PROCEEDINGS OF SCIENCE

# Machine Learning Estimation on the Trace of Inverse Dirac Operator using the Gradient Boosting Decision Tree Regression

**Benjamin J. Choi,**[a,*] **Hiroshi Ohno,**[a] **Takayuki Sumimoto**[b] **and Akio Tomiya**[c]

[a] *Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan*

[b] *FLECT Co., Ltd., 1-1-1 Shibaura, Minato-ku, Tokyo 105-0023, Japan*

[c] *Department of Information and Mathematical Sciences, Tokyo Woman's Christian University, 2-6-1 Zempukuji, Suginami-ku, Tokyo 167-8585, Japan*

*E-mail:* benchoi@ccs.tsukuba.ac.jp, hohno@ccs.tsukuba.ac.jp, akio@yukawa.kyoto-u.ac.jp

We present our preliminary results on the machine learning estimation of $\mathrm{Tr}\, M^{-n}$ from other observables with the gradient boosting decision tree regression, where $M$ is the Dirac operator. Ordinarily, $\mathrm{Tr}\, M^{-n}$ is obtained by linear CG solver for stochastic sources which needs considerable computational cost. Hence, we explore the possibility of cost reduction on the trace estimation by the adoption of gradient boosting decision tree algorithm. We also discuss effects of bias and its correction.

*The 41st International Symposium on Lattice Field Theory (LATTICE2024)*
*28 July - 3 August 2024*
*Liverpool, UK*

*Speaker





PoS(LATTICE2024)033

| ID | $L^3 \times T$ | $\beta$ | $\kappa$ | $c_{\rm SW}$ | $N_{\rm conf}$ |
|---|---|---|---|---|---|
| 0 | $16^3 \times 4$ | 1.60 | 0.13575 | 2.065 | 5500 |
| **1** | $\mathbf{16^3 \times 4}$ | **1.60** | **0.13580** | **2.065** | **5500** |
| 2 | $16^3 \times 4$ | 1.60 | 0.13585 | 2.065 | 5500 |

**Table 1:** Data used in this paper. These data are originally produced for Ref. [8]. Note that the first phase transition is observed in the ID-1 (the row with bold text).

## 1. Introduction

For Lattice QCD calculations on observables such as cumulants of the chiral order parameter, the trace of operators ($\mathrm{Tr}\, O$) is often necessary where $O$ represents an observable such as $O = M^{-1}$, inverse Dirac operator. Here, the trace of operator is often estimated with a stochastic source method [1]. Since $M = (\not{D} + m_0)$, the Dirac operator, is a large sparse matrix on the lattice, its inverse matrix obtained by a linear CG solver requires a considerable computational cost.

In this paper, we present our preliminary result on machine learning estimation of $\mathrm{Tr}\, M^{-n}$ from other observables such as $\mathrm{Tr}\, M^{-m}$ ($m < n$). As input observables, we also use plaquette and Polyakov loop obtained when generating gauge configurations with the hybrid Monte Carlo (HMC) algorithm [2–4]. Here, we use the gradient boosting decision tree regression method [5], based on the methodology of Yoon *et al.* [6]. Note that recently, a machine learning mapping between $\mathrm{Tr}\, M^{-1}$ at different quark masses and gradient flow times was studied in the similar way [7].

For this preliminary analysis, we use the data originally produced for Ref. [8] where Oakforest-PACS system [9] and BQCD program [10] are used. In this data, we use $N_f = 4$ Wilson clover action [11] and Iwasaki gauge action [12, 13] as explained in Ref. [8]. The data used in this work are given in Table 1. We choose 3 datasets which have the same lattice size of $16^3 \times 4$, the lattice coupling constant $\beta = 1.60$, the clover coefficient $c_{\rm SW} = 2.065$ and the number of gauge configurations $N_{\rm conf} = 5500$ but different $\kappa$ values. Especially, at the ID-1 in Table 1, a first phase transition is observed.

## 2. Analysis detail

Our notation and convention used in this work are summarized in Table 2. For the machine learning (ML) estimation, we perform the following two steps:

1. We train a model $f(X)$ using the data in the $S^X_{\rm TR}$ and $S^Y_{\rm TR}$.
2. We input $X \in S^X_{\rm UL}$ into the model $f(X)$ and obtain ML estimation $f(X) = Y^P \approx Y$.

