときと言い換えることができる。

超平面 実ベクトル空間 X に内積 $\langle \cdot, \cdot \rangle_X$ が定義されているものとする。(本稿で登場する関数空間 \mathcal{H} やユークリッド空間 \mathbb{R}^{n+1} が X に相当する。) 零元でない $x\in X$ が与えられたとき、 $H:=\{y\in X|\langle x, y \rangle_X=0\}$ は x と直交するベクトル全体の集合となる。 X が二次元平面であれば、 H は一次元の直線であり、 X が三次元空間であれば、 H は二次元の平面、一般に X が n 次元空間であれば、 H は $n-1$ 次元の超平面となる。更に、実数 c が与えられたとき、集合 $H_c:=\{y\in X|\langle x, y \rangle_X=c\}$ は H を平行移動した集合となり、これを一般に超平面という。

る値 $f(\mathbf{x})$ (微視的な挙動)からしばらく目を離し、関数のグラフ的なイメージを忘れて頂きたい。その代わりに、関数 f 自身の大域的な振舞い(巨視的な挙動)を見ていく。関数解析は、線形代数と同様の議論を関数へ拡張することを可能にする。具体的には、 U 上の関数 $f:U\rightarrow\mathbb{R}$ をベクトルとする線形空間 \mathcal{H} において、関数 $f, g\in\mathcal{H}$ の内積 $\langle f, g \rangle\in\mathbb{R}$ を定義する(零元を $\theta\in\mathcal{H}$ で表す)。内積は、関数の類似度を測る道具である。関数 $f, g\in\mathcal{H}$ 間の距離は、ノルム $\|f\|:=\sqrt{\langle f, f \rangle}$ を使って $\|f-g\|$ で測ることができる。

1. で述べたガウス核は、

$$\kappa(\mathbf{x}, \mathbf{y}) := \exp\left(-\frac{(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})}{2\sigma^2}\right)$$

$(\mathbf{x}, \mathbf{y}\in U, \sigma>0)$

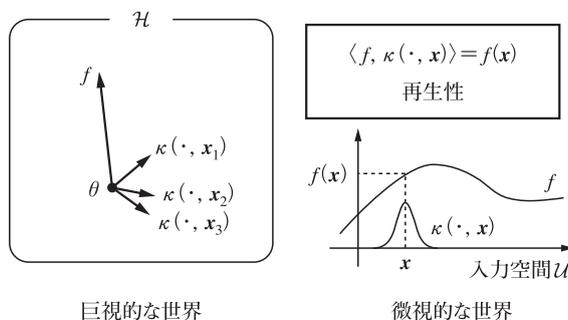


図1 再生性の概念図 再生核 κ の恩恵により、関数 f をベクトル空間上の点として扱うことで見えなくなった微視的な挙動、すなわち、ユークリッド空間上の各点 \mathbf{x} における値 $f(\mathbf{x})$ が内積 $\langle f, \kappa(\cdot, \mathbf{x}) \rangle = f(\mathbf{x})$ によって再現される。

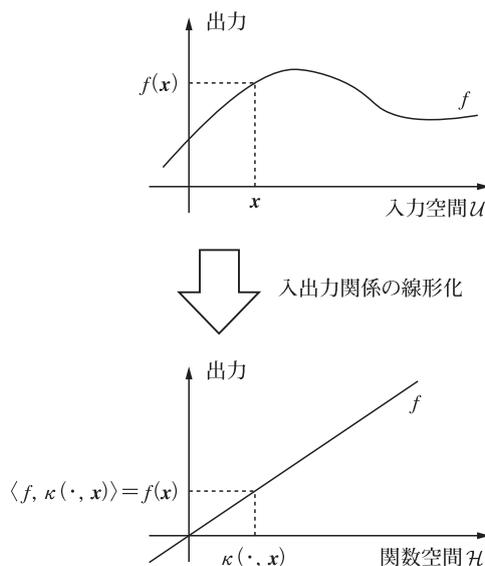


図2 再生核による入出力関係の線形化 再生核を使ってデータ \mathbf{x} を関数空間のベクトル $\kappa(\cdot, \mathbf{x})$ に写像することで、非線形な入出力関係を線形化することができる。

