

TR-H-039

0044

標準正則化理論の多価関数への拡張(2)
—多価正則化ネットワークとその学習アルゴリズム—

Masahiko SHIZAWA

志沢 雅彦

1993. 12. 1

ATR 人間情報通信研究所

〒619-02 京都府相楽郡精華町光台 2-2 ☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011

Facsimile: +81-7749-5-1008

標準正則化理論の多価関数への拡張 (2)

— 多価正則化ネットワークとその学習アルゴリズム —

志沢 雅彦

ATR 人間情報通信研究所

あらまし

本論文では、正則化ネットワーク (Regularization Network (RN)) を 1 対 h 写像、すなわち多価関数を近似する様に拡張した多価正則化ネットワーク (Multi-Valued Regularization Network (MVRN)) を提案する。多価関数の近似は、例から多重、多義的な入出力関係を学習するのに有用である。MVRN は、多価標準正則化理論 (Multi-Valued Standard Regularization Theory (MVSRT)) に基づいて導出される。MVSRT は、代数方程式による多価関数の直接的表現法に基づいている。未知関数の置き換えにより、線形の Euler-Lagrange 方程式が得られる。したがって、MVRN は、入出力サンプルデータの集合からのネットワーク結合重みパラメータの学習を連立一次方程式を解くだけで行うことができ、その際に、データのクラスタリングを行う必要がない。さらに、この連立方程式の次数は写像の多価度に依存せず、サンプルデータ数に等しくなることが示される。従来の正則化ネットワークの場合と同様に、MVRN を特殊化したり、サンプルデータ数よりも少ない基底関数で近似することによって、円形基底関数法 (Radial Basis Function (RBF))、一般化円形基底関数法 (Generalized RBF)、スプライン近似、HyperBF ネットワークなどを多価関数の近似に拡張することができる。本稿では、予備実験として、Gaussian RBF ネットワークを導出し、そのシミュレーション結果も報告する。さらに、多価ベクトル値関数を学習する MVRN の構成法について述べる。

キーワード:

正則化ネットワーク、標準正則化理論、多価関数、1 対多写像、写像学習、関数近似、フィードフォワードネットワーク

Extension of the Standard Regularization Theory into Multi-Valued Functions: Multi-Valued Regularization Networks and Learning Algorithms

Abstract

The regularization network (RN) is extended to approximate *multi-valued functions* so that one-to- h mapping, where h denotes multiplicity of the mapping, can be represented and learned from a finite number of input-output samples *without clustering operations* on the sample data set. Multi-valued function approximations are useful for learning ambiguous input-output relations from examples. This extension, which we call the Multi-Valued Regularization Network (MVRN), is derived from the Multi-Valued Standard Regularization Theory (MVSRT), which is an extension of standard regularization theory to multi-valued functions. MVSRT is based on a direct algebraic representation of multi-valued functions. By simple transformation of the unknown functions, we can obtain *linear* Euler-Lagrange equations. Therefore, the learning algorithm for MVRN is reduced to solving a linear system. It's rather surprising that the dimension of the linear system is invariant to the multiplicity h . The proposed theory can be specialized and extended into Radial Basis Function (RBF) Methods, Generalized RBF (GRBF), and HyperBF networks of multi-valued functions. We also describe how the vector-valued function approximations can be extended into the multi- and vector-valued function approximations.

Keywords: Regularization network, Standard regularization theory, Multi-valued functions, One-to-many mapping, Computational learning, Function approximation, Feedforward neural network.

もくじ

1	はじめに	4
2	正則化ネットワークの理論	5
2.1	正則化ネットワークの構成理論	5
2.2	正則化ネットワークの学習アルゴリズム	7
2.3	一般化正則化ネットワークの学習アルゴリズム	7
3	多価標準正則化理論 (MVSRT)	8
3.1	ベクトル値多価関数の一般表現	8
3.2	スカラー値多価関数の一般表現	9
3.3	h 価スカラー値関数の標準正則化理論	9
3.4	2 価 2 次元ベクトル値関数の標準正則化理論	9
4	多価正則化ネットワークと学習アルゴリズム	11
4.1	多価正則化ネットワークの構成理論	11
4.2	多価正則化ネットワークの学習アルゴリズム	12
4.3	2 価 2 次元ベクトル値の正則化ネットワーク	13
4.4	一般化多価正則化ネットワーク	13
5	シミュレーションによる理論の検証	14
6	多価正則化ネットワークによるビジョンのアルゴリズムの学習による実現	14
7	まとめ	15

1 はじめに

正則化ネットワーク (Regularization Network, RN) は、標準正則化理論 (Standard Regularization Theory) にもとづいて導出され、入出力写像のなめらかな関数による近似とその例からの学習を行う3層構造のフィードフォワードネットワークである [7] [8]。この正則化ネットワークを特殊化することにより、円形基底関数法 (Radial Basis Function Method, RBF)、スプライン近似法などの古典的な関数近似の方法がほとんどすべて導かれる。また、隠れ層ユニットの数を学習サンプル数に比較して少なくした一般化正則化ネットワーク (Generalized Regularization Network, GRN) を特殊化すると、一般化円形基底関数法 (Generalized RBF) などが導かれる。

正則化ネットワークを含め、いままで提案されてきた関数近似を行う (ニューラル) ネットワークとその学習アルゴリズムは、不連続をもつ関数や多価関数を直接扱うことを想定していない。これに関連する研究では、複数の並列的に配置されたネットワークの集合体として一価写像の区分的な近似を行う方法は提案されている (モジュラーネットワーク)[3]。しかし、この構造では、学習のためのアルゴリズムは、一般の非線形関数に対する最適化を用いなければならず、局所最適解をさけるためにかなりの工夫が必要である。一般に、1対多写像、すなわち、複数の異なる出力値の可能性を同一入力に対応させるようなネットワークの一般的な構成理論と適切な学習アルゴリズムは知られていない。

本論文では、前報告 [11] で提案した多価標準正則化理論 (Multi-Valued Standard Regularization Theory, MVSRT) にもとづいて、多価写像を近似する多価正則化ネットワーク (Multi-Valued Regularization Network, MVRN) の理論を展開する。この理論では、一般に、 n 次元入力空間から、 m 次元出力空間への h 価写像の入出力関係が、直接代数的に表現できることを利用する。また、未知関数の置き換えによって、この拘束条件が未知関数に関して線形形式に変換できる事実を用いて、2次元関数のエネルギー最小化問題、すなわち、標準正則化問題として定式化する。これによって、入出力サンプルデータからのネットワークの重み係数の計算、すなわち、学習のアルゴリズムが、従来の1価関数の正則化ネットワークと同様に連立1次方程式の解法に帰着される。

以下、まず、第2章では、多価正則化ネットワーク (MVRN) の理論を導く準備として、正則化ネットワークの理論の導出過程を Poggio & Girosi の論文 [7] にもとづいて簡単にレビューする。次に、第3章で、多価標準正則化理論 (MVSRT) の一般論を述べる。次に、第4章で、MVRN の理論を述べる。まず、スカラー値写像の MVRN とその学習アルゴリズムを導出する。次に、ベクトル値写像の MVRN とその学習アルゴリズムを、出力空間の次元数が、 $m = 2$ で多価度が $h = 2$ の場合について具体的に導出する。さらに、基底関数の数を学習サンプル数よりも少なくした一般化多価正則化ネットワーク (MVGRN) について簡単に述べる。第5章では、理論を検証する

