

ODAT-SEのチュートリアル

データ解析フレームワークODAT-SEミーティング

2025/5/08 13:00-14:00

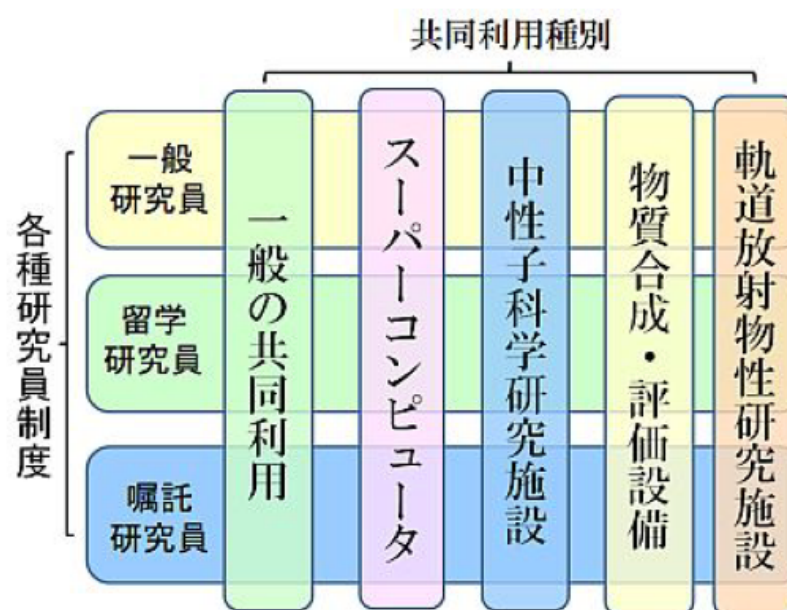
核融合科学研究所

東京大学物性研究所 青山龍美、本山裕一、吉見一慶

物性研究所の紹介



- 目的：物性科学の研究推進
～物性分野における世界最高水準の基礎研究の先導
- 1957年：東京大学附置の全国共同利用研究所として設立
- 2010年：全国共同利用・共同研究拠点に認定



共同利用スパコン

2023年度実績：

共同利用採択課題数 = 347

利用者数 = 1055

成果論文数 = 250

ソフトウェア開発・高度化プロジェクトの紹介

よりユーザーフレンドリーな
スーパーコンピュータの提供を目指して

毎年12月に公募 (2015年～)

- 1,2件の課題を選定
 - OSSとして年度末に公開
- 年間スケジュールイメージ



2023年度HPCIソフトウェア賞
普及部門賞の最優秀賞受賞！

主要メンバー (2025/2/27現在)



吉見 一慶

(2015/4～)



本山 裕一

(2018/4～)



青山 龍美

(2022/5～)

April



Kick-off meeting

May ~ Feb.



GitHub

Software development



Regular meeting/month

March ~



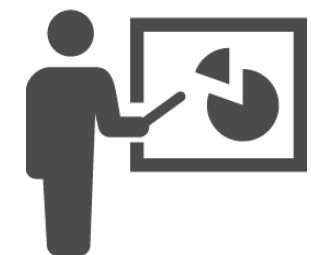
Website

MateriApps

Portal site



Preinstall



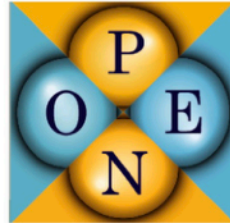
Lecture

これまで高度化されたソフトウェア一覧

First principles calculations and related applications



abICS



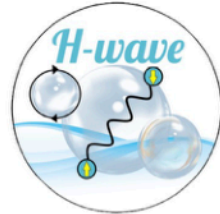
OpenMX



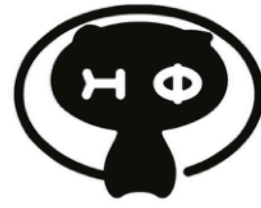
RESPACK

ESM-RISM

Quantum lattice model solvers



H-wave



H Φ



mVMC



DSQSS



DCore



TeNeS

FY 2023 HPCI software award
mVMC : Best performance award 、
H Φ : Outstanding performance award

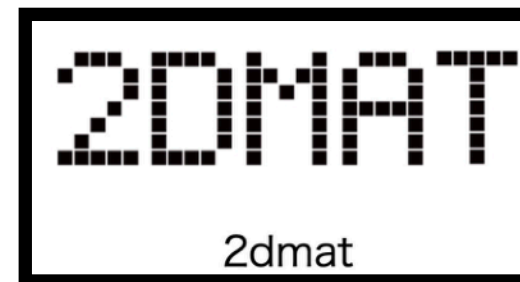
Mathematical libraries and others



PHYSBO



K ω



2dmat

MateriApps Installer

MateriApps

ODAT-SEの元プログラム (2020, 2021, 2024)

URL: <http://www.pasums.issp.u-tokyo.ac.jp>

本講演の流れ

- ODAT-SE開発の歴史

- ODAT-SEとは？

講演者 本山裕一、15分講演+5分質疑応答

- ODAT-SEを自分の問題に適用するには？

講演者：青山龍美、15分講演+5分質疑応答

- 今後の展望

講演者：吉見一慶、5分講演+5分質疑応答

ODAT-SE開発の歴史 (1)

2DMATの開発



星先生@NIFS

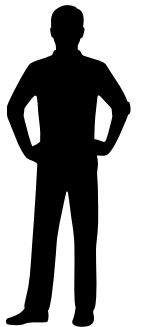
- 星グループで物性実験(**TRHEAPD**)に関するデータ解析を実施
 - 実験データとシミュレーションのズレを、レプリカ交換モンテカルロ法等を用いて網羅的に評価するプログラムの作成。
- **2020年度ソフトウェア開発高度化プロジェクト: プログラムの一般化**
 - プロトタイプのプログラムを整備・一般化し、**2DMAT**として公開。
- **2021年度ソフトウェア開発高度化プロジェクト: 実験シミュレータ+アルゴリズムの追加**
 - 物性実験用シミュレータの追加 (SXR, LEED)
 - モンテカルロ法にポピュレーションアニーリングモンテカルロ法を追加。
(並列数を増やす場合に効果的なモンテカルロ法)

ODAT-SE開発の歴史 (2)

ODAT-SEへの発展

- 問題点

- 2DMATではシミュレータと最適化モジュールが一体化しており、シミュレータ追加が行いにくい (本来は最適化モジュールと切り離しができるはず)。
- 2DMATという名称が2次元物質のみ対象としているように感じられる。



- **2024年度ソフトウェア開発高度化プロジェクト**

- 最適化モジュール(**ODAT-SE**)とシミュレータ部の切り離し
 - シミュレータ用ラッパー： <https://github.com/orgs/2DMAT/repositories>
- ポスト(主に可視化)ツールの整備
- 2024/11/25にODAT-SE v.3.0.0がリリース(2DMATのバージョンを引き継いでv.3からスタート)

本講演の流れ

- ODAT-SE開発の歴史

- ODAT-SEとは？

講演者 本山裕一、15分講演+5分質疑応答

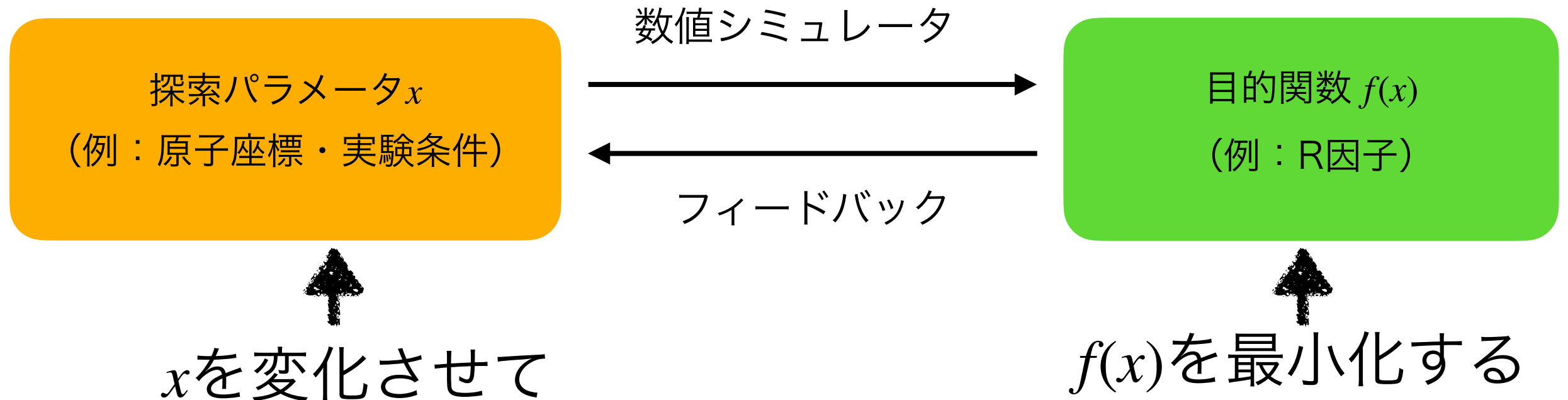
- ODAT-SEを自分の問題に適用するには？

講演者：青山龍美、15分講演+5分質疑応答

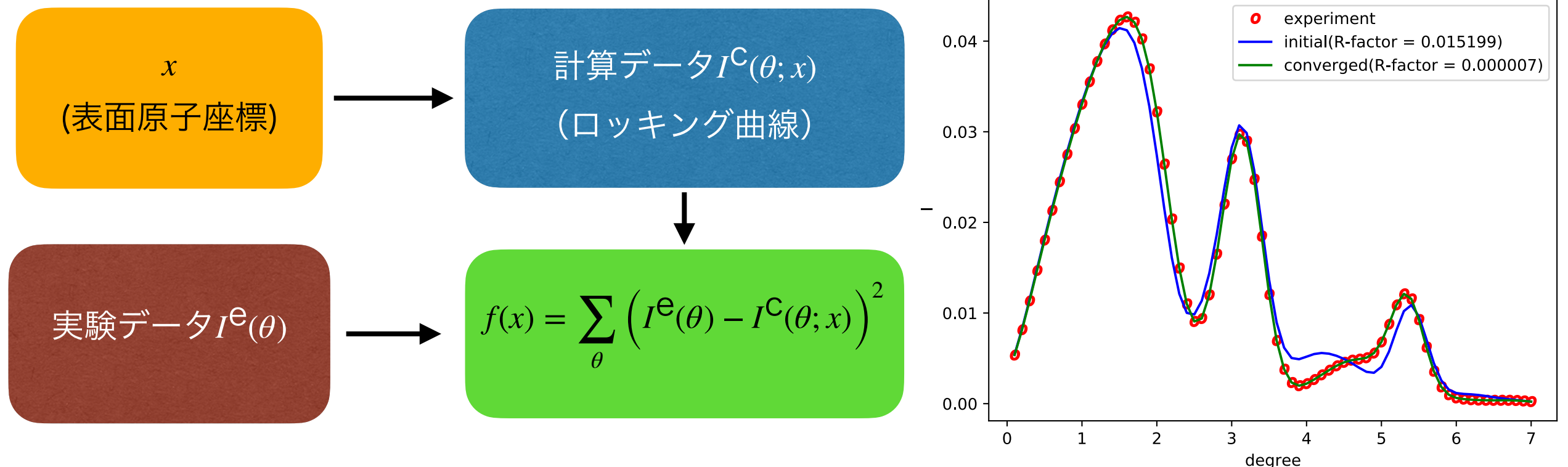
- 今後の展望

講演者：吉見一慶、5分講演+5分質疑応答

逆問題解析



数値シミュレータの例: TRHEPD シミュレータ (sim-trhepd-rheed)



