

最適化のための効率的なサンプリングに関する考察 Study on efficient design of sampling for optimization method

大阪大学 大学院情報科学研究科 土屋 耕治 (Koji Tsuchiya)
森田 浩 (Hiroshi Morita)

Graduate School of Information Science and Technology,
Osaka University

1 はじめに

最適化問題において、実用的には時間やコストによって採取するデータ数に制限がある場合が多いため、少ないデータ数で最適解を求めることが重要である。特に非線形問題に関しては計画的にサンプリングすることや適切な標本点を求めることは、最適解の探索を少ないデータ数で行うことと関係がある。Efficient Global Optimization (EGO) は Jones ら [3] が提案した大域的最適解を探索する手法である。EGO では応答曲面の予測を行い、予測から各点が最適解を取り得る期待値を推定する。新たに標本を採取する点を選択する指標として Expected Improvement (EI) を導入し、予測した値そのものを扱うだけでなく、予測の不確実性の大きさも考慮に入れている。EI に基づいて新たなサンプリングを行い、再び予測を行う。この繰り返しによって少ないデータ数で最適解を求めることが可能となる。

EGO の特徴の一つとして予測を用いていることが挙げられる。予測が不確実であると、最適解を得るために多くのデータ数が必要となるため、精度の高い予測を行わなければならない。一般に、予測値の精度は得られたデータの精度とそのデータ数に依存している。しかし、時間やコストの面において、この両者は相反する関係にある。精度の低い器具や装置から得られたデータは誤差が大きく、解析結果の信頼性は低下するが、データを一つ採取する際に費やすコストは低いため、多くのデータを得ることができる。その一方で、精度の高いデータは真値に近い値であると考えられるため信頼でき、さらにその数が多くあれば、解析結果はより妥当なものとなる。しかし、精度の高いデータを得るには、精度の高い装置や整った環境といった多くの時間やコスト等を必要とするため、これらに制限がある場合、多くのデータを得ることは困難である。Kennedy ら [5] は有限要素法から得られた精度の異なるデータを組み合わせた予測モデルの提案を行い、効率的なデータ採取と予測精度の向上を図った。このような異なる性質をもつ複数のデータを組み合わせた予測手法は多く行われている [1, 6]。

本研究では Kennedy らの予測モデルの拡張を行うとともに、EGO を用いた効率的なサンプリングによる最適化を行う。データに誤差が含まれる場合、実用面において厳密解を求めることは時間やコストの観点から困難な場合が多いが、十分に精度の高い近似解を得ることができれば問題ない。精度の異なるデータを組み合わせることによって、精度の低いデータのみによる最適化よりも高い精度の近似解を得るだけでなく、精度の高いデータのみによる非効率的なサンプリングによる最適化を改善することで、従来よりも少ない時間やコストによるサンプリングの可能性を探る。

以下 2 章で Kriging モデルによる予測と EGO の手法を述べる。3 章で精度の異なるデータを組み合わせるモデルを提案し、4 章で予測値とその不確実性を求め、EGO への適用を説明する。5 章で提案手法の数値実験とその結果を示し、最後に 6 章で本研究の考察を行う。

2 EGOによる最適化

2.1 Kriging モデル

本研究では応答曲面の予測に Kriging モデルを用いる。Kriging モデルは対象が非線形の場合でも大域的に高い予測精度を得ることができるとともに、標本点の相関関係を考慮したモデルであることから、計算機実験における実験計画に用いられる手法である [2, 7, 8]。

Kriging モデルでは、点 \mathbf{x} の予測変数 $Y(\mathbf{x})$ は回帰モデルとガウス過程から構成されると仮定し、

$$Y(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} + \eta(\mathbf{x}) + \epsilon \quad (1)$$

とおく。ここで、 ϵ は測定誤差である。 $\mathbf{h}(\cdot) = (h_1(\cdot), \dots, h_p(\cdot))^T$ は既知の関数からなるベクトル、 $\boldsymbol{\beta}_t = (\beta_1, \dots, \beta_p)^T$ は未知の回帰係数ベクトルを表している。 p は回帰を行う次数である。単回帰の場合、 $\mathbf{h}(\mathbf{x}) = (1, \mathbf{x})^T$ 、 $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ とおき、 $\mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} = \beta_0 + \beta_1 \mathbf{x}$ となる。また、 $\eta(\mathbf{x})$ は真のモデルと回帰モデルの差を表し、平均が 0 で、 $\eta(\mathbf{x})$ と $\eta(\mathbf{x}')$ 間の共分散関数が

$$\text{Cov}\{\eta(\mathbf{x}), \eta(\mathbf{x}')\} = \sigma_\eta^2 c(\mathbf{x}, \mathbf{x}') \quad (2)$$

で与えられる定常ガウス過程である。ここで、 σ_η^2 は分散、 $c(\mathbf{x}, \mathbf{x}')$ は相関関数を表している。この相関関数は 2 点 $\eta(\mathbf{x})$ 、 $\eta(\mathbf{x}')$ の距離のみに依存する関数で、様々な関数が提案されている [8]。

