

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年7月14日(14.07.2022)



(10) 国際公開番号

WO 2022/149240 A1

(51) 国際特許分類:
G06N 99/00 (2019.01)

(21) 国際出願番号: PCT/JP2021/000358

(22) 国際出願日: 2021年1月7日(07.01.2021)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 橋本 悠香 (HASHIMOTO, Yuka); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT知的財産センタ内 Tokyo (JP).

(74) 代理人: 伊東 忠重, 外 (ITOH, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1

番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).

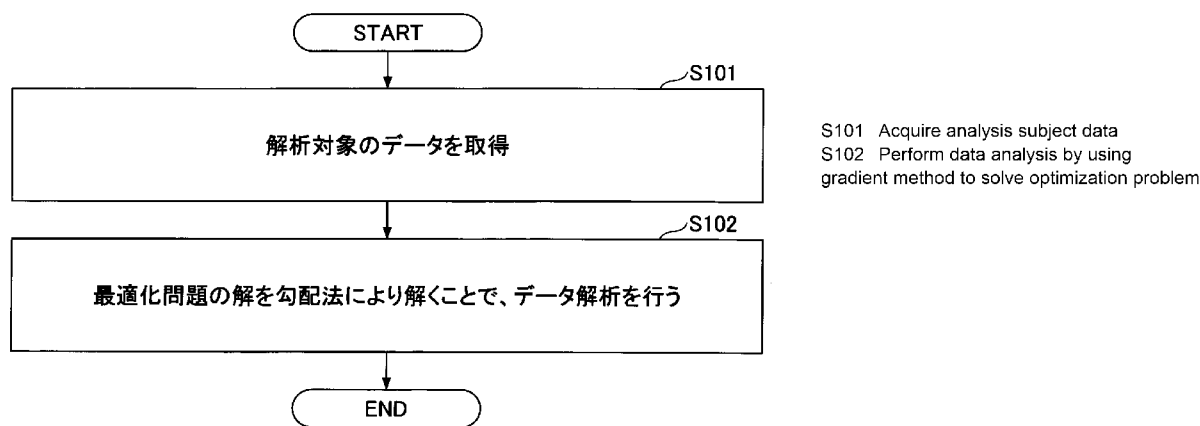
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,

(54) Title: ANALYSIS DEVICE, ANALYSIS METHOD, AND PROGRAM

(54) 発明の名称: 解析装置、解析方法及びプログラム

[図3]



(57) Abstract: This analysis device according to one embodiment comprises: an acquisition unit that acquires, as an analysis subject, a data set having a complex structure; and an optimization calculation unit that defines the data set in a space reproduced in a reproducing kernel Hilbert C*-module (RKHM) and calculates, using a method expanding the gradient method in a Hilbert space, a solution to an optimization problem related to a function taking a C*-algebra value.

(57) 要約: 一実施形態に係る解析装置は、複雑な構造を持つデータの集合を解析対象として取得する取得部と、前記データの集合をRKHM (reproducing kernel Hilbert C*-module) に写した空間上で定義され、かつ、C*-algebra値をとる関数に関する最適化問題の解を、Hilbert空間における勾配法を拡張した手法により計算する最適化計算部と、を有する。

[続葉有]

ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告（条約第21条(3)）

明 細 書

発明の名称： 解析装置、解析方法及びプログラム

技術分野

[0001] 本発明は、解析装置、解析方法及びプログラムに関する。

背景技術

[0002] 自然界や人間社会で現れるデータの多くは複雑な構造を持っている。例えば、時空間上の各点における気象データは連続的な関数から生成されているとみなすことができる。また、例えば、SNS (social networking service) におけるフォロー関係を表すデータは、複数の構成要素が相互に関係し合っている。このような複雑な構造を持つデータの解析は、統計学、機械学習、物理学、分子動力学等の分野における共通の課題である。これに対して、統計学や機械学習の分野では、元のデータを、RKHS (reproducing kernel Hilbert space) と呼ばれる無限次元の関数空間に写し、RKHS上で解析を行う技術が提案されている（非特許文献1及び2）。RKHSは元のデータを表現する空間よりも高次元であるため、豊富な表現力を持つ。このため、元のデータをRKHSに写すことで、複雑な振る舞いをするデータが単純な振る舞いをするデータに変換され、より高精度な解析が可能となる。また、RKHSはデータ解析における標準的な性質を持つHilbert空間と呼ばれる空間であるため、Hilbert空間における勾配法の1つである最急降下法（非特許文献3）等の標準的なアルゴリズムを適用することができる。

[0003] RKHSの一般化としてRKHM (reproducing kernel Hilbert C*-module) と呼ばれる空間も提案されており、物理学の分野においては、その理論的な解析が行われている（非特許文献4）。RKHMはC*-algebraと呼ばれる空間に値を持つ関数の空間である。C*-algebraとは複素数全体の集合や行列全体の集合等の一般化であり、共役とノルムの概念を持つ空間である。C*-algebraは複素数よりも複雑な数学的構造を持つため、複雑なデータを表現するために有用であると期待されている。

先行技術文献

非特許文献

[0004] 非特許文献1: K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, Kernel Mean Embedding of Distributions: A Review and Beyond. Foundations and Trends in Machine Learning. 10 (1-2), 1-141, 2017.

非特許文献2: Y. Hashimoto, I. Ishikawa, M. Ikeda, Y. Matsuo, and Y. Kawahara, "Krylov Subspace Method for Nonlinear Dynamical Systems with Random Noise," JMLR, 21 (172), 1-29, 2020.

非特許文献3: G. Smyrlis and V. Zisis, "Local convergence of the steepest descent method in Hilbert spaces," Journal of Mathematical Analysis and Applications, 300(2):436-453, 2004.

非特許文献4: Jaeseong Heo, "Reproducing kernel Hilbert C*-modules and kernels associated with cocycles," J. Math. Phys. 49, 103507, 2008.