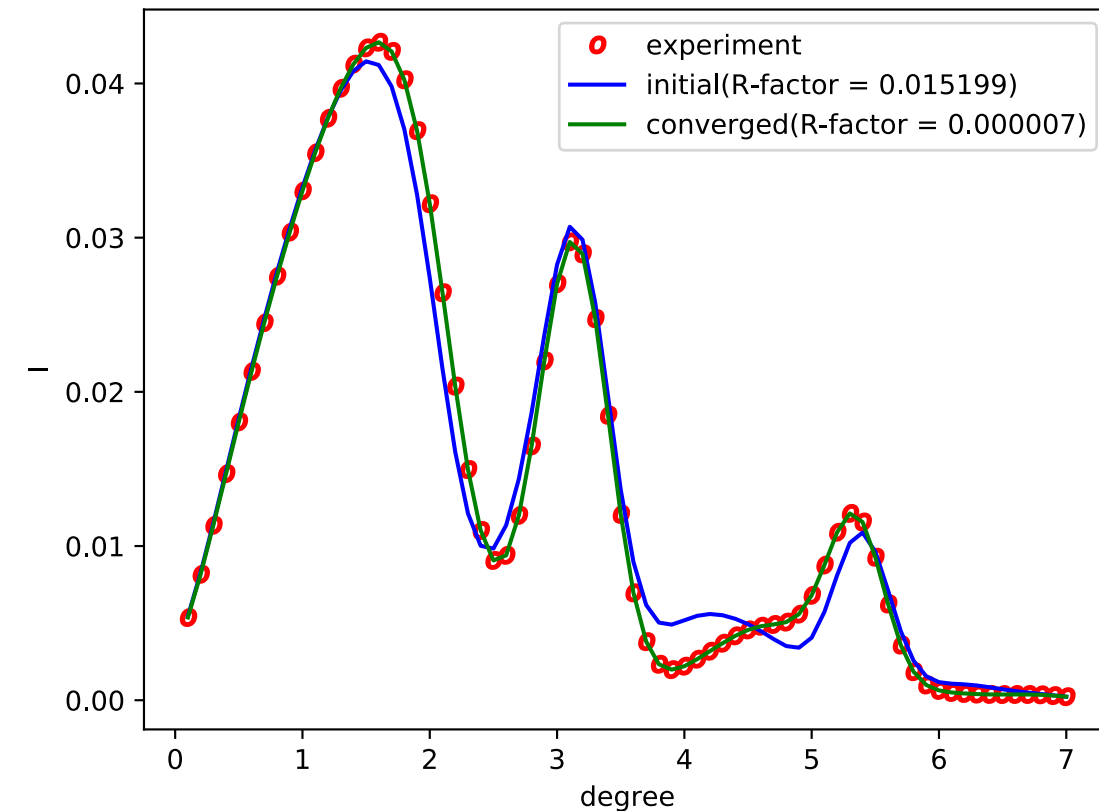
実験データと計算データの誤差を最小化する = 実験を再現する x を探す

ODAT-SEの構成

- ▶ ODAT-SEは大雑把に次の2つの主コンポーネントからなる

- ▶ 順問題**Solver**: 目的関数 $f(x)$ を計算する

- ▶ 実験データと直接比較可能な計算データを生成する
- ▶ 実験データとの差分として $f(x)$ を出す



- ▶ 逆問題**Algorithm**: $f(x)$ をもとに x の空間を探索する

- ▶ 最適化（ベイズ最適化など）
- ▶ サンプルング（モンテカルロ法など）

- ▶ Solverを切り替えることで**問題（実験手法）が変わっても同じように解析可能**
- ▶ Algorithmを切り替えることで**様々な数理手法で実験データを解析可能**

ODAT-SEのインストール

- ▶ ODAT-SEライブラリと、個別の順問題ソルバーモジュールをそれぞれインストールする
- ▶ ODAT-SE/順問題ソルバーはPython3で書かれている
 - ▶ pipコマンド一発でインストールできるように整備・公開している

```
python3 -m pip install ODAT-SE
```

- ▶ odatse moduleとodatseコマンドがインストールされる
- ▶ 順問題ソルバーについては次のパートを参照
- ▶ チュートリアル用のサンプルファイルやスクリプトが必要ならばGitHubからダウンロードする

```
git clone https://github.com/issp-center-dev/ODAT-SE
```

ODAT-SEの入力ファイル

- ▶ odatseコマンドはひとつの入力ファイル (input.toml) を引数に取る
- ▶ 入力ファイルはTOML形式
 - ▶ <https://toml.io/ja/>
 - ▶ 大雑把には、
 - ▶ [section] によるセクションと
 - ▶ key = value による定義
- ▶ 右の入力ファイルは
 - ▶ Himmelbrau 関数を ([solver])
 - ▶ Nelder-Mead 法で最適化する ([algorithm])

実行方法

```
$ odatse input.toml
```

入力ファイル例

```
$ cat input.toml
```

```
[base]
```

```
dimension = 2
```

```
output_dir = "output"
```

```
[algorithm]
```

```
name = "minsearch"
```

```
seed = 12345
```

```
[algorithm.param]
```

```
max_list = [5.0, 5.0]
```

```
min_list = [-5.0, -5.0]
```

```
[solver]
```

```
name = "analytical"
```

```
function_name = "himmelbrau"
```

Algorithm

探索手法

全探索

グリッドサーチ (mapper)

最適化

Nelder-Mead法 (minsearch)

ベイズ最適化 (bayes)

モンテカルロサンプリング

レプリカ交換法 (exchange)

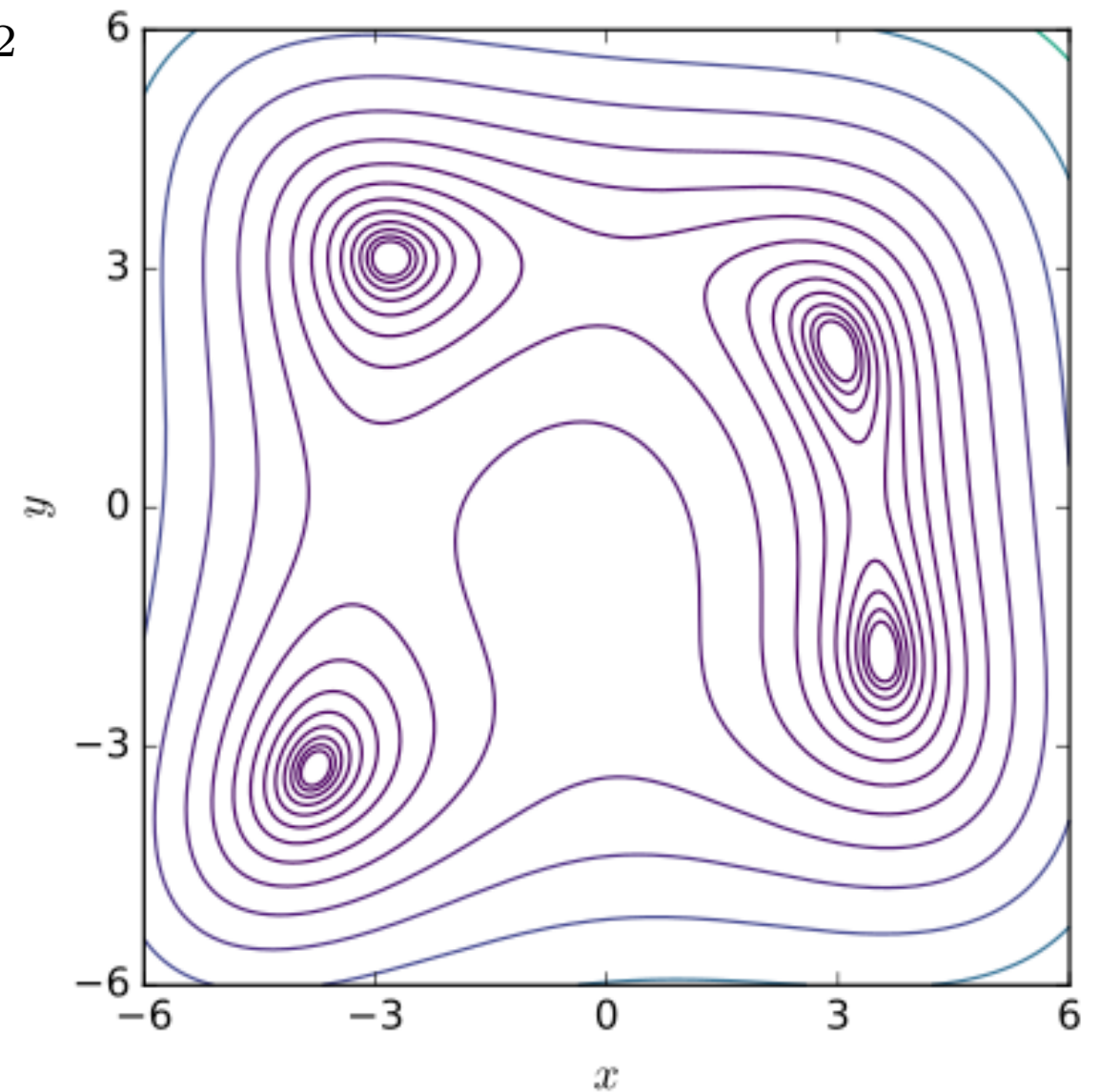
ポピュレーションアニーリング (pamc)