点 \mathbf{x}_0 における予測変数 $Y(\mathbf{x}_0)$ を考える。標本点を $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$ 、その標本の組を $\mathbf{Y} = (Y(\mathbf{x}^{(1)}), \dots, Y(\mathbf{x}^{(n)}))^T$ とする。また、各標本点における $\mathbf{h}(\cdot)$ からなる行列を

$$H = \begin{pmatrix} \mathbf{h}(\mathbf{x}^{(1)})^T \\ \vdots \\ \mathbf{h}(\mathbf{x}^{(n)})^T \end{pmatrix} \quad (3)$$

とおく。 $Y(\mathbf{x}_0)$ と各標本 $Y(\mathbf{x}^{(1)}), \dots, Y(\mathbf{x}^{(n)})$ の共分散からなる $n \times 1$ ベクトルを $\mathbf{r}(\mathbf{x}_0)$ 、各標本間の共分散からなる n 次正方行列を R とすると、 $\mathbf{r}(\mathbf{x}_0)$ の第 k 成分と R の第 (i, j) 成分はそれぞれ

$$[\mathbf{r}(\mathbf{x}_0)]_k = \sigma_\eta^2 c(\mathbf{x}_0, \mathbf{x}^{(k)}) \quad (4)$$

$$[R]_{i,j} = \sigma_\eta^2 c(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_\epsilon \delta_{i,j} \quad (5)$$

と表せる。ただし、 $\delta_{i,j}$ は $i=j$ のとき 1、 $i \neq j$ のとき 0 である。また、 σ_ϵ は測定誤差 ϵ の分散を表している。このとき、最良線形不偏予測量 (Best Linear Unbiased Predictor, BLUP) は

$$\hat{Y}(\mathbf{x}_0) = \mathbf{h}(\mathbf{x}_0)^T \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x}_0)^T R^{-1} (\mathbf{Y} - H \hat{\boldsymbol{\beta}}) \quad (6)$$

で与えられる。 $\hat{\boldsymbol{\beta}}$ は一般化最小二乗推定量で、 $\hat{\boldsymbol{\beta}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \mathbf{Y}$ となる。標本 \mathbf{Y} が与えられたときの予測変数を $Y^\dagger(\mathbf{x}_0) = [Y(\mathbf{x}_0) | \mathbf{Y}]$ とおくと、期待値 $E\{Y^\dagger(\mathbf{x}_0)\}$ は式 (6) に等しい。

また、予測量の平均二乗誤差 (Mean Squared Error, MSE) は、

$$s^2(\mathbf{x}_0) = \sigma_\eta^2 - \mathbf{r}(\mathbf{x}_0)^T R^{-1} \mathbf{r}(\mathbf{x}_0) + (\mathbf{h}(\mathbf{x}_0) - H^T R^{-1} \mathbf{r}(\mathbf{x}_0))^T (H^T R^{-1} H)^{-1} (\mathbf{h}(\mathbf{x}_0) - H^T R^{-1} \mathbf{r}(\mathbf{x}_0)) \quad (7)$$

となる。

2.2 Expected Improvement

大域的最小値の探索を考える。予測値が最小となる点付近をサンプリングすれば、より良い解が得られる可能性はあるが、大域的な最小値を得るためには予測した値そのものを考えるだけでは十分でない。Jones ら [3] は Kriging モデルによる予測をもとに、局所探索と大域探索のバランスを考慮にした新たな標本点を選択する指標として EI を提案した。EI は現在得られている標本の最小値 $Y_{min} = \min\{Y(\mathbf{x}^{(1)}), \dots, Y(\mathbf{x}^{(n)})\}$ に対して、サンプリングを行った場合に改善できる度合い、すなわち標本点 新しい最小値となり得る期待値を表している。この期待値は、予測の不確かさの大きさを考慮しているため、予測値が大きな点の付近も EI が大きくなる場合があり、局所探索のみにならない。

標本 Y が与えられたときの予測変数 $Y^\dagger(\mathbf{x})$ が正規分布 $N(\hat{Y}(\mathbf{x}), s^2(\mathbf{x}))$ に従うとする。 $\hat{Y}(\mathbf{x})$, $s^2(\mathbf{x})$ はそれぞれ BLUP, MSE である。点 \mathbf{x} におけるデータが、最小値より小さい値を取る確率は

$$I(\mathbf{x}) = \max\{Y_{min} - Y^\dagger(\mathbf{x}), 0\} \quad (8)$$

と表される。 $I(\mathbf{x})$ の期待値が点 \mathbf{x} での最小値を取り得る期待値となる。 $EI(\mathbf{x}) = E\{I(\mathbf{x})\}$ とおくと、

$$EI(\mathbf{x}) = (Y_{min} - \hat{Y}(\mathbf{x}))\Phi\left(\frac{Y_{min} - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right) + s(\mathbf{x})\phi\left(\frac{Y_{min} - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right) \quad (9)$$

