

ソフトウェア自動チューニングにおける 複数パラメータを同時推定する手法の効率化

多部田敏樹† 田中輝雄† 藤井昭宏† 関直人† 范谷瑛†

工学院大学†

1. はじめに

ユーザが手作業でプログラムを計算機環境に
適応させてチューニングすることは困難である。
実行性能に関係する要因の増え方がわずかでも
組み合わせの増え方が急激で、探索範囲が急増
するためである[1].

実行性能に関係する要因をパラメータ化して
性能パラメータとする。複数性能パラメータを
同時推定する手法として、一次元スプライン系
近似関数である **d-Spline** を利用した性能パラメータ
空間での反復次元探索が提案されている。
探索範囲が急増しても、探索回数の増え方は小
さいことがメリットである[2].

反復次元探索の課題には、メモリ使用量と
初期点に関わる[2][3]。メモリ使用量が多いほど、
計算機のリソースが減り、ユーザプログラムに
与える影響が大きくなる。初期点は適切ならば、
よい推定結果を得られるため、結果の有用性を
向上させながら探索回数を削減できる。

本研究は反復次元探索の効率化を目的とし、
メモリ使用量をより削減したアルゴリズムと、
適切な初期点を初期点候補の集合から取得する
手法を提案する。

2. メモリ使用量削減

2-1. 一次元 d-Spline による推定

d-Spline は $f = (f_1, f_2, f_3, \dots, f_j, \dots, f_n)^t$ で表される
離散的な近似関数である。 N 個の実測データを
 $y = (y_1, y_2, y_3, \dots, y_i, \dots, y_N)^t$ として、 y に追従した
 f を生成することで、最適な標本点の見当がつく
(本研究では $n = 3N + 2$ としている)。

y と f の差分や近似関数の滑らかさを考慮して
評価関数 $\min(\|y - Ef\|^2 + a^2\|Df\|^2)$ で評価する
(E は y と f の対応行列、 D は二階差分を示す行列、
 a は追従度係数)。

先行研究では、最小二乗問題に帰結した後、
QR 分解と **Givens** 変換を用いて、メモリ使用量を
削減している[4]。数式上での E と D は行列だが、
特殊な構造を利用することで、アルゴリズムの
メモリ使用量は $O(N)$ に抑えられている。

2-2. 推定のメモリ使用量削減

表 1 メモリ使用量の変化量(Byte)

	改良前	改良後
f ベクトル	$24N + 16$	$8N$
y ベクトル	$32N + 16$	$24N + 16$
E 行列	$24N + 16$	16
D 行列	$72N + 48$	$72N + 48$
合計	$152N + 96$	$104N + 80$

離散点 N 個での、一次元 **d-Spline** 推定のメモリ
使用量を、表 1 に示す。 f ベクトルと y ベクトルは、
プログラム上では数式より小さい行列サイズで
記述することができた。 E 行列は、疎行列なので、
零要素が排除されたアルゴリズムに改良した。
メモリ使用量は 30% 程度削減された。

3. 初期点候補の集合

3-1. 改良 LHD における初期点候補の集合

不規則に分散した点の集合を得る手法として
LHD が存在する[5]。 **N-Queen** 問題を条件とする
改良 **LHD** を利用して、初期点候補の集合を得る
手法を提案する。初期点候補の分布を図 1 に示す。
初期点候補数と集合パターン数の関係を、表 2 に
示す。

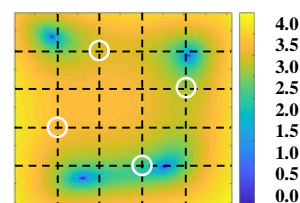


図 1 提案手法の初期点候補 (4 個の場合)
暗い点ほど最適解に近い

Efficiency Improvement of Method for Simultaneous
Estimation of Multiple Performance Parameters in Software
Automatic Tuning

Toshiki Tabet†, Teruo Tanaka†, Akihiro Fujii†, Naoto Seki†,
Guqing Fan†

†Kogakuin University

表2 初期点候補数と集合パターン数の関係

候補数	パターン数	候補数	パターン数
4	2	9	352
5	10	10	724
6	4	11	2680
7	40	12	14200
8	92	13	73712