例題：Himmelblau関数の最小値探索

- ▶ 以下ではHimmelblau関数をデモンストレーションに用いる

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

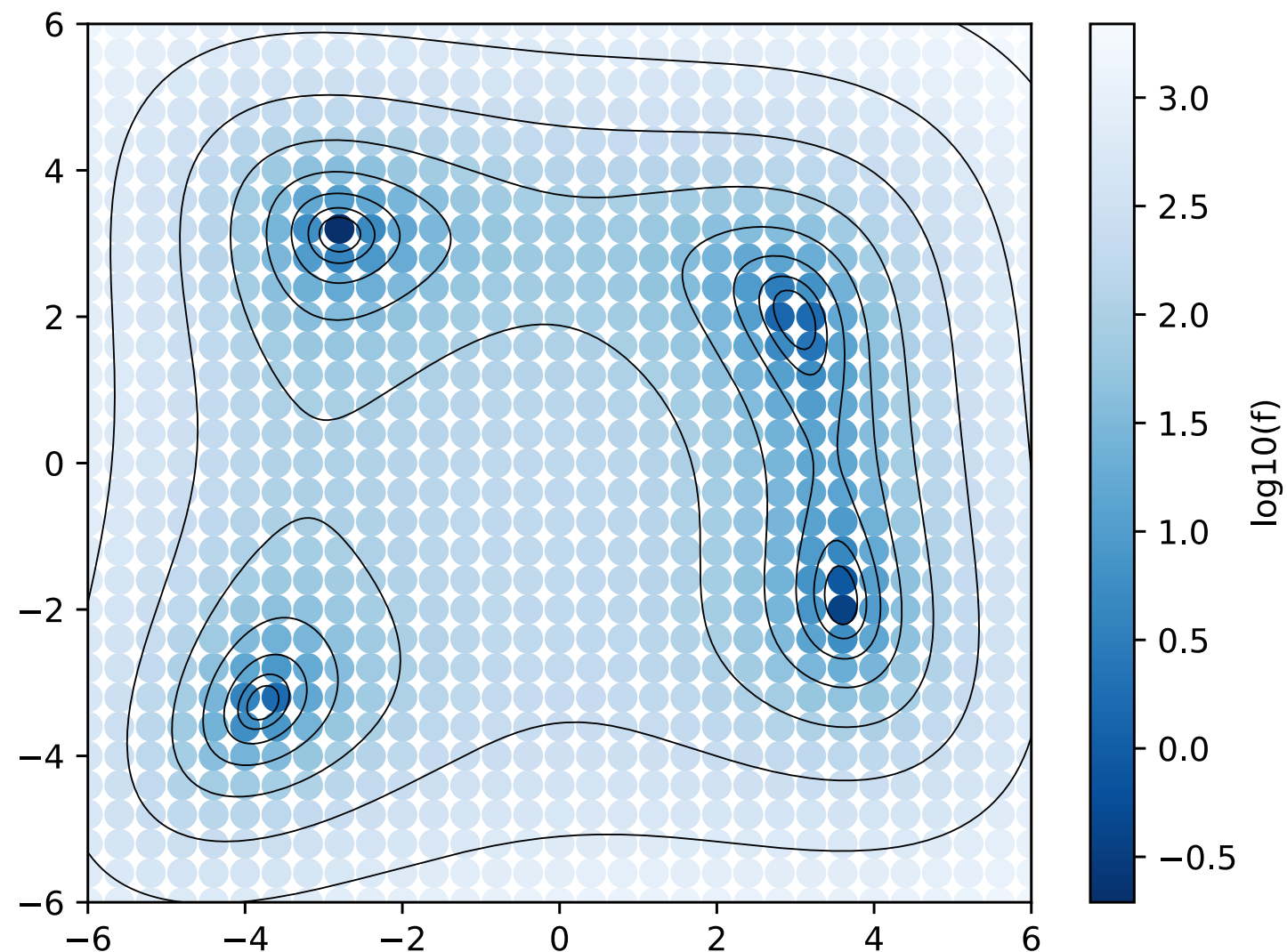
- ▶ 右図は $\log f$ の等高線プロット
- ▶ 4つの最小点 $f = 0$ を持つ
- ▶ ODAT-SEではSolverとしてこのようなデモンストレーション関数をいくつか実装してある



By Nicoguaro - Own work, CC BY 4.0,
<https://commons.wikimedia.org/w/index.php?curid=51597574>

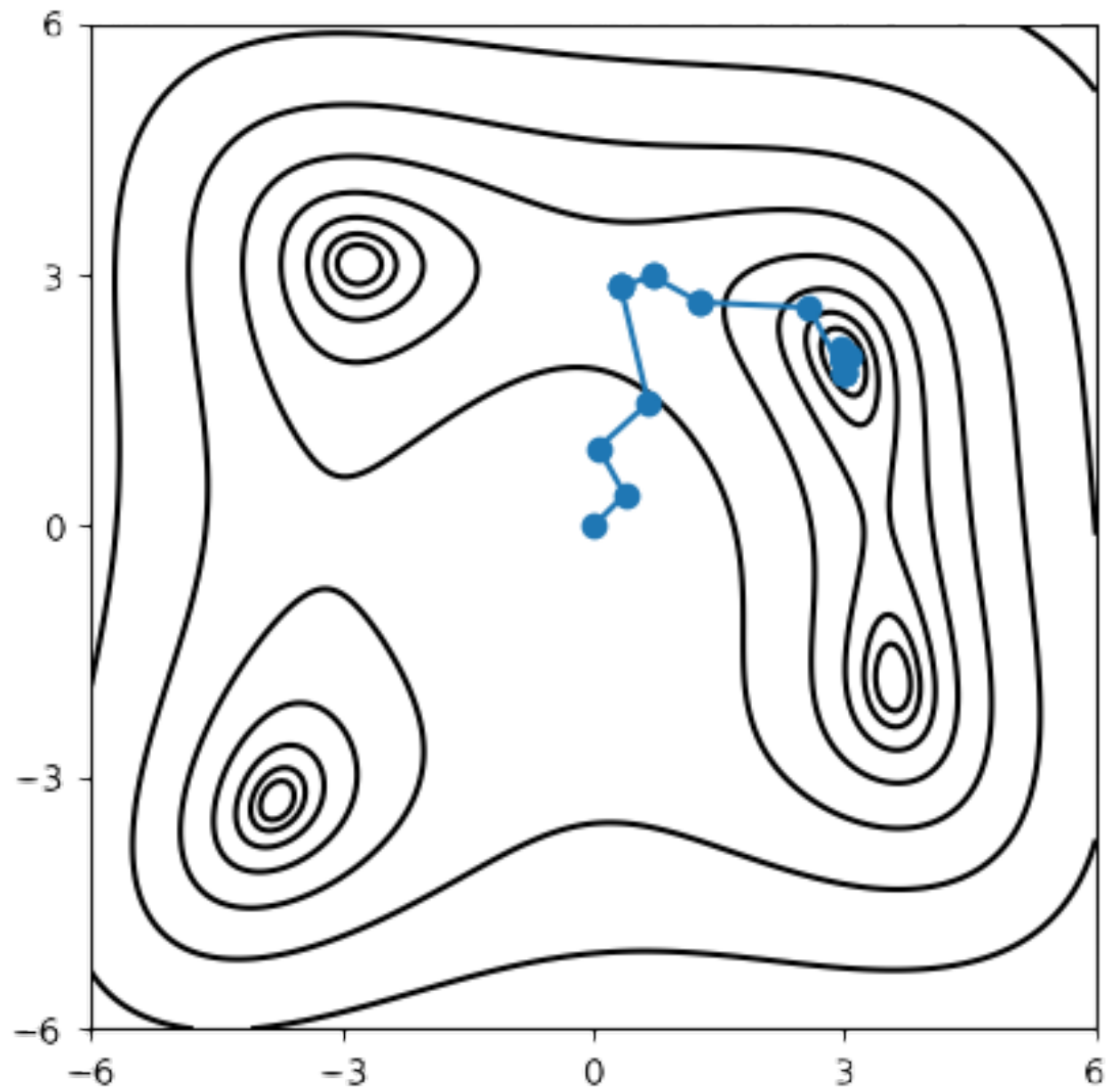
グリッドサーチ

- ▶ 探索空間をグリッド（格子点）に区切って、全点計算する
- ▶ 並列計算可能
 - ▶ 単純に担当箇所を分ける
- ▶ **大雑把な傾向を掴むのに便利**
 - ▶ 探索空間の範囲を変える
 - ▶ Nelder-Mead（次項）の初期値を作る
- ▶ パラメータ数（＝探索空間の次元）が増えると点数（＝計算コスト）も指数的に増えていくことに注意
 - ▶ 点数を抑えるとスカスカになる



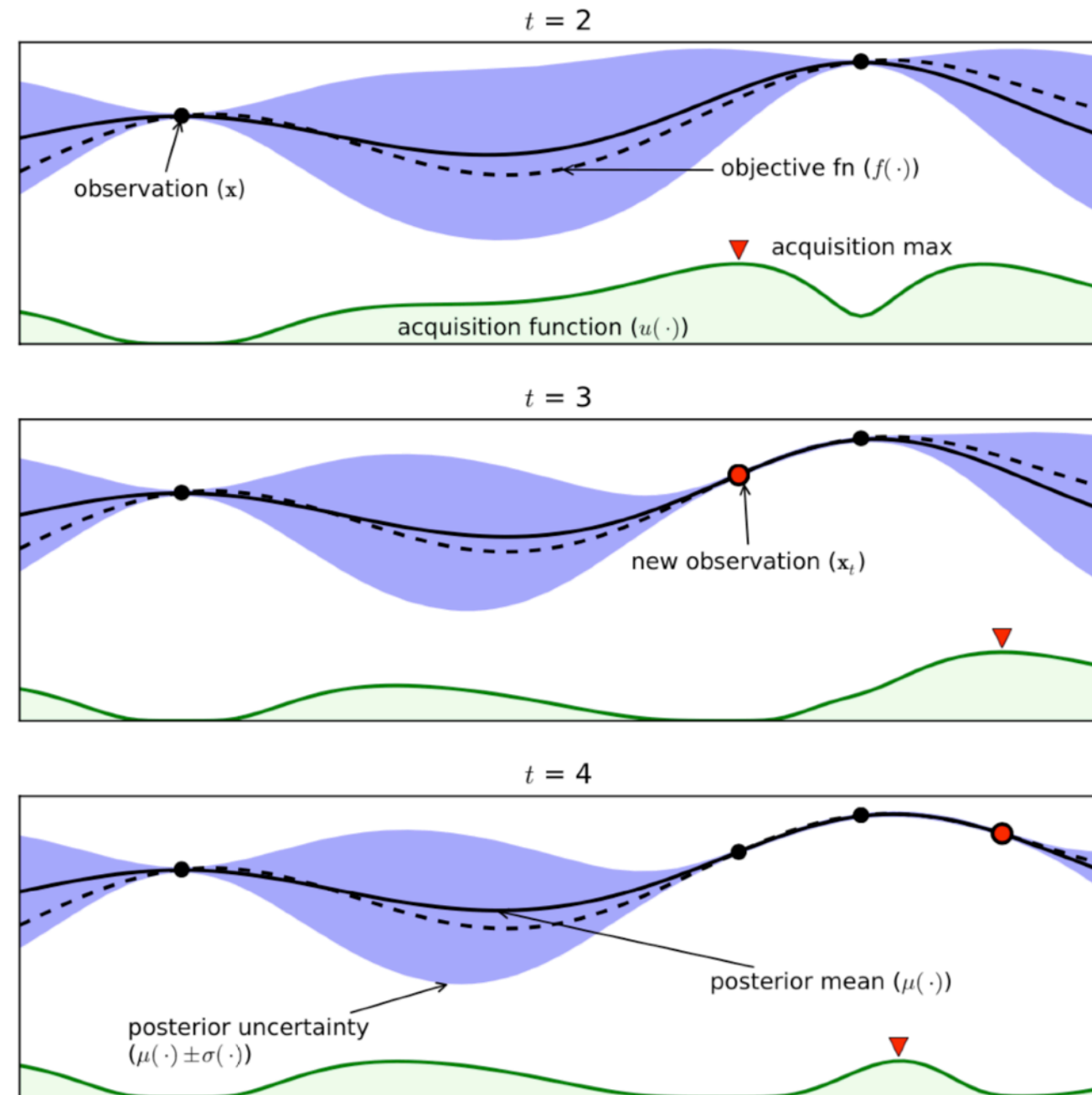
Nelder-Mead法

- ▶ Downhill Simplex法、アメーバ法などとも呼ばれる
- ▶ 連続空間の最適化手法
- ▶ 必要なのは目的関数の値 $f(x)$ のみ
(導関数は不要！)
- ▶ 初期値依存性がある
 - ▶ 右図では原点から始めた
 - ▶ 他の手法であたりを付けてから、最後に細かい結果を得たいときに使うとよい