となる。ここで、 $\Phi(\cdot)$ と $\phi(\cdot)$ はそれぞれ標準正規累積分布関数、標準正規確率密度関数である。期待値 EI が最大となる点における標本が最小値を取る可能性が高いので、新たな標本点は

$$\mathbf{x}^{(n+1)} = \arg \max\{EI(\mathbf{x})\} \quad (10)$$

で与える。

3 精度の異なる標本の組合せ

3.1 Kennedy の予測モデル

Kennedy ら [5] は有限要素法から得られた異なる精度のデータを組み合わせた。有限要素法は構造解析等に用いられ、対象を要素に分割して計算し、それらを足し合わせて対象全体の性質を求める手法である。この計算結果を一つのデータとすると、これは同じ入力の組に対して一意の出力となる確定的なデータである。有限要素法のような計算機実験から得られたデータの誤差は正規分布に従うような確率的な誤差ではなく、計算モデルから生じる確定的な誤差である。要素の区切りを細かくとれば、より真値に近い値を得るが、計算時間は長くなる。一方で、要素を大きくすると計算時間は短くなるが、データの精度は低くなる。データの精度と計算時間はトレードオフの関係にあるが、精度の異なるデータを組み合わせることで、予測精度の向上と効率的なデータ採取を行っている。

Kennedy らは精度の異なるデータ間に自己回帰モデルを構築した。分割要素の大きさが異なる s 個の有限要素モデルに対して、要素が大きい順にその出力を $q_1(\cdot), \dots, q_s(\cdot)$ とする。 $q_s(\cdot)$ が

最も精度の高いモデルからの出力である。また、 $t(t=1, \dots, s)$ 番目に分割要素が大きい有限要素モデル、すなわち t 番目に精度が低いモデルに対する入力を $\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(n_t)}$ とし、出力の組を $\mathbf{q}_t = (q_t(\mathbf{x}_t^{(1)}), \dots, q_t(\mathbf{x}_t^{(n_t)}))^T$ とする。これらの出力の組の間に自己回帰モデルを仮定すると

$$q_1(\mathbf{x}) = \zeta_1(\mathbf{x}) \quad (11)$$

$$q_t(\mathbf{x}) = \rho_{t-1} q_{t-1}(\mathbf{x}) + \zeta_t(\mathbf{x}) \quad (t=2, \dots, s) \quad (12)$$

となる。ここで、 ρ_{t-1} は自己回帰係数である。また、 $\zeta_t(\mathbf{x})$ は $q_{t-1}(\cdot), \dots, q_1(\cdot)$ に独立な Kriging モデルである。ただし、計算機実験から得られたデータを用いるので、測定誤差、すなわち、式 (1) に含まれる ϵ は考慮しない。

以上の仮定から、全標本 $\mathbf{q}_1, \dots, \mathbf{q}_s$ が与えられたときの最も精度の高い有限要素モデルに対する出力 $q_s(\cdot)$ の予測を行う。全標本数は $n_{\text{sum}} = \sum_{k=1}^s n_k$ となり、 $\zeta_t(\cdot)$ に含まれるガウス過程から、 $[q_s(\cdot), \mathbf{q}_s, \dots, \mathbf{q}_1]$ は $(1 + n_{\text{sum}})$ 次多変量正規分布となる。予測値は全標本を与えたときの条件付き分布 $[q_s(\cdot) | \mathbf{q}_s, \dots, \mathbf{q}_1]$ の期待値で与えられる。

このように Kennedy らは異なる有限要素モデルの出力間に式 (11), (12) の自己回帰モデルおよび Kriging モデルを仮定することで、データを得る計算時間と予測精度のトレードオフの解消を図った。

3.2 予測モデルの拡張

計算機実験とは異なり、測定や観測から得られた標本には不確実な要素が含まれているため、標本にはばらつきが存在する。本研究ではこのばらつきの大きさを標本の精度として取り扱う。真値を y 、標本を Y とすると、 $Y = y + \epsilon$ と表される。ここで、 ϵ は誤差を表し、正規分布 $N(0, \sigma^2)$ に従うとすると、標本のばらつきを表す尺度として誤差分散 σ^2 を用いることができる。精度を誤差分散の大きさで表し、誤差分散の小さい標本を精度が高い、誤差分散の大きい標本を精度が低いとする。

標本の精度にしたがって、標本のレベル化を行う。 $t(t=1, \dots, s)$ 番目に精度の低い標本の組がレベル t に属するとし、精度の高い標本がより高レベルに位置するように標本のレベル化を行う。レベル t に属する標本 $Y_t(\cdot)$ は $Y_t(\cdot) = y_0(\cdot) + \epsilon_t$ 、 $\epsilon_t \sim N(0, \sigma_t^2)$ と表される。ここで、 $y_0(\cdot)$ は真値、 ϵ_t は誤差、 σ_t^2 は誤差分散を表す。レベル化は誤差分散が

$$\sigma_1^2 > \sigma_2^2 > \dots > \sigma_s^2 \quad (t=1, \dots, s) \quad (13)$$

を満たすように行う。最も精度の低い標本の組はレベル 1、最も精度の高い標本の組はレベル s となる。

さらに本研究では、標本の誤差分散 σ_t^2 の値は未知であるが、異なる標本の組の誤差分散の大小関係を既知として、レベル化を行う。サンプリングの際に精度の高いデータを得るには多くの手間を要する場合がある。このとき、ある組に対しては多くの手間をかけるが、もう一方の組に対して手間を省くと、誤差分散の異なる標本が 2 組できる。この場合、データの精度の大小関係は分かるが、各組の誤差分散は未知である。また、高価な測定器を用いる回数が制限され、安価な測定器で得られたデータと組み合わせることを考えた場合も同様の状況である。本研究では、このような式 (13) が成り立つ条件下で予測値を求める。このレベル化により、精度の関係を 1 つの系列で見ることが可能になる。

