不確実性下における目的と手段の統合的探索に向けた連続腕バンディットの応用

三宅 悠介[1]　栗林 健太郎[1]

概要：不確実性の高い課題領域においては，その領域の目的に対する手段の有用性は，実際の行動を通じてのみ明らかになる．こうした状況で多くの候補から有効な手段を効率的に見極めるために，多腕バンディット問題として定式化する手法が用いられてきた．従来の多腕バンディットでは，探索の効率を高める実用的な簡略化として，固定された目的を前提とした定式化が行われてきた．しかし実際の課題では，目的自体も流動的であり，検討の過程で見直されることも少なくない．目的と手段は相互に依存する関係にあり，検討すべき組み合わせは多岐にわたるうえ，その対応関係も単純には捉えられない．このような制約のもとでも効率的に目的と手段の有用性を見極めるためには，行動による評価結果を他の目的にも横断的に活かす知識の伝達に加え，許容できないリスクを伴う組み合わせを適切に回避する仕組みが求められる．本報告では，目的と手段の双方を探索対象とした，不確実性下における意思決定の枠組みを提案する．具体的には，両者の特徴量空間を統合した空間上で，ガウス過程モデルに基づく連続腕バンディットにより有用な組み合わせを効率的に探索し，推定の不確実性に基づくリスク制御を組み込むことで，実行可能性を高める．評価では，高次元空間における最適化問題を対象とし，既存手法との比較を通じて，探索精度と計算効率の両立を確認した．その結果，提案手法が高次元設定にも適用可能であることが示唆された．

Applying Continuous-Armed Bandits to Integrated Exploration of Goals and Means under Uncertainty

Abstract: Only actions reveal how effective a means is for achieving a goal in uncertain domains. Prior work has modeled such problems using multi-armed bandits, often assuming a fixed goal to simplify exploration. In practice, goals may shift, and their relation to means is complex and interdependent. Effective decision-making requires models that transfer knowledge across goals and avoid risky combinations. This paper proposes a framework that jointly explores goals and means under uncertainty. It embeds both into a shared feature space and applies a continuous-armed bandit with a Gaussian process to identify promising pairs. The model incorporates risk control based on predictive uncertainty. Experiments on high-dimensional optimization tasks compare the proposed method with standard approaches. Results suggest that it balances accuracy and efficiency and scales to high-dimensional settings.

1. はじめに

情報システムの運用やサービス提供の現場では，目的に対する有効な手段が明確でないまま，手探りで意思決定を行う状況が多く見られる．たとえば，システムの信頼性を高めるための設定調整，EC サイトにおけるユーザーの興味を引く施策の検討といった場面では，どのような手段が有効であるかが事前には分からず，試行にはコストやリスクが伴う．さらに，こうした試行の結果が，目的そのものの見直しにつながることもあり，目的と手段は固定的な関係ではなく，相互に影響し合う動的な関係にある．

このような不確実性下で，有効な目的と手段を限られた試行の中から見出すには，結果を柔軟に学習に取り込みながら，将来の選択肢に対して過度なリスクを避ける枠組みが求められる．連続腕バンディットは，連続的な選択肢の中から試行を通じて報酬の傾向を学び，探索と活用のバランスを取りつつ有望な選択肢を効率的に特定する手法であり，このような意思決定課題に対する有力なアプローチである．従来の応用では目的を固定し，手段の探索に特化していたが，本報告では，目的も候補の一部とみなして横断

[1] GMO ペパボ株式会社 ペパボ研究所
Pepabo R&D Institute, GMO Pepabo, Inc.

的に探索することで，より柔軟かつ汎用的な適応を目指す．

本報告では，目的と手段の特徴量を連結した高次元空間を定義し，ガウス過程モデルに基づく連続腕バンディット方策を用いて有望な組合せを逐次的に選択する．また，予測の不確実性を活用して，将来的なリスクを抑制する仕組みを導入する．空間の高次元性に伴う計算負荷や精度劣化の問題に対しては，乱択化フーリエ特徴（Random Fourier Features, RFF）と正則化を導入することで対処し，さらにハイパーパラメータ推定においては，理論的に知られた式変形を応用して解析的に導出することで，逐次的な観測下でも高速な更新を実現している．

評価シミュレーションでは，高次元性の影響が生じる次元数での連続腕バンディット問題を対象に，各時点におけるこれまで選択された候補の中で最も良い解と真の最適解との差（累積リグレット）および次候補の選定時間を測定した．その結果，提案手法はベースライン手法と比較して，より迅速に最適解に近づく傾向を示し，ハイパーパラメータ推定の高速化により，選定時間を約 4 分の 1 に短縮できることを確認した．

本報告では，まず関連する背景と先行研究を整理した上で，提案する意思決定枠組みと構成要素の設計について述べる．続いて，シミュレーションによる評価実験を通じて，その有効性と今後の展開可能性を検討する．

2. 目的と手段の統合的探索の課題

実運用される情報システムにおいては，その利用者や利用場面は多岐にわたり，それぞれに応じて求められる目的や利用可能な手段も多種多様である．たとえば，目的として「どのようなユーザーやターゲットに向けて」，手段として「どのような施策やアイテムを提示するか」を考えると，異なるユーザー群に対して異なる応答を設計する必要があり，その組合せは膨大となる．さらに，例えば EC においては，閲覧数，回遊率，購入率など成果指標の選択自体が状況によって変化しうるが，これら自体も目的空間の一部として扱うことで，「閲覧数を目的としたときに有効だった施策」が購入率など他の目的に対しても有効かどうかといった転移可能性の探索も可能となる．このように，目的と手段の組合せは多次元かつ膨大であり，それらを個別に取り扱うことは現実的ではない．したがって，目的や手段をいくつかの抽象的な要素に分解し，その組み合わせとして表現するアプローチが必要となる [1].