ためのシミュレーションとして、スカラー値の MVRN について、基底関数が等方的 Gaussian の場合 (Gaussian RBF) の場合のシミュレーション結果を示す。次に、第 6 章では、MVRN が、どのような形で実際問題に応用されるのかをビジョンの場合を例にとって考察する。第 7 章で本論文をまとめる。

2 正則化ネットワークの理論

本章では、まず、正則化ネットワークの理論を Poggio & Girosi[7] の論文にもとづいて本論文の展開に必要な範囲でレビューする。

正則化ネットワークは、「学習=関数近似」という立場から、ビジョンにおけるなめらかな表面復元の計算理論である標準正則化理論を拡張して、 n 次元実数空間 R^n から m 次元実数空間 R^m への滑らかな写像をまばらに与えられた学習サンプル点から構成する超曲面の生成問題と考えることにより得られる。以下では、多価正則化ネットワーク導出の準備として、 $m = 1$ の場合、すなわち、スカラー値写像の場合について、正則化ネットワークの理論をまとめておく。

2.1 正則化ネットワークの構成理論

n 次元空間におけるスカラー値関数の正則化問題は、次のエネルギー汎関数の最小化問題として定式化される。

$$E[f] = \sum_{i=1}^N (y_{(i)} - f(\mathbf{x}_{(i)}))^2 + \lambda \|Sf\|^2 \quad (1)$$

ここで、 $y = f(\mathbf{x})$ が求める写像で、 N 個のデータ点が $\{(\mathbf{x}_{(i)}, y_{(i)}) | i = 1, 2, \dots, N\}$ の様に与えられているとした。 S は正則化作用素であり、一般に微分作用素、あるいは、擬微分作用素が用いられる。 $\| \cdot \|$ は、関数空間の 2 乗ノルムで、一般に関数 $g(\mathbf{x})$ に対して、

$$\|g\|^2 = \int_{R^n} \{g(\mathbf{x})\}^2 d\mathbf{x} \quad (2)$$

と定義される。積分領域は、 R^n の全空間とする。 λ は正定数で、正則化項の相対的な重みを表す。汎関数 (1) の Euler-Lagrange 方程式は次式である。

$$-\sum_{i=1}^N \left[\{y_{(i)} - f(\mathbf{x})\} \delta(\mathbf{x} - \mathbf{x}_{(i)}) \right] + \lambda \hat{S} S f(\mathbf{x}) = 0 \quad (3)$$

または、

$$\hat{S} S f(\mathbf{x}) = \frac{1}{\lambda} \sum_{i=1}^N \{y_{(i)} - f(\mathbf{x})\} \delta(\mathbf{x} - \mathbf{x}_{(i)}) \quad (4)$$

とかくことができる。ここで、 $\delta(\cdot)$ は、Dirac のデルタ関数で、ベクトル $\mathbf{x} = (x_1, x_2, \dots, x_n)$ に対しては、

$$\delta(\mathbf{x}) = \delta(x_1)\delta(x_2)\cdots\delta(x_n) \quad (5)$$

で定義する。作用素 \hat{S} は、 S の随伴 (adjoint) 作用素である。一般に、実定数係数の線形偏微分作用素

$$T = \sum_i a_i \frac{\partial^{i_1+\dots+i_n}}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \quad (6)$$

の随伴作用素は、

$$\hat{T} = \sum_i (-1)^{i_1+\dots+i_n} a_i \frac{\partial^{i_1+\dots+i_n}}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} \quad (7)$$

で定義される ($i = (i_1, i_2, \dots, i_n)$)。方程式 (4) は、関数 $f(\mathbf{x})$ に関する偏微分方程式であり、次の関数方程式によって定義される作用素 $\hat{S}S$ の Green 関数 $K(\mathbf{x}, \mathbf{x}')$ を用いて解くことができる。¹

$$\hat{S}S K(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \quad (8)$$

この場合、 $\hat{S}S$ が自己随伴作用素であることから、Green 関数は、対称、すなわち $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})$ である。正則化問題の解 $f(\mathbf{x})$ は、次の様に書くことができる。

$$f(\mathbf{x}) = \frac{1}{\lambda} \sum_{i=1}^N \{y_{(i)} - f(\mathbf{x}_{(i)})\} K(\mathbf{x}, \mathbf{x}_{(i)}) \quad (9)$$

この式は、関数 $f(\mathbf{x})$ が Green 関数の線形結合で表されることを示している。重みパラメータ r_i を

$$r_i = \{y_{(i)} - f(\mathbf{x}_{(i)})\} / \lambda \quad (10)$$

で定義すると、関数 $f(\mathbf{x})$ は、次の様に表される。

$$f(\mathbf{x}) = \sum_{i=1}^N r_i K(\mathbf{x}, \mathbf{x}_{(i)}) \quad (11)$$

この Green 関数の線形和による任意関数 $f(\mathbf{x})$ の表現は正則化ネットワーク と呼ばれる 図 3 の様な 3 層ネットワークで表現できる。このネットワークは、中間層の数が、学習サンプルデータの数 N に等しい場合であり、厳密な意味での正則化問題の解が得られる。実際、式 (11) は、各学習サンプル点 $\mathbf{x}_{(i)}$ を中心とする「基底関数」 $K(\mathbf{x}, \mathbf{x}_{(i)})$ の線形結合で関数 $f(\mathbf{x})$ を表現したものになっている。正則化作用素 S の設定により、基底関数と学習される関数の性質が規定される。本論文では、基底関数が Gaussian 円形基底関数 (Radial Basis Function) になる場合だけを示す。この場合は、 S を次の様に定義される擬微分作用素とおいた場合である。

$$\|Sf\|^2 = \sum_{p=0}^{\infty} \frac{\sigma^{2p}}{p!2^p} \sum_{\substack{1 \leq i_1, \dots, i_p \leq n, \\ i_1 + \dots + i_p = p}} \int_{R^n} \left(\frac{\partial^p f(\mathbf{x})}{\partial x_{i_1} \dots \partial x_{i_p}} \right)^2 d\mathbf{x} \quad (12)$$

¹この関数は、本来は、基本解と呼ばれ、Green 関数を決めるには偏微分方程式 (4) において境界条件が必要になるが、本論文では、主に Green 関数が Gaussian になる正則化作用素を用いるので、境界条件は必要ない。したがって、Poggio & Girosi の論文 [7] におけると同様にこの関数をはじめから Green 関数と呼ぶ。

このとき、グリーン関数は、次の等方的 Gauss 関数になる。

$$K(\mathbf{x}, \mathbf{x}') = G(\|\mathbf{x} - \mathbf{x}'\|) = A \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (13)$$

ここで、 A は定数。

2.2 正則化ネットワークの学習アルゴリズム

ネットワーク (11) の重みパラメータ r_i の学習は、次の N 元線形連立方程式を解くことによって実現される。

$$\mathbf{K}\mathbf{r} + \mathbf{z} = \mathbf{0} \quad (14)$$

ここで、

$$\begin{aligned} \mathbf{K} &= (K_{ij}) = \left(K(\mathbf{x}_{(i)}, \mathbf{x}_{(j)}) + \lambda \delta_{ij} \right), \\ \mathbf{r} &= (r_1, r_2, \dots, r_N)^T, \\ \mathbf{z} &= (y_{(1)}, y_{(2)}, \dots, y_{(N)})^T \end{aligned} \quad (15)$$

である。ここで、 δ_{ij} は、Kronecker のデルタである。この方程式は、式 (11) の \mathbf{x} を $\mathbf{x}_{(i)}$ に、添字 i を j にそれぞれ置き換えた

$$f(\mathbf{x}_{(i)}) = \sum_{j=1}^N r_j K(\mathbf{x}_{(i)}, \mathbf{x}_{(j)}) \quad (16)$$