レベル間の関係には Kennedy らと同様に自己回帰モデルを採用する。レベル t において、点 \mathbf{x} における目的変数を $Y_t(\mathbf{x})$ とすると、

$$Y_1(\mathbf{x}) = Z_1(\mathbf{x}) \quad (14)$$

$$Y_t(\mathbf{x}) = \rho_{t-1}Y_{t-1}(\mathbf{x}) + Z_t(\mathbf{x}) \quad (t = 2, \dots, s) \quad (15)$$

となる。 ρ_{t-1} は自己回帰係数、 $Z_t(\cdot)$ は Kriging モデルを仮定し、

$$Z_t(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta}_t + \eta_t(\mathbf{x}) + \epsilon_t \quad (16)$$

とおく。 $\boldsymbol{\beta}_t$, $\eta_t(\cdot)$, ϵ_t はそれぞれレベル t における回帰係数ベクトル、ガウス過程、測定誤差を表している。また、 $\eta_t(\cdot)$ は平均 0 で、共分散関数は

$$\text{Cov}\{\eta_t(\mathbf{x}), \eta_t(\mathbf{x}')\} = \sigma_{\eta,t}^2 c_t(\mathbf{x}, \mathbf{x}') \quad (17)$$

$$c_t(\mathbf{x}, \mathbf{x}') = \exp\{-\gamma_t(\mathbf{x} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}')\} \quad (18)$$

と表される。 $\sigma_{\eta,t}^2$, $c_t(\mathbf{x}, \mathbf{x}')$ はそれぞれレベル t における分散、相関関数である。また、相関関数は一般的に用いられている式 (18) で表される指数相関関数を仮定する [2, 5, 6, 7]。 $\boldsymbol{\beta}_t$, $\sigma_{\eta,t}^2$, ϵ_t , γ_t , ρ_{t-1} および測定誤差の分散 $\sigma_{\epsilon,t}^2$ がレベル t に属するパラメータである。

Kriging モデルを用いることにより、各レベルにおいて予測値を求めることができるだけでなく、予測の不確実さを推定することができる。よって、各レベルにおいて EI を利用することが可能になる。

4 予測分布

4.1 BLUP

拡張したモデルに対する予測分布を考え、レベル化した標本に対して、標本を加えるごとに予測値の更新を行えるアルゴリズムを導く。レベル t における n_t 個のサンプリング点 $\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(n_t)}$ とし、その標本の組を $\mathbf{Y}_t^{n_t} = (Y_t(\mathbf{x}_t^{(1)}), \dots, Y_t(\mathbf{x}_t^{(n_t)}))^T$ とする。

レベル 1 における予測は Kriging モデルによる予測と等しい。標本 \mathbf{Y}_1 が与えられたときの予測変数 $Y_1^\dagger(\mathbf{x})$ の期待値、すなわち $Y_1(\mathbf{x})$ の BLUP を予測値とする。これは式 (6) で与えられる。 $Y_1^\dagger(\mathbf{x})$, $Y_1^\dagger(\mathbf{x}')$ 間の共分散関数は式 (17), (18) と同様に 2 点 \mathbf{x} , \mathbf{x}' に依存する関数となり、 $C_1(\mathbf{x}, \mathbf{x}') = \text{Cov}\{Y_1^\dagger(\mathbf{x}), Y_1^\dagger(\mathbf{x}')\}$ とおくと

$$C_1(\mathbf{x}, \mathbf{x}') = \sigma_{\eta,1}^2 c_1(\mathbf{x}, \mathbf{x}') - \mathbf{r}_1(\mathbf{x})^T R_1^{-1} \mathbf{r}_1(\mathbf{x}') \\ + (\mathbf{h}(\mathbf{x}) - H_1^T R_1^{-1} \mathbf{r}_1(\mathbf{x}))^T (H_1^T R_1^{-1} H_1)^{-1} (\mathbf{h}(\mathbf{x}') - H_1^T R_1^{-1} \mathbf{r}_1(\mathbf{x}')) \quad (19)$$

と表せる。ここで、 H_1 , \mathbf{r}_1 , R_1 はそれぞれ式 (3), (4), (5) に等しい。また、 $\mathbf{x} = \mathbf{x}'$ のとき、式 (19) はレベル 1 における MSE となり、 $s_1^2(\mathbf{x}) = C_1(\mathbf{x}, \mathbf{x})$ である。

次に、レベル $t (\geq 2)$ における予測を考える。レベル t における標本点は $\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(n_t)}$ であり、これらの標本点における下のレベル $t-1$ の予測は $\mathbf{Y}_{t-1}^\dagger = (Y_{t-1}^\dagger(\mathbf{x}_t^{(1)}), \dots, Y_{t-1}^\dagger(\mathbf{x}_t^{(n_t)}))^T$ と表すことができる。新たにサンプリングした点において、下のレベルの予測 \mathbf{Y}_{t-1}^\dagger と新たな