3-2. 提案手法の評価実験

提案手法を用いた初期点候補の集合について評価実験を行う。実験方法を以下に記す。

1. 初期点候補の集合を生成
2. 集合から最も適切な離散点を初期点に決定
3. 反復次元探索によって最適解を推定

1~3を二次元データ5種類[6]に対し各6000回行った。ランダムな一点から始めた反復次元探索の標本点数に対する、集合の最適な点から始めた反復次元探索の標本点数の倍率（5種類の平均）を図2に示す。初期点候補がランダムな点の最適解導出回数に対する、提案手法の最適解導出回数の倍率（5種類の平均）を図3に示す。

図2より、ランダムな点の集合と提案手法は反復次元探索の標本点数を同程度削減する。図3より、提案手法はランダムな点の集合より最適解の導出回数が全体的に多い。

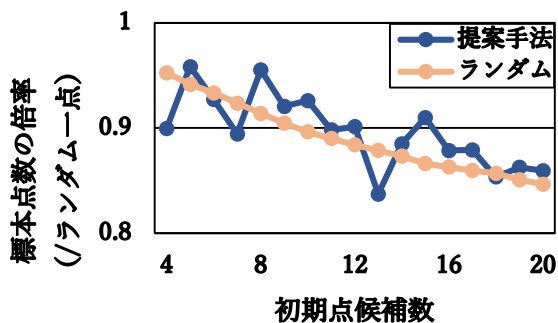


図2 初期点候補数と標本点数の関係

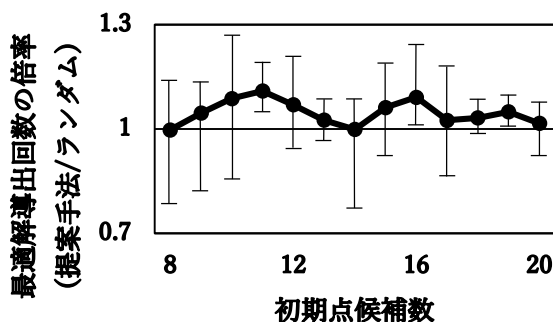


図3 初期点候補数と最適解導出回数との関係

提案手法は最適解導出回数を増やししながら、反復次元探索の標本点を削減できる。ただし候補数が多ければ最適解導出倍率は収束する。総合的な標本点数を抑えるため、初期点候補は適度な個数にする必要がある。

4. おわりに

本研究では性能パラメータ空間の反復次元探索に2つの項目で効率化を行った。第1に、メモリ使用量を30%程度削減したことで、ユーザプログラムへの影響を軽減した。第2に、最適解を導出する確率を高めつつ標本点を減らしたことで、結果の有用性を向上させながら探索回数を削減した。

今後の課題としては、高次元空間に拡張した初期点候補の集合が挙げられる。本研究の提案手法は、表2と図3より、集合パターンが少ない場合、導出回数が安定しない。高次元空間ほど集合パターンが減少するため不安定になりやすい。そのため初期点候補の集合については、他の手法も考慮に入れる必要がある。

謝辞

本研究の一部はJSPS 科研費JP17K00164, JP16H02823, JP18K19782, JP18K11340の助成を受けて行われた。

参考文献

- [1] 片桐孝洋, ソフトウェア自動チューニング-数値計算ソフトウェアへの適用とその可能性, 慧文社(2004)
- [2] 望月大義, 藤井昭宏, 田中輝雄, ソフトウェア自動チューニングにおける複数同時性能パラメータ探索手法の提案と評価, 情報処理学会論文誌, Vol.11, No.2, pp.1-16, 2018
- [3] 関直人, 望月大義, 范谷瑛, 藤井昭宏, 田中輝雄, 片桐孝洋, ソフトウェア自動チューニングにおける反復次元探索による複数性能パラメータ推定法の実装と評価, 第80回全国大会講演論文集, Vol.2018, No.1, pp.33-34, 2018
- [4] Teruo Tanaka, Ryo Otsuka, Akihiro Fujii, Takahiro Katagiri, Toshiyuki Imamura, "Implementation of d-Spline-based incremental performance parameter estimation method with ppOpen-AT", Scientific Programming, Vol.22, No.4, pp.299-307, 2014
- [5] 三輪恵律也, 望月大義, 藤井昭宏, 田中輝雄, Latin Hypercube Design の3次元空間への拡張, 第80回全国大会講演論文集, Vol.2018, No.1, pp.31-32, 2018
- [6] Virtual Library of Simulation Experiments: [https://www.sfu.ca/~ssurjano/\(2018/12/25\)](https://www.sfu.ca/~ssurjano/(2018/12/25))