を、式 (10) に代入して整理すると得られる。

2.3 一般化正則化ネットワークの学習アルゴリズム

方程式 (14) は、学習サンプル数 N が多くなると計算量が約 N^3 に比例して大きくなり、かつ、方程式が悪条件になりやすくなる。そこで、中間層 = 基底関数の数を N より少なくし、基底関数の中心点を $t_{(j)}$ ($j = 1, 2, \dots, M < N$) とおいたものは一般化正則化ネットワーク (Generalized Regularization Network, GRN) と呼ばれ、これによって正則化問題の近似解が得られる。

$$\tilde{f}(\mathbf{x}) = \sum_{j=1}^M \tilde{r}_j K(\mathbf{x}, t_{(j)}) \quad (17)$$

このときの重みパラメータ \tilde{r}_j は、次の M 元連立 1 次方程式を解いて得られる。

$$\tilde{\mathbf{K}}\tilde{\mathbf{r}} + \tilde{\mathbf{z}} = \mathbf{0}. \quad (18)$$

ここで、

$$\begin{aligned} \tilde{\mathbf{K}} &= (D^T D + \lambda J), \\ \tilde{\mathbf{r}} &= (\tilde{r}_1, \dots, \tilde{r}_M)^T, \end{aligned}$$

$$\begin{aligned}
\tilde{z} &= D^T z, \\
D &= (D_{ij}) = (K(\mathbf{x}_i, t_j)), \\
J &= (J_{ij}) = (K(t_i, t_j)),
\end{aligned} \tag{19}$$

とおいた。

3 多価標準正則化理論 (MVSRT)

著者は、前回の報告 [11] において、代数的手法を用いた多価関数の直接表現を用いて、多価関数の正則化問題を標準正則化問題、すなわち、未知関数に関して 2 次の汎関数のエネルギー最小化に帰着させる理論 (多価標準正則化理論, MVSRT) [11] を提案し、ビジョンにおける透明視問題の基本アルゴリズムである、ノイズを含んだ疎なデータからの多重表面復元の超並列アルゴリズムを導いた。

本論文で提案する多価正則化ネットワークも以下の様にこの定式化から導かれる。アルゴリズムのもとになる原理は同一のものであるが、導かれる結果は全く異なるものである。すなわち、正則化による表面復元の場合には、正則化された関数の各点における値そのものを全領域にわたって計算するアルゴリズムであったが、正則化ネットワークの場合には、値ではなく入出力の写像をあらわすネットワークとその重み係数を求めることになり、そこに本質的な違いがある。

3.1 ベクトル値多価関数の一般表現

m 次元ベクトル値関数 $f(x)$ による入出力写像 $y = f(x)$ ($(x, y) \in R^n \times R^m$) を考えよう。ベクトル値関数を多価関数に拡張する場合には、それぞれベクトルの各要素ごとに多価に拡張するだけでは十分でなく、ベクトル値の異なる成分の間で、複数の値の候補の間の対応関係を表せなければならない。このことを数学的に表現するためには、テンソル積を用いた多価関数の表現が必要になる。一般の h 価 m 次元ベクトル値関数の場合には、次式で表される。

$$(y - f_1(x)) \otimes (y - f_2(x)) \otimes \cdots \otimes (y - f_h(x)) = 0, \tag{20}$$

ここで、 \otimes は、テンソル積 (Kronecker 積) を表す。この入出力関係の「基本関係式」は、各学習サンプルデータが、関数 $y = f_k(x)$ ($k = 1, 2, \dots, h$) のいずれか少なくとも一つを満たすという関係を数学的に厳密に表現している。この拘束式は、 m^h 個の連立方程式になる。しかし、問題の本来の意味から、任意の 2 個の関数 $f_i(x)$ と $f_j(x)$ を入れ換えても原問題の意味は変化しないはずである。このことを数学的に表現するには、テンソル積の対称化の操作が必要になる。対称化により、独立な方程式の数は、 $m H_h$ 個になる。これらの点に関する一般論は複雑になるので省略するが、 $m = 2$ 、 $h = 2$ の場合については後述する。

3.2 スカラー値多価関数の一般表現

$R^n \times R$ 上の h 個の超曲面 $y = f_i(\mathbf{x})$ ($i = 1, 2, \dots, h$) を同時に表現する基本関係式は、次の様に表される。

$$\begin{aligned}\Lambda^{(h)}(\mathbf{x}, y) &\stackrel{\text{def}}{=} \prod_{k=1}^h (y - f_k(\mathbf{x})) \\ &= F_1^{(h)}(\mathbf{x}) + F_2^{(h)}(\mathbf{x})y + \dots + F_h^{(h)}(\mathbf{x})y^{h-1} + y^h \\ &= 0\end{aligned}\quad (21)$$

ここで、関数 $F_k(\mathbf{x})$ ($k = 1, 2, \dots, h$) は、関数 $(-f_j(\mathbf{x}))$ ($j = 1, 2, \dots, h$) の基本対称式によって定義される次の関数系である。

$$\begin{aligned}F_1^{(h)}(\mathbf{x}) &\stackrel{\text{def}}{=} (-1)^h f_1(\mathbf{x})f_2(\mathbf{x}) \cdots f_h(\mathbf{x}) \\ &\quad \dots, \\ F_{h-1}^{(h)}(\mathbf{x}) &\stackrel{\text{def}}{=} \sum_{i_1=1}^h \sum_{i_2=i_1+1}^h f_{i_1}(\mathbf{x})f_{i_2}(\mathbf{x}), \\ F_h^{(h)}(\mathbf{x}) &\stackrel{\text{def}}{=} -\sum_{i=1}^h f_i(\mathbf{x})\end{aligned}\quad (22)$$

この関数系の変換によって、方程式 (21) の様に未知関数に関して線形の方程式が得られる。関数系の変換 (22) の逆変換は、 y に関する h 次の代数方程式

$$\Lambda^{(h)}(\mathbf{x}, y) = 0 \quad (23)$$

の h 個の実数解を \mathbf{x} の関数と見なしたものとして定義される。

3.3 h 価スカラー値関数の標準正則化理論

データの基本関係式 (21) を用いて、 h 価関数の標準正則化問題を次のエネルギー汎関数の最小化問題として定式化する。

$$E^{(h)}[F_1^{(h)}, F_2^{(h)}, \dots, F_h^{(h)}] = \sum_{i=1}^N \left\{ \Lambda^{(h)}(\mathbf{x}_{(i)}, y_{(i)}) \right\}^2 + \sum_{k=1}^h \lambda_k \left\| S_k^{(h)} F_k^{(h)} \right\|^2. \quad (24)$$

ここで、 S_k および λ_k は、関数 $F_k^{(h)}$ の正則化作用素、および正則化パラメータを表すが、以下では簡単のため、これらを、すべての未知関数に共通とした、 $S_k = S$, $\lambda_k = \lambda$ の場合を考察する。

3.4 2 価 2 次元ベクトル値関数の標準正則化理論

ベクトル値関数の MVSRT を $m = 2$, $h = 2$ の場合について述べる。この場合、入出力関係は、次式で拘束される。

$$(\mathbf{y} - f_1(\mathbf{x})) \otimes (\mathbf{y} - f_2(\mathbf{x})) = \mathbf{0}, \quad (25)$$