発明の概要

発明が解決しようとする課題

[0005] ところで、RKHSは無限次元の空間であるため、元のデータが単純な構造であれば（つまり、低次元の空間で表現されるデータであれば）、RKHSは元の空間よりも豊富な表現力を持つ。しかしながら、元のデータが複雑な構造を持つ場合（つまり、高次元又は無限次元の空間で表現されるデータの場合）、RKHSの表現力では不十分である。実際、データ間の近さ（内積）は1つの数値（スカラー値）であるため、元のデータが高次元である場合に内積を計算すると、データを構成する各要素同士の近さの情報は失われてしまう。

[0006] そこで、高次元又は無限次元のデータを表現するために、RKHSよりも更に大きな空間が必要となる。RKHMはRKHSを一般化した空間であるため、RKHSよりも大きな空間で豊富な表現力を持つ。このため、高次元又は無限次元のデータはRKHMを用いて表現することで、より高精度な解析が可能になることが期待される。しかしながら、RKHMをデータ解析に応用した例はこれまでにほ

とどなく、RKHMをデータ解析に応用する新たな枠組みが必要であると考えられる。

[0007] 本発明の一実施形態は、上記の点に鑑みてなされたもので、RKHM上のデータ解析を実現することを目的とする。

課題を解決するための手段

[0008] 上記目的を達成するため、一実施形態に係る解析装置は、複雑な構造を持つデータの集合を解析対象として取得する取得部と、前記データの集合をRKHM (reproducing kernel Hilbert C*-module) に写した空間上で定義され、かつ、C*-algebra値をとる関数に関する最適化問題の解を、Hilbert空間における勾配法を拡張した手法により計算する最適化計算部と、を有する。

発明の効果

[0009] RKHM上のデータ解析を実現することができる。

図面の簡単な説明

[0010] [図1]本実施形態に係る解析装置のハードウェア構成の一例を示す図である。

[図2]本実施形態に係る解析装置の機能構成の一例を示す図である。

[図3]本実施形態に係る解析処理の流れの一例を示すフローチャートである。

[図4A]提案手法による主成分分析の結果を可視化した場合の一例を示す図である。

[図4B]既存手法による主成分分析の結果を可視化した場合の一例を示す図である。

[図5]提案手法により相互作用を解析した場合の密度関数値の一例を示す図である。

発明を実施するための形態

[0011] 以下、本発明の一実施形態について説明する。本実施形態では、複雑な構造を持つデータ（例えば、多変量データや関数データ等）を対象として、RKHM上のデータ解析を実現することができる解析装置10について説明する。特に、本実施形態では、C*-algebraに値を持つ関数に関する最適化問題を解く

必要があるデータ解析を想定し、Hilbert空間における勾配法を拡張することで、この最適化問題を解析装置10により解く場合について説明する。これにより、複雑な構造を持つデータの性質を保存しながら必要な情報を抽出（例えば、データ間の関係性を表す情報等を抽出）したり、データを可視化したりする等の様々なデータ解析の実現が可能となる。

[0012] また、本実施形態では、C*-algebraに値を持つ関数に関する最適化問題を解く必要があるデータ解析への応用例として、主成分分析（PCA：principal component analysis）とデータの構成要素間の相互作用の解析についても説明する。

[0013] <理論的構成>

まず、本実施形態の理論的構成について説明する。

[0014] <<準備>>

データが属する空間を X として、 $\{x_0, x_1, \dots\} \subseteq X$ を与えられたデータの集合とする。 A をC*-algebraとして、 A 値positive definite kernel $k : X \times X \rightarrow A$ を考える。ただし、写像 $k : X \times X \rightarrow A$ が A 値positive definite kernel であるとは、以下の条件1及び条件2を満たすことをいう。

[0015] （条件1）任意の $x, y \in X$ に対して、 $k(x, y) = k(y, x)^*$ （ $*$ は共役を表す）

（条件2） m を任意の自然数として、任意の $x_0, x_1, \dots, x_{m-1} \in X$ と任意の $c_0, c_1, \dots, c_{m-1} \in A$ に対して、

[0016] [数1]

$$\sum_{t=0}^{m-1} \sum_{s=0}^{m-1} c_t^* k(x_t, x_s) c_s$$

はpositive

ここで、positiveはC*-algebraで正定値であることを意味し、全ての固有値が0以上であるエルミート行列（つまり、エルミート正定値）等の一般化である。

[0017] A値positive definite kernel k が与えられたとき、 X から A 値関数への写像 ϕ を、 $\phi(x) = k(\cdot, x)$ により定義する。この写像 ϕ はfeature mapとも呼ばれる。

[0018] 自然数 m と、 $x_0, x_1, \dots, x_{m-1} \in X$ と、 $c_0, c_1, \dots, c_{m-1} \in A$ に対して、

[0019] [数2]

$$\sum_{t=0}^{m-1} \phi(x_t) c_t$$

全体から構成される空間を $M_{k,0}$ とする。また、 m, m' を自然数、 $x_0, x_1, \dots, x_{m-1}, y_0, y_1, \dots, y_{m'-1} \in X$ 、 $c_0, c_1, \dots, c_{m-1}, d_0, d_1, \dots, d_{m'-1} \in A$ として、 $M_{k,0}$ に以下で定義される演算 $\langle \cdot, \cdot \rangle_k$ を定義する。

[0020] [数3]

$$\left\langle \sum_{t=0}^{m-1} \phi(x_t) c_t, \sum_{t=0}^{m'-1} \phi(y_t) d_t \right\rangle_k = \sum_{t=0}^{m-1} \sum_{s=0}^{m'-1} c_t^* k(x_t, y_s) d_s$$

このように定義された演算 $\langle \cdot, \cdot \rangle_k$ は A 値内積の性質を持つ。すなわち、 $u, v, w \in M_{k,0}$ と $c, d \in A$ に対して、以下の4つの性質を有する。

- [0021] ・ $\langle u, v \rangle_k = \langle v, u \rangle_k^*$
 ・ $\langle u, u \rangle_k$ は positive
 ・ $\langle u, u \rangle_k = 0$ と $u = 0$ は同値
 ・ $\langle u, v c + w d \rangle_k = \langle u, v \rangle_k c + \langle u, w \rangle_k d$
 この内積 $\langle \cdot, \cdot \rangle_k$ を用いて、複素数値のノルム

[0022] [数4]

$$\|v\|_k = \|\langle v, v \rangle_k\|^{\frac{1}{2}}$$

を定めることができる。このノルムに関して $M_{k,0}$ を完備化した空間を M_k と表し、 k に関する reproducing kernel Hilbert C^* -module (RKHM) と呼ぶ。

M_k は一意に構成することができる。また、 M_k には A 値の大きさ $|\cdot|_k$ も、

[0023] [数5]

$$|v|_k = \langle v, v \rangle_k^{\frac{1}{2}}$$

により定義することができる。

[0024] 《 A 値関数に対する最急降下法 》

A を可換な C^* -algebra とする。このとき、 $f : M \rightarrow A_r$ を Hilbert C^* -module M から A_r への写像 (関数) とし、 $f(v)$ を最小にする $v \in M$ を求めることを考える。ただし、 $A_r = \{c - d \mid c, d \in A_+\}$ 、 A_+ は positive な要素全体から成る集合である。

[0025] f の $v \in M$ における $u \in M$ 方向の微分 $D f_v(u)$ は、

[0026]

[数6]

$$\frac{d}{dt}f(v + tu)|_{t=0}$$

で定義される。 Df_v は M から A への写像であるが、もし Df_v が A 線形であれば、ある $w \in M$ が一意に存在し、 $Df_v(u) = \langle w, u \rangle$ を満たす。この $w \in M$ を ∇f_v と表し、 f の v における勾配と呼ぶ。ただし、写像 $L : M \rightarrow A$ が A 線形とは、 $u, v \in M, c \in A$ に対して $L(u + v) = L(u) + L(v)$ と $L(uc) = L(u)c$ とが成り立つことである。