で与えられる。 $\kappa(\mathbf{x}, \mathbf{y})$ は、 $\mathbf{y} \in \mathcal{U}$ を任意に固定した定数ベクトルとすることで、変数 $\mathbf{x} \in \mathcal{U}$ の関数とみることができる。厳密には $\kappa(\mathbf{x}, \mathbf{y})$ は実数であり、これを関数と呼ぶと誤解が生じるため、 $\kappa(\mathbf{x}, \cdot)$ または $\kappa(\cdot, \mathbf{x})$ と表記するのが慣例である（対称性 $\kappa(\mathbf{x}, \mathbf{y}) = \kappa(\mathbf{y}, \mathbf{x})$ に注意）。再生核の名前は、関数解析の立場を取ることで見失われた微視的な挙動 $f(\mathbf{x})$ が、関数の内積によって再生されることに由来する。具体的には、(i) $\kappa(\cdot, \mathbf{x}) \in \mathcal{H}$ が全ての $\mathbf{x} \in \mathcal{U}$ に対して成立し、更に (ii) 再生性 $\langle f, \kappa(\cdot, \mathbf{x}) \rangle = f(\mathbf{x})$ が全ての $\mathbf{x} \in \mathcal{U}$ と $f \in \mathcal{H}$ に対して成り立つ^(注2) (図1, 2)。様々な再生核が知られているが、本稿では近似問題との親和性が高いガウス核の使用を主に想定して論じる。

3. 再生核適応フィルタとアルゴリズム

3.1 非線形多変数関数の適応推定問題

$\mathcal{U} \subset \mathbb{R}^L$ を入力空間、 $\phi: \mathcal{U} \rightarrow \mathbb{R}$ を非線形関数とする。適応信号処理では、一般に ϕ が時間変動を伴う状況も想定するが、ここでは簡単のため、時間変動は考えない。入力ベクトル列 $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots) \subset \mathcal{U}$ とそれに対応する出力信号列 $(d_0, d_1, d_2, \dots) \subset \mathbb{R}$ が順次、観測される。ここで、 $d_n := \phi(\mathbf{u}_n)$ ($n=0, 1, 2, \dots$) とする。時々刻々観測される入出力データ (\mathbf{u}_n, d_n) を基に、未知の非線形関数 ϕ を逐次的に推定する問題を考える。

3.2 再生核適応フィルタ

再生核適応フィルタでは、推定したい非線形関数 ϕ が、あらかじめ設計した再生核 κ と入力空間 \mathcal{U} から定まる関数空間 \mathcal{H} の元であると仮定する。この \mathcal{H} は、全ての $\kappa(\cdot, \mathbf{x})$ ($\mathbf{x} \in \mathcal{U}$) とそれらの線形結合（更にその極限）を含む線形空間である。再生核適応フィルタは、

$$\phi_n(\mathbf{u}) := \sum_{j=0}^n h_j \kappa(\mathbf{u}, \mathbf{x}_j), \quad h_j \in \mathbb{R}, \quad \mathbf{u}, \mathbf{x}_j \in \mathcal{U} \quad (1)$$

と書き表される。ここで、 \mathbf{x}_j は定数ベクトルである。 $\kappa(\mathbf{u}, \mathbf{x}_j)$ は \mathbf{u} の関数とみることができ、関数の集合 $\{\kappa(\cdot, \mathbf{x}_j)\}_{j=0}^n$ を辞書という。辞書は、 \mathcal{H} の部分空間の基底のような役割を果たすが、一次独立性^(用語)が一般に保証されない点に注意されたい。以下では、「辞書は、関数空間 \mathcal{H} の中で実際に探索する部分空間を決定するものである」ことを心にとどめて、読み進めて頂きたい。再生核適応アルゴリズムは、「辞書の設計法」と「係数

$h_{j,n}$ の更新規則」から成る。これらを具体的に見ていこう。

3.3 Kivinen らのアイデア⁽⁸⁾

出力 $\phi(\mathbf{u}_n)$ が所望信号 d_n に一致する（つまり、瞬時誤差を零とする）超平面^(用語)

$$\Pi_n := \{\phi \in \mathcal{H} | \langle \phi, \kappa(\cdot, \mathbf{u}_n) \rangle = \phi(\mathbf{u}_n) = d_n\} \quad (2)$$

を定義する。Kivinen らは、超平面 Π_n への直交射影に基づくアルゴリズム

$$\phi_{n+1} := \phi_n + \mu(P_{\Pi_n}(\phi_n) - \phi_n), \quad \mu \in (0, 2) \quad (3)$$

を提案した^(注3)（初期ベクトルは $\phi_0 := \theta$ ）。ここで、直交射影 $P_{\Pi_n}(\phi_n)$ とは、 Π_n 上で ϕ_n から最も近い点として定義され、

$$P_{\Pi_n}(\phi_n) = \phi_n + \frac{d_n - \langle \phi_n, \kappa(\cdot, \mathbf{u}_n) \rangle}{\|\kappa(\cdot, \mathbf{u}_n)\|^2} \kappa(\cdot, \mathbf{u}_n) \quad (4)$$