対称化により、 $f_1(\mathbf{x})$ と $f_2(\mathbf{x})$ の役割を同等にすると、

$$\frac{1}{2} \{(\mathbf{y} - f_1(\mathbf{x})) \otimes (\mathbf{y} - f_2(\mathbf{x})) + (\mathbf{y} - f_2(\mathbf{x})) \otimes (\mathbf{y} - f_1(\mathbf{x}))\} = \mathbf{0}, \quad (26)$$

が 2 価の場合の拘束方程式である。各関数の成分関数を次の様におく。

$$\begin{aligned} \mathbf{y} &= f_1(\mathbf{x}) = (f_{1,1}(\mathbf{x}), f_{1,2}(\mathbf{x}))^T, \\ \mathbf{y} &= f_2(\mathbf{x}) = (f_{2,1}(\mathbf{x}), f_{2,2}(\mathbf{x}))^T. \end{aligned} \quad (27)$$

すると、式 (26) は、次の 3 個の方程式と等価である。

$$\begin{aligned} y_1^2 - (f_{1,1}(\mathbf{x}) + f_{2,1}(\mathbf{x}))y_1 + f_{1,1}(\mathbf{x})f_{2,1}(\mathbf{x}) &= 0, \\ y_2^2 - (f_{1,2}(\mathbf{x}) + f_{2,2}(\mathbf{x}))y_2 + f_{1,2}(\mathbf{x})f_{2,2}(\mathbf{x}) &= 0, \\ 2y_1y_2 - (f_{1,2}(\mathbf{x}) + f_{2,2}(\mathbf{x}))y_1 - (f_{1,1}(\mathbf{x}) + f_{2,1}(\mathbf{x}))y_2 \\ + (f_{1,1}(\mathbf{x})f_{2,2}(\mathbf{x}) + f_{1,2}(\mathbf{x})f_{2,1}(\mathbf{x})) &= 0. \end{aligned} \quad (28)$$

ここで、この初めの 2 個の式は、出力空間の各成分について独立に、スカラー値の 2 価関数の表現をしたものになっている。この 2 個の式だけでは、

$$\begin{aligned} \mathbf{y} &= f_1(\mathbf{x}) = (f_{1,1}(\mathbf{x}), f_{2,2}(\mathbf{x}))^T, \\ \mathbf{y} &= f_2(\mathbf{x}) = (f_{2,1}(\mathbf{x}), f_{1,2}(\mathbf{x}))^T. \end{aligned} \quad (29)$$

という出力空間の成分が入れ違いになった偽の解が生じる。3 番目の式によって、この偽の解が禁止され、本来の 2 価ベクトル値関数の表現として正確なものになる。この様に、一般に、式 (20) における様な、テンソル積の形で多価ベクトル値関数を表現した場合に表れる本来の自由度に比較して冗長な方程式は、多価ベクトル値関数の表現としての一意性を保証するために必要なものである。

式 (28) を未知関数に関して線形の形式に変換すると、次の関係式が得られる。

$$\begin{aligned} y_1^2 + G_1(\mathbf{x})y_1 + F_1(\mathbf{x}) &= 0, \\ y_2^2 + G_2(\mathbf{x})y_2 + F_2(\mathbf{x}) &= 0, \\ 2y_1y_2 + G_2(\mathbf{x})y_1 + G_1(\mathbf{x})y_2 + H(\mathbf{x}) &= 0. \end{aligned} \quad (30)$$

ここで、5 個の未知関数は次の様に定義される。

$$\begin{aligned} F_1(\mathbf{x}) &= f_{1,1}(\mathbf{x})f_{2,1}(\mathbf{x}), \\ F_2(\mathbf{x}) &= f_{1,2}(\mathbf{x})f_{2,2}(\mathbf{x}), \\ G_1(\mathbf{x}) &= -(f_{1,1}(\mathbf{x}) + f_{2,1}(\mathbf{x})), \\ G_2(\mathbf{x}) &= -(f_{1,2}(\mathbf{x}) + f_{2,2}(\mathbf{x})), \\ H(\mathbf{x}) &= f_{1,1}(\mathbf{x})f_{2,2}(\mathbf{x}) + f_{1,2}(\mathbf{x})f_{2,1}(\mathbf{x}). \end{aligned} \quad (31)$$

さて、この場合の正則化問題は、式 (30) の 3 個の式を用いて以下の様に 3 個の標準正則化問題に帰着できる。まず、次の 2 個の標準正則化問題より、 $H(\mathbf{x})$ 以外の関

数を求めることができる。

$$E^{(2,2a)}[F_1, G_1] = \sum_{i=1}^N \left\{ \left(y_{1(i)} \right)^2 + G_1(\mathbf{x}(i)) y_{1(i)} + F_1(\mathbf{x}(i)) \right\}^2 + \lambda \|SF_1(\mathbf{x})\|^2 + \lambda \|SG_1(\mathbf{x})\|^2, \quad (32)$$

$$E^{(2,2b)}[F_2, G_2] = \sum_{i=1}^N \left\{ \left(y_{2(i)} \right)^2 + G_2(\mathbf{x}(i)) y_{2(i)} + F_2(\mathbf{x}(i)) \right\}^2 + \lambda \|SF_2(\mathbf{x})\|^2 + \lambda \|SG_2(\mathbf{x})\|^2. \quad (33)$$

これらは、スカラー値の2価正則化理論を入力空間の各座標成分に関して独立に用いたものにほかならない。

これらの正則化問題に基づいて得られる関数 $F_1(\mathbf{x})$, $F_2(\mathbf{x})$, $G_1(\mathbf{x})$, $G_2(\mathbf{x})$ をそれぞれ $\bar{F}_1(\mathbf{x})$, $\bar{F}_2(\mathbf{x})$, $\bar{G}_1(\mathbf{x})$, $\bar{G}_2(\mathbf{x})$ とすると、関数 $H(\mathbf{x})$ に関する正則化問題は次の様に定式化される。

$$E^{(2,2c)}[H] = \sum_{i=1}^N \{ 2y_{1(i)}y_{2(i)} + \bar{G}_2(\mathbf{x}(i))y_{1(i)} + \bar{G}_1(\mathbf{x}(i))y_{2(i)} + H(\mathbf{x}(i)) \}^2 + \lambda \|SH(\mathbf{x})\|^2 \quad (34)$$

これは、 $y'_{(i)} = 2y_{1(i)}y_{2(i)} + \bar{G}_2(\mathbf{x}(i))y_{1(i)} + \bar{G}_1(\mathbf{x}(i))y_{2(i)}$, $f'(\mathbf{x}) = -H(\mathbf{x})$ と置き換えてみればわかる様に、汎関数 (1) と全く同様のスカラー値関数 $f'(\mathbf{x})$ に関する標準正則化問題である。

4 多価正則化ネットワークと学習アルゴリズム

4.1 多価正則化ネットワークの構成理論

以下、 n 次元実数空間 R^n 上の h 価スカラー値関数を近似する正則化ネットワークを MVSRT にもとづいて求める。

エネルギー汎関数 (24) の最小化問題における、平衡条件 (Euler-Lagrange 方程式) は、各未知関数 $F_k^{(h)}$ ($k = 1, 2, \dots, h$) に関する変分を零とおくことにより求められる。

$$\frac{\delta E^{(h)}[F_1^{(h)}, F_2^{(h)}, \dots, F_h^{(h)}]}{\delta F_k^{(h)}} = 0 \quad (35)$$