[0027] Hilbert C^* -module M から可換な C^* -algebra A への関数 f に対する最急降下法のアルゴリズムは以下の手続き1及び2で構成される。

[0028] 手続き1： $v_0 \in M$ をランダムに選ぶ。また、パラメータ $\varepsilon > 0$ を選ぶ。

[0029] 手続き2： $t = 0, 1, \dots$ に対して、

[0030] [数7]

$$v_{t+1} = v_t + \varepsilon \nabla f_{v_t}$$

により v_t の更新を行う。そして、 $v_{t+1} - v_t$ が十分に小さくなったら（例えば、予め決めた閾値未満となったら）更新を終了する。

[0031] ε が十分に小さければ、上記の手続き2で求めた v_{t+1} に対する $f(v_{t+1})$ は $f(v_t)$ よりも小さくなることが示せる。これにより、 $f(v)$ を最小にする $v \in M$ が求まる。

[0032] ここで、上記のアルゴリズムに確率的な作用を追加して確率的勾配降下法に拡張する等、一般的な勾配法に関するテクニックは本実施形態でも同様に

適用可能である。また、 ∇f_v を低次元の空間に射影して近似的に計算する等の工夫を行えば、計算量を削減することも可能である。

[0033] なお、上記では、 M はHilbert C^* -moduleであるものとして関数 f に対する最急降下法のアルゴリズムを説明したが、 $M=M_k$ 、つまり M はRKHMとしても上記の最急降下法のアルゴリズムは当然に適用可能である。

[0034] <解析装置10のハードウェア構成>

次に、本実施形態に係る解析装置10のハードウェア構成について、図1を参照しながら説明する。図1は、本実施形態に係る解析装置10のハードウェア構成の一例を示す図である。

[0035] 図1に示すように、本実施形態に係る解析装置10は一般的なコンピュータ又はコンピュータシステムで実現され、入力装置101と、表示装置102と、外部I/F103と、通信I/F104と、プロセッサ105と、メモリ装置106とを有する。これらの各ハードウェアは、それぞれがバス107を介して通信可能に接続されている。

[0036] 入力装置101は、例えば、キーボードやマウス、タッチパネル等である。表示装置102は、例えば、ディスプレイ等である。なお、解析装置10は、入力装置101及び表示装置102のうちの少なくとも一方を有していてもよい。

[0037] 外部I/F103は、記録媒体103a等の外部装置とのインタフェースである。記録媒体103aとしては、例えば、CD (Compact Disc)、DVD (Digital Versatile Disk)、SDメモリカード (Secure Digital memory card)、USB (Universal Serial Bus) メモリカード等がある。

[0038] 通信I/F104は、解析装置10を通信ネットワークに接続するためのインタフェースである。プロセッサ105は、例えば、CPU (Central Processing Unit) やGPU (Graphics Processing Unit) 等の各種演算装置である。メモリ装置106は、例えば、HDD (Hard Disk Drive) やSSD (Solid State Drive)、RAM (Random Access Memory)、ROM (Read Only Memory)、フラッシュメモリ等の各種記憶装置である。

[0039] 本実施形態に係る解析装置 10 は、図 1 に示すハードウェア構成を有することにより、後述する解析処理を実現することができる。なお、図 1 に示すハードウェア構成は一例であって、解析装置 10 は、他のハードウェア構成を有していてもよい。例えば、解析装置 10 は、複数のプロセッサ 105 を有していてもよいし、複数のメモリ装置 106 を有していてもよい。

[0040] <解析装置 10 の機能構成>

次に、本実施形態に係る解析装置 10 の機能構成について、図 2 を参照しながら説明する。図 2 は、本実施形態に係る解析装置 10 の機能構成の一例を示す図である。

[0041] 図 2 に示すように、本実施形態に係る解析装置 10 は、取得部 201 と、解析部 202 とを有する。これら各部は、例えば、解析装置 10 にインストールされた 1 以上のプログラムがプロセッサ 105 に実行させる処理により実現される。

[0042] また、本実施形態に係る解析装置 10 は、記憶部 203 を有する。記憶部 203 は、例えば、メモリ装置 106 により実現される。なお、記憶部 203 は、例えば、解析装置 10 と通信ネットワークを介して接続される記憶装置（データベースサーバ等）により実現されていてもよい。

[0043] 記憶部 203 には、複雑な構造を持つデータの集合（例えば、 $\{x_0, x_1, \dots, x_{m-1}\} \subseteq X$ ）が記憶されている。

[0044] 取得部 201 は、記憶部 203 に記憶されているデータの集合を解析対象のデータとして取得する。

[0045] 解析部 202 は、取得部 201 により取得されたデータの集合に対して RKHM 上で様々なデータ解析を行う。すなわち、解析部 202 は、取得部 201 により取得されたデータの集合を写像 ϕ により RKHM M_k 上に写した上で、RKHM M_k 上で様々なデータ解析を行う。ここで、解析部 202 には最適化計算部 210 が含まれており、データ解析の際に必要な最適化問題（つまり、 C^* -algebra に値を持つ関数に関する最適化問題）の解を、最適化計算部 210 が上記の手続き 1 及び 2 を実行することで計算する。なお、最適化問題はどの

ようなデータ解析を行うかによって、そのデータ解析に応じて定式化される。

[0046] なお、図2に示す解析装置10の機能構成は一例であって、他の機能構成であってもよい。例えば、最適化計算部210が解析部202に含まれるのではなく、解析部202の外にあってもよい。

[0047] <解析処理の流れ>

次に、本実施形態に係る解析処理の流れについて、図3を参照しながら説明する。図3は、本実施形態に係る解析処理の流れの一例を示すフローチャートである。

[0048] まず、取得部201は、記憶部203に記憶されているデータの集合を解析対象のデータとして取得する（ステップS101）。

[0049] 次に、解析部202は、上記のステップS101で取得したデータをRKHM M_k 上で解析する（ステップS102）。このとき、解析部202は、最適化計算部210により、当該データ解析に応じて定式化された最適化問題の解を上記の手続き1及び2によって計算する。これにより、複雑な構造を持つデータの解析をRKHM上で実現することができる。

[0050] <データ解析への応用例>

次に、C*-algebraに値を持つ関数に関する最適化問題を解く必要があるデータ解析への応用例として、主成分分析とデータの構成要素間の相互作用の解析について説明する。

[0051] <<主成分分析>>

取得部201によって取得されたデータを $x_1, \dots, x_n$ とする。ここでは特に、データ $x_1, \dots, x_n$ は高次元又は関数に値をとるような無限次元のデータであるとする。このとき、以下の式(1)に示す最適化問題を解くことにより、データの性質を最もよく表す空間の軸 $p_1, \dots, p_r$ と、軸 p_j における x_i の座標 $\langle p_j, \phi(x_i) \rangle_k$ とを求める。