で与えられる。図3を見てみよう。 Π_n ($n=0, 1, 2$) へ向かう矢印はベクトル $P_{\Pi_n}(\phi_n) - \phi_n$ を表しており、式(4)から、これは $\kappa(\cdot, \mathbf{u}_n)$ の定数倍である。実際に、超平面 Π_n は、 \mathcal{H} 上で $\kappa(\cdot, \mathbf{u}_n)$ と直交するベクトル全体の集合を平行移動したものである。射影を繰り返すことで、 ϕ_n が ϕ に単調に近づいていく様子が分かる。線形フィルタの学習同定法⁽¹⁶⁾と原理は同じである^(注4)。

さて、時刻 n の再生核適応フィルタは、どのような辞書で表現されるだろうか。式(3), (4)から、時刻 n における辞書は $\{\kappa(\cdot, \mathbf{u}_j)\}_{j=0}^n$ である。時間が経過して n

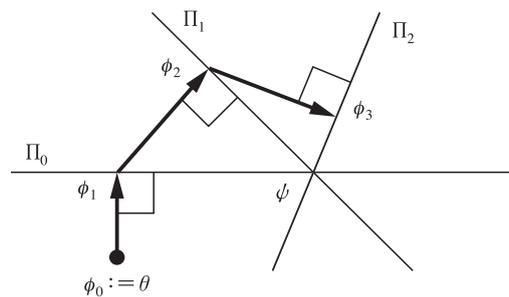


図3 Kivinen らの手法（ステップサイズ $\mu := 1$ ）射影を繰り返すことで、超平面 Π_0, Π_1, Π_2 の交点に位置する非線形関数 ϕ に単調に近づく様子が分かる。

(注3) 正確には、Kivinen らは式(4)の分母を1で置き換えた形を確率的勾配降下法から導いた。ガウス核の場合、 $\|\kappa(\cdot, \mathbf{u}_n)\|^2 = \kappa(\mathbf{u}_n, \mathbf{u}_n) = 1$ であるから、式(3)は厳密な意味でも Kivinen らの手法に一致する。

(注4) 単調性は、ピタゴラスの定理⁽¹⁴⁾から厳密に保証される。ここで、 ϕ_n ($n=0, 1, 2, \dots$) の単調性は ϕ への収束を保証するものでないことを注意しておく(文献(17), Example 1(b)).

(注2) 再生核はヒルベルト空間に対して定義される。正確な定義を理解しなくても本稿の大筋を追えるが、興味のある読者は文献(15)を参照されたい。ヒルベルト空間と工学への応用に興味のある読者には、文献(14)がお薦めである。

が大きくなるとどのようなことが起こるだろうか。再生核適応フィルタの出力を計算するためには、データ $\{\mathbf{u}_j\}_{j=0}^n$ と係数 $\{h_{j,n}\}_{j=0}^n$ が必要である。つまり、あらかじめ決められた時間内でのタスクであれば必要な容量を決定できるが、そうでなければ、(仮想的に) 無限の容量が必要ということになり、非現実的である。Kivinen らは、この問題を避けるために、切断規則を提案している。具体的には、辞書サイズ (辞書の要素数) の上限を事前に決めておき、その上限を超えたら、新しいデータを辞書に加える際に最も古いデータを捨てる。

このアプローチの妥当性を考察してみよう。辞書を規定するデータ \mathbf{u}_j はガウス関数の中心点を決定する (図 4)。ガウス関数は中心点から遠ざかると指数関数的に減衰するため、辞書の各要素 $\kappa(\cdot, \mathbf{u}_j)$ は、微視的に見れば、各点 \mathbf{u}_j 付近での関数の挙動を近似する役割を担う。入力ベクトルが現れる領域で関数を良好に近似したいので、端的に言えば、この領域をカバーするだけの中心点を取ればよいということになる。しかし、この領域を事前に知ることは一般に困難であるため、Kivinen らの方法のように、入力データを中心点とする $\kappa(\cdot, \mathbf{u}_j)$ を辞書に加えていく方法が広く用いられている。以上の議論から、Kivinen らのアプローチの欠点が浮かび上がる。一つは、現在の辞書でカバーされる領域のデータが入ってきた場合でも辞書に追加されてしまうため、冗長性の高い辞書が構築される可能性があるという点である。もう一つは、推定において中心的な役割を果たすデータを辞書から削除してしまう危険性がある点である。以下に紹介する手法では、辞書でカバーされない領域のデータ \mathbf{u}_n が観測されたときのみ、新しい要素 $\kappa(\cdot, \mathbf{u}_n)$ を辞書に加える選択的な辞書構築法を取っている。このようにして、辞書サイズを削減することを「辞書のスパース化 (sparsification)」という。