Note that only $Z \in S^Z_{\rm LB} \subset S^Z$ ($Z = X, Y$) can be used for the model training, which means that there may exist any kind of unwanted fluctuations or weird bias due to the *partial* data usage, in principle. Therefore, using the methodology given in Ref. [6], we do not use all $S^Z_{\rm LB}$ for the model training but use only $S^Z_{\rm TR} \subset S^Z_{\rm LB}$, the training set. We use remaining $S^Z_{\rm BC} = S^Z_{\rm LB} \setminus S^Z_{\rm TR}$ for the bias correction. Then we calculate following two estimations:

$$\bar{Y}_{\mathcal{P}1} = \frac{1}{N_{\rm UL}} \sum_{Y_i \in S^Y_{\rm UL}} Y^P_i + \frac{1}{N_{\rm BC}} \sum_{Y_j \in S^Y_{\rm BC}} \left( Y_j - Y^P_j \right), \tag{1}$$

| Symbol | Description |
|---|---|
| $X, Y$ | observables used as input ($X$) and output ($Y$) (*e.g.*, $X = \mathrm{Tr}\, M^{-1}$, $Y = \mathrm{Tr}\, M^{-2}$) |
| $S^Z$ | the total dataset of $Z = X, Y$ where $S^Z = S^Z_{\mathrm{LB}} \cup S^Z_{\mathrm{UL}}$ |
| $S^Z_{\mathrm{LB}}$ | the labeled set of $Z = X, Y$ where $S^Z_{\mathrm{LB}} = S^Z_{\mathrm{TR}} \cup S^Z_{\mathrm{BC}}$ |
| $S^Z_{\mathrm{TR}}$ | the training set of $Z = X, Y$ |
| $S^Z_{\mathrm{BC}}$ | the bias correction set of $Z = X, Y$ |
| $S^Z_{\mathrm{UL}}$ | the unlabeled set of $Z = X, Y$ |
| $N$ | the number of elements of $S^Z$ where $N = \left\|S^X\right\| = \left\|S^Y\right\|$ |
| $N_{\mathrm{LB}}$ | the number of elements of $S^Z_{\mathrm{LB}}$ where $N_{\mathrm{LB}} = \left\|S^X_{\mathrm{LB}}\right\| = \left\|S^Y_{\mathrm{LB}}\right\|$ |
| $N_{\mathrm{TR}}$ | the number of elements of $S^Z_{\mathrm{TR}}$ where $N_{\mathrm{TR}} = \left\|S^X_{\mathrm{TR}}\right\| = \left\|S^Y_{\mathrm{TR}}\right\|$ |
| $N_{\mathrm{BC}}$ | the number of elements of $S^Z_{\mathrm{BC}}$ where $N_{\mathrm{BC}} = \left\|S^X_{\mathrm{BC}}\right\| = \left\|S^Y_{\mathrm{BC}}\right\|$ |
| $N_{\mathrm{UL}}$ | the number of elements of $S^Z_{\mathrm{UL}}$ where $N_{\mathrm{UL}} = \left\|S^X_{\mathrm{UL}}\right\| = \left\|S^Y_{\mathrm{UL}}\right\|$ |
| $f(X)$ | the model trained with $S^X_{\mathrm{TR}}$ and $S^Y_{\mathrm{TR}}$ |
| $Y^P$ | the machine learning estimation on $Y$ |

**Table 2:** Notation and convention used in this paper for the explanation of our work.

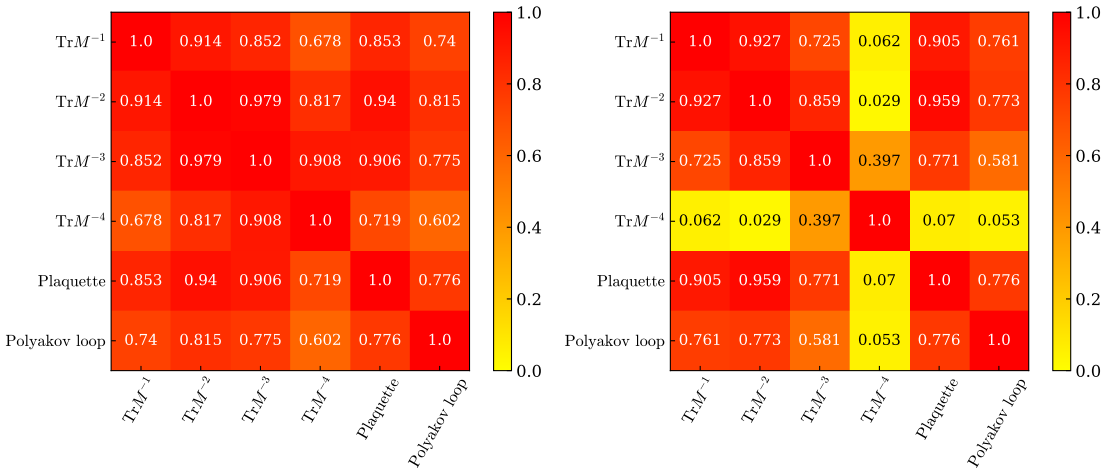
$$\bar{Y}_{\mathcal{P}2} = \frac{N_{\mathrm{UL}}}{N}\bar{Y}_{\mathcal{P}1} + \frac{N_{\mathrm{LB}}}{N}\bar{Y}_{(\mathrm{LB})} \quad \text{where} \quad N = N_{\mathrm{LB}} + N_{\mathrm{UL}}\,, \quad \bar{Y}_{(\mathrm{LB})} = \frac{1}{N_{\mathrm{LB}}}\sum_{Y_k \in S^Y_{\mathrm{LB}}} Y_k\,. \tag{2}$$