ベイズ最適化

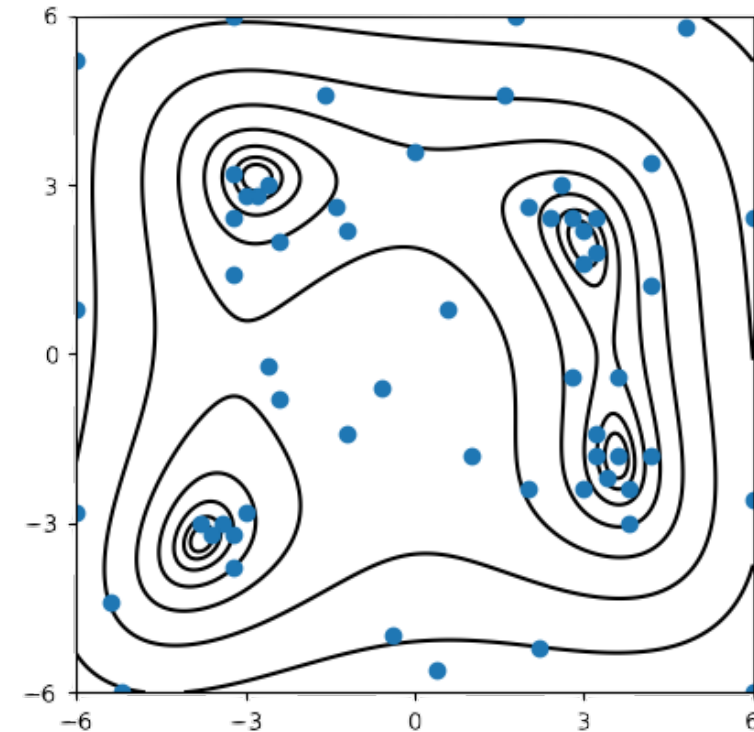
- ▶ 目的関数 $f(x)$ の評価が難しい（コストがかかる）ときに有用
- ▶ $f(x)$ をサロゲートモデル $g(x)$ で近似する
 - ▶ $g(x)$ という関数をベイズ推論の事後分布として表現し最適化する = 「ベイズ」最適化
- ▶ 破線が $f(x)$ 、黒丸・赤丸が測定値
- ▶ 実線が $g(x)$ の期待値で、青がその不確かさ
- ▶ 緑が獲得関数
 - ▶ 獲得関数は、 $g(x)$ の期待値と分散とから定義される、候補点の「スコア」
- ▶ 三角が獲得関数の最大
 - ▶ 次にこの x で $f(x)$ を評価し、 $g(x)$ を更新



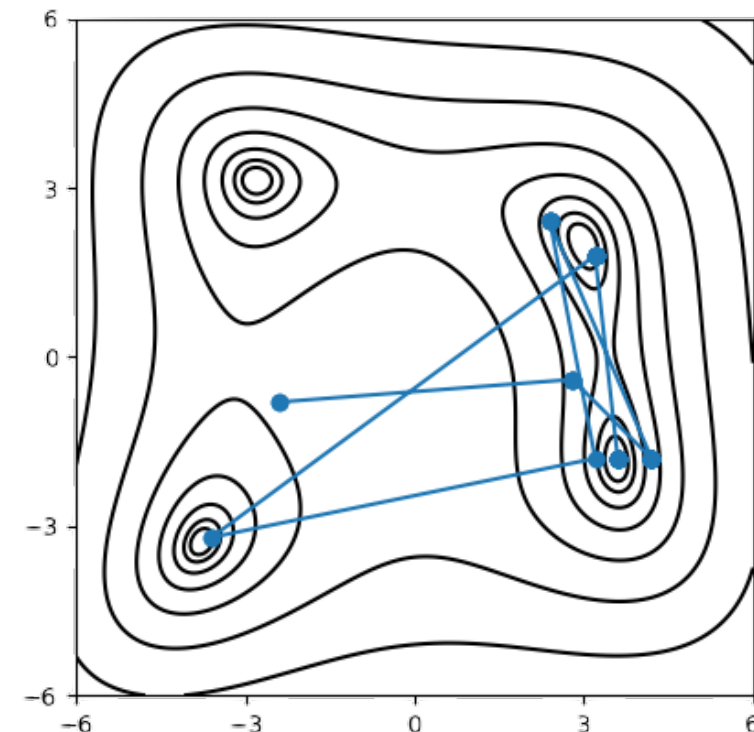
taken from E. Brochu, et al., arXiv:1012.2599

ベイズ最適化

- ▶ odat-se はベイズ最適化に **PHYSBO** ライブラリを利用している
 - ▶ <https://www.pasums.issp.u-tokyo.ac.jp/physbo/>
 - ▶ これも物性研ソフトウェア開発・高度化プロジェクトで支援している
- ▶ 探索空間は離散化されている（最初に離散化した点の集合を与える）
 - ▶ 解像度や探索コストをコントロール可能
 - ▶ 気になる場合は最適化結果を初期値にして Nelder-Mead をする
 - ▶ 獲得関数の計算で並列計算可能
 - ▶ 探索空間中で担当箇所を分ける



探索した点



最小値の変遷

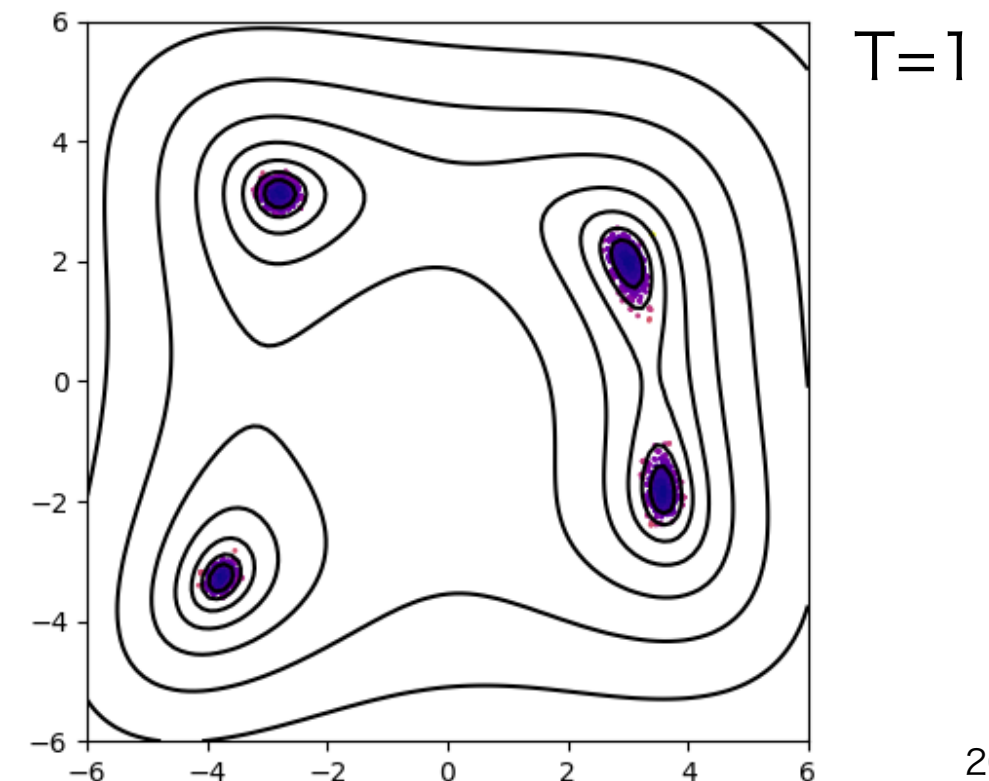
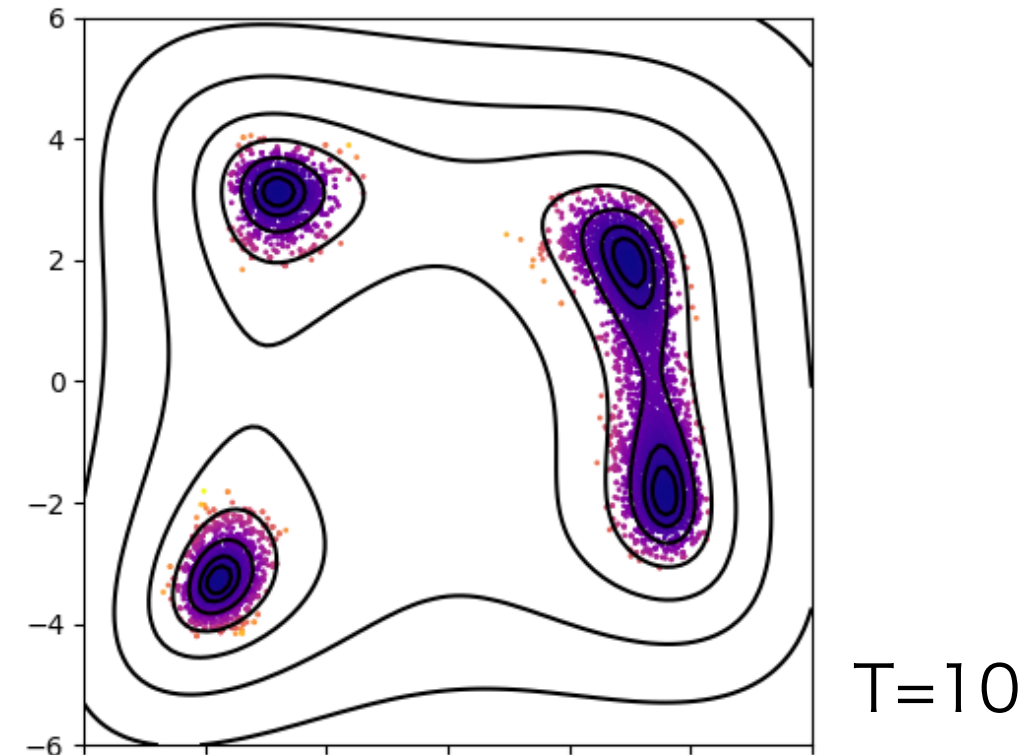
モンテカルロサンプリング