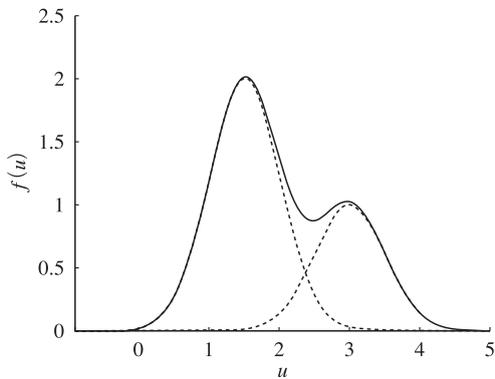


図 4 ガウス核 $\kappa(x, y) := \exp\left(-\frac{(x-y)^2}{2 \times 0.5^2}\right)$ ($x, y \in \mathbb{R}$) を用いた非線形関数 $2\kappa(u, 1.5) + \kappa(u, 3)$ ($u \in \mathbb{R}$) (i) 中心点の位置 (y の値), (ii) 中心点の個数, (iii) 各ガウス関数の係数 (山の高さ) を変えることによって、様々な形状の非線形関数を生成することができる。

3.4 Richard らのアイデア⁽¹²⁾

時刻 n における辞書を $\{\kappa(\cdot, \mathbf{u}_j)\}_{j \in \mathcal{J}_n}$ ($\mathcal{J}_n := \{j_1^{(n)}, j_2^{(n)}, \dots, j_{r_n}^{(n)}\} \subset \{0, 1, 2, \dots, n\}$) とする。ここで、 r_n は辞書サイズを表す。あるしきい値 $\delta \in (0, 1)$ と辞書 $\{\kappa(\cdot, \mathbf{u}_j)\}_{j \in \mathcal{J}_n}$ に対して、 $\kappa(\cdot, \mathbf{u}_n)$ がコヒーレンス基準

$$\max_{j \in \mathcal{J}_n} \left| \frac{\kappa(\mathbf{u}_n, \mathbf{u}_j)}{\sqrt{\kappa(\mathbf{u}_n, \mathbf{u}_n) \kappa(\mathbf{u}_j, \mathbf{u}_j)}} \right| \leq \delta \quad (5)$$

を満たす場合、入力データ \mathbf{u}_n が辞書でカバーされない領域にあると判断し、 $\kappa(\cdot, \mathbf{u}_n)$ を辞書に追加する。ここで、式 (5) を満たすことは、 $\kappa(\cdot, \mathbf{u}_n)$ と似た関数が辞書に含まれないことを意味する。 ϕ_{n+1} を表現する辞書 $\{\kappa(\cdot, \mathbf{u}_j)\}_{j \in \mathcal{J}_{n+1}}$ の張る部分空間を辞書部分空間と呼び、これを $M_n (\subset \mathcal{H})$ と表記する。入力 \mathbf{u}_n に対する出力は $\phi_{n+1}(\mathbf{u}_n) = \sum_{j \in \mathcal{J}_{n+1}} h_{j, n+1} \kappa(\mathbf{u}_n, \mathbf{u}_j) = \mathbf{h}_{n+1}^T \mathbf{k}_n$ とベクトル表現できる。この表現に基づいて、ユークリッド空間 $\mathbb{R}^{r_{n+1}}$ の超平面

$$H_n := \{\mathbf{h} \in \mathbb{R}^{r_{n+1}} \mid \mathbf{h}^T \mathbf{k}_n = d_n\} \quad (6)$$

への直交射影に基づくアルゴリズム

$$\mathbf{h}_{n+1} := \mathbf{h}_n + \mu(P_{H_n}(\mathbf{h}_n) - \mathbf{h}_n), \quad \mu \in (0, 2) \quad (7)$$

を定義する。これが、Richard らの基本的なアイデアである⁽¹²⁾。ここで、辞書の初期設定は $\mathcal{J}_{-1} := \emptyset$ とする。新しいデータが追加されて辞書が更新された場合 ($r_{n+1} = r_n + 1$ の場合)、ベクトル \mathbf{h}_n は一つ前の時刻での更新で得られた ($r_n \times 1$) の係数ベクトルに零を新しい成分として追加した ($r_{n+1} \times 1$) のベクトルとする。コヒーレンスは比較的シンプルな判別基準であり、低演算量で計算できることが大きな利点である。非線形適応フィルタの概念図とフィルタ更新のチャート図をそれぞれ図 5, 6 に示す。