Here, $\bar{Y}_{\mathcal{P}1}$ in Eq. (1) is the statistical average on $Y^P$ with the bias correction. To improve the statistical precision, we further introduce $\bar{Y}_{\mathcal{P}2}$ in Eq. (2), the weighted average of $\bar{Y}_{\mathcal{P}1}$ and the original CG results in the labeled set, $\bar{Y}_{(\mathrm{LB})}$. To obtain statistical errors, we use the bootstrap resampling method with $N_{\mathrm{BS}} = 10,000$ where $N_{\mathrm{BS}}$ represents the number of bootstrap resamples.

For the ML estimation on $Y$ using $X$, the strong correlation between $X$ and $Y$ is often required. In the Fig. 1, we report the correlation between $\mathrm{Tr}\, M^{-n}$ ($n = 1, 2, 3, 4$), plaquette and Polyakov loop from two datasets: `ID-0` (the heaviest quark) and `ID-2` (the lightest quark). The red (yellow) color represents strong (weak) correlation between observables. Note that we need to be careful with $Y = \mathrm{Tr}\, M^{-4}$ case, due to the weak correlation with other observables such as $X = \mathrm{Tr}\, M^{-m}$ ($m = 1, 2, 3$), $X =$ Plaquette and $X =$ Polyakov loop. Except for it, we observe strong correlations in most cases. Hence, using the overall tendency of strong correlation, we perform the ML estimation. We also check the results for $Y = \mathrm{Tr}\, M^{-4}$ and compare it with the other analysis results for $Y = \mathrm{Tr}\, M^{-n}$ ($n = 1, 2, 3$).

For the economic efficiency of the ML estimation, that is, to reduce the computational cost from the linear CG solver, we should use the least possible $S^Z_{\mathrm{LB}}$ ($Z = X, Y$). Here, note that we need to use $Y$ as well as $Y^P$ for the bias correction. This means that we should grant sufficient statistics to the $S^Z_{\mathrm{BC}}$ among the $S^Z_{\mathrm{LB}}$. In summary, we monitor the following two factors in this work:

1. We find out minimal $\mathcal{R}_{\mathrm{LB}} \equiv \dfrac{N_{\mathrm{LB}}}{N}$ where $N = N_{\mathrm{LB}} + N_{\mathrm{UL}}$ and $\mathcal{R}_{\mathrm{LB}} = 5,\ 10,\ \cdots,\ 50\%$.
2. We find out maximal $\mathcal{R}_{\mathrm{TR}} \equiv \dfrac{N_{\mathrm{TR}}}{N_{\mathrm{LB}}}$ where $N_{\mathrm{LB}} = N_{\mathrm{TR}} + N_{\mathrm{BC}}$ and $\mathcal{R}_{\mathrm{TR}} = 10,\ 20,\ \cdots,\ 90\%$.




(a) $\kappa = 0.13575$, `ID-0` (the heaviest quark) (b) $\kappa = 0.13585$, `ID-2` (the lightest quark)

**Figure 1:** Correlation between physical observables.

Here, we also monitor $\mathcal{R}_{\rm TR} = 0\%$ ($S^Z_{\rm TR} = \emptyset$ for $Z = X, Y$) to check that $S^Y_{\rm LB}$ itself approaches to the true answer along the increase of $\mathcal{R}_{\rm LB}$, that is, we do not perform ML estimation at $\mathcal{R}_{\rm TR} = 0\%$ but observe only the statistical average and error of the $S^Y_{\rm LB}$. We also monitor $\mathcal{R}_{\rm TR} = 100\%$ ($S^Z_{\rm BC} = \emptyset$) to check what happens when we do not use the bias correction method in the ML estimation.