深層ニューラルネットワークの実用化に伴い，このような複雑な対象を効率的に扱うために，低次元の連続ベクトルとして意味的特徴を保持しながら表現する埋め込みが主流となっている [2]．目的や手段も同様に，離散的なラベルや構造ではなく，連続的な特徴量空間上のベクトルとして埋め込み表現することが妥当である．

このような連続値ベースの表現に基づく選択肢の探索には，連続腕バンディットの枠組みが適している [3], [4]．特に，報酬の予測モデルを組み込んだモデルベース型のアプローチは有用であり，過去の観測をもとに将来の報酬を予測しつつ，探索と活用のバランスを取ることができる．中でも，複数の連続的特徴量の相互作用をとらえ，かつ確率的に将来の選択肢を評価できる手法として，ガウス過程モデル [5] と Thompson Sampling[6] の組合せは有力である．一方で，本報告が対象とするように目的と手段の統合空間を探索対象とする場合，入力の次元数が増えることで．予測精度の低下やモデル学習・推論の計算負荷の増大といった課題が顕在化する [7].

これに対し，入力空間の構造に関する知識が事前に得られていれば，次元を低次元の潜在空間へ写像する手法や，入力空間をいくつかの部分構造に分解し，加法的な関係として近似する構造を仮定する手法によって，処理負荷を軽減することが可能である [8], [9]．しかし，実運用環境ではそのような仮定を明示的に置くことが難しい．したがって，本報告では，事前知識に依存せず，高次元空間をそのまま扱う前提に立ち，以下の技術的課題に対応する必要がある．

まず，入力次元が高くなることで，必要な観測データ数が増加し，それに伴ってモデル学習および次候補の評価に要する計算時間も増大する．これに対しては，RFF を導入することで，ガウス過程モデルの学習および予測の計算量を削減する手法が有効である [5], [10]．また，高次元におけるモデルの汎化性能の低下を防ぐためには，モデルの構造に適切な正則化を導入することが有効であることが報告されている [11].

ただし，これらの工夫を施したとしても，ガウス過程モデルの性能に大きく影響を与えるカーネル関数のハイパーパラメータ推定が，依然として計算コスト上のボトルネックとなる．特に，本報告のように観測が逐次的に得られる状況では，各段階でハイパーパラメータを更新する必要があり，その処理時間が全体の実行効率に与える影響は小さくない．既存研究においても，しばしば $\mathbb{R}^{N\times N}$ の大規模な行列に対してコレスキー分解を直接適用する実装にとどまっており [5]，特に尤度やその勾配の導出における計算効率化を明示的に扱う研究は限られている．

たとえば，RFF を用いて逐次的にガウス過程回帰モデルを更新しバンディット方策に適用する研究 [12] では，ハイパーパラメータは事前に固定されたものを用いており，逐次更新の対象とはしていない．これは，ハイパーパラメータの変更がカーネル関数の構造全体に影響を及ぼすため，差分的な更新が困難であるという構造的な制約によるものである．

以上のように，本報告で対象とする「目的と手段の統合的探索」には，実用上のスケーラビリティや逐次学習への対応といった技術的課題が含まれており，それらを解決す

る手法設計が求められる．以降では，これらの課題に対応可能な枠組みを構築し，逐次的な最適化過程においてその有効性を検証する．

3. 提案手法

本節では，まず本報告が想定する目的と手段の統合的探索に関する枠組みの全体像を示す．次に，この枠組みにおいて採用する連続腕バンディットの方策を説明する．続いて，データ数および次元数の増加に対する対処法について述べ，最後にハイパーパラメータ推定の高速化手法を示す．

3.1 目的と手段の統合的探索

本報告では，目的と手段の組合せ空間における探索問題として，連続腕バンディットの枠組みを用いて，目的と手段を統合的に探索する方策の設計を目指す．

まず，目的の集合 $\mathcal{G}$ と手段の集合 $\mathcal{A}$ の各要素を，あらかじめ学習された汎用的な埋め込みモデルを用いて，それぞれ埋め込みベクトル $\mathbf{g} \in \mathbb{R}^{D_g}$，$\mathbf{a} \in \mathbb{R}^{D_a}$ として表現する．次に，これらのベクトルを結合し，探索空間上の点 $\mathbf{x} = (\mathbf{g}^\top, \mathbf{a}^\top)^\top \in \mathbb{R}^D, D = D_g + D_a$ として定義することで，目的と手段の関係性を統一的に扱えるようにする．この空間 $\mathcal{X} = \mathbb{R}^D$ における探索を連続腕バンディット問題とみなし，報酬関数 $f(\mathbf{x})$ を最大化するような点，すなわち目的と手段の組み合わせを求める．

この探索では，観測データに基づき報酬の予測値やその不確実性を推定し，活用と探索のバランスを取るバンディット方策を適用する．さらに，予測される不確実性に基づいて，リスクの高い目的と手段の組み合わせを避けるといった制御も可能である．本枠組みのアルゴリズムを Algorithm 1 に示す．ここで，`FILTER` は予測された報酬値およびその不確実性 $\tilde{f}(\mathbf{x}_t^*), \sigma(\mathbf{x}_t^*)$ に基づき，リスク制御を目的として，選択された腕 $\mathbf{x}_t^*$ を受容・拒否する任意の関数である．