3.5 辞書スパース化技術の光と影

コヒーレンス以外にも、様々な基準が提案されている⁽¹⁰⁾。Platt 基準は、 \mathbf{u}_n が辞書でカバーされない領域にあり、かつ推定誤差が大きいときに限り、 $\kappa(\cdot, \mathbf{u}_n)$ を辞書に追加する。ALD 基準 (Approximate Linear Dependency) は、 $\kappa(\cdot, \mathbf{u}_n)$ を仮に辞書に加えた場合の線形従属の度合いを近似的に計算して、辞書に追加するか否かを決定する。関数空間 \mathcal{H} での射影に基づく Kivinen らの手法に、これらの基準を取り込んだらどうだろうか。式 (3), (4) から、 ϕ_n の更新方向は $\pm \kappa(\cdot, \mathbf{u}_n)$ で与えられ、辞書に $\kappa(\cdot, \mathbf{u}_n)$ が含まれない場合、一般には、この更新方向ベクトルが辞書部分空間 M_n から飛び出してしまうことになる。すなわち、関数空間 \mathcal{H} における

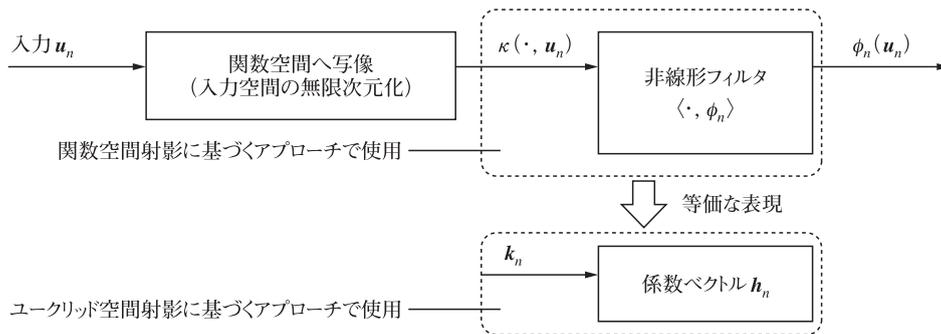


図5 再生核適応フィルタの概念図 $\kappa(\cdot, \mathbf{u}_n)$ は関数空間のベクトルであり、 ϕ_n との内積によって出力 $\phi_n(\mathbf{u}_n)$ が得られる。この表現を使うのが関数空間で射影するアプローチ (3.3, 3.6) であり、これと等価な表現 $\phi(\mathbf{u}_n) = \mathbf{h}_n^T \mathbf{k}_n$ を使うのがユークリッド空間で射影するアプローチ (3.4) である。



図6 新しいデータ (\mathbf{u}_n, d_n) が観測されたときの再生核適応フィルタ ϕ_n 更新のチャート図 辞書更新と係数更新によって ϕ_n が更新される。

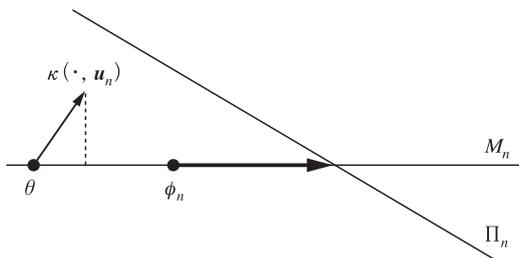


図7 Yukawa らのアイデア：辞書部分空間 M_n に沿った Π_n への射影 Π_n の法線ベクトル $\kappa(\cdot, \mathbf{u}_n)$ の M_n への射影を用いることで、 $\kappa(\cdot, \mathbf{u}_n)$ が辞書に含まれない場合でも新しいデータを生かした更新が実現できる。

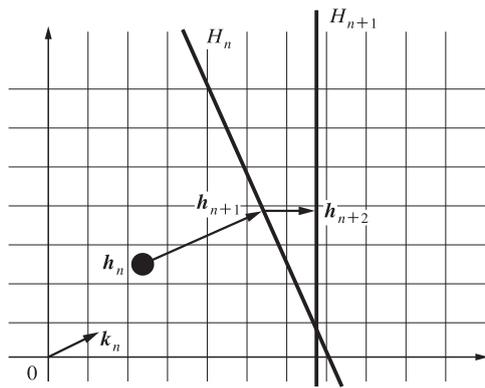
射影に基づくアプローチは、辞書が更新されないと再生核適応フィルタを更新できないことになる。このことは、ユークリッド空間での射影に基づく Richard らの手法が全ての観測データを利用して辞書部分空間内の近似精度を向上していくことと対照的であり、関数空間 \mathcal{H} での射影に基づくアプローチに内在する最大の問題点であった。以下、この問題を解決するためのアイデアを紹介し、Richard らの手法との関係を明らかにする。