For the ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ ($n = 1, 2, 3, 4$), we use the gradient boosting decision tree regression method [5]. To be specific, we use `LightGBM` [14] framework via `JuliaAI/MLJ.jl` [15]. We use 40 boosting stages of depth-3 trees with learning rate of 0.1 and sub-sampling of 0.7.

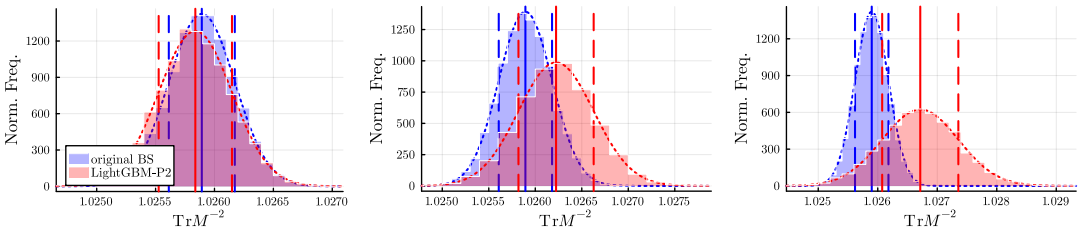
To check the usefulness of this method, we compare the ML estimation, $\{\bar{Y}_Q, \sigma_Q\}$ ($Q = \mathcal{P}_1, \mathcal{P}_2$), with the original CG result, $\{\bar{Y}_{\rm Orig}, \sigma_{\rm Orig}\}$. We prepare following two evaluation criteria (**EC-x**).

**EC-1** We check whether statistical average of ML estimation and original CG results are close to each other or not.

(a) If both of $\bar{Y}_{\rm Orig}$ and $\bar{Y}_Q$ agree with 1 $\sigma$ level each other, then we grant score 2. (Fig. 2(a))

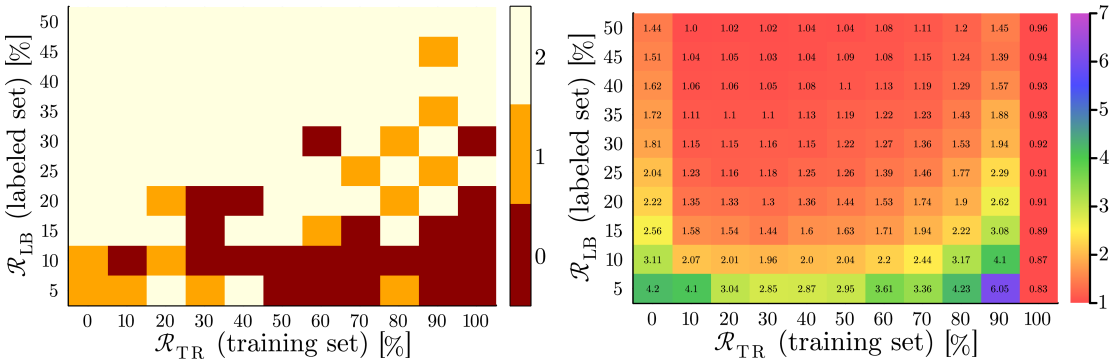
(b) If only one of them is included in the other's 1 $\sigma$ error, then we grant score 1. (Fig. 2(b))

(c) If $\bar{Y}_{\rm Orig}$ and $\bar{Y}_Q$ do not agree with 1 $\sigma$ level each other, then we grant score 0. (Fig. 2(c))




(a) Score 2 (b) Score 1 (c) Score 0

**Figure 2:** Histograms on the three possible cases can be occurred where blue (red) histogram represents original CG result (ML estimation). Here, data for $X = \mathrm{Tr}\, M^{-1}$ and $Y = \mathrm{Tr}\, M^{-2}$ are used. Vertical solid (dotted) lines represent statistical averages (1 $\sigma$ errors).