3.2 ガウス過程モデルを用いた連続腕バンディット

連続腕バンディット問題では，未知の報酬関数 $f(\mathbf{x})$ に基づいて観測される報酬 y を得る状況を考える．「腕」と呼ばれる候補 $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$ は D 次元の連続実数空間上の点であり，この空間から逐次的に腕を選択して観測された報酬を最大化することが目的である．すなわち，$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$ として期待報酬 $\mathbb{E}[y \mid \mathbf{x}] = f(\mathbf{x})$ を最大化するような腕 $\mathbf{x}^*$ を効率的に求めることが課題となる．このような逐次的選択問題では，既知の報酬期待値の高い腕を選択する「活用」と，未知の腕の情報を獲得する「探索」のバランスを取る方策が重要となる．

連続腕バンディットにおける代表的な方策として，Thompson Sampling が知られている．この方策では，これまでの観測データに基づいて学習された報酬関数 $f(\mathbf{x})$ の事後分布から，関数サンプル $\tilde{f}$ を予測分布として得て，そのもとでの最大化を行うことで次の腕を選択する．事後分布のモデルとしては，ガウス過程を用いることで，柔軟な関数近似と，未観測点における不確実性を伴う予測分布の導出が可能となる．

Algorithm 1: 目的手段空間に対する統合的探索

Input: 初期観測データ $\mathcal{D}_0 = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_0}$

for $t \in \{1, \ldots, T\}$ **do**

- // ハイパーパラメータと GP の事後分布を更新
- $\boldsymbol{\theta}_{t-1} \leftarrow \arg\max_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{t-1} \mid \boldsymbol{X}_{t-1}, \boldsymbol{\theta})$
- $\mathcal{GP}$ を再構築
- // ガウス過程モデルより関数サンプル $\tilde{f}$ を生成
- $\tilde{f} \sim \mathcal{GP}$
- // Thompson Sampling による腕選定
- $\mathbf{x}_t^* \leftarrow \arg\max_{\mathbf{x}\in\mathcal{X}} \tilde{f}(\mathbf{x})$
- // リスク制御によるフィルタリング（必要に応じて）
- $\mathbf{x}_t^* \leftarrow \mathrm{FILTER}(\mathbf{x}_t^*, \tilde{f}, \sigma(\mathbf{x}_t^*))$
- // 選択した腕を実行し，報酬を観測
- $y_t \leftarrow f(\mathbf{x}_t^*) + \varepsilon_t \quad (\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2))$
- // 観測データを追加
- $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(\mathbf{x}_t^*, y_t)\}$

ガウス過程は，任意の入力点集合 $\mathbf{x}_1, \ldots, \mathbf{x}_N$ に対して，対応する関数値 $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))^\top$ が多変量正規分布 $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, K)$ に従うという仮定に基づく．ここで $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ は与えられたカーネル関数 k によって構成されるカーネル行列である．このカーネル関数により得られるカーネル行列 K は，予測点集合に対するガウス過程の予測分布の平均ベクトル $\boldsymbol{\mu}_*$ と共分散行列 Σ_* の計算に利用され，予測分布 $\mathbf{f}_* \mid \mathcal{D}_N \sim \mathcal{N}(\boldsymbol{\mu}_*, \Sigma_*)$ が得られる．

連続腕バンディットにおける Thompson Sampling では，この予測分布に従って，候補点集合 $\{\mathbf{x}_*^{(1)}, \ldots, \mathbf{x}_*^{(M)}\}$ に対して関数サンプル $\tilde{f}$ を 1 つ生成し，その中で最大値を与える点 $\tilde{\mathbf{x}} = \arg\max_{\mathbf{x}_*^{(m)}} \tilde{f}(\mathbf{x}_*^{(m)})$ を次の選択腕とする．

ただし，この方策には以下のような課題がある．

(1) 観測数 N に対してカーネル行列 $K \in \mathbb{R}^{N\times N}$ の逆行列計算が必要となり，計算量は $\mathcal{O}(N^3)$ に達する．
(2) 高次元空間 $\mathbb{R}^D$ においては，候補点集合の構築や分布計算の精度の確保のために点数 M を多く取る必要があり，予測分布の計算負荷が増大する．
(3) 高次元ではガウス過程モデルの予測性能が低下しやすく，信頼性のある候補選定が困難になる．

3.3 データ数と次元数の増加への対処

本報告では，前述の課題 (1) に対処するために，2 節で述べたカーネル関数の近似として RFF を導入する [5]．RFF は，シフト不変なカーネル関数に対し以下のような近似を与える．

$$k(\mathbf{x},\mathbf{x}') \approx \mathbf{z}(\mathbf{x})^\top \mathbf{z}(\mathbf{x}'), \quad \mathbf{z}(\mathbf{x}) = \sqrt{\frac{2}{R}}\cos(\Omega^\top \mathbf{x} + \mathbf{b}) \tag{1}$$

ここで $\Omega \in \mathbb{R}^{D\times R}$ の各列は $\Omega_{:,r} \sim \mathcal{N}(\mathbf{0}, \sigma_k^{-2} I)$ から独立にサンプリングされ，$\mathbf{b} \in \mathbb{R}^R$ は $[0, 2\pi]$ 上の一様分布からサンプリングされる．本報告では，以下のような一般化されたガウスカーネル関数 k を用いる．

$$k(\mathbf{x},\mathbf{x}') = \sigma_w^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma_k^2}\right) + \sigma_\varepsilon^2 \delta_{\mathbf{x},\mathbf{x}'}$$