3.6 Yukawa らのアイデア⁽¹⁸⁾

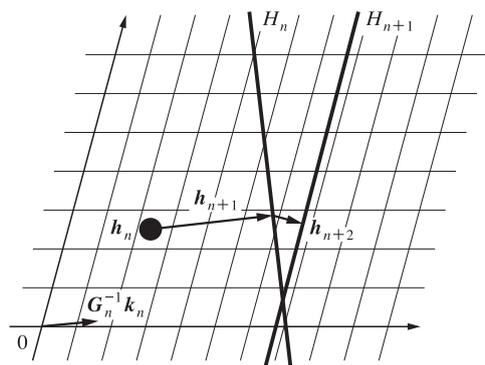
基本的なアイデアは、「辞書部分空間 M_n に沿って超

平面 Π_n に射影する」ことである^(注5) (図7)。言い換えると、 M_n と Π_n の共通部分に射影する。この射影は、 ϕ_n を「 $\kappa(\cdot, \mathbf{u}_n)$ から M_n への射影」の方向（または、その逆方向）に動かすことで得られる。さて、ここで辞書が一次独立であると仮定しよう。ガウス核でコヒーレンス基準を採用した場合には、一次独立性は自動的に保証される。このとき、正規方程式を解く際に現れるグラム行列 \mathbf{G}_n ((p, q) 成分は $\kappa(\mathbf{u}_p^{(n)}, \mathbf{u}_q^{(n)})$) は正定値行列になる。ユークリッド空間 $\mathbb{R}^{r_{n+1}}$ に \mathbf{G}_n -内積 $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{G}_n} := \mathbf{x}^T \mathbf{G}_n \mathbf{y}$ を定義すると、これは辞書部分空間 M_n と同型である。したがって、 $\phi_n \in \mathcal{H}$ から共通部分 $\Pi_n \cap M_n$ への射影は、 $\mathbf{h}_n \in \mathbb{R}^{r_{n+1}}$ から H_n への (\mathbf{G}_n -内積の意味での) 射影に対応する。直感的には、斜めに傾いた方眼紙の上で直交射影を計算していると捉えて頂きたい (図8)。 \mathbf{G}_n -内積の意味での射影を得るには、 $\mathbf{G}_n^{-1} \mathbf{k}_n$ の計算が必要である。 \mathbf{G}_n は $r_n \times r_n$ 行列であるため、辞書サイズ r_n が大きければ、 $\mathbf{G}_n^{-1} \mathbf{k}_n$ の計算を避けたいこともあるだろう。そこで、辞書 $\{\kappa(\cdot, \mathbf{u}_j)\}_{j \in \mathcal{J}_n}$ から $\kappa(\cdot, \mathbf{u}_n)$ とのコヒーレンスが高いものを選抜し、選ばれた $\kappa(\cdot, \mathbf{u}_i)$ だけで張られる部分空間 \tilde{M}_n を定義する。これを平行移動した $V_n := \tilde{M}_n + \phi_n (\subset M_n)$ に沿って Π_n へ射影することで、 M_n に沿って射影した場合と比べたときの性能劣化を最小限に抑えながら、演算量を効果的に削減できる (V_n はアフィン部分空間と呼ばれる)。これら二つのアイデアを統合した一般的なアルゴリズムを HYPASS 法 (HYperplane Projection along Affine SubSpace)⁽¹⁸⁾ と呼ぶ。 $\kappa(\cdot, \mathbf{u}_n)$ が辞書に追加されないときでも ϕ_n を更新できる

(注5) この更新方向は、劣勾配射影の立場からも説明できることが知られている (文献(17), Example 5)。



(a)



(b)

図8 二つのアプローチの本質的な相違点 (a)Richardらの手法(標準的なユークリッド計量を利用)と(b)Yukawaらの手法(グラム行列 G_n で決まる計量を利用). 本質的な違いは計量の取り方だけ. グラム行列の利用により, 多くの場合, 収束特性が改善される.

ため, HYPASS法はKivinenらの手法の拡張となっている. $\kappa(\cdot, \mathbf{u}_n)$ が辞書に追加されたときは, $\kappa(\cdot, \mathbf{u}_n) \in M_n$ であるから, Kivinenらの更新規則(式(3))に一致する. HYPASS法は, Dodd-Kadirkamanathan-Harrison (2003)の手法⁽¹⁹⁾とChen-Zhao-Zhu-Principe (2012)のQKLMS法(Quantized Kernel Least Mean Square)⁽²⁰⁾を特別な場合として含む.