- ▶ ここまでは $f(x)$ が最小値をとる点 x を探索した（点推定）
- ▶ 結果の**不確かさ**を見積もりたい、あるいは x の**確率分布** $P(x|D)$ を見積もりたい
 - ▶ D は与えられた情報、例えば再現すべき実験データ
- ▶ **ベイズの定理**より、
$$P(x|D) = \frac{P(D|x)P(x)}{P(D)} = \frac{P(D|x)P(x)}{\int dx P(D|x)P(x)}$$
- ▶ 事前分布 $P(x)$ は探索空間上の一様分布 $P(x) = 1/\Omega_x$ とする(Ω_x は探索空間の体積)
- ▶ 条件付き確率 $P(D|x)$ として**カノニカル分布**（ボルツマン分布）を用いる
 - ▶ ハイパーパラメータ T を導入して $P(D|x; T) \propto \exp[-f(x; D)/T]$ とする
 - ▶ T は f に対する**精度・解像度**（統計物理のアナロジーでは**温度**）
- ▶ 全体の $P(x|D; T)$ も同じカノニカル分布に従う（規格化因子のみ異なる）
 - ▶ x を**マルコフ連鎖モンテカルロ法**でサンプリングする
 - ▶ T をどうやって設定するかが問題(→**レプリカ交換法・ポピュレーションアニーリング**)

レプリカ交換モンテカルロ法

K. Hukushima and K. Nemoto, J. Phys. Soc. Jan.. **65**, 1604 (1996)

- ▶ マルコフ連鎖モンテカルロ法では、現在の x から確率的に次の x' を生成する
 - ▶ 普通は x の近くに x' を配置する
 - ▶ 高さ T ぐらいの山なら乗り越えられる
 - ▶ T が高すぎると構造 (山とか谷) が見えなくなる
 - ▶ $T = \infty$ では重み w が一定になる
 - ▶ T が小さすぎると局所解 (穴) から出てこられなくなる
- ▶ **レプリカ交換モンテカルロ法**
 - ▶ 複数の**レプリカ**を用意
 - ▶ 各レプリカは異なる T で並列してモンテカルロサンプリングを行う
 - ▶ 時々 T を交換する

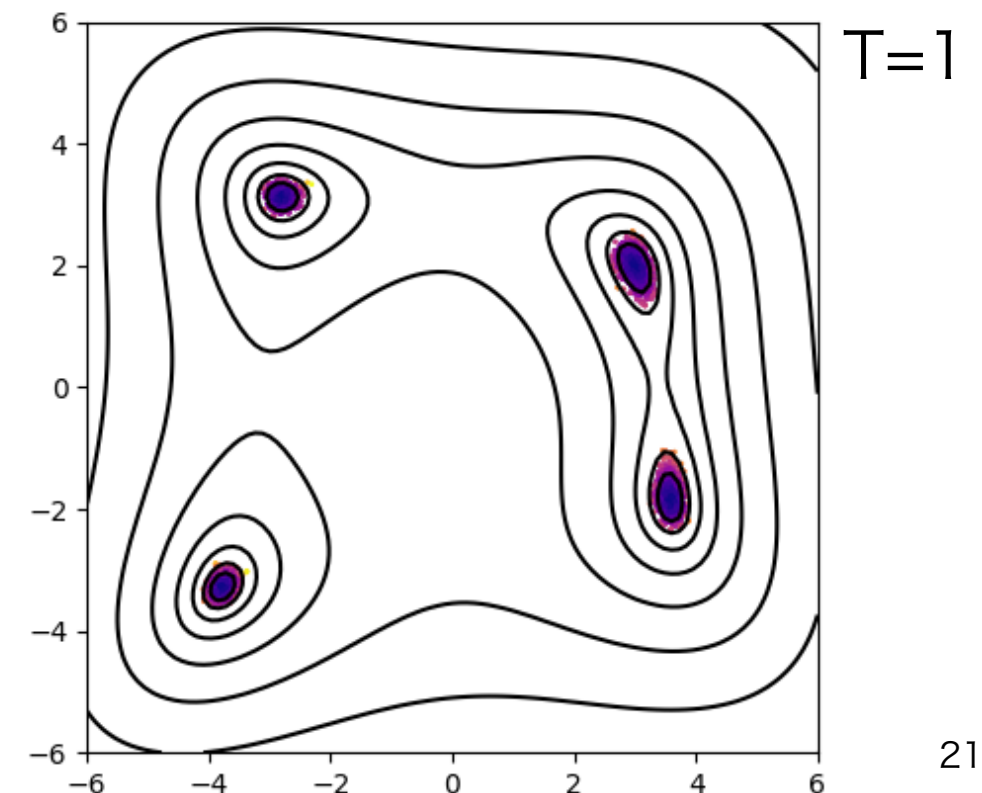
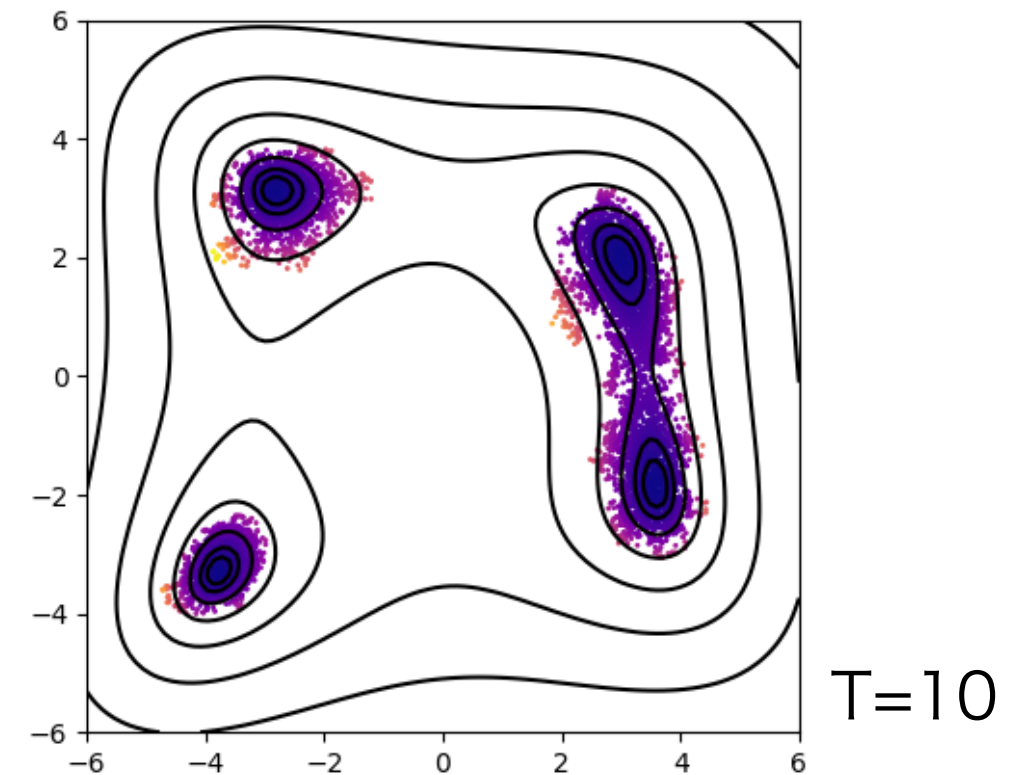


ポピュレーションアニーリング法(PAMC)

K. Hukushima and Y. Iba, AIP Conf. Proc. **690**, 200 (2003)

- ▶ 交換モンテカルロ法では複数の T を並列に計算していた
 - ▶ 並列化は容易だが、あまり並列度をあげられない
 - ▶ プロセス数が少ないと T の数を取りづらい
- ▶ 共通の T で大量のレプリカを並列計算して、同時に T を下げていく（アニーリング）のが
ポピュレーションアニーリング

- ▶ 小さい T で局所解にトラップされても数の暴力で押し切る
 - ▶ 望み薄なレプリカを消去し、うまくいっているものを分裂させる
- ▶ 原理的には好きなだけ並列数を増やせる
 - ▶ 並列数を増やせばレプリカ数が増え、ひいては統計精度が増す