$\delta_{\mathbf{x},\mathbf{x}'}$ は，$\mathbf{x} = \mathbf{x}'$ のときに 1，それ以外では 0 を取るクロネッカーのデルタ関数である．この近似を用いることで，カーネル行列は次のように近似できる．

$$K \approx \sigma_w^2 Z Z^\top + \sigma_\varepsilon^2 I, \quad Z = (\mathbf{z}(\mathbf{x}_1), \ldots, \mathbf{z}(\mathbf{x}_N))^\top \in \mathbb{R}^{N\times R}$$

$R \ll N$ とすることで，$ZZ^\top \in \mathbb{R}^{N\times N}$ に対しても，Woodbury の恒等式[*1]などの行列恒等式を用いて $R \times R$ の小さな行列の計算に変形可能であり，カーネル行列の構築や逆行列の計算にかかる計算量を大幅に削減できる．この結果，データ数 N の増加に対しても，計算効率を保った推論が可能となる．

次に，前述の課題 (2) に対応するため，2 節で述べた候補点集合の構築を介さずに関数サンプルを評価する手法 [10] を採用する．RFF の導入により，ガウス過程の予測分布 $\mathcal{N}(\boldsymbol{\mu}_*, \Sigma_*)$ による関数サンプルの生成は，観測データから得られた重みベクトルの事後分布 $\tilde{\mathbf{w}} \sim \mathcal{N}(\boldsymbol{\mu}_w, \Sigma_w)$ を用いて，任意の入力点に対して $\tilde{f}(\mathbf{x}) = \mathbf{z}(\mathbf{x})^\top \tilde{\mathbf{w}}$ のように評価できる．ここで，$\boldsymbol{\mu}_w$ および Σ_w は，観測データから導出された事後分布のパラメータであり，候補点集合には依存しない．この形式により，$\tilde{f}$ の最大化（Thompson Sampling における次の腕の選定）は，候補点集合の構築を介さずに，$\tilde{f}$ の $\mathbf{x}$ に関する勾配を用いた最適化として可能となる．

最後に，前述の課題 (3) への対応として，2 節で述べた正則化手法 [11] を導入する．本手法では，カーネルのスケールパラメータ σ_k^2 に対して対数正規分布を事前分布として仮定し，次元数 D の増加に伴う汎化性能の低下を抑制する効果を期待する．このとき，σ_k^2 に対する負の対数事前分布は，以下のように与えられる．

$$\log p(\sigma_k^2) = -\log \sigma_k^2 - \frac{1}{2\sigma_{k0}^2}\left(\log \sigma_k^2 - \mu_{k0} - \frac{1}{2}\log D\right)^2 + \text{const.} \tag{2}$$

ここで，第一項の $-\log \sigma_k^2$ は，σ_k^2 が過剰に大きな値を取ることを抑制し，モデルの過学習を防ぐ役割を担う．一方，第二項の二乗項は，$\log \sigma_k^2$ が $\mu_{k0} + \frac{1}{2}\log D$ に近づくよう促す効果を持ち，次元数 D に応じた適切なスケールパラメータの選択を誘導する．これらの項を尤度の勾配に正則化項として加えることで，高次元におけるハイパーパラメータ推定の安定性と予測精度の向上が期待できる．

[*1] $(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$

3.4 乱択化フーリエ特徴を用いたハイパーパラメータ推定の高速化

本報告で用いる方策では，カーネルのスケールやノイズ分散などのハイパーパラメータの推定が性能に大きく影響する．この推定は対数尤度の最適化として行われ，その繰り返し計算において尤度と勾配の高速な評価が求められる．従来手法では $N \times N$ カーネル行列の逆行列や行列式の計算がボトルネックとなるが，本報告では RFF によるカーネル近似の構造を活かし，提案方策に適した高速な導出を行った．

以下では，RFF と正則化を用いた場合における，ガウス過程モデルの対数尤度およびハイパーパラメータ $(\sigma_k^2, \sigma_w^2, \sigma_\epsilon^2) \in \boldsymbol{\theta}$ に対する勾配の導出を示す．まず，対数尤度 $\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ は次のように変形される．

$$\begin{aligned}\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) &\propto -\log|K| - \mathbf{y}^\top K^{-1}\mathbf{y} + 2\log p(\sigma_k^2) \\ &= -N\log\sigma_\varepsilon^2 - \log|B| \\ &\quad - \frac{1}{\sigma_\varepsilon^2}\left(\|\mathbf{y}\|^2 - \frac{1}{\sigma_\varepsilon^2}\|Z^\top\mathbf{y}\|^2_{A^{-1}}\right) \\ &\quad - 2\log\sigma_k^2 - \frac{1}{\sigma_{k0}^2}\left(\log\sigma_k^2 - \mu_{k0} - \frac{1}{2}\log D\right)^2\end{aligned}$$

ここで，$A = \frac{1}{\sigma_w^2}I + \frac{1}{\sigma_\varepsilon^2}Z^\top Z$，$B = \sigma_w^2 A = I + \frac{\sigma_w^2}{\sigma_\varepsilon^2}Z^\top Z$ と定義される．なお，ノルム記号 $\|\cdot\|^2$ は，対応するベクトルのユークリッドノルムの二乗を表す．第 2 項の行列式の変形には Weinstein–Aronszajn の恒等式[*2]を用いている．また，第 4 と第 5 項は式 (2) より求めた．この変形によって，観測データ数 N に依存する大規模なカーネル行列 $K \in \mathbb{R}^{N\times N}$ の逆行列計算や行列式計算を回避し，R に依存した小さな行列の計算で尤度を効率的に評価することが可能となる．