[0052]

[数8]

$$\inf_{\{p_j\}_{j=1}^r \subseteq \mathcal{M}_k: \text{ONS}} \sum_{i=1}^n \left| \phi(x_i) - \sum_{j=1}^r p_j \langle p_j, \phi(x_i) \rangle_k \right|_k^2 \quad (1)$$

ただし、ONSとは正規直交系を表す。

[0053] これにより、データ x_i を、 $\langle p_j, \phi(x_i) \rangle_k \in A$ に縮約できる。例えば、 x_i が多変数関数である場合、 A を 1 変数関数の空間に選べば、多変数関数を 1 変数関数に縮約できる。

[0054] 上記の式 (1) に示す最適化問題の解が存在するとすれば、その解は、ある $c_i \in A$ に対して

[0055] [数9]

$$p_j = \sum_{i=1}^n \phi(x_{i,j}) c_{i,j}$$

と表せることが示せる。これより、上記の式 (1) に示す最適化問題の解は、 A が可換な場合と可換でない場合とのそれぞれで以下のように計算できる。

[0056] ・ A が可換な場合

A が可換な場合、上記の式 (1) に示す最適化問題は、以下の式 (2) のように変形できる。

[0057]

[数10]

$$\inf_{\{\sqrt{G}c_j\}_{j=1}^r \subseteq A^n: \text{ONS}} - \sum_{j=1}^r c_j^* G^2 c_j \quad (2)$$

ただし、 $c_j = [c_{1,j}, \dots, c_{n,j}]^T \in A^n$ であり、 G は (i, j) 成分が $k(x_i, x_j) \in A$ の $n \times n$ 行列である。

[0058] そして、 $r = 1$ のとき、上記の式(2)に示す最適化問題を以下の式(3)に示す制約付き最適化問題に変換し、最適化計算部210が上記の手続き1及び2によって解を計算する。

[0059] [数11]

$$\inf_{\sqrt{G}c \in A^n} (-c^* G^2 c + \lambda(c^* G c - I)^*(c^* G c - I)) \quad (3)$$

ただし、 $\lambda > 0$ はパラメータ、 I は恒等写像である。

[0060] $r > 1$ のときは、 $r - 1$ までに得られた解を $c_1, \dots, c_{r-1}$ とし、 A^n の中で $\sqrt{G}c_1, \dots, \sqrt{G}c_{r-1}$ と直交する空間上において、上記の式(3)に示す最適化問題を最適化計算部210が上記の手続き1及び2によって解くことで、解 c_r を計算する。これを $r = 2, \dots$ に対して繰り返すことで、解 $c_1, \dots, c_s$ （ただし、 $s \leq r$ ）を得る。なお、各 r に対して、 $\sqrt{G}c_r$ は、 $(\sqrt{G}) \times c_r$ のことを意味する。

[0061] 具体的には、 $r = 2$ のときは、 $r = 1$ のときに得られた解を c_1 とし、 A^n の中で $\sqrt{G}c_1$ と直交する空間上において、上記の式(3)に示す最適化問題を最適化計算部210が解くことで解 c_2 を計算する。次に、 $r = 3$ のときは、 A^n の中で $\sqrt{G}c_1, \sqrt{G}c_2$ と直交する空間上において、上記の式(3)に

示す最適化問題を最適化計算部 210 が解くことで解 c_3 を計算する。次に、 $r = 4$ のときは、 A^n の中で $\sqrt{G}c_1$, $\sqrt{G}c_2$, $\sqrt{G}c_3$ と直交する空間上において、上記の式 (3) に示す最適化問題を最適化計算部 210 が解くことで解 c_4 を計算する。以降も同様に、 $r = 5, \dots, s-1$ に対して、これを繰り返す。

[0062] ・ A が可換でない場合

A が可換でない場合は、 A の要素は、ある Hilbert 空間 W 上の有界線形作用素と同一視できることが知られている。そこで、 $p_j \in M_k$ は任意の $x \in X$ に対して $p_j(x)$ がランク t の線形作用素になると仮定し、上記の式 (1) に示す最適化問題のトレースをとった以下の問題を考える。

[0063] [数12]

$$\inf_{\{\sqrt{G}c_j\}_{j=1}^r \subseteq W^n: \text{ONS}} \sum_{j=1}^r \text{tr} \left(-c_j^* G^2 c_j + \lambda (c_j^* G c_j - I)^* (c_j^* G c_j - I) \right)$$

この問題は Hilbert 空間 W^n 上の実数値関数に対する最小化問題であるため、Hilbert 空間における既知の最急降下法を用いて解くことができる。したがって、この場合は、最適化計算部 210 が、Hilbert 空間における既知の最急降下法によって上記の最小化問題を解けばよい。

[0064] ≪ データの構成要素間の相互作用の解析 ≫

区間 $[0, 1]$ 上で定義され、かつ、 X に値を取る 2 変数の連続関数により記述されるデータを $x_1, \dots, x_n$ とする。このようなデータとしては、例えば、空間上に連続的に分布したデータや観測点が変化するデータ等が挙げられる。また、実数又は複素数値のデータを $y_1, \dots, y_n$ とする。これらのデータ $x_1, \dots, x_n, y_1, \dots, y_n$ は取得部 201 によって取得される。

[0065] また、 x_i と y_i はある関数 f により、

[0066] [数13]

$$y_i = \int_0^1 \int_0^1 f(s, t, x_i(s, t)) ds dt$$

と表される関係にあるものとする。

[0067] A を、 $[0, 1]$ 上のある関数に対する有界線形作用素全体から成る空間とする。また、 x_i に対して、

[0068] [数14]

$$\mu_{x_i}$$

を、 A 値測度で、

[0069] [数15]

$$\langle 1_s, \mu_{x_i} 1_t \rangle_{L^2} = \delta_{x_i(s, t)}$$

を満たすものとする。ただし、

[0070]

[数16]

$$\langle \cdot, \cdot \rangle_{L^2}$$

は、 $[0, 1]$ 上の関数 u, v に対して

[0071] [数17]

$$\langle u, v \rangle_{L^2} = \int_0^1 u(t)v(t) dt$$

により定義される内積である。 1_s は点 s において 1、それ以外の点において 0 をとる関数である。また、 $x \in X$ に対して、 δ_x は、 x における Dirac 測度である。

[0072] 更に、 Φ を、 A 値測度を RKHM M_k に写す kernel mean embedding とする。なお、 A 値測度と kernel mean embedding に関する定義については、例えば、参考文献「Y. Hashimoto, I. Ishikawa, M. Ikeda, F. Komura, and Y. Kawahara, "Kernel Mean Embeddings of Von Neumann-Algebra-Valued Measures," arXiv: 2007.14698.」等を参照されたい。