4. 多核適応フィルタと辞書の適応的精練技術⁽¹³⁾

筆者が提唱する「多核適応フィルタ(multikernel adaptive filter)」と「辞書の適応的精練技術」について述べる. 「良い」再生核とは? この問いから始めよう. 広いクラスの関数を表現できるのが「良い」再生核と言えるだろうか. (ここで, クラスとは関数を分類するものと思って頂きたい. 例えば「二乗可積分関数のクラス」など.) 二つの分散 $0 < \sigma_1^2 < \sigma_2^2$ を持つガウス核の規定する空間 $\mathcal{H}_{\sigma_1^2}$ と $\mathcal{H}_{\sigma_2^2}$ が入れ子構造 $\mathcal{H}_{\sigma_1^2} \supset \mathcal{H}_{\sigma_2^2}$ を持つことが知られている⁽²¹⁾. したがって, 一見すると, 小さい分散を持つガウス核が「良い」と思える. しかし, 見

落としてはいけないのが辞書サイズである. 現実的には, 辞書サイズの上限を決めておく必要があり, この場合, ガウス核の分散は適切に定めなければならない. 直感的には, 分散が小さすぎると(限られた辞書サイズでは)大きい領域をカバーできず, 逆に分散が大きすぎると非線形関数 ψ の高周波成分を精度良く捉えられない. さて, 事前に分からない ψ に適した再生核をどうやって設計したらよいだろうか.

この素朴な疑問から, 多核適応フィルタの研究が始まった. 「複数の再生核」に対応した「複数の関数空間」を定義し, それらの空間上のベクトル和を全て含む大きな空間(和空間という)を考える. この和空間の元として, 適応フィルタ ϕ_n をモデル化する. 文献(13)では, 分散の異なる複数のガウス核を用いた場合を想定しており, 辞書に用いるデータ \mathbf{u}_j を全てのガウス核で共通とすることで計算量の増加を抑えている. 辞書を一般化する際, 各再生核に対する適切な辞書の選別法とメモリ容量・計算量増加への対策が重要である. 最もシンプルな多核適応アルゴリズムは, Richardらの手法をこの和空間モデルに拡張することで得られる. 多核適応フィルタは ψ の低周波成分と高周波成分を各々に適した再生核で表現することができるため, 小さい辞書サイズで高精度な関数推定が可能となる.

もう一つのアイデア「辞書の適応的精練技術」は, スパース最適化(圧縮センシングなど)で広く用いられる l_1 ノルム最小化原理(劣決定系のスパース解が, ある条件下, l_1 ノルム最小化によって求まるという原理)から導かれる. 多核適応フィルタの係数は, 辞書と再生核の両方に依存する. ガウス核であれば, 各中心点と各分散に対して一つの係数が割り当てられると考えればよい. これらの係数を辞書の要素ごとに(中心点ごとに)ブロック化したブロック l_1 ノルム(各ブロックに対応する部分ベクトルのユークリッドノルムの総和)を定義し, これを正則化項とする. この正則化項を損失関数に加えたコスト関数(正確には, 時間変動するコスト関数列)に対して, 「微分可能な凸関数(ここでは損失関数)に対する最急降下法」と「微分できない凸関数(ここではブロック l_1 ノルム)の近接写像」を交互に繰り返すMurakami-Yamagishi-Yukawa-Yamada (2010)の適応アルゴリズム⁽²²⁾を適用する. ブロック l_1 ノルムによる正則化のおかげで, 係数がブロック単位でスパース化され, 推定に役立たない中心点に対応する係数が自動的に零となる. このような中心点に対応する要素を辞書から削除することで, 適応的に辞書を精練する. Richardらのグループによって, 再生核数が1の場合における適応的精練技術の解析が与えられた⁽²³⁾. また, Tobarらにより, $\phi(\mathbf{u})$ がベクトル値を取る場合における多核適応フィルタが提案された⁽²⁴⁾. 筆者らは, 多数の再生核から非線形関数 ψ に適合するものを自動抽出する「再

生核の適応的精練技術」を EUSIPCO 2013 で提案し、独国グループとの共同研究によって移動体通信の伝送損失分布図オンライン推定における有効性が実証された⁽²⁵⁾。その他、最新情報は、Web (<http://www.ykw.elec.keio.ac.jp/yukawa/>) で確認頂ける。

5. おわりに

非線形適応信号処理技術の新潮流である再生核適応フィルタの基本原理解から多核適応フィルタと辞書適応的精練技術までを解説した。「微視的な眼」とともに、「巨視的な眼 (関数解析^{(14), (15)}による)」から関数を見ることの有用性が理解頂けたと思う。本稿が、再生核適応フィルタ研究の発展と実用化への一助となることを願う。

再生核適応フィルタのツールボックスが Machine Learning Open Source Software (<http://mloss.org/software/view/498/>) からダウンロードできる。Kivinenらの手法から筆者の多核適応フィルタや Φ -PASS 法 (HYPASS 法の拡張) まで、2014 年 4 月時点で 18 種類のアプローチが含まれている。本稿では説明を割愛させて頂いたが、RLS 型アルゴリズム^{(9), (26)}、適応射影劣勾配法に基づくアルゴリズム⁽²⁷⁾、辞書サイズ固定型アルゴリズム⁽²⁶⁾、複素数・Cayley-Dickson 数 (超複素数) への拡張^{(28), (29)}、音響通信システムへの応用⁽⁶⁾など、様々な方向へ研究が進展を続けている。