次に，各ハイパーパラメータに対する勾配計算の高速化について述べる．RFF を適用する前の一般的なガウス過程モデルにおいて，各ハイパーパラメータ $\theta_i \in \boldsymbol{\theta}$ に対する対数尤度の勾配は，以下で与えられる．

$$\begin{aligned}\frac{\partial}{\partial\theta_i}\log\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) \propto &-\operatorname{tr}\left(K^{-1}\frac{\partial K}{\partial\theta_i}\right) + \mathbf{y}^\top K^{-1}\frac{\partial K}{\partial\theta_i}K^{-1}\mathbf{y} \\ &+ \frac{\partial}{\partial\sigma_k^2}\log p(\sigma_k^2)\end{aligned}$$

この表現では K^{-1} の計算コストが高く，特に観測数 N が大きい場合には非現実的である．RFF を適用することで，この勾配の構造を近似によって簡略化できる．

ノイズ分散パラメータ σ_ε^2 と重み分散パラメータ σ_w^2 に関しては，それぞれ $\frac{\partial K}{\partial\sigma_\varepsilon^2} = I$ と $\frac{\partial K}{\partial\sigma_w^2} = K$ であることを利用して以下のように勾配を求めることができる．

[*2] $\det(I + AB) = \det(I + BA)$ for $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times n}$

$$\frac{\partial}{\partial \sigma_\varepsilon^2}\log \mathcal{L}(\boldsymbol{\theta}\mid \mathbf{y}) = -\left(\frac{N}{\sigma_\varepsilon^2} - \frac{\sigma_w^2}{\sigma_\varepsilon^2}\operatorname{Tr}(B^{-1}Z^\top Z)\right) + \frac{1}{\sigma_\varepsilon^4}\Big(\|\mathbf{y}\|^2 - 2\sigma_w^2 \mathbf{y}^\top Z B^{-1} Z^\top \mathbf{y} + \sigma_w^4 \left\|ZB^{-1}Z^\top \mathbf{y}\right\|^2\Big)$$

$$\frac{\partial}{\partial \sigma_w^2}\log \mathcal{L}(\boldsymbol{\theta}\mid \mathbf{y}) = -\operatorname{Tr}\left(\frac{1}{\sigma_\varepsilon^2}Z^\top Z - \frac{\sigma_w^2}{\sigma_\varepsilon^2}Z^\top Z\cdot B^{-1}\cdot Z^\top Z\right) + \frac{1}{\sigma_\varepsilon^4}\Big((Z^\top \mathbf{y})^\top Z^\top \mathbf{y} - 2\sigma_w^2 (Z^\top \mathbf{y})^\top Z^\top Z B^{-1} Z^\top \mathbf{y} + \sigma_w^4 \left\|Z^\top Z B^{-1} Z^\top \mathbf{y}\right\|^2\Big)$$

これらの導出においては，K^{-1} に対して Woodbury の恒等式[*1]を適用した．

パラメータ σ_k^2 に関しては，RFF の構成に直接影響を与えるため，対数尤度の勾配計算において特別な配慮が必要となる．式 (1) より，$\boldsymbol{\Omega}\in\mathbb{R}^{R\times D}$ は，σ_k に依存する．したがって，σ_k を変更するたびに $\boldsymbol{\Omega}$ を再サンプリングする必要が生じるが，これはハイパーパラメータ最適化の不安定化をもたらす．そこで本報告では，再パラメータ化の方式を採用し，あらかじめ $\widetilde{\boldsymbol{\Omega}}\sim\mathcal{N}(\mathbf{0},I)$ を固定して利用する．

$$\boldsymbol{\Omega} = \frac{1}{\sigma_k}\widetilde{\boldsymbol{\Omega}} \quad\Rightarrow\quad \mathbf{z}(\mathbf{x}) = \sqrt{\frac{2}{R}}\cos\left(\frac{1}{\sigma_k}\widetilde{\boldsymbol{\Omega}}\mathbf{x} + \mathbf{b}\right)$$

このとき，σ_k は明示的に式中に残るため，微分可能な形で勾配を導出することができる．σ_k^2 に関する勾配を導出すると，次のようになる．

$$\frac{\partial}{\partial \sigma_k^2}\log \mathcal{L}(\boldsymbol{\theta}\mid \mathbf{y}) = -2\sigma_w^2\left\{\frac{1}{\sigma_\varepsilon^2}\operatorname{Tr}\left(Z^\top \frac{\partial Z}{\partial \sigma_k^2}\right) - \frac{\sigma_w^2}{\sigma_\varepsilon^2}\operatorname{Tr}\left(Z^\top Z\cdot B^{-1}\cdot Z^\top \frac{\partial Z}{\partial \sigma_k^2}\right)\right\} + 2\sigma_w^2\left\{(Z^\top \boldsymbol{\alpha})^\top\left(\frac{\partial Z}{\partial \sigma_k^2}\right)^\top \boldsymbol{\alpha}\right\} - \frac{2}{\sigma_k^2}\left(1 + \frac{1}{\sigma_{k0}^2}\left(\log\sigma_k^2 - \mu_{k0} - \frac{1}{2}\log D\right)\right)$$