(a) Result of **EC-1** (b) Result of **EC-2**

**Figure 3:** Results on (a) **EC-1** and (b) **EC-2** of $\mathcal{P}2$ estimation using $(X\,,\,Y) = (\text{Plaquette}\,,\,\text{Tr}\,M^{-3})$ for `ID-0` dataset where the heaviest quark is used. White, orange, red color of (a) represents score 2, 1, 0, respectively. Here, the magnitude of $\mathcal{R}_\sigma$ (Eq. (3)) in (b) is represented using the color of rainbow. For example, if $\mathcal{R}_\sigma \approx 7$ for certain point, then we observe purple-like color there.

| $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ | Plaquette | Polyakov loop | $\text{Tr}\,M^{-1}$ | $\text{Tr}\,M^{-2}$ | $\text{Tr}\,M^{-3}$ |
|---|---|---|---|---|---|
| $\text{Tr}\,M^{-1}$ | {**40**, 50} | {**40**, 40} | | | |
| $\text{Tr}\,M^{-2}$ | {**30**, 40} | {**35**, 30} | {**30**, 40} | | |
| $\text{Tr}\,M^{-3}$ | {**30**, 40} | {**40**, 40} | {**35**, 40} | {**25**, 70} | |
| $\text{Tr}\,M^{-4}$ | {**45**, 40} | {**45**, 40} | {**45**, 40} | {**35**, 40} | {**25**, 40} |

**Table 3:** Results on $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ of $\mathcal{P}2$ estimation for `ID-0` dataset where the heaviest quark is used. Here, the column and row represent the $X$ (input) and $Y$ (output), respectively. For example, $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\} = \{\mathbf{30}, 40\}$ represents that ML estimation got consistent score 2 at **EC-1** and showed $\mathcal{R}_\sigma \leq 1.2$ at **EC-2** in the $\mathcal{R}_{\rm LB} \gtrsim 30\%$ and $\mathcal{R}_{\rm TR} \lesssim 40\%$ region.

**EC-2** We check whether

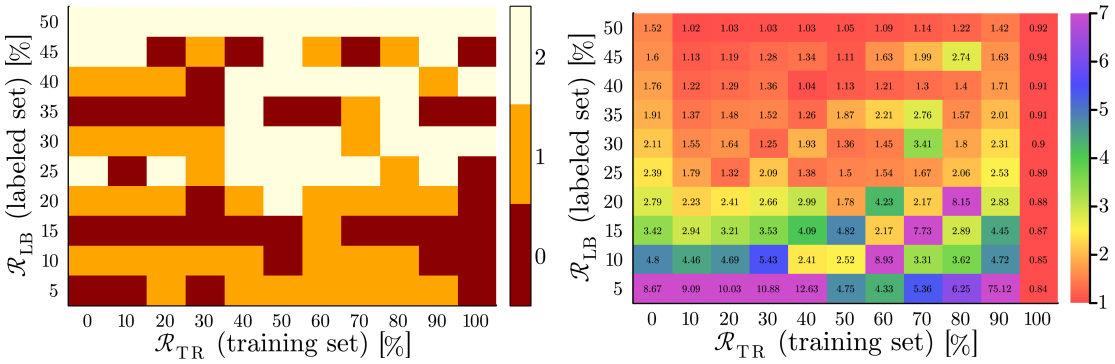
$$\mathcal{R}_\sigma \equiv \frac{\sigma_Q}{\sigma_{\rm Orig}} \approx 1\,, \tag{3}$$

that is, the statistical error of ML estimation, $\sigma_Q$, is close to that of original CG result, $\sigma_{\rm Orig}$, or not. Currently, we are finding unambiguous explicit criterion for this. In this paper, we use $\mathcal{R}_\sigma \leq 1.2$ which is tentatively determined empirically, monitoring our preliminary results.

Therefore, if an ML estimation got score 2 at the **EC-1** and turned out that $\mathcal{R}_\sigma \approx 1$ (tentatively $\mathcal{R}_\sigma \leq 1.2$ in this paper) at the **EC-2**, then this ML estimation can be thought that it imitates its original CG result as well as possible.

## 3. Results

Here we show our preliminary results on the ML estimation of $Y = \text{Tr}\,M^{-n}$. Note that we use single $X$ for the ML estimation of $Y$ in this analysis. As examples, we show results on **EC-1** and




(a) Result of **EC-1** (b) Result of **EC-2**

**Figure 4:** Results on (a) **EC-1** and (b) **EC-2** of $\mathcal{P}2$ estimation using $(X\,,\,Y) = (\text{Plaquette}\,,\,\text{Tr}\,M^{-3})$ for `ID-2` dataset where the lightest quark is used. We use same notation as in Fig. 3.

| $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ | Plaquette | Polyakov loop | $\text{Tr}\,M^{-1}$ | $\text{Tr}\,M^{-2}$ | $\text{Tr}\,M^{-3}$ |
|---|---|---|---|---|---|
| $\text{Tr}\,M^{-1}$ | {**30**, 40} | {**35**, 30} | | | |
| $\text{Tr}\,M^{-2}$ | {**40**, 80} | {**40**, 40} | {**30**, 50} | | |
| $\text{Tr}\,M^{-3}$ | N.A. | N.A. | N.A. | N.A. | |
| $\text{Tr}\,M^{-4}$ | N.A. | N.A. | N.A. | N.A. | N.A. |

**Table 4:** Results on $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ of $\mathcal{P}2$ estimation for `ID-2` dataset where the lightest quark is used. We use same notation as in Table 3. Here, "N.A." means that we cannot find consistent good region for **EC-1** and **EC-2**.