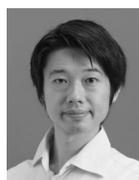
謝辞 本稿の執筆にあたり多くの方々から貴重な御助言を頂き、心から深謝申し上げる。本稿に関する研究調査は KDDI 財団の助成による。

文 献

- (1) N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337-404, May 1950.
- (2) V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- (3) B. Schölkopf and A.J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2001.
- (4) K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181-202, March 2001.
- (5) 福水健次, *カーネル法入門*, 朝倉出版, 東京, 2010.
- (6) J.M. Gil-Cacho, M. Signorello, T. van Waterschoot, M. Moonen, and S.H. Jensen, "Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm," *IEEE Trans. Audio, Speech and Language Processing*, vol. 21, no. 9, pp. 1867-1878, Sept. 2013.
- (7) I. Steinwart, "On the influence of the kernel on the consistency of support vector machines," *J. Mach. Learn. Res.*, vol. 2, pp. 67-93, 2001.
- (8) J. Kivinen, A.J. Smola, and R.C. Williamson, "Online learning with kernels," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2165-2176, Aug. 2004.
- (9) Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275-2285, Aug. 2004.

- (10) W. Liu, J. Principe, and S. Haykin, *Kernel Adaptive Filtering*, Wiley, New Jersey, 2010.
- (11) K. Slavakis, P. Bouboulis, and S. Theodoridis, "Online learning in reproducing kernel Hilbert spaces," *The E-Reference Signal Processing*, Elsevier, 2014, to appear.
- (12) C. Richard, J. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1058-1067, March 2009.
- (13) M. Yukawa, "Multikernel adaptive filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4672-4682, Sept. 2012.
- (14) 山田 功, *工学のための関数解析*, 数理工学社, 東京, 2009.
- (15) 小川英光, *工学系の関数解析*, 森北出版, 東京, 2010.
- (16) J. Nagumo and J. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282-287, 1967.
- (17) I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numer. Funct. Anal. Optim.*, vol. 25, no. 7 & 8, pp. 593-617, 2004.
- (18) M. Yukawa and R. Ishii, "An efficient kernel adaptive filtering algorithm using hyperplane projection along affine subspace," in *Proc. EUSIPCO*, pp. 2183-2187, 2012.
- (19) T.J. Dodd, V. Kadiramanathan, and R.F. Harrison, "Function estimation in Hilbert space using sequential projections," in *IFAC Conf. Intell. Control Syst. Signal Process.*, pp. 113-118, 2003.
- (20) B. Chen, S. Zhao, P. Zhu, and J.C. Principe, "Quantized kernel least mean square algorithm," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22-32, Jan. 2012.
- (21) A. Tanaka, H. Imai, M. Kudo, and M. Miyakoshi, "Theoretical analyses on a class of nested RKHS's," in *Proc. IEEE ICASSP*, pp. 2072-2075, 2011.
- (22) Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in *Proc. IEEE ICASSP*, pp. 3734-3737, 2010.
- (23) W. Gao, J. Chen, C. Richard, and J. Huang, "Online dictionary learning for kernel LMS," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2765-2777, June 2014.
- (24) F.A. Tobar, S.-Y. Kung, and D.P. Mandic, "Multikernel least mean square algorithm," *IEEE Trans. Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 265-277, Feb. 2014.
- (25) M. Kasparick, R.L.G. Cavalcante, S. Valentin, S. Stańczak, and M. Yukawa, "Kernel-based adaptive online reconstruction of coverage maps with side information," *IEEE Trans. Veh. Technol.*, 2014, submitted: arXiv:1404.0979 [cs. NI].
- (26) S. Van Vaerenbergh, M. Lázaro-Gredilla, and I. Santamaría, "Kernel recursive least-squares tracker for time-varying regression," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1313-1326, Aug. 2012.
- (27) K. Slavakis, S. Theodoridis, and I. Yamada, "Online kernel-based classification using adaptive projection algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2781-2796, July 2008.
- (28) P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 964-978, March 2011.
- (29) T. Mizoguchi and I. Yamada, "An algebraic translation of Cayley-Dickson linear systems and its applications to online learning," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1438-1453, March 2014.

(平成 26 年 4 月 30 日受付 平成 26 年 6 月 16 日最終受付)



湯川 正裕 (正員)

平 14 東工大・工・電気電子卒。平 18 同大学院博士課程了。現在、慶大・理工・電子・専任講師。主として、信号処理工学、情報通信工学の研究に従事。工博。平 17 年度本会論文賞、平 25 年度電気通信普及財団賞テレコムシステム技術賞、平 26 年度文部科学大臣表彰若手科学者賞各受賞。