ここで $\boldsymbol{\alpha} = \frac{1}{\sigma_\varepsilon^2}\mathbf{y} - \frac{\sigma_w^2}{\sigma_\varepsilon^2}ZB^{-1}Z^\top \mathbf{y}$ と定義される．最終項は正則化項に対応する．なお，$\frac{\partial Z}{\partial \sigma_k^2}\in\mathbb{R}^{N\times R}$ は以下のように解析的に与えられる．

$$\frac{\partial Z}{\partial \sigma_k^2} = \sqrt{\frac{2}{R}}\cdot\sin\left(\frac{1}{\sigma_k}X\widetilde{\boldsymbol{\Omega}}^\top + \mathbf{b}\right)\odot\left(\frac{X\widetilde{\boldsymbol{\Omega}}^\top}{2\sigma_k^3}\right)$$

この勾配式により，RFF の構造を維持したまま σ_k^2 の影響を解析的に評価できるため，効率的かつ安定にハイパーパラメータの更新が可能となる．

なお，ハイパーパラメータ $\boldsymbol{\theta}$ の各成分 θ_i は，いずれも 0 より大きい値が期待されるため，最適化においては対数空間 $\eta(\theta_i)=\log\theta_i$ 上での推定が適切となる．このとき連鎖律より，$\eta(\theta_i)$ に関する勾配は，これまで導出した θ_i に関する勾配に θ_i 自身を乗じることで得られる．

4. 評価

4.1 評価方法

3 節で導入した逐次的な最適化手法の有用性を検証するため，シミュレーションによる評価実験を実施した．本実験では，連続腕バンディットにおいて高次元性が性能に与える影響が現れる状況として，32 次元の探索空間を対象とした．目的関数には，各次元における真の最適値からの距離の二乗和として以下のように定義される shifted sphere 関数を用いた．

$$f_r(\mathbf{x}) = \sum_{i=1}^{D}(x_i - r_i)^2$$
$$\text{where}\quad \mathbf{x} = (x_1, x_2, \ldots, x_D)^\top \in [-3.0, 3.0]^D,$$
$$\mathbf{r} = (r_1, r_2, \ldots, r_D),\quad r_i \sim \mathcal{U}(-3.0, 3.0)$$

ここで，$\mathbf{x}$ は探索対象の入力ベクトルであり，各成分は $[-3.0, 3.0]$ の範囲に制限されている．$\mathbf{r}$ は各次元におけるランダムなシフト値ベクトルであり，それぞれの r_i は同じく $[-3.0, 3.0]$ の一様分布から独立にサンプリングされる．観測値には平均 0，標準偏差 0.01 の正規分布に従う誤差を加えており，ノイズを含む状況下で各 r_i を推定することが最適化の目標となる．なお，本設定では連続腕バンディットの枠組みにおける期待報酬の最大化は，この関数値の最小化に対応する．

この問題設定において，各方策の観測数の増加に伴う計算コストの増大や過学習の回避性能を評価するため，以下の 2 つの指標を測定対象とした．

- 各時点におけるこれまでに選択された候補の中で最も良いものと，真の最適解との差の累積（累積リグレット）
- 次候補の選定に要する計算時間（推論時間）

各手法において 200 回の候補選定を実施し，逐次的な性能を評価した．ただし，各手法の選定前には事前情報として 1600 回分のランダムな候補点に対応する観測結果を与えている．これは高次元空間において，初期段階から安定した予測性能を確保するためには一定量の観測データが不可欠であること，ならびにこの前提のもとでの計算効率の改善が現実的な要請となるためである．なお，乱数を用いた確率の計算結果を平均化するために，異なる乱数シードを用いてシミュレーションを 10 回行い，この平均を結果として用いた．

比較対象として，以下の選択方策を評価に含めた．

- ランダムサンプリング（Random）: 候補を一様分布に従ってランダムに選定する方策．
- Tree-structured Parzen Estimator サンプリング（TPE）[13]: カーネル密度推定した分布の比を用いる方策．
- ガウス過程モデル + Thompson Sampling: 提案方策

を含む以下の 4 つの方策を比較対象とした．

– 3 節で導入した方策（GP-RFF）
– 3 節の方策から正則化機能を除く（GP-RFF-No-Prior）
– 3 節の方策からハイパーパラメータ探索の高速化機能を除く（GP-RFF-Naive-Gradient）
– 標準的なガウス過程モデル（GP）

各方策について，Random は愚直なベースライン方策として採用した．TPE は，観測データを目的値に基づいて上位群と下位群に二分し，それぞれに対してカーネル密度推定により構築した分布の比に基づいて次候補を生成する最適化手法である．Optuna などの広く用いられている最適化ライブラリにおいて標準的に採用されていることから，提案手法の実用性を検証する目的で比較対象とした．なお，Random および TPE の実装には，評価時点での最新バージョンである Optuna 4.3.0 [3] を用いた．ガウス過程モデル + Thompson Sampling については，各要素技術の効果を検証することを目的に，提案方策およびその派生方策を評価対象とした．

ガウス過程モデルについて推定対象とするハイパーパラメータは σ_k^2 と σ_ε^2 とし，σ_w^2 は 1 と固定した．これは予備実験において σ_w^2 を可変にすると σ_k^2 への正則化の効果を打ち消すような値を取る特性が確認されたことへの対処である．この挙動の解明は今後の課題である．なお，正則化のハイパーパラメータには予備実験で有効性が確認された値 $\mu_{0k} = 0, \sigma_{0k}^2 = 0.005$ を採用した．