標本 $Y_t^{n_t}$ を比較するとともに、これらを用いて予測値の更新を行う。式(15)より、予測変数は $Y_t(\mathbf{x}_0) = \rho_{t-1}Y_{t-1}(\mathbf{x}_0) + \mathbf{h}(\mathbf{x}_0)\beta_t + \epsilon$ と表すことができる。 $Y_t(\mathbf{x}_0)$ の BLUP を求める。

$$\hat{Y}_t(\mathbf{x}_0) = \mathbf{a}(\mathbf{x}_0)^T \mathbf{Y}_t^{n_t} \quad (20)$$

と仮定し、不偏性の条件下で $\text{MSE}\{\hat{Y}_t(\mathbf{x}_0)\}$ を最小にする $\mathbf{a}(\mathbf{x}_0)$ を求める。

$$\begin{cases} \min & E\{(\mathbf{a}(\mathbf{x}_0)^T \mathbf{Y}_t^{n_t} - Y_t(\mathbf{x}_0))^2\} \\ \text{s.t.} & E\{\mathbf{a}(\mathbf{x}_0)^T \mathbf{Y}_t^{n_t} - Y_t(\mathbf{x}_0)\} = 0 \end{cases} \quad (21)$$

$Y_t(\mathbf{x}_0)$ と各標本 $Y_t(\mathbf{x}_t^{(1)}), \dots, Y_t(\mathbf{x}_t^{(n_t)})$ からなる共分散ベクトルを $\mathbf{r}_t(\mathbf{x}_0)$ 、各標本間の共分散からなる行列を R_t とすると、 $\mathbf{r}_t(\mathbf{x}_0)$ の第 k 成分および R_t の (i, j) 成分はそれぞれ

$$[\mathbf{r}_t(\mathbf{x}_0)]_k = \rho_{t-1}^2 C_{t-1}(\mathbf{x}_0, \mathbf{x}_t^{(k)}) + \sigma_t^2 c_t(\mathbf{x}_0, \mathbf{x}_t^{(k)}) \quad (22)$$

$$[R_t]_{i,j} = \rho_{t-1}^2 C_{t-1}(\mathbf{x}_t^{(i)}, \mathbf{x}_t^{(j)}) + \sigma_t^2 c_t(\mathbf{x}_t^{(i)}, \mathbf{x}_t^{(j)}) + \sigma_\epsilon \delta_{i,j} \quad (23)$$

となる。また、レベル t の標本点 $(\mathbf{x}^{(1)}), \dots, (\mathbf{x}^{(n_t)})$ におけるレベル $t-1$ の BLUP の組を $\hat{\mathbf{Y}}_{t-1}^{n_t} = (\hat{Y}_{t-1}(\mathbf{x}^{(1)}), \dots, \hat{Y}_{t-1}(\mathbf{x}^{(n_t)}))^T$ として、 $\mathbf{g}_t(\mathbf{x}_0) = (\hat{\mathbf{Y}}_{t-1}(\mathbf{x}_0), \mathbf{h}(\mathbf{x}_0)^T)^T$ 、 $F_t = (\hat{\mathbf{Y}}_{t-1}^{n_t}, H_t)$ とおくと、式(21)は

$$\begin{cases} \min & \mathbf{a}(\mathbf{x}_0)^T R_t \mathbf{a}(\mathbf{x}_0) - 2\mathbf{a}(\mathbf{x}_0)^T \mathbf{r}(\mathbf{x}_0) \\ \text{s.t.} & F_t^T \mathbf{a}(\mathbf{x}_0) - \mathbf{g}_t(\mathbf{x}_0) = 0 \end{cases} \quad (24)$$

と表すことができる。これより、 $\mathbf{a}(\mathbf{x}_0)$ が求まり、レベル t における BLUP は

$$\hat{Y}_t(\mathbf{x}_0) = \mathbf{g}_t(\mathbf{x}_0)^T \mathbf{b} + \mathbf{r}_t(\mathbf{x}_0)^T R_t^{-1} (\mathbf{Y}_t^{n_t} - F_t \mathbf{b}) \quad (25)$$

が得られる。ここで、 $\mathbf{b} = (F_t^T R_t^{-1} F_t)^{-1} F_t^T R_t^{-1} \mathbf{Y}_t^{n_t}$ である。

また、共分散関数を $C_t(\mathbf{x}, \mathbf{x}') = \text{Cov}\{Y_t^\dagger(\mathbf{x}), Y_t^\dagger(\mathbf{x}')\}$ とおくと、

$$\begin{aligned} C_t(\mathbf{x}_0, \mathbf{x}') &= \rho_{t-1}^2 C_{t-1}(\mathbf{x}, \mathbf{x}') + \sigma_t^2 c_t(\mathbf{x}, \mathbf{x}') - \mathbf{r}_t(\mathbf{x})^T R_t^{-1} \mathbf{r}_t(\mathbf{x}') \\ &\quad + (\mathbf{g}_t(\mathbf{x}) - F_t^T R_t^{-1} \mathbf{r}_t(\mathbf{x}))^T (F_t^T R_t^{-1} F_t)^{-1} (\mathbf{g}_t(\mathbf{x}') - F_t^T R_t^{-1} \mathbf{r}_t(\mathbf{x}')) \end{aligned} \quad (26)$$

となる。 $\mathbf{x} = \mathbf{x}'$ のとき、式(26)はレベル t における MSE で、 $s_t^2(\mathbf{x}) = C_t(\mathbf{x}, \mathbf{x})$ となる。