これを実際に計算すると、次の h 個の方程式が得られる ($k = 1, 2, \dots, h$)。

$$\sum_{i=1}^N \left(y_{(i)} \right)^{k-1} \Lambda^{(h)}(\mathbf{x}, y_{(i)}) \delta(\mathbf{x} - \mathbf{x}(i)) + \lambda \hat{S} S F_k^{(h)}(\mathbf{x}) = 0 \quad (36)$$

この方程式は、

$$\hat{S} S F_k^{(h)}(\mathbf{x}) = -\frac{1}{\lambda} \sum_{i=1}^N \left(y_{(i)} \right)^{k-1} \Lambda^{(h)}(\mathbf{x}, y_{(i)}) \delta(\mathbf{x} - \mathbf{x}(i)) \quad (37)$$

と変形できる。正則化ネットワークの場合と同様にして式 (8) で定義される Green 関数を用いてこの偏微分方程式を解くと次の様に関数 $F_k^{(h)}(\mathbf{x})$ が表される。

$$F_k^{(h)}(\mathbf{x}) = -\frac{1}{\lambda} \sum_{i=1}^N (y_{(i)})^{k-1} \Lambda^{(h)}(\mathbf{x}_{(i)}, y_{(i)}) K(\mathbf{x}, \mathbf{x}_{(i)}) \quad (38)$$

この表現は、各関数 $F_k^{(h)}(\mathbf{x})$ が、Green 関数 $K(\mathbf{x}, \mathbf{x}_{(i)})$ の線形結合で表されることを示している。ここで、

$$r_i^{(h)} = -\frac{1}{\lambda} \Lambda^{(h)}(\mathbf{x}_{(i)}, y_{(i)}) \quad (39)$$

とおくと、式 (38) は、次の形に表現されることがわかる。

$$F_k^{(h)}(\mathbf{x}) = \sum_{i=1}^N r_i^{(h)} (y_{(i)})^{k-1} K(\mathbf{x}, \mathbf{x}_{(i)}) \quad (40)$$

ここで、各 k に関して、係数 $r_i^{(h)}$ が共通になっていることに注意。これは、 N 個の未知係数 $r_i^{(h)}$ によって、 h 個の関数すべての重み係数が決定されることを示している。

図 4 は、以上の結果をもとに図示した h 個関数を近似する正則化ネットワークである。式 (23) において示した様に、関数系の逆変換を行う分解ネットワークには、 h 次の代数方程式を解く機能を与えられればよい。したがって、この部分は、前稿 [11] [12] の多重表面復元でも述べた Durand-Kerner 法にもとづいて、この代数方程式の h 個の解を同時に求める数値解法をリカレント結合のネットワークで実現したり、あるいは、別の正則化ネットワークを用いて逆変換の写像を近似的に実現してしまう方法が考えられる。後者の方法では、「計算」は、フィードフォワード型で、分解ネットワークの重み係数の学習は、もとの問題とは独立に行うことができ、また、解の退化の判定条件なども学習させておくことが可能であるというメリットがある。

4.2 多価正則化ネットワークの学習アルゴリズム

重みパラメータ $r_i^{(h)}$ を決める学習アルゴリズムは、式 (40) において、 \mathbf{x} を $\mathbf{x}_{(i)}$ に、添字 i を j にそれぞれ置き換えた

$$F_k^{(h)}(\mathbf{x}_{(i)}) = \sum_{j=1}^N r_j^{(h)} (y_{(j)})^{k-1} K(\mathbf{x}_{(i)}, \mathbf{x}_{(j)}) \quad (41)$$

を式 (39) に代入して整理すると得られ、次の N 元連立 1 次方程式になる。

$$K^{(h)} \mathbf{r}^{(h)} + \mathbf{z}^{(h)} = 0 \quad (42)$$

ここで、各行列とベクトルは次の様に定義される。

$$\begin{aligned} K^{(h)} &= \left(K_{ij}^{(h)} \right) = \left(\left\{ \sum_{k=1}^h (y_{(i)} y_{(j)})^{k-1} \right\} K(\mathbf{x}_{(i)}, \mathbf{x}_{(j)}) + \lambda \delta_{ij} \right), \\ \mathbf{r}^{(h)} &= \left(r_1^{(h)}, r_2^{(h)}, \dots, r_N^{(h)} \right)^T, \end{aligned}$$

$$z^{(h)} = \left(\{y_{(1)}\}^h, \{y_{(2)}\}^h, \dots, \{y_{(N)}\}^h \right)^T \quad (43)$$

ここで、連立1次方程式の次数が、多価度 h に依存しないという著しい性質をもつことがわかる。連立1次方程式の計算量は、一般に、次数の3乗に比例して増加するので、この性質のメリットは大きい。

4.3 2価2次元ベクトル値の正則化ネットワーク

本節では、2価2次元ベクトル値写像の正則化ネットワーク表現をベクトル値の多価標準正則化理論から導く方法を述べる。

正則化問題 (32), (33) は、スカラー値の2価標準正則化問題と同形式なので、これまで述べてきた2価スカラー値関数のMVRNでネットワークの構成と重みパラメータの学習を行うことができる。 $H(x)$ を求める正則化問題 (34) は、1価関数の標準正則化問題であるから、従来の正則化ネットワークの理論に基づいてネットワーク構成と学習アルゴリズムを導くことができる。

ベクトル値の多価正則化ネットワークの場合には、式 (29) の様な偽の解を取り除くため、式 (30) の第3の拘束

$$2y_1y_2 + G_2(x)y_1 + G_1(x)y_2 + H(x) = 0 \quad (44)$$

に基づいた判定を行うネットワーク構造を付け加える必要がある。

4.4 一般化多価正則化ネットワーク

従来の一般化正則化ネットワーク (GRN) と同様に中間層 = 基底関数の数を N より少なくした一般化されたMVGRNを考えることができる。このネットワークは、基本構造はMVRNと同じであるが、ネットワークの重みの決め方が異なる。1価関数の場合と同様に、基底関数の中心の集合を t_j ($j = 1, 2, \dots, M < N$) とおき、 $F_k(x)$ に対応する近似解を $\tilde{F}_k(x)$ として、

$$\tilde{F}_k(x) = \sum_{j=1}^M \tilde{r}_{k,j} K(x, t_{(j)}) \quad (45)$$

とおく。ここで、残念ながらMVGRNの場合には、関数間で重みパラメータの共有はできず、 hM 個の未知パラメータを決定する必要がある。パラメータの学習アルゴリズムを求めるためには、この式をエネルギー汎関数 (19) に代入して (ただし、 $S_k = S$, $\lambda_k = \lambda$ とした。)、 $\tilde{r}_{k,j}$ ($k = 1, 2, \dots, h; j = 1, 2, \dots, M$) に関して偏微分し、零とおくことにより、次の hM 元連立1次方程式が導かれる。

$$\tilde{K}^{(h)} \tilde{r}^{(h)} + \tilde{z}^{(h)} = 0, \quad (46)$$

ここで、

$$\begin{aligned}
 K^{(h)} &= \begin{bmatrix} D_1^T D_1 + \lambda J & D_1^T D_2 & \dots & D_1^T D_h \\ D_2^T D_1 & D_2^T D_2 + \lambda J & \dots & D_2^T D_h \\ \dots & \dots & \dots & \dots \\ D_h^T D_1 & D_h^T D_2 & \dots & D_h^T D_h + \lambda J \end{bmatrix}, \\
 \tilde{r}^{(h)} &= (\tilde{r}_{1,1}, \dots, \tilde{r}_{1,M}, \tilde{r}_{2,1}, \dots, \tilde{r}_{2,M}, \dots, \tilde{r}_{h,1}, \dots, \tilde{r}_{h,M})^T, \\
 \tilde{z}^{(h)} &= \left((z^{(h)})^T D_1, (z^{(h)})^T D_2, \dots, (z^{(h)})^T D_h \right)^T, \\
 (D_k)_{ij} &= (y_{(i)})^{k-1} K(x_{(i)}, t_{(j)}), \\
 (J)_{ij} &= K(t_{(i)}, t_{(j)}), \\
 z^{(h)} &= \left(\{y_{(1)}\}^h, \{y_{(2)}\}^h, \dots, \{y_{(N)}\}^h \right)^T \tag{47}
 \end{aligned}$$