また，RFF を用いない標準的なガウス過程モデル（GP）では，RFF に依存するハイパーパラメータ探索の高速化機能および，候補点集合を介さないサンプリング近似は使用していない．このため，Thompson Sampling による次候補の選定に際しては，本来であれば探索空間全体に対してガウス過程の事後分布を構成すべきところ，32 次元空間ではその計算が現実的でないため，探索範囲から一様にサンプリングした 1600 点を候補集合とし，その中から事後分布に基づいて選定を行う直接的な手法を採用している．この候補数は空間全体を網羅するには不十分であるが，評価環境の計算資源の制約内で現実的に動作可能な範囲として設定している．

4.2 評価結果

図 1 に，累積リグレットの推移を示す．なお，縦軸は対数スケールであることに注意されたい．本実験において，最も累積リグレットが少なかったのは，提案方策である GP-RFF であり，累積値は 4384.9 であった．目的関数の最適値が 0 であるのに対し，この方策では最良点の値が 18.3 まで近づいており，高次元空間における有効な候補の選定が可能であることが示された．これは，本方策で採用した一連の要素技術が，高次元性に対して有効に機能したことを示唆している．

Random は，累積リグレットが継続的に増加しており，この設定では事前に与えられた 1600 点および選定点 200 点の合計 1800 点の範囲では，最適解に到達することが極めて困難であることを示している．一方で，Random を除くすべての方策が，より小さいリグレットを示しており，探索方策の導入によって一定の改善効果が得られることが確認された．

TPE は，今回の実験設定においては最適解への収束が比較的遅く，累積リグレットも大きい結果となった．TPE は，初期観測の分布が群の分割と候補生成に大きな影響を与える．本評価では前提条件を揃えるため，事前点としてランダムな 1600 点を与えたが，参考として，1800 点全体を TPE による探索フェーズとして使用した場合の結果を確認したところ，最良値は 24.8，累積リグレットは 5048.9 まで改善し，TPE 自身が試行を通じて分布推定を洗練していく方が効果的であることが示唆された．このことから，事前観測データの有無や性質に応じて，TPE と提案方策を組み合わせて運用することが有効である可能性も示された．

標準的な GP（RFF を用いない方策）は，Random を除く中では最もリグレットが大きくなった．これは，Thompson Sampling において 1600 点のランダムサンプリング候補の中に，最適解近傍の点が含まれていなかったことによると考えられる．この結果は，高次元空間においては候補集合の質と量が精度に直結すること，および，提案方策で採用した候補点集合の構築を介さずに関数サンプルを評価する手法が現実的な有効策であることを示している．

また，GP-RFF-Naive-Gradient（ハイパーパラメータ推定高速化なし）は，GP-RFF とほぼ同等のリグレットを示した．この二者の違いは，ハイパーパラメータ推定における計算効率にあり，尤度および勾配の理論的値は同一であるため，出力結果にも差は出にくい．ただし，数値計算上の誤差や収束挙動の違いにより，わずかに性能に差が生じたと考えられる．

GP-RFF-No-Prior（正則化なし）は，GPR-RFF と比較して累積リグレットが増加しており，正則化がハイパーパラメータ推定を通じて探索性能の向上に寄与したことが示される．この点については，図 2 に示す σ_k^2 の推移からも確認できる．正則化によって σ_k^2 の過度な上昇が抑えられ，次元数に応じて適切な範囲に保たれた結果，汎化性能が維持され，有用な次候補が選定されていたと考えられる．

ただし，この効果は正則化の強さ，特に事前分布のパラメータ σ_{k0}^2 の設定に強く依存していた．たとえば $\sigma_{k0}^2 = 1.0$ の場合，正則化の効果はほとんど見られなかった一方で，$\sigma_{k0}^2 = 0.001$ のように強すぎる設定では，初期の探索において σ_k^2 がほぼ 0 に抑え込まれ，結果として性能が著しく低下するケースも観測された．正則化項のハイパーパラメータは性能に大きく影響しうる要素であり，慎重な選定が求められる．

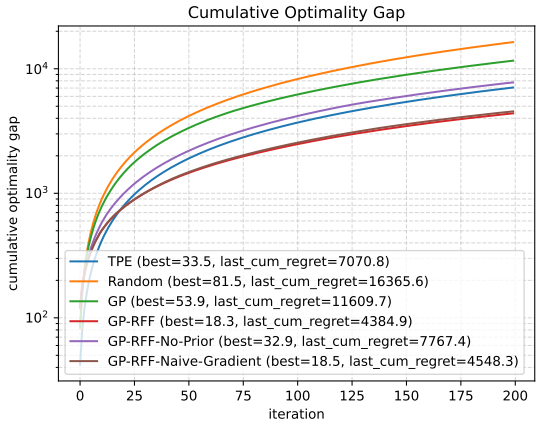

図 1 各時点における最良解と真の最適解との差（累積リグレット）比較

Fig. 1 Comparation of cumulative regrets.

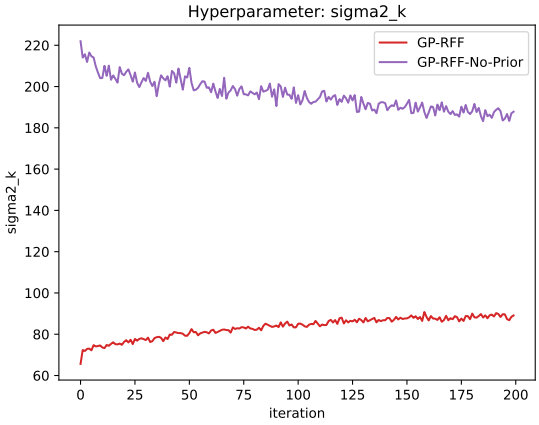

図 2 σ_k^2 の比較

Fig. 2 Comparation of σ_k^2.