4.2 パラメータの推定

パラメータの推定に関して、Kennedy ら [5] と同様の仮定を導入する。レベル t におけるパラメータを推定する際に2つの仮定を与える。1つ目は、レベル $1, \dots, t-1$ のパラメータはすべて固定して取り扱う。レベル t のパラメータはより低いレベルのパラメータから受ける寄与は小さいと考える。2つ目は、レベル $t+1$ 以上に属する標本の情報は使用しないことである。この2つの仮定により、各レベルごとにパラメータを分離するとともに、下のレベルのパラメータから順に各レベルごとに推定を行っていく。また、パラメータの推定には最尤推定法を用いる。

レベル1におけるパラメータは $\beta_1, \sigma_{\eta,1}^2, \sigma_{\epsilon,1}^2, \gamma_1$ である. $\Sigma_1^2 = \sigma_{\eta,1}^2 + \sigma_{\epsilon,1}^2, \tau_1 = \frac{\sigma_{\eta,1}^2}{\Sigma_1^2}, R_1^* = \frac{1}{\Sigma_1^2} R_1$ とおくと, R_1^* は対角成分が1, その他が $\tau_1 c_1(x^{(i)}, x^{(j)})$ の行列となる. レベル1における対数尤度関数は,

$$l_1(\beta_1, \Sigma_1^2, \tau_1, \gamma_1) \propto -n_1 \log \Sigma_1^2 - \log |R_1^*| + \frac{1}{\Sigma_1^2} (Y_1^{n_1} - H_1 \beta_1)^T R_1^{*-1} (Y_1^{n_1} - H_1 \beta_1) \quad (27)$$

となり, β_1 と τ_1 の最尤推定量はそれぞれ $\hat{\beta}_1 = (H_1^T R_1^{-1} H_1)^{-1} H_1^T R_1^{-1} Y_1^{n_1}, \hat{\Sigma}_1^2 = \frac{1}{n_1} (Y_1^{n_1} - H_1 \beta_1)^T R_1^{*-1} (Y_1^{n_1} - H_1 \beta_1)$ となる. 式(27)を最大にすることで, τ_1, γ_1 の推定量を得る. すなわち, $\{\hat{\tau}_1, \hat{\gamma}_1\} = \arg \min_{0 < \tau_1 < 1, \gamma_1 > 0} \{n_1 \log \hat{\tau}_1 + \log |R_1^*|\}$ を求めればよい.

レベル t におけるパラメータは, $\beta_t, \sigma_{\eta,t}^2, \sigma_{\epsilon,t}^2, \gamma_t, \rho_{t-1}$ である. $T_t = Y_t^{n_t} - \rho_{t-1} M_{t-1}$ とおくと, 対数尤度関数は

$$l_t(\beta_t, \sigma_{\eta,t}^2, \sigma_{\epsilon,t}^2, \gamma_t, \rho_{t-1}) \propto -\log |R_t| + (T_t - H_t \beta_t)^T R_t^{-1} (T_t - H_t \beta_t) \quad (28)$$

となる. β_t の最尤推定量は $\hat{\beta}_t = (H_t^T R_t^{-1} H_t)^{-1} H_t^T R_t^{-1} T_t$ となる. 式(28)を最大にすることで $\sigma_{\eta,t}^2, \sigma_{\epsilon,t}^2, \gamma_t, \rho_{t-1}$ の推定量を得る. すなわち, $\{\hat{\sigma}_{\eta,t}^2, \hat{\sigma}_{\epsilon,t}^2, \hat{\gamma}_t, \hat{\rho}_{t-1}\} = \arg \min \{\log |R_t| - (T_t - H_t \hat{\beta}_t)^T R_t^{-1} (T_t - H_t \hat{\beta}_t)\} (\sigma_{\eta,t}^2, \sigma_{\epsilon,t}^2, \gamma_t > 0, 0 \leq \rho_{t-1} \leq 1)$ を求めればよい.

4.3 EGO への適用

大域的最小値の探索を考える. 各レベルにおいて, 初期データ数 $n_t^{(ini)}$ および追加するデータ数 $n_t^{(add)}$ を定めておく. ただし, 初期データ数はレベルにおけるパラメータ数よりも多いものとする. 標本点とその点で得られたデータの値からなる集合 $D_t = \{(x_t^{(i)}, Y_t(x_t^{(i)})) | i = 1, \dots, n_t\}$ とする.

レベル1におけるEGOを示す.

- (1.1) $j = 0$ とする. 初期データの標本点 $x_1^{(1)}, \dots, x_1^{(ini)}$ を決め, $D_1^{(0)} = \{(x_1^{(i)}, y_1(x_1^{(i)})) | i = 1, \dots, n_1^{(ini)}\}$ を得る.
- (1.2) $D_1^{(j)}$ からすべての点における $\hat{Y}_1(x), s_1^2(x)$ を求め, $EI(x)$ を得る. $j > n_1^{(add)}$ ならば次のレベルへ. $j \leq n_1^{(add)}$ ならば(1.3)へ.
- (1.3) 点 $x_1^{new} = \arg \max \{EI(x)\}$ でサンプリングを行う. $D_1^{(j+1)} = D_1^{(j)} \cup \{(x_1^{new}, Y(x_1^{new}))\}$, $j = j + 1$ として, (1.2)へ.