[0073] このとき、 V を、

[0074] [数18]

$$\Phi(\mu_{x_1}), \dots, \Phi(\mu_{x_n})$$

により張られる空間とし、 V 上の A 線形作用素 P_f を以下により定める。

[0075] [数19]

$$P_f \Phi(\mu_{x_i}) = \int_0^1 \int_0^1 f(s, t, x_i(s, t)) ds dt I$$

ただし、 I は恒等作用素である。

[0076] 通常、与えられたデータ $x_1, \dots, x_n$ にはノイズが含まれていたり、類似したデータが含まれていたりするため、実際には V の中の低次元の部分空間によって P_f の振る舞いは決定すると考えられる。そこで、

[0077] [数20]

$$\Phi(\mu_{x_1}), \dots, \Phi(\mu_{x_n})$$

に対して、上記で説明した主成分分析と同様に、 V の中の低次元の部分空間を求める。具体的には、 $r < n$ に対して、以下の最適化問題を考える。

[0078] [数21]

$$\inf_{\{p_j\}_{j=1}^r \subseteq \mathcal{M}_k: \text{ONS}} \sum_{i=1}^n \left| \Phi(\mu_{x_i}) - \sum_{j=1}^r p_j \langle p_j, \Phi(\mu_{x_i}) \rangle_k \right|_k^2$$

この場合、 A は可換でないため、上記で説明した主成分分析と同様に、最適化計算部210が上記の手続き1及び2によって解を計算する。これによ

り求めた解 $p_1, \dots, p_r$ に対して、 $Q = [p_1, \dots, p_r]$ とおく。 P_f を $p_1, \dots, p_r$ の張る空間に射影した作用素は、 $P_f Q Q^*$ と表せる。

[0079] V 上の作用素 $P_f Q Q^*$ により最も大きく変化する方向は、

[0080] [数22]

$$v := Q Q^* P_f | Q Q^* P_f |_k^{-1} \in V$$

であることが示せる。空間 V は A 値測度により構成される空間であるため、 v も A 値測度に対応する。つまり、 v に対応する A 値測度で x_i をサンプルすれば、 y_i を最も大きく変化させることができる。これにより、 x_i をどのように動かせば y_i へ大きく影響するかがわかる。

[0081] <実験>

次に、上記の応用例で説明した主成分分析とデータの構成要素間の相互作用の解析とを実験した結果について説明する。

[0082] 《主成分分析》

以下の3種類の $[0, 1]$ 上の2変数関数から、ランダムに3種類のサンプルセットを生成した。

[0083] $y_1(s, t) = e^{10(s-t)}$

$y_2(s, t) = 10st$

$y_3(s, t) = \cos(10(s-t))$

各サンプルセット i は関数 y_i を $[0, 1]$ 上の11点で離散化してノイズを加えた20個のサンプルから成るものとする。ただし、ノイズは、平均0、標準偏差0.3のGaussian分布から生成される乱数とした。

[0084] このとき、これらのサンプルに対して、上記の応用例における「主成分分析」の「 A が可換な場合」で説明した手法（以下、「提案手法」ともいう。

）により上記の（３）に示す最適化問題を解いた。ここで、パラメータは $\lambda = 0.1$ 、 $\varepsilon = 0.01$ とした。また、 k は 1 変数関数に値を持つカーネル
 [0085] [数23]

$$(k(x_1, x_2))(t) = \int_0^1 \int_0^1 (t - x_1(s_1, s_2))(t - x_2(s_1, s_2)) ds_1 ds_2$$

と設定した。

[0086] また、提案手法との比較のため、上記の離散化した点をベクトル化したデータに対して RKHS を用いた既存手法により主成分分析も行った。

[0087] 提案手法による主成分分析の結果を可視化したものを図 4 A に、既存手法による主成分分析の結果を可視化したものを図 4 B にそれぞれ示す。既存手法では各サンプルを点として表現するため、関数としての特徴を捉えられず、３種類のサンプルセットをうまく分離できていない。一方で、提案手法では各サンプルを関数として表現するため、３種類のサンプルセットをうまく分離できていることがわかる。

[0088] 《データの構成要素間の相互作用の解析》

$x_1, \dots, x_n$ を、区間 $[0, 1]$ 上で定義された関数

[0089] [数24]

$$x_i(s, t) = e^{\omega_{i,1}s} e^{\omega_{i,2}t}$$

とする。ただし、 $n = 30$ とし、 $\omega_{1,1}, \dots, \omega_{n,1}, \omega_{1,2}, \dots, \omega_{n,2}$ は $[-1, 0]$ 上の一様分布に従う乱数である。

[0090] また、 $i = 1, \dots, n$ に対して、

[0091] [数25]

$$y_i = \int_0^1 \int_0^1 x_i(s, t)^{2-5|s-t|} ds dt$$

とする。

[0092] 更に、 k は有界線形作用素を値にとるカーネル

$$k(x_1, x_2) = (x_1^3 + x_1^2 + x_1)(x_2^3 + x_2^2 + x_2) \mid$$

と設定した。

[0093] このとき、上記の応用例における「データの構成要素間の相互作用の解析」で説明した手法により上記の数21を解いた。すなわち、数21に示す最適化問題を、上記の応用例における「主成分分析」の「 A が可換な場合」で説明した手法により解いた。ここで、パラメータは $\lambda = 0.1$ 、 $\varepsilon = 0.01$ とした。

[0094] このとき、 y_i を最も大きく変化させるベクトルを求め、それに対応する測度 λ に対して

[0095] [数26]

$$\langle 1_s, \lambda 1_t \rangle_{L^2}$$

を求めた。ここでは、 $s = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ とした。

[0096]

[数27]

$$\mu_{x_i}$$

の定義により、

[0097] [数28]

$$\langle 1_s, \lambda 1_t \rangle_{L^2}$$

により $x_i(s, t)$ をサンプルすれば、 y_i が大きく動くということを表している。測度

[0098] [数29]

$$\langle 1_s, \lambda 1_t \rangle_{L^2}$$

に対する密度関数 p を図5に示す。図5に示すように、 $s = 0.1$ 、 $t = 0.2$ のときは、大きな値における密度関数の値は大きくなっている。実際に、 $2 - 5 |s - t|$ は正になるため、 $x_i(s, t)$ が大きい方が y_i は大きくなる。一方で、 $s = 0.1$ 、 $t = 0.8$ のときは、小さな値における密度関数の値は大きくなっている。実際に、 $2 - 5 |s - t|$ は負になるため、 $x_i(s, t)$ が小さい方が y_i は大きくなる。

[0099] このように、 $x_i(s, t)$ と y_i の関係性が、 x_i 及び y_i の情報のみから

抽出できている。

[0100] 本発明は、具体的に開示された上記の実施形態に限定されるものではなく、請求の範囲の記載から逸脱することなく、種々の変形や変更、既知の技術との組み合わせ等が可能である。

