International workshop on numerical methods and simulations for materials design and strongly correlated quantum matters @ RIKEN

# Estimation of effective models by machine learning

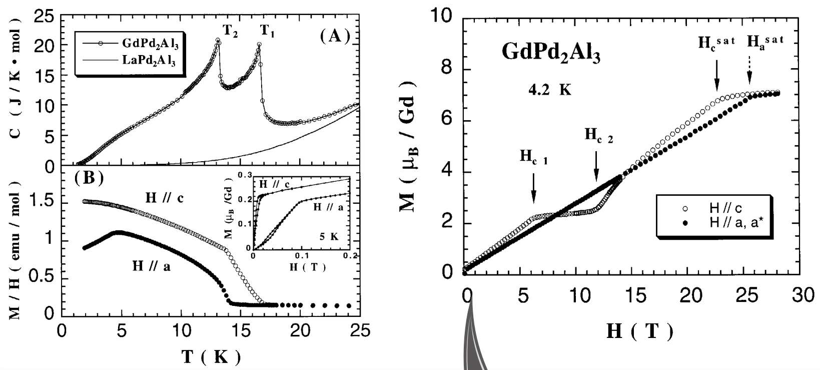

NIMS　Ryo Tamura

Collaborator:
U. Tokyo NIMS　Koji Hukushima


25/Mar/2017

# *Motivation*

## Experimental results




## Candidate models

$b_{ij}(\mathbf{s}_i \cdot \mathbf{s}_j)^2$

$J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j$ $\quad \mathbf{d}_{ij} \cdot [\mathbf{s}_i \times \mathbf{s}_j]$ $\quad D_i(s_i^z)^2$

$$\frac{\mathbf{s}_i \cdot \mathbf{s}_j}{r_{ij}^3} - 3\frac{(\mathbf{s}_i \cdot \mathbf{r}_{ij})(\mathbf{s}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5}$$

input

## Model selection

***Machine learning***

input

L1 regularization
L2 regularization
Full search
+
Cross validation

output

**Plausible effective model for experimental results (selection of model parameters in candidate model)**

# *As the first stage*

**To estimate the spin Hamiltonian from data of magnetic materials by machine learning**

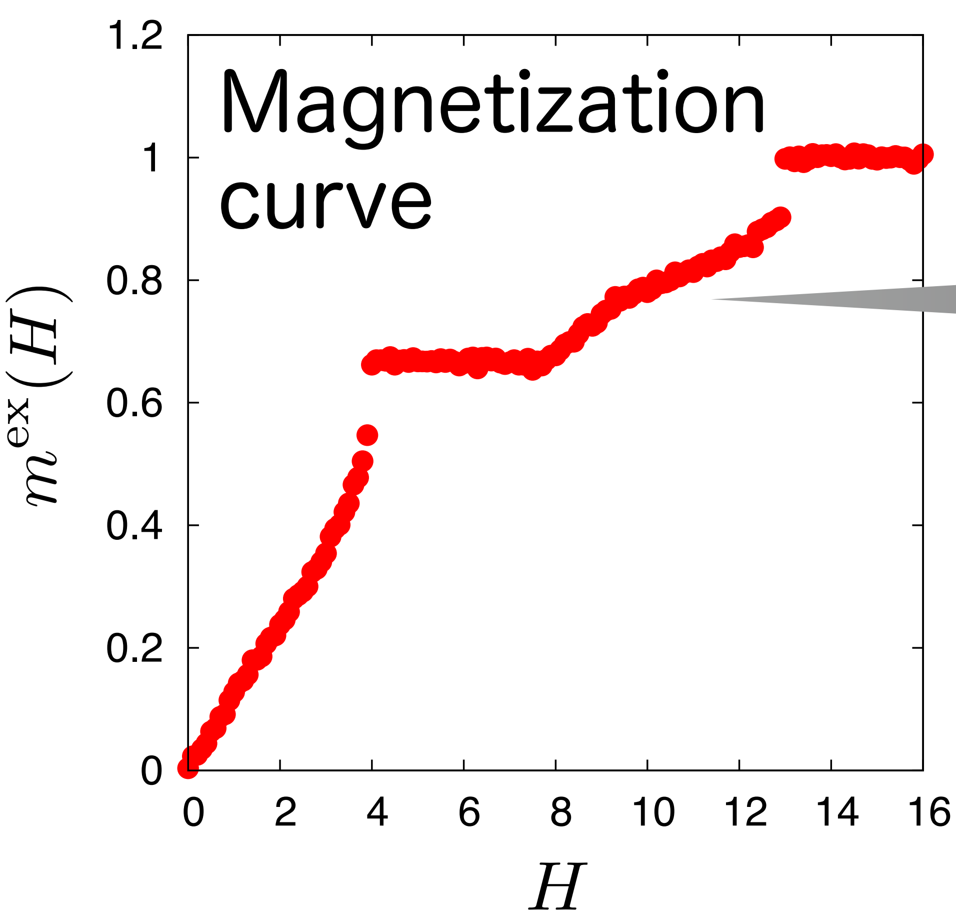
**If we can estimate spin Hamiltonian ..**

- Expect the **spin snapshot, magnetic structure, and structure factor**.
- Expect the properties which **cannot be observed directly** such as magnetic specific heat and magnetic entropy.
- Expect the properties **in extreme environments** such as super high magnetic field and super low temperature.