レベル $t (\geq 2)$ におけるEGOを示す.

- (t.1) $j = 0$ とする. レベル $t-1$ で得られた $EI(x)$ を適合度関数とし, ルーレット選択を用いて標本点 $x_t^{(1)}, \dots, x_t^{(ini)}$ を選択し, $D_t^{(0)} = \{(x_t^{(i)}, Y_t(x_t^{(i)})) | i = 1, \dots, n_t^{(ini)}\}$ を得る.
- (t.2) $D_t^{(j)}$ および Y_{t-1}^{\dagger} を用いて, すべての点における $\hat{Y}_t(x), s_t^2(x)$ を求め, $EI(x)$ を得る. $j > n_t^{(add)}$ ならば次のレベルへ. $j \leq n_t^{(add)}$ ならば(t.3)へ.
- (t.3) 点 $x_t^{new} = \arg \max \{EI(x)\}$ でサンプリングを行う. $D_t^{(j+1)} = D_t^{(j)} \cup \{(x_t^{new}, Y(x_t^{new}))\}$, $j = j + 1$ として, (t.2)へ.

5 数値実験

精度の異なる標本の組に提案手法を適用した数値実験を行い、得られた解の精度および効率性を評価する。

1 変量関数 $f(x) = x^2 - 10 \cos(2\pi x) + 10$ ($-5.12 \leq x \leq 5.12$), (最適解 $x^* = 0, f(x^*) = 0$) に対して, 2つの異なる精度の標本を用いて, 最適解の探索を行う. 精度の低い標本の組をレベル 1 に属する標本, 精度の高い標本の組をレベル 2 に属する標本とする. 誤差分散はそれぞれ $\sigma_1^2 = 2^2, \sigma_2^2 = 1^2$ とするが, 数値実験を行う上では未知の値とする. 回帰関数を $h(x)^T \beta = \beta$ (定数) とする. サンプルングを行う点および予測を行う点は, 区間 $[-5.12, 5.12]$ を等間隔に分けた 101 点で, $x^{(1)} = -5.12, \dots, x^{(101)} = 5.12$ とする.

レベル 1, 2 におけるそれぞれの初期データ数 $n_1^{(ini)}, n_2^{(ini)}$ と追加データ数 $n_1^{(add)}, n_2^{(add)}$ を表 1 に示すようにあらかじめ定めておく. レベル 1 における初期データの標本点は無作為に抽出する. 無作為抽出によって初期データの標本点を変え, 各サンプルング法 (A1, A2, A3) に対してそれぞれ 50 回の数値実験を行った.

実験結果の評価には探索終了時まで得られたデータの最小値

$$y_{\min} = \min\{y(x_1^{(1)}), \dots, y(x_1^{(n_1)}), y(x_2^{(1)}), \dots, y_1(x_2^{(n_2)})\} \quad (29)$$

を用いる. 50 回の実験で得られたデータの最小値 $y_{\min}^{(1)}, \dots, y_{\min}^{(50)}$ の平均 $E\{y_{\min}\}$ と分散 $V\{y_{\min}\}$ を求める. 各サンプルング法に対する結果を表 1 に示す.

精度の低い標本のみ, 精度の高い標本のみを用いた最適解の探索も同様に行う. 初期データ数 $n^{(ini)}$, 追加データ数 $n^{(add)}$ をあらかじめ定める. 実験結果として, 得られたデータの最小値 y_{\min} の平均と分散を求める. 精度の低いデータによるサンプルング法 (B1, B2, B3, B4) とその結果を表 2 に, 精度の高いデータによるサンプルング法 (C1, C2, C3, C4) とその結果を表 3 にそれぞれ示す. また, 表 1 の (A2), 表 2 の (B2), 表 2 の (C2) のサンプルング法で得られた $y_{\min}^{(i)}$ ($i = 1, \dots, 50$) を図 1, 2, 3 にそれぞれ示す. 各図において $y_{\min}^{(i)}$ はそれぞれ \bullet で, 関数 $f(x)$ を破線で表している.

表 1: Evaluation of y_{\min} by different prediction data

	$n_1^{(ini)}$	$n_1^{(add)}$	$n_2^{(ini)}$	$n_2^{(add)}$	$E\{y_{\min}\}$	$V\{y_{\min}\}$
(A1)	15	5	7	3	2.485	12.03
(A2)	15	5	8	2	1.654	10.78
(A3)	15	5	10	1	1.048	13.12

表 2: Evaluation of y_{\min} by low prediction data

	$n^{(ini)}$	$n^{(add)}$	$E\{y_{\min}\}$	$V\{y_{\min}\}$
(B1)	10	30	1.019	13.25
(B2)	15	25	0.941	17.25
(B3)	20	20	0.680	15.47
(B4)	25	15	0.022	12.91

表 3: Evaluation of y_{\min} by high prediction data

	$n^{(ini)}$	$n^{(add)}$	$E\{y_{\min}\}$	$V\{y_{\min}\}$
(C1)	8	12	4.414	22.52
(C2)	10	10	2.468	16.93
(C3)	12	8	2.517	15.89
(C4)	15	5	2.520	20.44