符号の説明

[0101]	1 0	解析装置
	1 0 1	入力装置
	1 0 2	表示装置
	1 0 3	外部 I / F
	1 0 3 a	記録媒体
	1 0 4	通信 I / F
	1 0 5	プロセッサ
	1 0 6	メモリ装置
	1 0 7	バス
	2 0 1	取得部
	2 0 2	解析部
	2 0 3	記憶部
	2 1 0	最適化計算部

請求の範囲

- [請求項1] 複雑な構造を持つデータの集合を解析対象として取得する取得部と、
- 、
- 前記データの集合をRKHM (reproducing kernel Hilbert C*-module) に写した空間上で定義され、かつ、C*-algebra値をとる関数に関する最適化問題の解を、Hilbert空間における勾配法を拡張した手法により計算する最適化計算部と、
- を有する解析装置。
- [請求項2] 前記データの集合をRKHMに写した空間をM、前記関数を f 、前記関数 f の $v_t \in M$ における勾配を ∇f_{v_t} として、
- 前記最適化計算部は、
- 前記Hilbert空間における勾配法を拡張した手法として、 $v_0 \in M$ をランダムに選択すると共に $\varepsilon > 0$ を予め決められた値に設定した上で、 $t = 0, 1, \dots$ に対して、 $v_{t+1} - v_t$ が予め決められた閾値未満となるまで $v_{t+1} = v_t + \varepsilon \nabla f_{v_t}$ により $v_t \in M$ を更新することで、前記最適化問題の解を計算する、請求項1に記載の解析装置。
- [請求項3] 前記C*-algebraを A として、 A は可換であり、かつ、前記関数 f の $v \in M$ における $u \in M$ 方向の微分 $D f_v(u)$ が A 線形である場合、前記勾配 ∇f_v は、 $\langle \cdot, \cdot \rangle$ を前記RKHM上の内積として、 $D f_v(u) = \langle w, u \rangle$ を満たす $w \in M$ である、請求項2に記載の解析装置。
- [請求項4] 前記複雑な構造を持つデータは、高次元若しくは無限次元の空間で表現される多変量データ、又は、複数の変数をとる所定の連続関数により記述されるデータのいずれかである、請求項1乃至3の何れか一項に記載の解析装置。
- [請求項5] 前記最適化計算部により計算された解を用いて、前記データの集合に対する主成分分析、又は、前記データと所定の関係にある他のデータとの相互作用、を前記RKHM上で解析する解析部を有する、請求項1

乃至4の何れか一項に記載の解析装置。

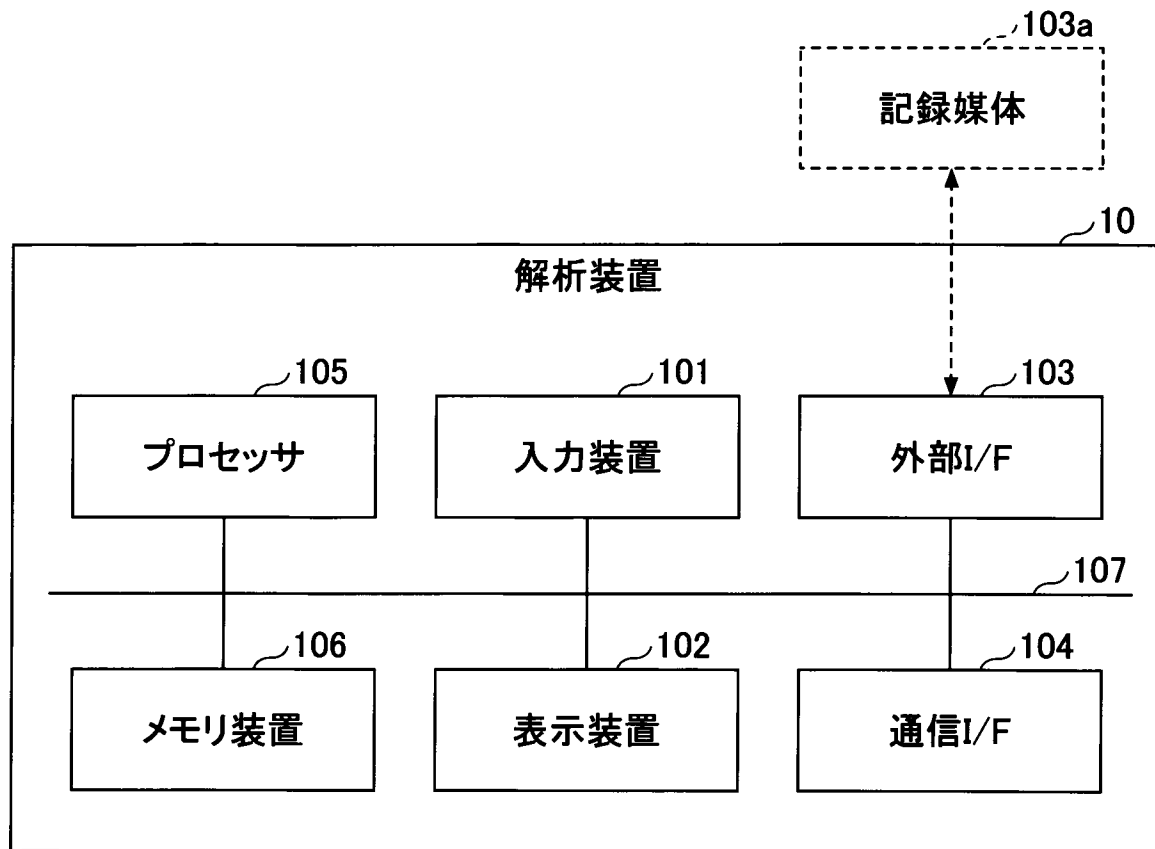
[請求項6] 複雑な構造を持つデータの集合を解析対象として取得する取得手順と、

前記データの集合をRKHM (reproducing kernel Hilbert C*-module) に写した空間上で定義され、かつ、C*-algebra値をとる関数に関する最適化問題の解を、Hilbert空間における勾配法を拡張した手法により計算する最適化計算手順と、