図 3 は，各方策における次候補の選定時間の推移を示している．結果から，最も高速であったのは Random であり，次に TPE が続いた．TPE は探索範囲全体を網羅的に走査する必要がなく，また次元ごとに独立した分布からサンプリングする方策であるため，次候補の生成に要する計算量が抑えられていると考えられる．

続いて，提案手法である GP-RFF および GP-RFF-No-Prior が比較的短い選定時間を示した．両者の間で選定時間に差が見られなかったことから，正則化の有無は実行時間に大きな影響を及ぼさないと考えられる．加えて，これらの方策は試行回数の増加に伴っても選定時間がほぼ一定であり，観測データ数の増加に対しても計算コストが安定していることが確認された．

一方で，GP および GP-RFF-Naive-Gradient は，尤度や勾配の計算において観測データ数に比例して計算負荷が増加するため，初期 1600 点の影響もあり，提案方策と比較して大幅に時間を要する結果となった．さらに，特に GP においては，次候補の選定に際して探索空間における複数点への事後分布の構築と評価が必要となるため，RFF ベースの方策と比べて処理時間が大きくなった．

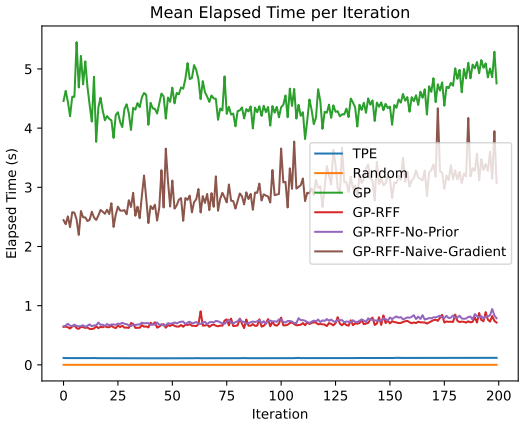

図 3 次候補の選定時間の比較

Fig. 3 Comparation of elapsed time.

200 試行目の選定時間を見ると，GP-RFF は約 0.72 秒であるのに対し，GP-RFF-Naive-Gradient は約 3.07 秒であり，提案するハイパーパラメータ推定の式変形により約 4.3 倍の高速化が実現できた．

以上より，本評価設定において，提案方策は精度を維持しながら実行時間を大幅に抑えることができ，実用的な高次元最適化手法として有効であることが示された．

5. おわりに

本報告では，目的と手段を共通の埋め込み空間上で統合的に扱い，その空間上での探索を連続腕バンディット問題として定式化する枠組みを提案した．さらに，RFF を用いたガウス過程モデルを方策に導入することで，非線形性と不確実性の推定を両立しつつ，計算効率の高い意思決定を実現する構成を示した．また，ハイパーパラメータ推定においても，RFF の構造を活かした尤度およびその勾配の高速計算により，逐次的な最適化を効率化した．

今後は，提案手法をより高次元かつ複雑な目的・手段空間に適用するためのスケーラビリティの確保に加え，実環境下におけるリスク制約への対応や，近似精度と実行速度の両立 [14] に取り組む予定である．これらの課題に対しては，実データを用いた実験により実証的な検証を進めていく．

参考文献

[1] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

[2] Derek Zhiyuan Cheng, Ruoxi Wang, Wang-Cheng Kang, Benjamin Coleman, Yin Zhang, Jianmo Ni, Jonathan Valverde, Lichan Hong, and Ed Chi. Efficient data representation learning in google-scale systems. In *Proceedings of the 17th ACM Conference on Recommender*

Systems, pp. 267–271, 2023.

[3] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[4] Yuto Nakashima, Mingzhe Yang, and Yukino Baba. Swipeganspace: Swipe-to-compare image generation via efficient latent space exploration. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pp. 675–685, 2024.

[5] Miguel Lázaro-Gredilla, Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Aníbal R Figueiras-Vidal. Sparse spectrum gaussian process regression. *The Journal of Machine Learning Research*, Vol. 11, pp. 1865–1881, 2010.

[6] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pp. 127–135, 2013.

[7] Miguel González-Duque, Richard Michael, Simon Bartels, Yevgen Zainchkovskyy, Søren Hauberg, and Wouter Boomsma. A survey and benchmark of high-dimensional bayesian optimization of discrete sequences. *arXiv preprint arXiv:2406.04739*, 2024.

[8] Kirthevasan Kandasamy, Jeff Schneider, and Barnabás Póczos. High dimensional bayesian optimisation and bandits via additive models. In *International conference on machine learning*, pp. 295–304. PMLR, 2015.

[9] Ziyu Wang, Frank Hutter, Masrour Zoghi, David Matheson, and Nando De Feitas. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, Vol. 55, pp. 361–387, 2016.

[10] Sattar Vakili, Henry Moss, Artem Artemev, Vincent Dutordoir, and Victor Picheny. Scalable thompson sampling using sparse gaussian process models. *Advances in neural information processing systems*, Vol. 34, pp. 5631–5643, 2021.

[11] Carl Hvarfner, Erik Orm Hellsten, and Luigi Nardi. Vanilla bayesian optimization performs great in high dimensions. *arXiv preprint arXiv:2402.02229*, 2024.

[12] Yusuke Miyake, Ryuji Watanabe, and Tsunenori Mine. Online nonstationary and nonlinear bandits with recursive weighted gaussian process. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 11–20. IEEE, 2024.

[13] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, Vol. 24, , 2011.

[14] Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, et al. Rethinking attention with performers. *arXiv preprint arXiv:2009.14794*, 2020.