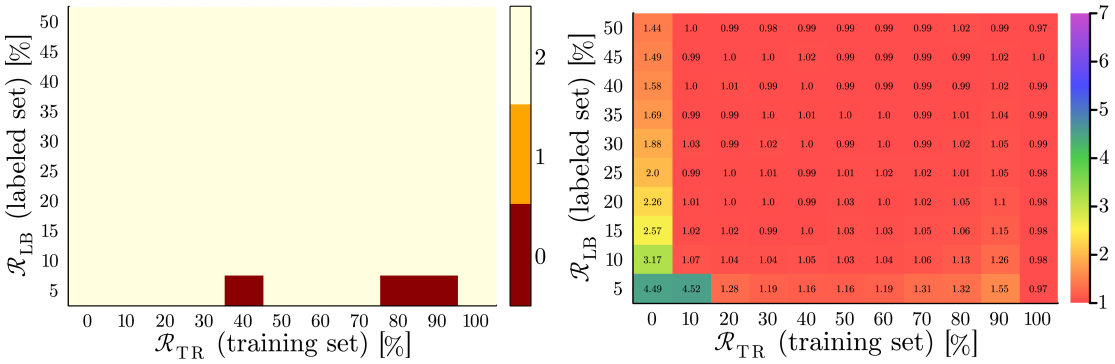
**EC-2** using $(X\,,\,Y) = (\text{Plaquette}\,,\,\text{Tr}\,M^{-3})$ for `ID-0` dataset (Fig. 3), `ID-1` dataset (Fig. 5), and `ID-2` dataset (Fig. 4).

We report our results on `ID-0` dataset where the heaviest quark is used in the measurement. In Fig. 3(a), we observe that we obtain score 2 consistently in the region of $\mathcal{R}_{\rm LB} \gtrsim 30\%$ and $\mathcal{R}_{\rm TR} \lesssim 50\%$ (**EC-1**). In Fig. 3(b), we observe that we obtain $\mathcal{R}_\sigma \leq 1.2$ in the region of $\mathcal{R}_{\rm LB} \gtrsim 30\%$ and $\mathcal{R}_{\rm TR} \lesssim 40\%$ (**EC-2**). As we can see from $\mathcal{R}_{\rm TR} = 100\%$ column ($S^Z_{\rm BC} = \emptyset$ for $Z = X, Y$) of Fig. 3(a), we need $\mathcal{R}_{\rm LB} \gtrsim 35\%$, slightly larger $\mathcal{R}_{\rm LB}$ than those of $S^Z_{\rm BC} \neq \emptyset$ case.

In Table 3, we report our analysis on all possible cases with single input ($X$) and output ($Y$) for `ID-0` dataset. We found that the ML estimation on $Y = \text{Tr}\,M^{-n}$ works well for $\mathcal{R}_{\rm LB} \gtrsim 40\%$ even when we use $X =$ Plaquette and $X =$ Polyakov loop. We also found that the ML estimation on $Y = \text{Tr}\,M^{-4}$ needs more ratio of labeled set ($\mathcal{R}_{\rm LB}$) than $Y = \text{Tr}\,M^{-n}$ ($n = 1, 2, 3$). However, we need $\mathcal{R}_{\rm LB} \gtrsim 25\%$ when $(X\,,\,Y) = (\text{Tr}\,M^{-2}\,,\,\text{Tr}\,M^{-3})$ and $(X\,,\,Y) = (\text{Tr}\,M^{-3}\,,\,\text{Tr}\,M^{-4})$.

Next, we report our results on `ID-2` dataset where the lightest quark is used in the measurement. We cannot find consistent score-2 region (**EC-1**) from Fig. 4(a). We also cannot find consistent $\mathcal{R}_\sigma \leq 1.2$ region (**EC-2**) from Fig. 4(b). The ML estimation on $Y = \text{Tr}\,M^{-3}$ using $X =$ Plaquette does not work well for `ID-2` dataset.