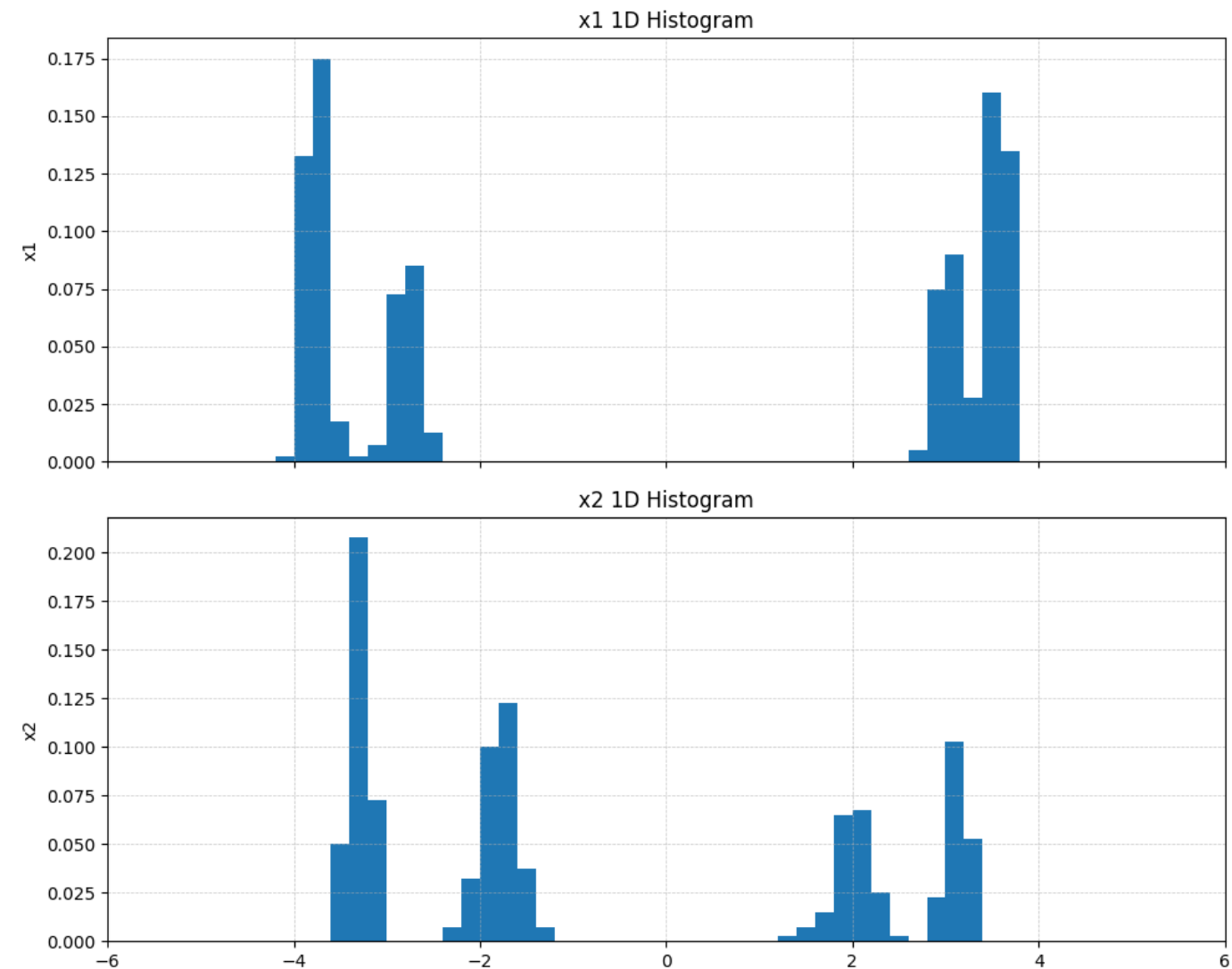
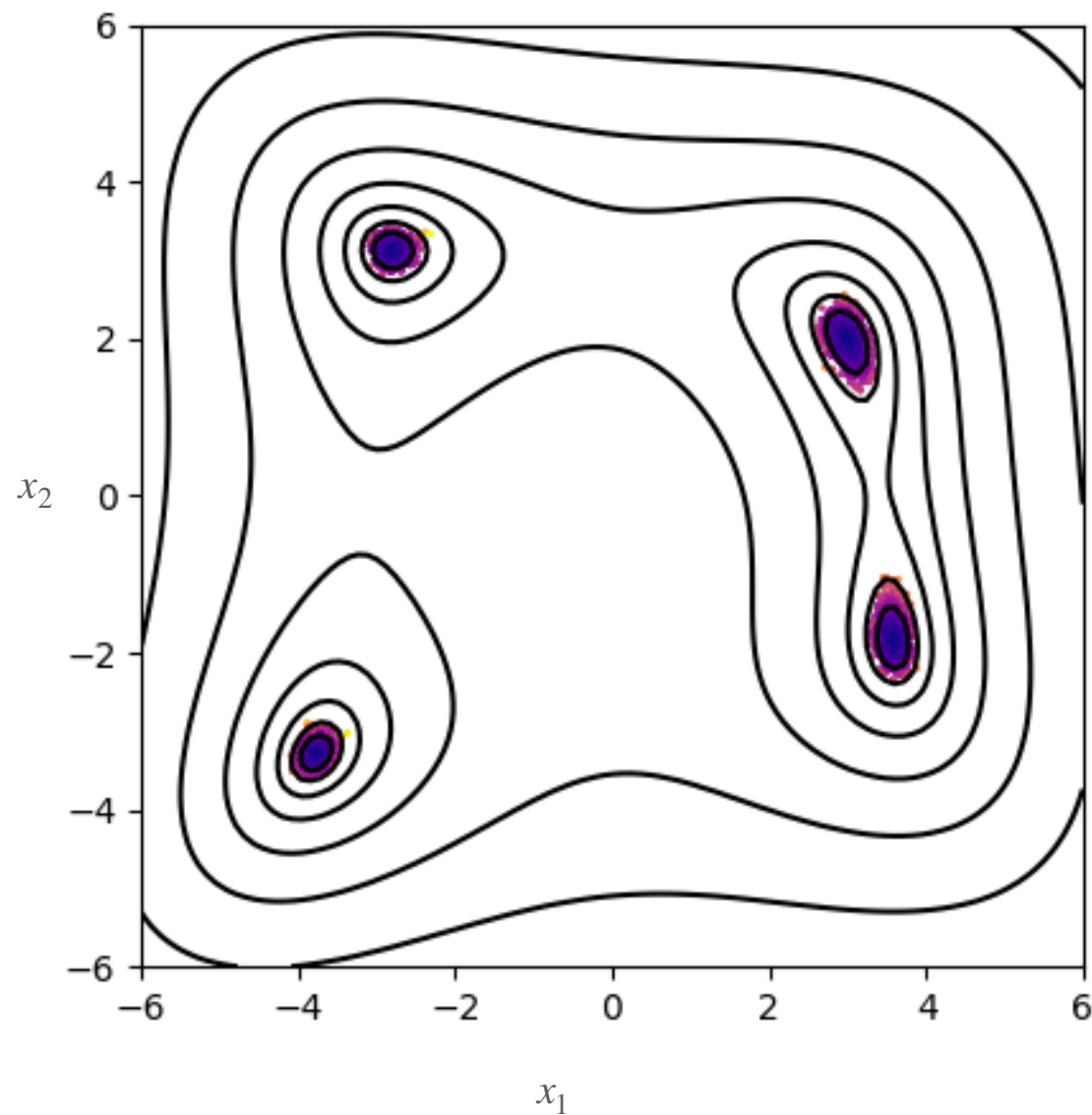


後処理ツール (PAMC)

- ▶ PAMCについてはサンプリングの後処理ツールがいくつか用意されている
- ▶ `separateT.py`
 - ▶ ログファイルを各 T に分割する
- ▶ `summarize_each_T.py`
 - ▶ T ごとに各レプリカ最後のスナップショット $(x, f(x))$ をまとめる
- ▶ `plt_1D_histogram.py`, `plt_2D_histogram.py`
 - ▶ `summarize_each_T.py` でまとめた x のスナップショットから1次元、2次元の周辺分布を計算・プロットする
- ▶ `plt_model_evidence.py`
 - ▶ T ごとの**model evidence** $P(D; T)$ を計算する
 - ▶ model evidenceを最大化するような T （や他のハイパーパラメータ）が最も尤もらしいモデル

後処理ツール (PAMC)

1次元周辺分布 by plt_1D_histogram.py



3パラメータ以上の**多次元解析**の可視化で効果的！

まとめと展望

逆問題解析パート

- ▶ ODAT-SE
 - ▶ 一般的な最適化・サンプリングのためのフレームワークです
 - ▶ グリッドサーチ
 - ▶ Nelder-Mead
 - ▶ ベイズ最適化
 - ▶ レプリカ交換モンテカルロ法
 - ▶ ポピュレーションアニーリング
 - ▶ 手元のパソコンから大規模計算機まで使えます
 - ▶ Nelder-Mead以外はMPI並列化対応済み！
 - ▶ mpi4pyのインストールが必要
 - ▶ 順問題＝解きたい問題をどう定義するか？→次パートのテーマ

本講演の流れ

- ODAT-SE開発の歴史

- ODAT-SEとは？

講演者 本山裕一、15分講演+5分質疑応答

- ODAT-SEを自分の問題に適用するには？

講演者：青山龍美、15分講演+5分質疑応答

- 今後の展望

講演者：吉見一慶、5分講演+5分質疑応答

本講演の流れ

- ODAT-SE開発の歴史
- ODAT-SEとは？
講演者 本山裕一、15分講演+5分質疑応答
- ODAT-SEを自分の問題に適用するには？
講演者：青山龍美、15分講演+5分質疑応答
- 今後の展望
講演者：吉見一慶、5分講演+5分質疑応答
- まとめ

今後の展望：生成AIの活用 (1)