5 シミュレーションによる理論の検証

Gaussian RBF の場合について、CM-5 上にアルゴリズムをインプリメントし、シミュレーションを行った。分解ネットワークには、Durand-Kerner 法のリカレントネットワークを用いた。サンプルデータは、説明のために 図 1 に示したものを用了。これは、3本の直線上にそれぞれ 500 個のデータ点を平均 0、標準偏差 0.02 の Gaussian ノイズを加えてランダムに発生させたものである。図 5 は、図 1 のデータに対して、 x から y への 3 価写像を近似する正則化ネットワークを構成し、学習によって重み係数を求め、できた写像をプロットしたものである。基底関数には、 $\sigma = 0.1$ の Gaussian を用いた。正則化パラメータは、 $\lambda = 10.0$ とした場合を示している。

6 多価正則化ネットワークによるビジョンのアルゴリズムの学習による実現

ビジョンやパターン認識の多くの問題は、画像から、ある空間への写像を構成する問題と考えられる。従って、本稿で提案した正則化ネットワークは、これらにおいて、一般に多価写像が必要になる場合に適用することができる。以下、ビジョンの場合について考察する。

例えば、コンピュータビジョンでは、適当な仮定のもとで導かれる画像生成過程の光学的、幾何学的基本拘束条件と「滑らかさ」などの事前知識を導入し、適当な近似手法や最適化手法を用いて解析的あるいは数値的手段でこの写像を求めることによって必要なアルゴリズムを構築する。しかし、

1. この基本拘束条件を厳密にあるいは近似的に求めることは、かなり単純化された状況以外では一般に難しい。
2. それにもまして、この基本拘束条件にもとづいて逆問題を解く解析的、あるいは数値的アルゴリズムを求めることは、一般に非常に困難な問題である。

3. 解析的アルゴリズムでは、一般に、データに誤差が含まれない場合には厳密解が得られる。しかし、多くのビジョンのタスクでは、厳密解が得られる必要はかならずしもない。

一方、例からの学習によって、ビジョンの問題を解く写像を関数近似を行うネットワークなどを用いて求めればよいという立場も存在する。これは、基本拘束条件から出発する方法をあきらめて、ビジョンの逆問題を解く写像を近似的に構成できればよいと考える。この写像の構成による方法には、一般に、次の様な問題が存在する。

1. 一般に逆問題では、たとえ「滑らかさ」などの事前知識による拘束を用いても最終的な解が複数組（有限個）存在する。これは、多義的知覚、多義的解釈、透明視、複数物体が存在する場合などである。従って、逆問題の写像は、一般に多価写像となる。
2. 例からの学習は不良設定問題であり、写像の「滑らかさ」の様な事前知識が必要である。しかし、一般に逆問題の写像は、至るところ滑らかであることは無く、特異点や不連続が存在する。このような写像の多価性、特異点、不連続の存在は、一般に解析的手段で知ることができるものであり、解析的アルゴリズムでは、方程式が退化する場合や、符号の判定などによる場合分けに相当するものである。問題が複雑になるにつれてこの種の解析は非常に繁雑になる。

本報告の多価写像の近似理論によって、解析的手段によらずに逆問題における多義的解釈、多重属性や不連続を学習によってフィードフォワードネットワークに実現することができると考えられる。

7 まとめ

標準正則化理論に基づいた、関数近似 = 学習の理論である正則化ネットワーク (RN) の理論を拡張し、多価関数を近似することができる多価正則化ネットワーク (MVRN) の理論を構築した。MVRN は、もともとは透明視における多重表面復元モデルのために、標準正則化理論を拡張して得られた多価標準正則化理論 (MVSRT) に基づいて構築された。MVRN によると、学習サンプルデータ数と同一の次元の連立1次方程式を解くだけで多価関数の近似を行う多層フィードフォワードネットワークが構成できる。また、中間層 = 基底関数の数を学習サンプル数よりも少なくした一般化正則化ネットワーク (GRN) の多価関数への拡張である多価一般化正則化ネットワーク (MVGRN) の理論も導いた。これにより、ビジョンをはじめ、一般に、逆問題が多価写像になる様な（生体）情報処理課題における逆モデルの学習による実現が容易になるものと考えられる。

今回は基底関数の中心とノルムが固定された場合の理論を展開したが、これらも学習可能にすると、より高精度な写像が、少ない基底関数で学習できる様になる。この拡

張は、Poggio & Girosi [7] における方法と同様にして実現できると考えられるが今後の課題としたい。

MVRN では、関数系の置き換えによって、一つの正則化ネットワークで多価関数を近似したが、目的によっては、別々のネットワークに学習させたい場合も存在する。この場合にも基本関係式 (20) を学習アルゴリズムの導出に用いることができる。すなわち、この基本関係式の 2 乗残差を誤差関数とすることにより、学習アルゴリズムを導くことができる。このアルゴリズムは、係数が相互に依存しあった h 組の連立 1 次方程式群となり、単純な数値計算では、最適解が得られる保証がなくなるが、従来のモジュラーネットワークよりも見通しのよい計算になる。また、これと同様にすれば、多価関数の基本関係式 (20) は、正則化ネットワークに限らず、多層パーセプトロンなど従来のフィードフォワードネットワークを用いたモジュラーネットワークによる多価写像の学習にも用いることができるであろう。

謝辞 本研究を始めるにあたり有益な討論をしていただいた、MIT 人工知能研究所 Tomaso Poggio 教授、Federico Girosi 博士に感謝します。また、ご討論頂いた ATR 人間情報通信研究所 佐藤雅昭博士、安藤広志博士、水谷博之氏に感謝します。シミュレーションプログラム作成にご協力頂いた、株式会社 CSK 吉川徹氏に感謝します。最後に、本研究の機会を与えていただいた ATR 人間情報通信研究所 東倉洋一社長、赤松茂室長に感謝します。

参考文献

- [1] Ferdinando A, Mussa-Ivaldi FA (1992) From basis functions to basis fields: vector field approximation from sparse data. *Biol Cybern* 567:479-489
- [2] Girosi F, Poggio T (1990) Networks and the best approximation property. *Biol Cybern* 63: 169-176
- [3] Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE (1991) Adaptive mixtures of local experts. *Neural Comput.* 3: 79-87
- [4] 並木美喜雄 (1982) デルタ関数と微分方程式、岩波
- [5] Poggio T, Torre V, Koch C (1985) Computational vision and regularization theory. *Nature* 317:314-319
- [6] Poggio T (1990) A theory of how the brain might work. *Proc Cold Harbor Symp on Quantitative Biology IV*: 899-910
- [7] Poggio T, Girosi F (1990) Networks for approximation and learning. *Proc IEEE* 78(9): 1481-1497
- [8] Poggio T, Girosi F (1990) Regularization algorithms that are equivalent to multilayer networks. *Science* 247: 978-982
- [9] Poggio T, Edelman S (1990) A network that learns to recognize 3D Objects. *Nature* 343: 263-266
- [10] Poggio T, Edelman S, Fahle M (1992) Learning of visual modules from examples: A framework for understanding adaptive visual performance. *CVGIP: Image Understanding* 56(1): 22-30
- [11] 志沢雅彦 (1993.9) 「標準正則化理論の多価関数への拡張と多重表面の復元」信学会ニューロコンピューティング研究会、NC93-34
- [12] Shizawa M (1993.10) Multi-Valued Standard Regularization Theory (1): Global Reconstruction of Multiple Transparent Surfaces via Massively Parallel Relaxation Algorithms. ATR technical report TR-H-037
- [13] Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. W.H.Winston Washington DC