# *As the first stage*

**To estimate the spin Hamiltonian from data of magnetic materials by machine learning**

To estimate the spin Hamiltonian from magnetization curve by machine learning.




***Machine learning***

$$\mathcal{H} = \cdots$$

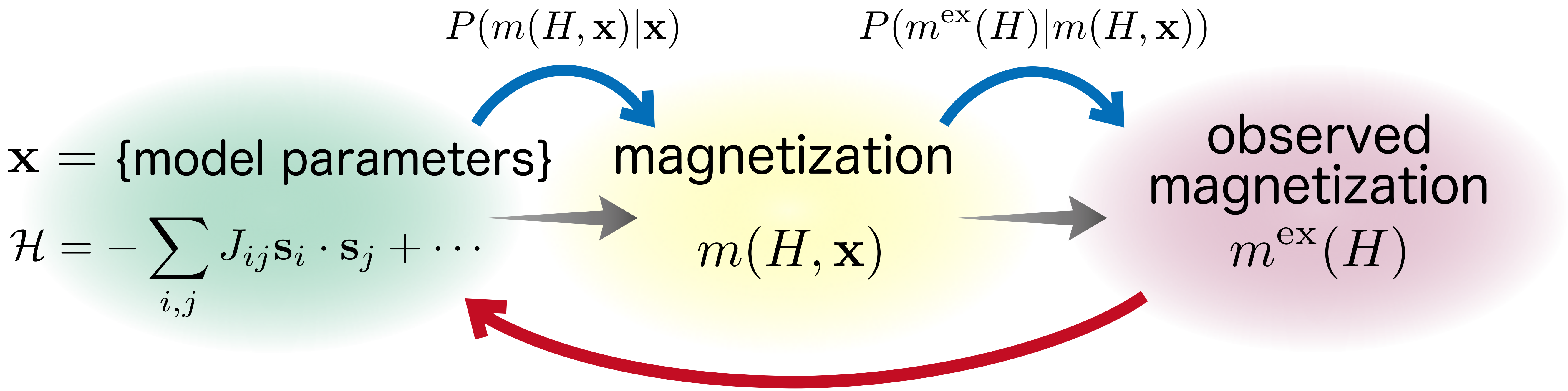
$$b_{ij}(\mathbf{s}_i \cdot \mathbf{s}_j)^2$$

$$J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j \quad \mathbf{d}_{ij} \cdot [\mathbf{s}_i \times \mathbf{s}_j] \quad D_i(s_i^z)^2$$

$$\frac{\mathbf{s}_i \cdot \mathbf{s}_j}{r_{ij}^3} - 3\frac{(\mathbf{s}_i \cdot \mathbf{r}_{ij})(\mathbf{s}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5}$$

# *Forward modeling and Bayes modeling*

## Forward modeling




$$P(\mathbf{x}|m^{\mathrm{ex}}(H)) = \frac{P(m^{\mathrm{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\mathrm{ex}}(H))}$$

## Bayes modeling

$P(B|A)$ : Conditional probability of event $B$ given event $A$
(Posterior distribution : 事後分布)

# *Thermal average - forward modeling -*

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\quad$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x}=$ {model parameters}

$$\mathcal{H}=-\sum_{i,j}J_{ij}\mathbf{s}_i\cdot\mathbf{s}_j+\cdots$$

→ magnetization $m(H,\mathbf{x})$ → observed magnetization $m^{\mathrm{ex}}(H)$

Definition of magnetization as thermal average of spins

$$\langle\mathbf{s}_i\rangle_{H,\mathbf{x}}=\frac{\mathrm{Tr}\mathbf{s}_i\mathrm{e}^{-\beta\mathcal{H}}}{\mathrm{Tr}\mathrm{e}^{-\beta\mathcal{H}}}\quad\longrightarrow\quad m(H,\mathbf{x})=\left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle\mathbf{s}_i\rangle_{H,\mathbf{x}}\right|$$

Conditional probability of $m(H,\mathbf{x})$ given $\mathbf{x}$

$$P(m(H,\mathbf{x})|\mathbf{x})=\delta\left(m(H,\mathbf{x})-\left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle\mathbf{s}_i\rangle_{H,\mathbf{x}}\right|\right)$$

Magnetization is uniquely obtained when the model parameters are given.