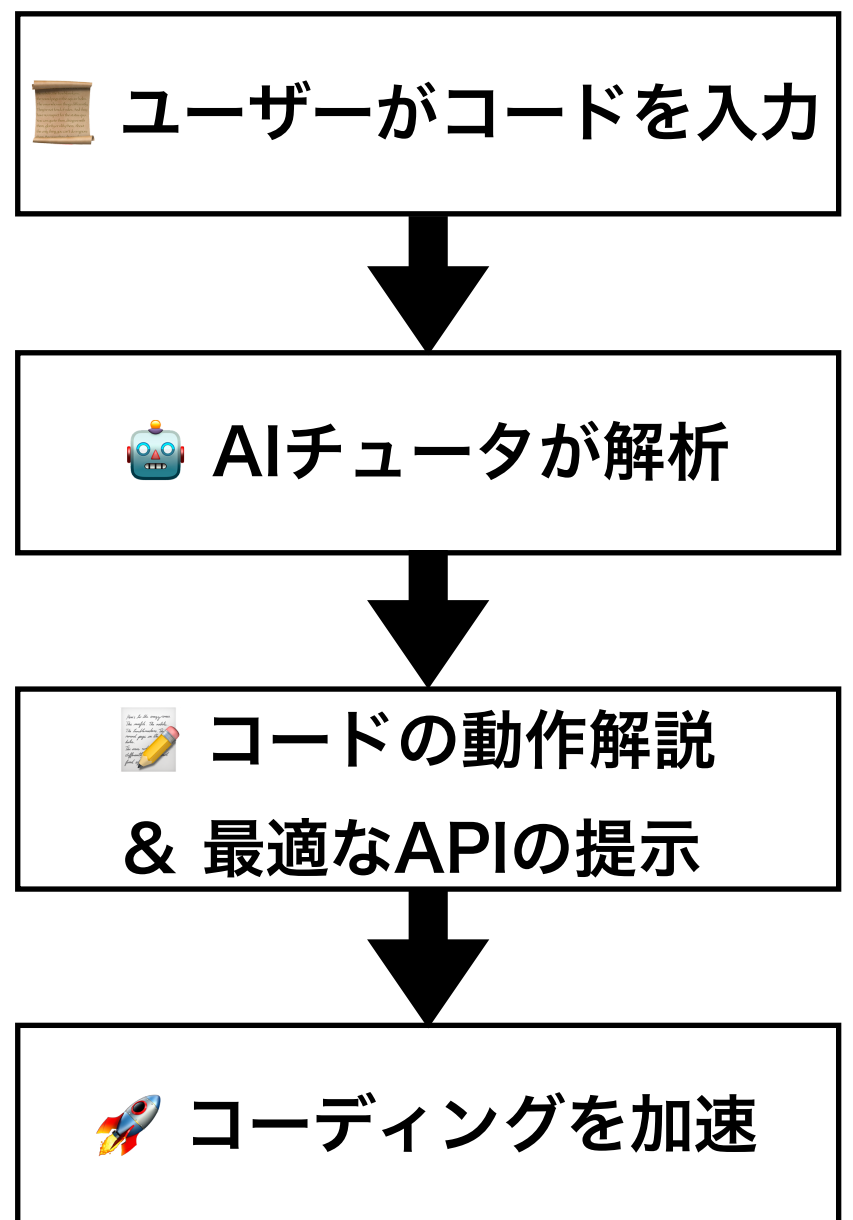
AIチュータの進化と自動コード解説・最適な呼び出し方の提案

AIチュータの役割の変化

- AIがコードの動作を解析し、自動で解説。
- 適切なAPIや関数の利用方法を提示。
- 例: 「このコードは何をしているのか？」 → AIが解説。

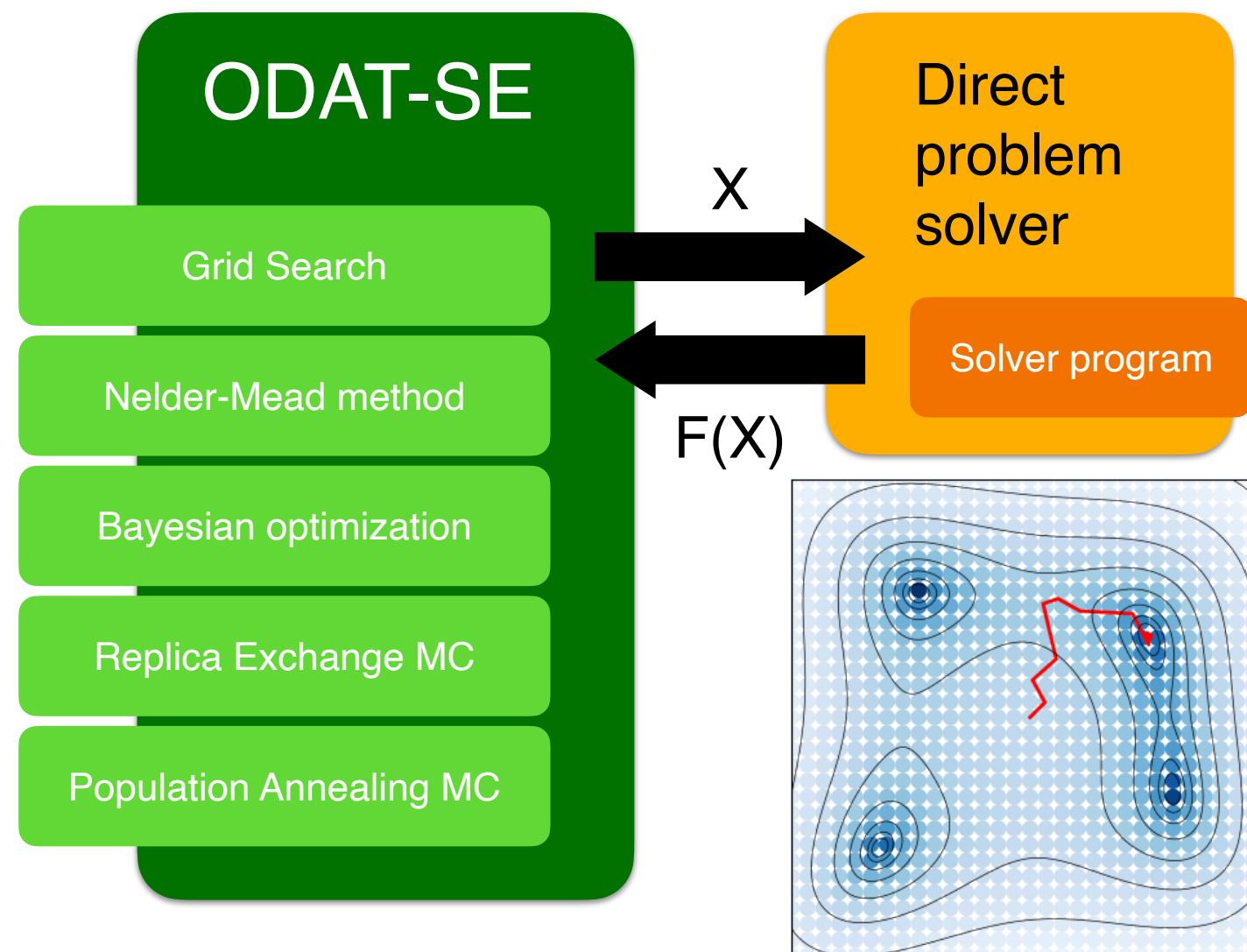
取り組むべき課題

- AIによるコード解析を活かすために、開発者が**どのようにコードを書けばAIが理解しやすいかを考慮**する必要がある。
- AIチュータを活用するために、開発者が**適切な質問をし、適切な出力を得るスキル**を身につける必要がある。
- **AIによる解説が正しいかどうかを判断**し、適切に修正・補完できる知識が求められる。



今後の展望：生成AIの活用 (1)

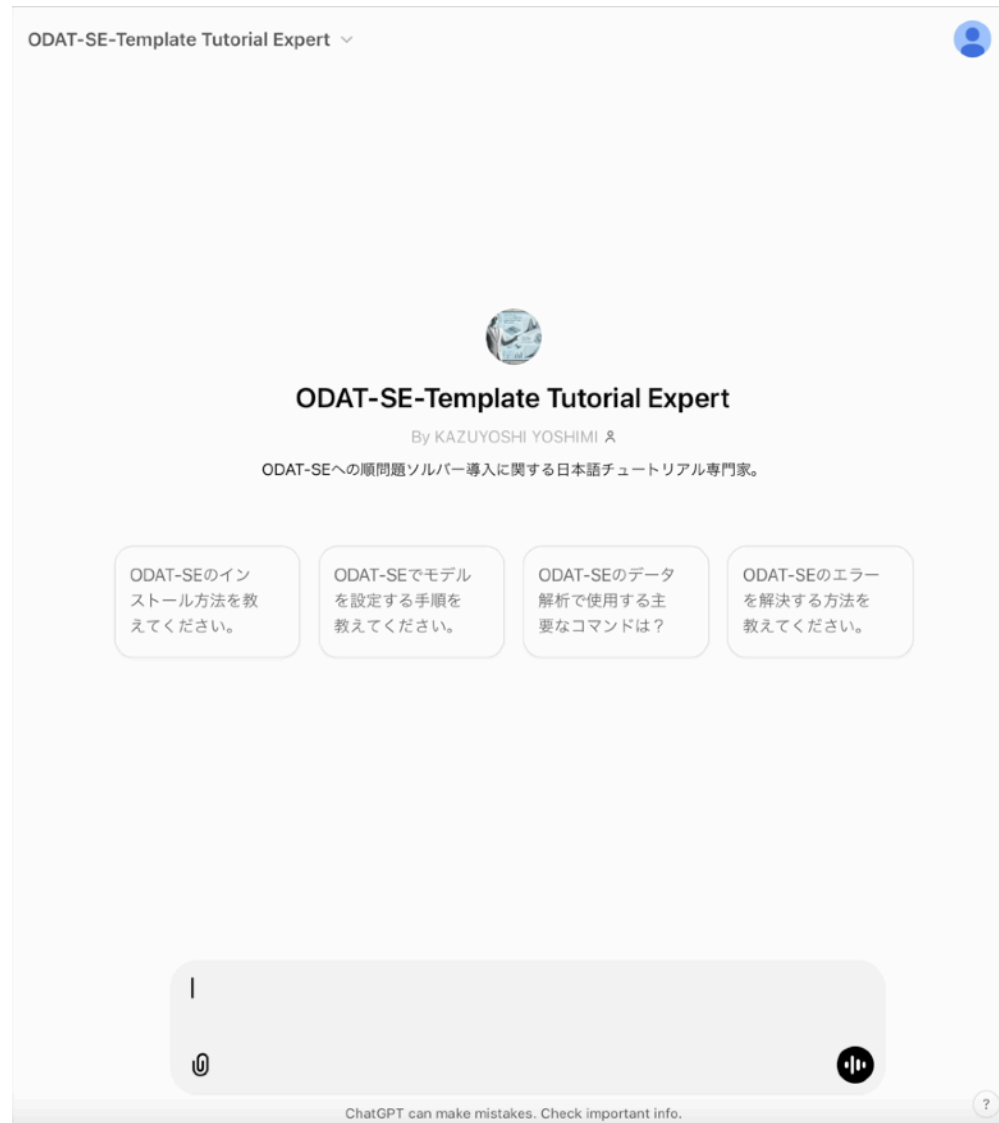
AIチュータの進化と自動コード解説・最適な呼び出し方の提案



マニュアル・チュートリアル of 整備・作成

今後の展望：生成AIの活用 (1)

ODAT-SE template Tutorial Expertの試提供



チュートリアルに従って実行した際に、エラーが出た場合には、そのエラーをメッセージに貼り付けると解決方法を提案してくれる。

実際、今の場合、input.tomlに
[solver]
type = "FourierSolver"
を追加していないため、エラーが出たが、これを追加すると実行できた。

CCMSハンズオン講習会 2024/12/2開催でテスト運用

<https://ccms.issp.u-tokyo.ac.jp/event/7002>

今後の展望：生成AIの活用 (2)

AIが複数のレガシーコードを統合し、最適なシステムを提案？

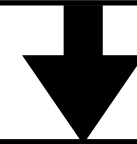
🔗 モジュールの関連性をAIが解析

- レガシーコードのモジュール間の関係を理解し、再構成。
- 自動で「推奨モジュール」を選択し、組み合わせを提案。
- 例: 「〇〇の機能を実装したい」 → AIが関連コードを検索し、最適なモジュールを提示。

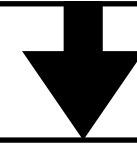
⚠️ 取り組むべき課題

- AIが推奨するコードの組み合わせが適切かどうかを評価できる能力を身につける必要がある。
- 開発者がAIに対してどのようなコードを提供するかによって、解析結果が変わるため、入力コードの整理が重要になる。
- レガシーコードのモジュール化を進め、AIが解析しやすい環境を整備する必要がある。