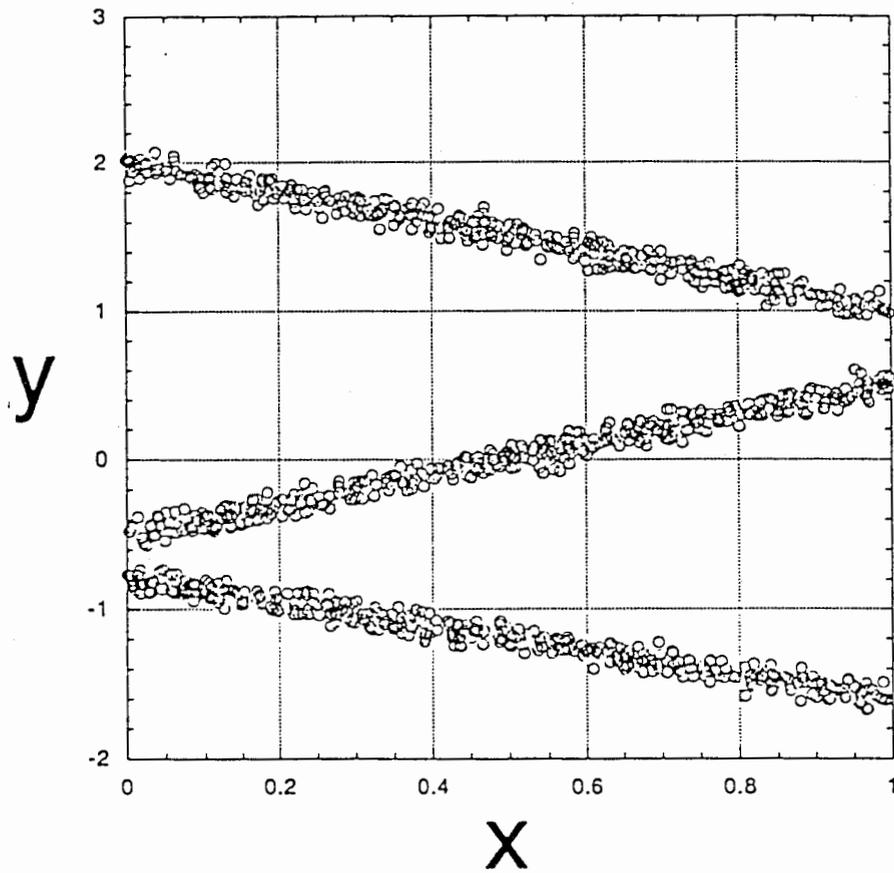


図 1: 3 値写像によってモデル化されるべき入出力学習データの例

Fig.1 Input-output learning sample data which is to be modeled by a three-valued mapping.

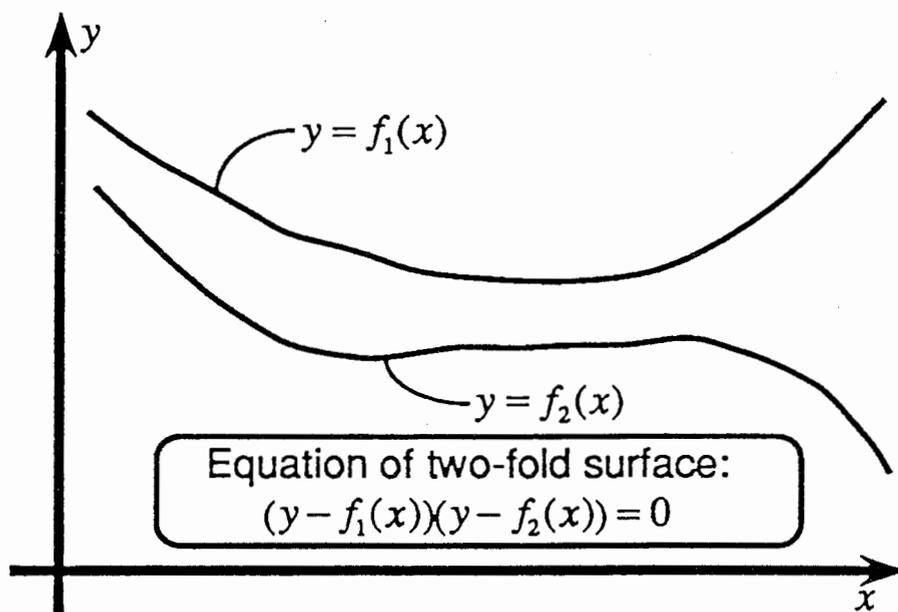


図 2: 多価写像の代数式による直接表現

Fig.2 Direct representation for multi-valued mapping using an algebraic equation.

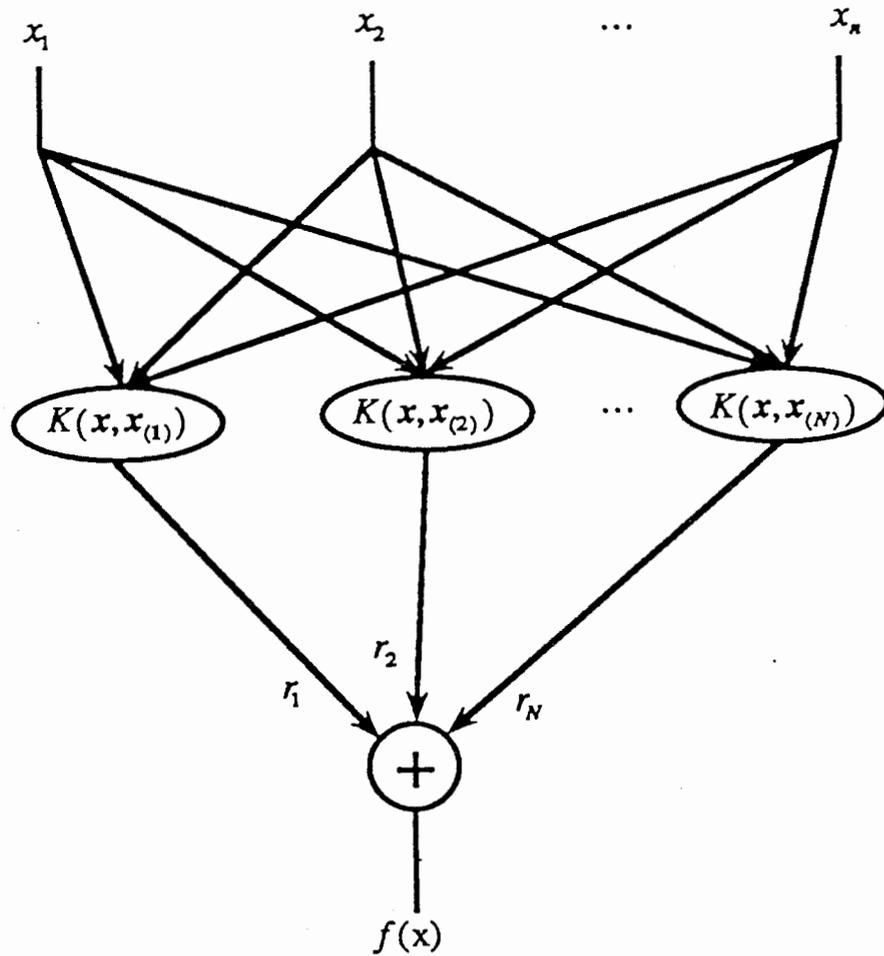


図 3: (1 価) 正則化ネットワーク
 Fig.3 (Single-valued) regularization network.