In Table 4, we report our analysis on all possible cases with single input ($X$) and output ($Y$) for `ID-2` dataset. We found that only the ML estimation on $Y = \text{Tr}\,M^{-n}$ ($n = 1, 2$) works well for




(a) Result of **EC-1** (b) Result of **EC-2**

**Figure 5:** Results on (a) **EC-1** and (b) **EC-2** of $\mathcal{P}2$ estimation using $(X\,,\,Y) = (\text{Plaquette}\,,\,\text{Tr}\,M^{-3})$ for `ID-1` dataset where the first order phase transition observed. We use same notation as in Fig. 3.

| $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ | Plaquette | Polyakov loop | $\text{Tr}\,M^{-1}$ | $\text{Tr}\,M^{-2}$ | $\text{Tr}\,M^{-3}$ |
|---|---|---|---|---|---|
| $\text{Tr}\,M^{-1}$ | {**10**, 80} | {**10**, 70} | | | |
| $\text{Tr}\,M^{-2}$ | {**10**, 90} | {**10**, 70} | {**10**, 80} | | |
| $\text{Tr}\,M^{-3}$ | {**10**, 80} | {**10**, 70} | {**10**, 70} | {**10**, 90} | |
| $\text{Tr}\,M^{-4}$ | {**40**, 70} | {**40**, 80} | {**40**, 80} | {**40**, 80} | {**40**, 80} |

**Table 5:** Results on $\{\mathcal{R}^{\gtrsim}_{\rm LB}, \mathcal{R}^{\lesssim}_{\rm TR}\}$ of $\mathcal{P}2$ estimation for `ID-1` dataset where the first order phase transition is observed. We use same notation as in Table 3.

$\mathcal{R}_{\rm LB} \gtrsim 40\%$. The results on **EC-1** and **EC-2** for $Y = \text{Tr}\,M^{-n}$ $(n = 3, 4)$ are all similar with Fig. 4.

Finally, we report our results on `ID-1` dataset where the first order phase transition is observed. In Fig. 5(a), we observe that we obtain score 2 consistently in the region of $\mathcal{R}_{\rm LB} \gtrsim 10\%$ and $\mathcal{R}_{\rm TR} \lesssim 90\%$ (**EC-1**). In Fig. 5(b), we observe that we obtain $\mathcal{R}_\sigma \leq 1.2$ in the region of $\mathcal{R}_{\rm LB} \gtrsim 10\%$ and $\mathcal{R}_{\rm TR} \lesssim 80\%$ (**EC-2**).

In Table 5, we report our analysis on all possible cases with single input $(X)$ and output $(Y)$ for `ID-1` dataset. We found that the ML estimation on $Y = \text{Tr}\,M^{-n}$ $(n = 1, 2, 3)$ works well for $\mathcal{R}_{\rm LB} \gtrsim 10\%$ even when we use $X =$ Plaquette and $X =$ Polyakov loop. On the other hand, the ML estimation on $Y = \text{Tr}\,M^{-4}$ needs more ratio of labeled set $(\mathcal{R}_{\rm LB})$ than $Y = \text{Tr}\,M^{-n}$ $(n = 1, 2, 3)$, *i.e.*, $\mathcal{R}_{\rm LB} \gtrsim 40\%$.

## 4. Summary and to-do list

We performed preliminary analysis on ML estimation on $Y = \text{Tr}\,M^{-n}$ $(n = 1, 2, 3, 4)$ using $X = \text{Tr}\,M^{-m}$ $(m < n)$, $X =$ Plaquette and $X =$ Polyakov loop. Here, we used the gradient boosting decision tree regression method [5].

With the heaviest quark (`ID-0` dataset), we observed that the ML estimation of $Y = \text{Tr}\,M^{-n}$ $(n = 1, 2, 3)$ showed consistently good results at $\mathcal{R}_{\rm LB} \gtrsim 40\%$ where $Y = \text{Tr}\,M^{-4}$ required slightly more $\mathcal{R}_{\rm LB}$ than $n = 1, 2, 3$ cases. With the lightest quark (`ID-2` dataset), we observed that only the

ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ $(n = 1, 2)$ works well at $\mathcal{R}_{\mathrm{LB}} \gtrsim 40\%$. On the other hand, the `ID-1` dataset where the first order phase transition is observed, we observed that the ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ $(n = 1, 2, 3)$ showed consistently good results at $\mathcal{R}_{\mathrm{LB}} \gtrsim 10\%$. However, $Y = \mathrm{Tr}\, M^{-4}$ still required $\mathcal{R}_{\mathrm{LB}} \gtrsim 40\%$ even in this dataset.