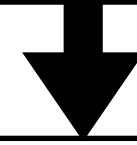
💡 ユーザーが開発目的を入力



🏗️ AIが関連する
レガシーコードを解析



🔗 AIが最適なコードの
組み合わせを提案

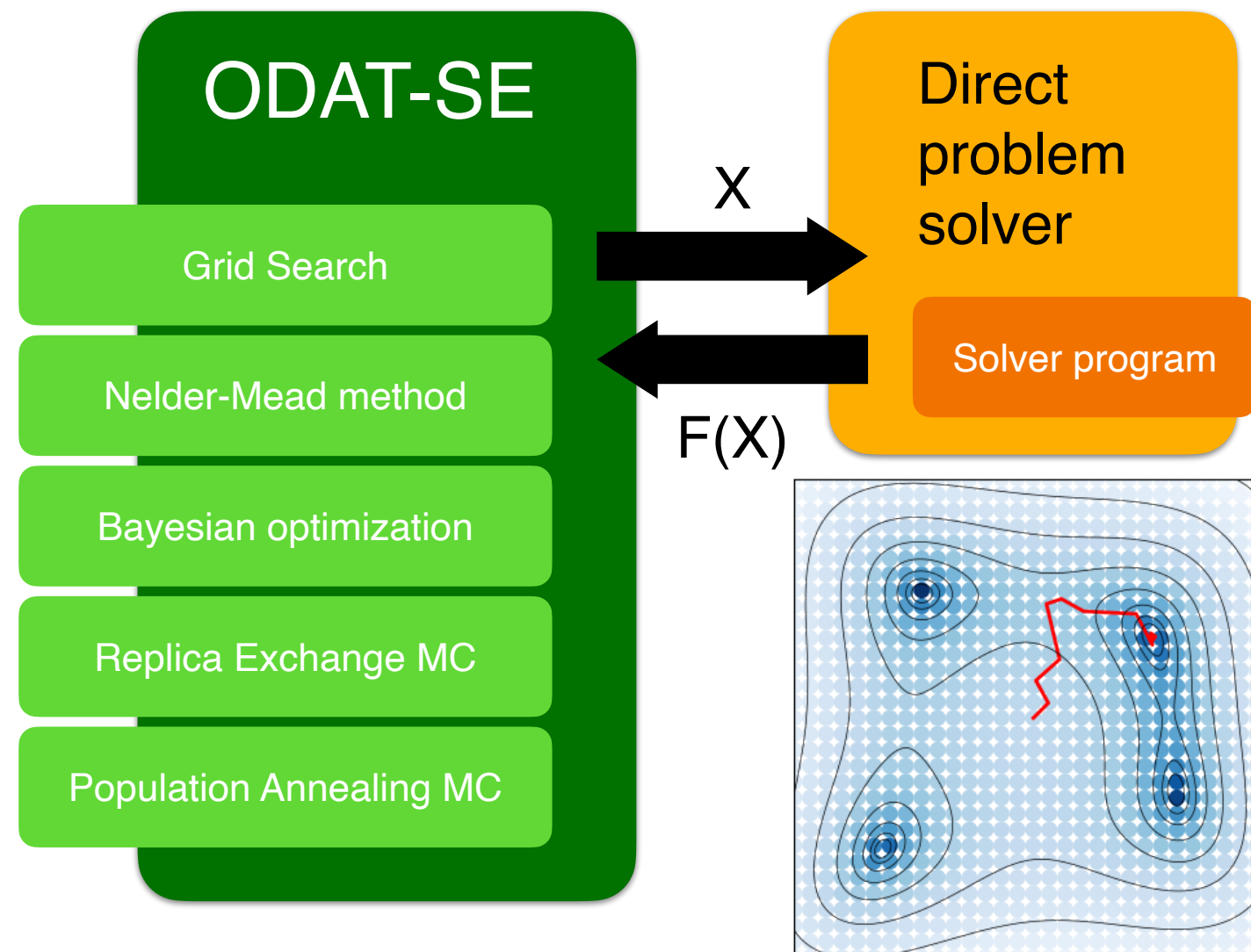


🚀 開発プロセスが加速

今後の展望：生成AIの活用 (2)

AIが複数のレガシーコードを統合し、最適なシステムを提案？

マニュアル・チュートリアルを整備・作成



超次元状態エンジニアリングによる未来予測型デジタルシステム

Project manager

星 健夫

自然科学研究機構
核融合科学研究所 教授



プロジェクト概要

本プロジェクトでは、核融合反応を発電や様々な場面で利用するフュージョンエネルギーシステムの設計や性能試験をデジタル空間上で行えるデジタルプラットフォームおよびバーチャルラボラトリ（仮想実験室：Vラボ）を構築します。

デジタル空間上でフュージョンエネルギーシステムにおけるプラズマ（電離したガス）の状態や構成機器の複雑性と時空間スケール（時間的・空間的な広がり）を再現するために、時間軸・空間（座標）軸・速度軸・物理量などが組み合わされた「超次元データ空間」（図 1）を新たに定義し、その特異な空間の特性を組み込んだ計算手法を確立します。さらに、その計算手法を「超次元状態エンジニアリング」として応用できるようにするために、これまでにない革新的 AI/データ駆動科学技術を開発します。

これにより、フュージョンエネルギーシステムの実験をデジタル空間上で実施できる V ラボを構築し、デジタル空間上での未来（次世代）のフュージョンエネルギーシステムの要素技術やシステム全体の性能予測（未来予測）を可能にします（図 2）。本プロジェクトによって、リアル空間では多大な時間とコストを要する試行錯誤（試作機を開発し性能試験を行うこと）のプロセスを大幅に低減させることで、多様なフュージョンエネルギーシステムの早期の社会実装やコスト削減を実現し、フュージョンエネルギーが供給される社会を目指します。

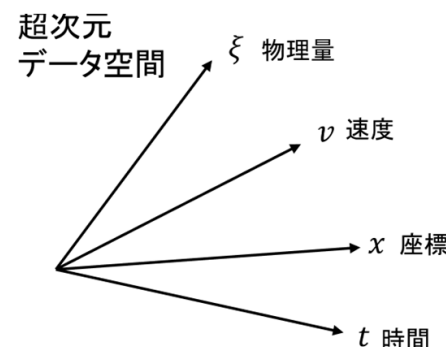


図 1. 「超次元データ空間」の模式図

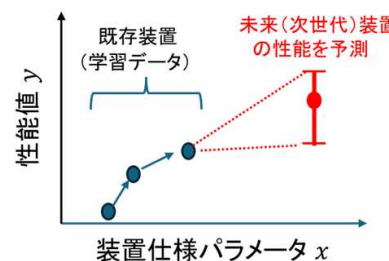


図 2. 未来予測技術の模式図

2034 年までのマイルストーン

デジタル空間上に構築した 3 種 V ラボ（①磁場閉じ込め方式、②慣性閉じ込め方式などの発電向けフュージョンエネルギーシステムおよび③中性子照射などの材料実験施設）において、リアル空間の既存装置の実験結果を使ってベンチマークを行い、3 つ全ての V ラボがリアル空間の既存装置の実験を再現することを実証します。さらに、V ラボがフュージョンエネルギーシステムの未来予測機能を持ち、その未来予測機能は汎用性があることを実証します。

2029 年までのマイルストーン

デジタル空間上に構築した 3 種 V ラボにおいて、リアル空間の既存装置の実験結果を使ってベンチマークを行い、3 つの V ラボのうち 1 つがリアル空間の既存装置の実験を再現することを実証します。

プロジェクト内の研究開発テーマ構成

私たちは、次の 4 つの研究開発項目を中心に進めています。

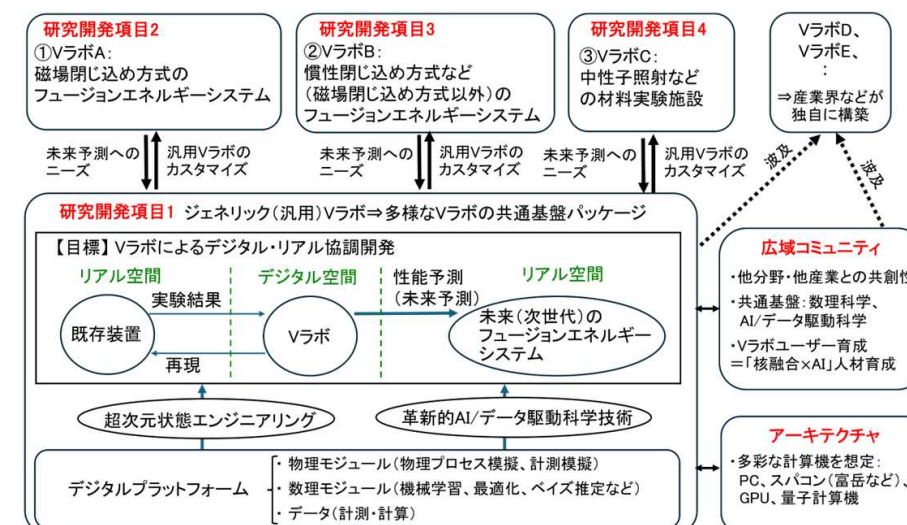


図 3. プロジェクトの全体構成図

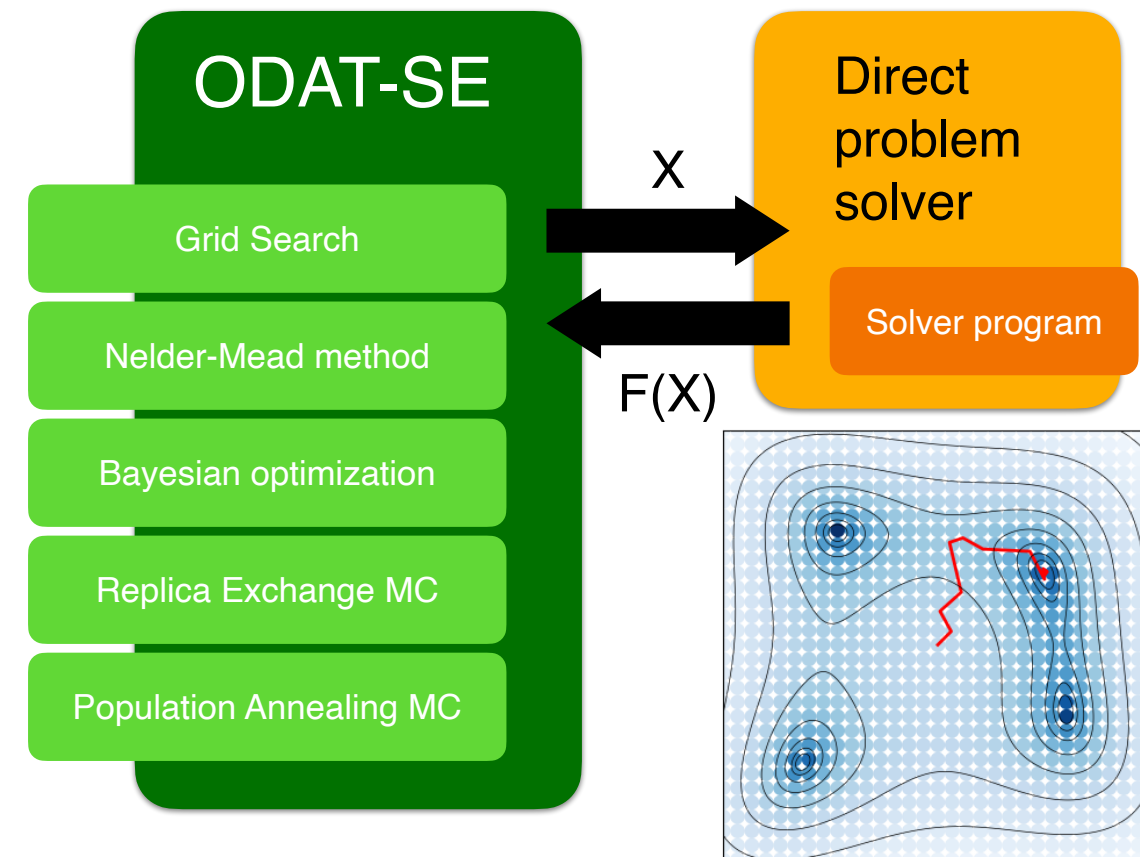
2025年4月より研究開発項目1-2 「デジタルプラットフォームの構築と有用性実証」で参画

PI 川島直輝、主な研究参加者 吉見一慶、青山龍美

まとめ

- ODAT-SE

- ▶ 順問題**Solver**: 目的関数 $f(x)$ を計算する
- ▶ 逆問題**Algorithm**: $f(x)$ をもとに x の空間を探索する
- ▶ 特徴
 - ▶ Solverを切り替えることで**問題（実験手法）が変わっても同じように解析可能**
 - ▶ Algorithmを切り替えることで**様々な数理手法で実験データを解析可能**
- ▶ 問い合わせ先
 - ▶ 2dmat-dev@issp.u-tokyo.ac.jp



- 明日5/9 10:00-12:00にODAT-SEに関して個別相談会を開催予定！
奮ってご参加ください。