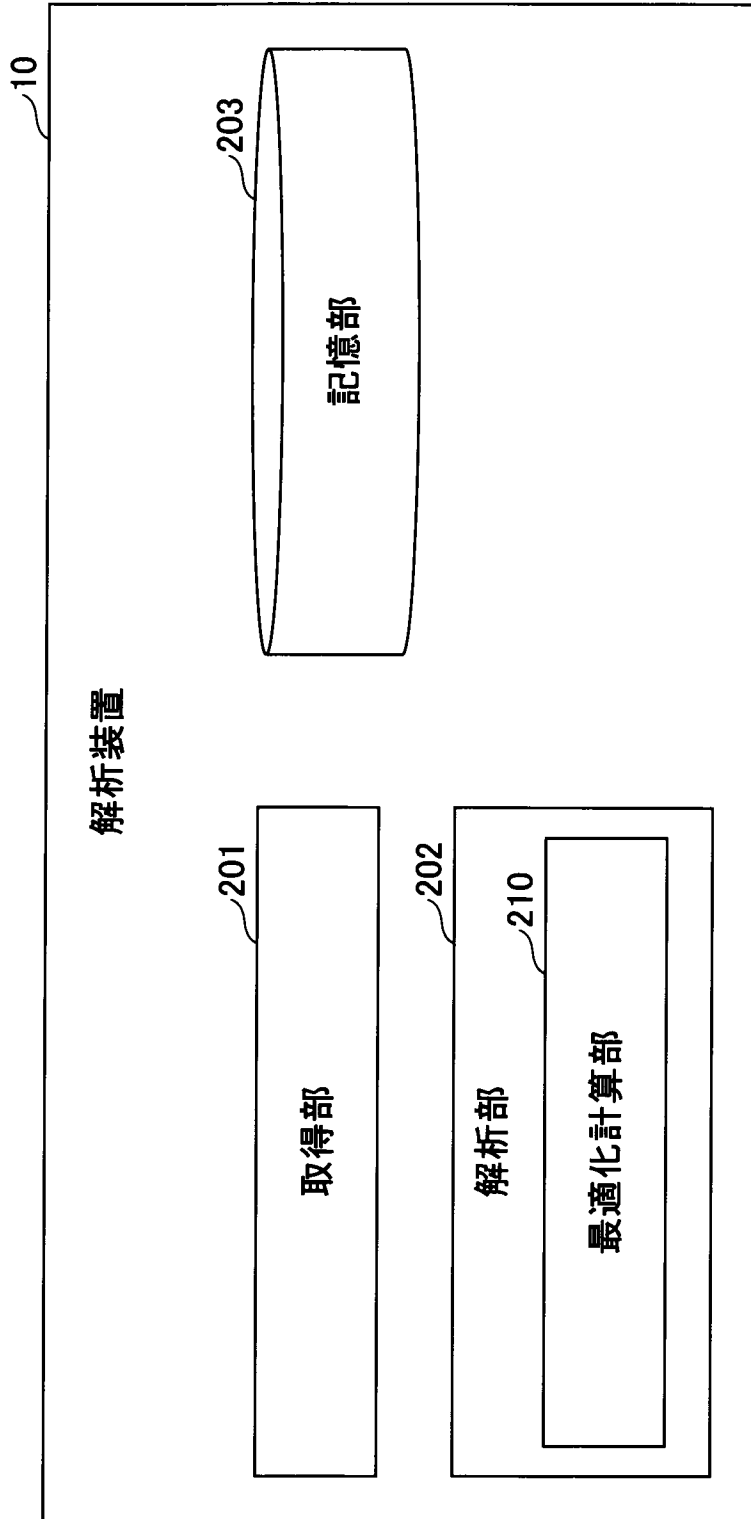
を実行する解析方法。

[請求項7] コンピュータを、請求項1乃至5の何れか一項に記載の解析装置として機能させるプログラム。

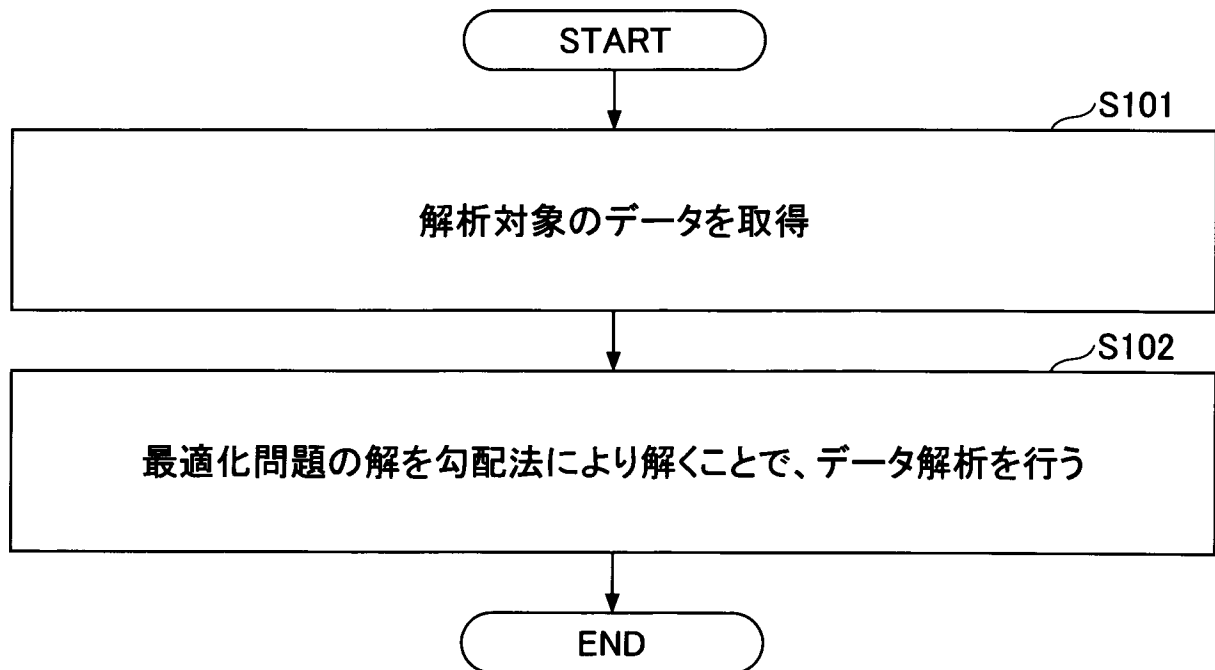
[図1]