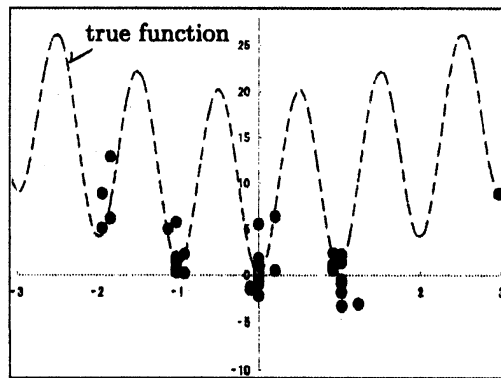


図 1: y_{\min} by different precision data

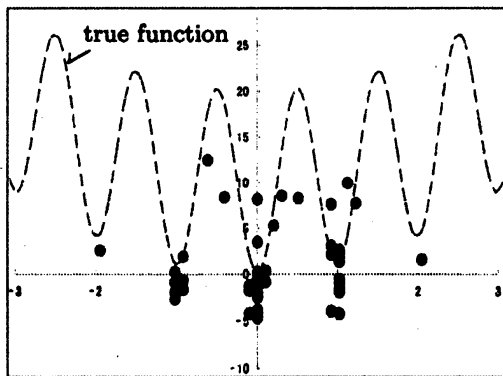


図 2: y_{\min} by low precision data

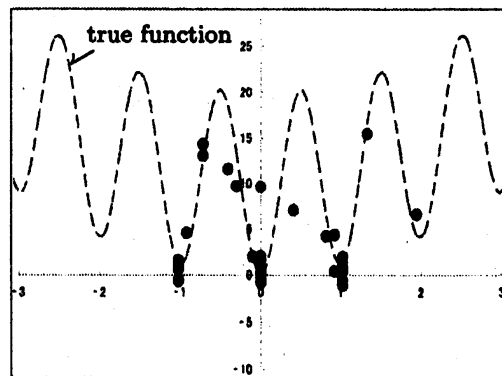


図 3: y_{\min} by high precision data

図 1 は図 2, 3 と比較して, y_{\min} が大域的最適解もしくは局所解に集まっている. 図 2, 3 は大域的最適解 $x^* = 0$ の付近に点が多くがあるが y_{\min} の値は大きなものもある. また, 表 2 より, 精度の低いデータによる $E\{y_{\min}\}$ の値は最も小さい. しかし, $V\{y_{\min}\}$ は異なる精度を組み合わせたデータによるものが最も良い. また, データの総数が同じでも初期データ数と追加データ数の割合によって結果が大きく異なる.

精度の低いデータは多くの点でサンプリングすることができるが, ばらつきが大きくなる可能性もある. 精度の高いデータはより良い解を得ようとすると多くのデータ数が必要であることがわかる. 精度の異なるデータを組み合わせることで, 高い精度の解を得るとともに, 効率的なサンプリングも実現できた.

6 まとめ

本研究では, EGO を用いて最適化のための効率的なサンプリングに関する考察を行った. EGO は予測を用いた最適化手法あり, Kriging モデルを用いることで異なる精度のデータを組み合わせることが可能となるため, これらを用いた精度の異なるデータによる最適化手法を提案した. さらに数値実験により, 精度の異なるデータを組み合わせることで, 単一精度のデータを用いた最適化よりも高い精度の解を得ることが可能であるとともに, 効率的なサンプリングも実現できることを検証した.

今後の課題としては多くの最適化問題に適用し, 有効性を検証していくことが考えられる. ま

た、サンプリングの効率性を考えるうえで、実用的な観点を欠かすことはできない。時間やコストによる制約、サンプリングを行う点に対する制約などが課された問題への取り組みが挙げられる。

参考文献

- [1] D. Hirst, G. Storvik, and A. R. Syversveen : A hierarchical modelling approach to combining environmental data at different scales, *Journal of the Royal Statistical Society, Series C*, **52**, 377-390 (2003).
- [2] D. den Hertog, J.P.C. Kleijnen and A.Y.D. Siem The correct Kriging variance estimated by bootstrapping, *Journal of the Operational Research Society*, **57**, 400-409 (2006).
- [3] D. R. Jones, M. Schonlau and W. J. Welch : Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* , **13**, 455-492 (2000).
- [4] D. R. Jones : A taxonomy of global optimization methods based on response surfaces, *Journal of Global Optimization* , **21**, 345-383 (2001).
- [5] M. C. Kennedy and A. O'Hagan : Predicting and output from a complex computer code when fast approximations are available, *Biometrika* , **87**, 1-13 (2000).
- [6] M. C. Kennedy and A. O'Hagan : Bayesian calibration of computer models, *Journal of the Royal Statistical Society, Series B*, **63**, 425-464 (2001).
- [7] J. Sacks, W. J. Welch, T. J. Mitchell and H. P. Wynn : Design and analysis of computer experiments, *Statistical Science*, **4**, 409-435 (2001).
- [8] T. J. Santner , B. J. Williams and W. I. Notz : The Design and Analysis of Computer Experiments, Springer, New York (2003)
- [9] A. Söbester, S. Leary and A. Keane : On the design of optimization strategies based on global response surface approximation models *Journal of Global Optimization* , **33**, 31-59 (2005).