In this preliminary result with three datasets, we observed that ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ works well with heavier quark mass. Especially, we observed that the ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ works quite well when the first order phase transition is observed at the dataset. However, the quality of $Y = \mathrm{Tr}\, M^{-4}$ estimation is not good comparing with $Y = \mathrm{Tr}\, M^{-n}$ $(n = 1, 2, 3)$. To get better $Y = \mathrm{Tr}\, M^{-4}$ estimation, we need to use $X = \mathrm{Tr}\, M^{-3}$, for example (except for `ID-2` dataset where the lightest quark is used).

In this paper, we used single input $(X)$ for the ML estimation of $Y = \mathrm{Tr}\, M^{-n}$ $(n = 1, 2, 3, 4)$. We need to check whether we get better results when we use multiple inputs: for example, we use $X = \{\text{Plaquette}\,, \mathrm{Tr}\, M^{-1}\}$ for the ML estimation of $Y = \mathrm{Tr}\, M^{-4}$. This work is in progress.

We need to check whether we can obtain reliable cumulants of chiral order parameters such as susceptibility, skewness, kurtosis [8] using ML estimation. This work is also in progress.

## Acknowledgments

The work of A. T. was partially supported by JSPS KAKENHI Grant Numbers 20K14479, 22H05111 and 22K03539. A. T. and H. O. were partially supported by JSPS KAKENHI Grant Number 22H05112. This work was partially supported by MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Grant Number JPMXP1020230411, JPMXP1020230409).

## References

[1] S.-J. Dong and K.-F. Liu, *Stochastic estimation with Z(2) noise*, *Phys. Lett. B* **328** (1994) 130 [hep-lat/9308015].

[2] S. Duane and J.B. Kogut, *Hybrid Stochastic Differential Equations Applied to Quantum Chromodynamics*, *Phys. Rev. Lett.* **55** (1985) 2774.

[3] S. Duane and J.B. Kogut, *The Theory of Hybrid Stochastic Algorithms*, *Nucl. Phys. B* **275** (1986) 398.

[4] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.

[5] J.H. Friedman, *Greedy function approximation: A gradient boosting machine.*, *The Annals of Statistics* **29** (2001) 1189 .

[6] B. Yoon, T. Bhattacharya and R. Gupta, *Machine Learning Estimators for Lattice QCD Observables*, *Phys. Rev. D* **100** (2019) 014504 [1807.05971].

[7] J. Kim, G. Pederiva and A. Shindler, *Machine learning mapping of lattice correlated data*, *Phys. Lett. B* **856** (2024) 138894 [2402.07450].

[8] H. Ohno, Y. Kuramashi, Y. Nakamura and S. Takeda, *Continuum extrapolation of the critical endpoint in 4-flavor QCD with Wilson-Clover fermions*, *PoS* **LATTICE2018** (2018) 174 [1812.01318].

[9] T. Boku, K.-I. Ishikawa, Y. Kuramashi and L. Meadows, *Mixed Precision Solver Scalable to 16000 MPI Processes for Lattice Quantum Chromodynamics Simulations on the Oakforest-PACS System*, in *5th International Workshop on Legacy HPC Application Migration: International Symposium on Computing and Networking*, 9, 2017 [1709.08785].

[10] Y. Nakamura and H. Stuben, *BQCD - Berlin quantum chromodynamics program*, *PoS* **LATTICE2010** (2010) 040 [1011.0199].

[11] B. Sheikholeslami and R. Wohlert, *Improved Continuum Limit Lattice Action for QCD with Wilson Fermions*, *Nucl. Phys. B* **259** (1985) 572.

[12] Y. Iwasaki, *Renormalization group analysis of lattice theories and improved lattice action: Two-dimensional non-linear O(N) sigma model*, *Nucl. Phys. B* **258** (1985) 141.

[13] Y. Iwasaki, *Renormalization Group Analysis of Lattice Theories and Improved Lattice Action. II. Four-dimensional non-Abelian SU(N) gauge model*, 1111.7054.

[14] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma *et al.*, *LightGBM: A highly efficient gradient boosting decision tree*, *Advances in Neural Information Processing Systems* **30** (2017) .

[15] A.D. Blaom, F. Kiraly, T. Lienart, Y. Simillides, D. Arenas and S.J. Vollmer, *MLJ: A julia package for composable machine learning*, *Journal of Open Source Software* **5** (2020) 2704.