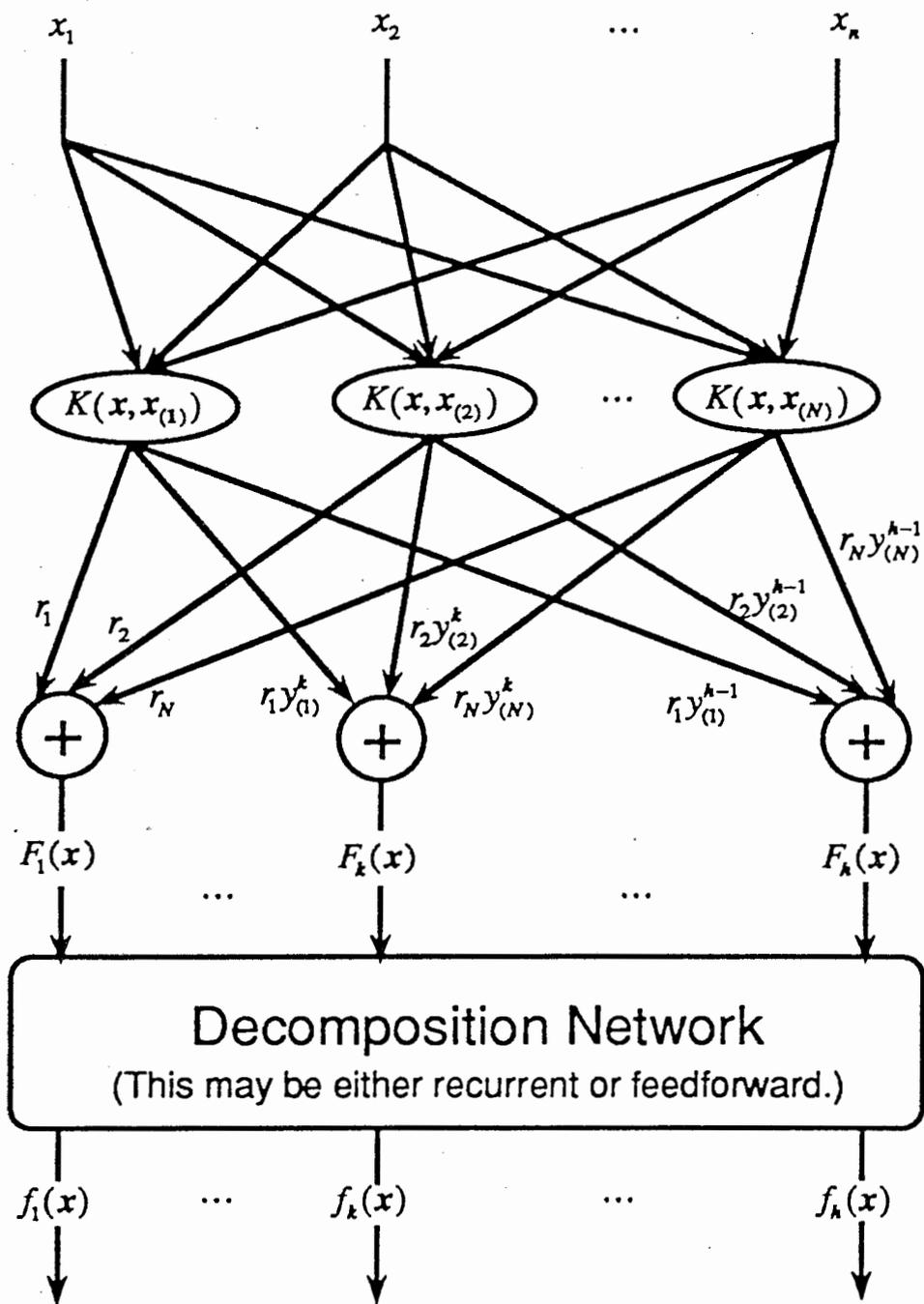


図 4: h 価正則化ネットワーク
 Fig.4 h -valued regularization network.

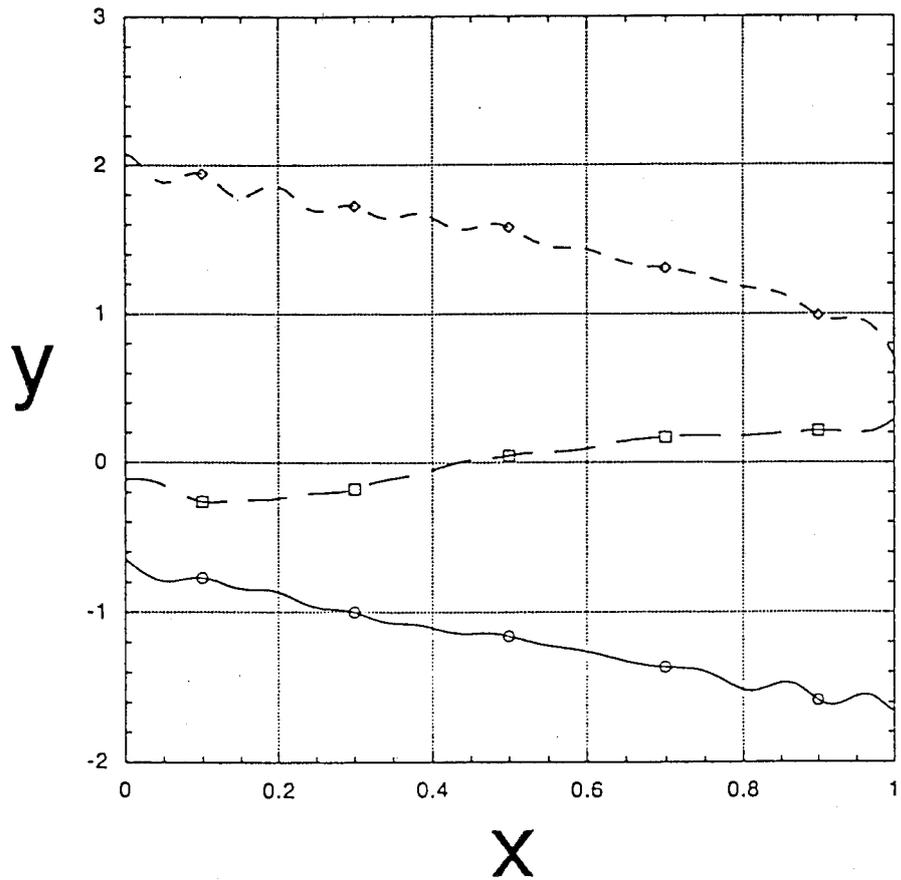


図 5: シミュレーション結果の例 (3 価正則化ネットワークによる近似)

Fig.5 Simulation results of three-valued regularization network.