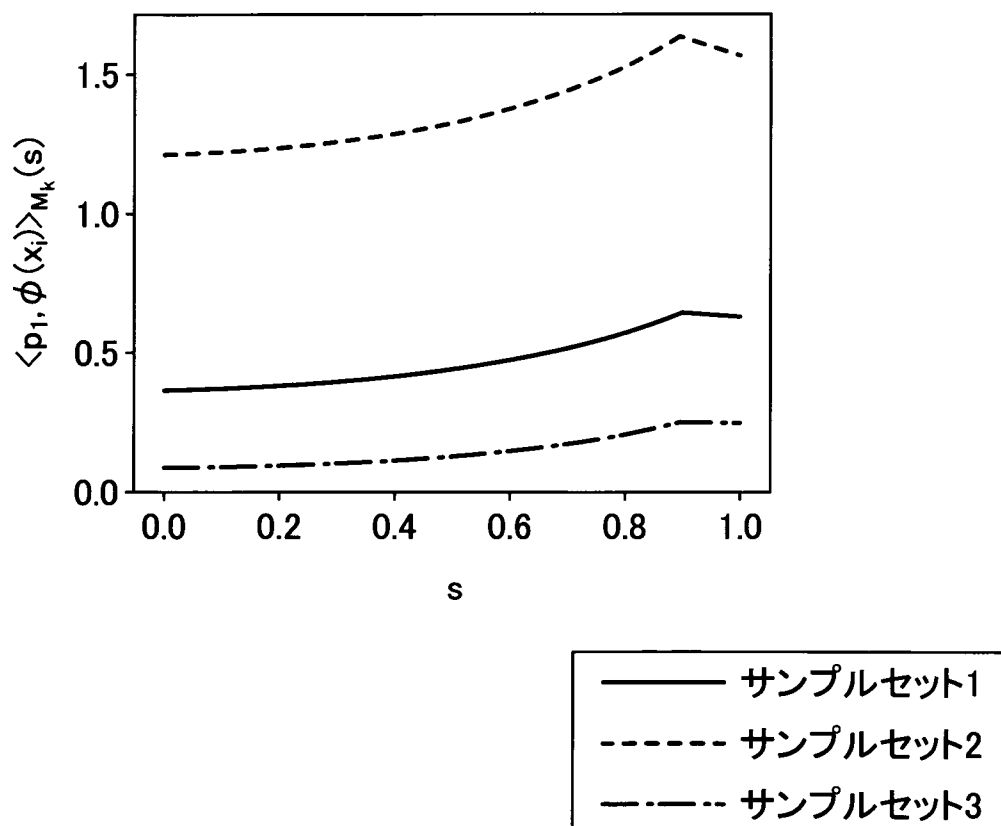
[図2]



[図3]

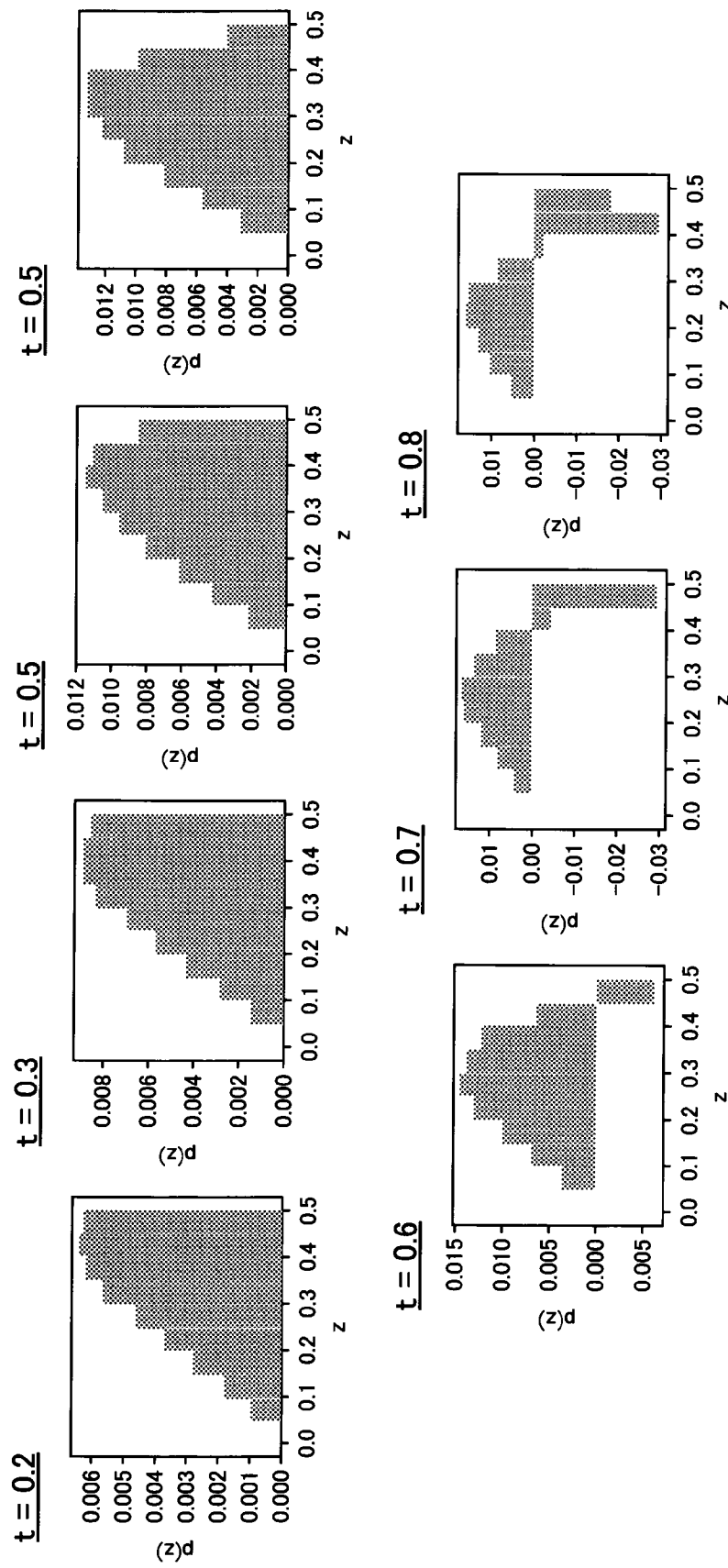


[図4A]



- サンプルセット1
- サンプルセット2
- サンプルセット3

[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/000358

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06N99/00 (2019.01) i

FI: G06N99/00180

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G06N3/00-99/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HASHIMOTO, Yuka et al., Analysis via orthonormal systems in reproducing kernel hilbert C*-modules and applications, arXiv [online], 02 March 2020 [retrieved on: 05 April 2021], <URL:https://arxiv.org/abs/2003.00738>, entire text	1-7

☐ Further documents are listed in the continuation of Box C.	☐ See patent family annex.
---	---

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 05 April 2021	Date of mailing of the international search report 13 April 2021
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 99/00(2019.01)i FI: G06N99/00 180										
B. 調査を行った分野										
調査を行った最小限資料（国際特許分類（IPC）） G06N3/00-99/00										
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2021年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2021年	日本国実用新案登録公報	1996 - 2021年	日本国登録実用新案公報	1994 - 2021年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2021年									
日本国実用新案登録公報	1996 - 2021年									
日本国登録実用新案公報	1994 - 2021年									
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）										
C. 関連すると認められる文献										
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号								
A	HASHIMOTO, Yuka, et al., Analysis via Orthonormal Systems in Reproducing Kernel Hilbert C*-Modules and Applications, arXiv [online], 2020.03.02 [検索日: 2021.04.05], <URL: https://arxiv.org/abs/2003.00738> 全文	1-7								
☐ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。										
<table border="0"> <tr> <td style="vertical-align: top;"> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 </td> <td style="vertical-align: top;"> “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献 </td> </tr> </table>			* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献						
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献									
国際調査を完了した日 05.04.2021	国際調査報告の発送日 13.04.2021									
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 杉浦 孝光 5B 5287 電話番号 03-3581-1101 内線 3545									