# *Observation noise - forward modeling -*

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $P(m^{\rm ex}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters} → magnetization → observed magnetization

$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$ $m(H,\mathbf{x})$ $m^{\rm ex}(H)$

Existence of observation noise in $m^{\rm ex}(H)$

$$m^{\rm ex}(H) = m(H,\mathbf{x}) \underline{+\,\varepsilon}$$

observation noise

Assumption : $P(\varepsilon) \propto \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$

Conditional probability of $m^{\rm ex}(H)$ given $m(H,\mathbf{x})$

$$P(m^{\rm ex}(H)|m(H,\mathbf{x})) \propto \exp\left(-\frac{1}{2\sigma^2}(m^{\rm ex}(H) - m(H,\mathbf{x}))^2\right)$$

# *Conditional probability - forward modeling -*

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters} → magnetization → observed magnetization

$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$ → $m(H,\mathbf{x})$ → $m^{\mathrm{ex}}(H)$

Conditional probability of $m^{\mathrm{ex}}(H)$ given $\mathbf{x}$

$$P(m^{\mathrm{ex}}(H)|\mathbf{x}) \propto \int dm(H,\mathbf{x}) P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))P(m(H,\mathbf{x})|\mathbf{x})$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(m^{\mathrm{ex}}(H) - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle \mathbf{s}_i\rangle_{H,\mathbf{x}}\right|\right)^2\right]$$

$m^{\mathrm{ex}}(H)$ where $P(m^{\mathrm{ex}}(H)|\mathbf{x})$ is maximize. → observed magnetization

# Bayes modeling

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\qquad$ $P(m^{\rm ex}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters}

$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$

magnetization

$m(H,\mathbf{x})$

observed magnetization

$m^{\rm ex}(H)$

$$P(\mathbf{x}|m^{\rm ex}(H)) = \frac{P(m^{\rm ex}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\rm ex}(H))}$$

## Bayes modeling

$\mathbf{x}$ where $P(\mathbf{x}|m^{\rm ex}(H))$ is maximize. → plausible model parameters

# *Prior distribution - Bayes modeling -*

**Bayes modeling**

$$P(\mathbf{x}|m^{\mathrm{ex}}(H)) = \frac{P(m^{\mathrm{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\mathrm{ex}}(H))}$$

$\mathbf{x} =$ {model parameters}

$$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$$

→ magnetization $m(H, \mathbf{x})$ → observed magnetization $m^{\mathrm{ex}}(H)$

$P(\mathbf{x})$ : Prior distribution (prior knowledge about model parameters)
(事前分布)

- If prior knowledge does not exist, $P(\mathbf{x}) \propto \mathrm{const.}$
- If $\mathbf{x}$ is sparse (number of model parameters is small),

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right)$$

$\lambda$ : amplitude of regularization (hyperparameter)

$K$ : number of model parameters

# *Posterior distribution - Bayes modeling -*

**Bayes modeling**

$$P(\mathbf{x}|m^{\mathrm{ex}}(H)) = \frac{P(m^{\mathrm{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\mathrm{ex}}(H))}$$

$\mathbf{x} =$ {model parameters}

$$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$$

→ magnetization $m(H, \mathbf{x})$ → observed magnetization $m^{\mathrm{ex}}(H)$

We assume that each magnetization is independently obtained in magnetization curve.

Assumption：$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l=1,\cdots,L}) = \prod_{l=1}^{L} P(\mathbf{x}|m^{\mathrm{ex}}(H_l))$

**Posterior distribution**

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l=1,\cdots,L}) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{l=1}^{L}\left(m^{\mathrm{ex}}(H_l) - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle\mathbf{s}_i\rangle_{H_l,\mathbf{x}}\right|\right)^2 - \lambda\sum_{k=1}^{K}|x_k|\right]$$

observed magnetization curve

# *How to determine hyperparameter*

**L1 regularization**

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right)$$

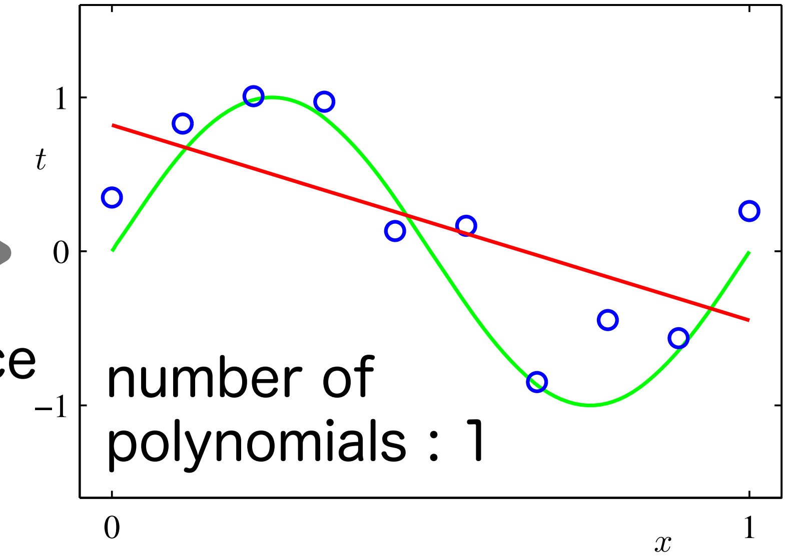
→

Large $\lambda$ : # of model parameters becomes small.

Small $\lambda$ : # of model parameters becomes large.

For large $\lambda$ case,
the observed magnetization curve
will not be fitted.

⟷ Correspondence




from “pattern recognition and machine learning”

# How to determine hyperparameter

**L1 regularization**

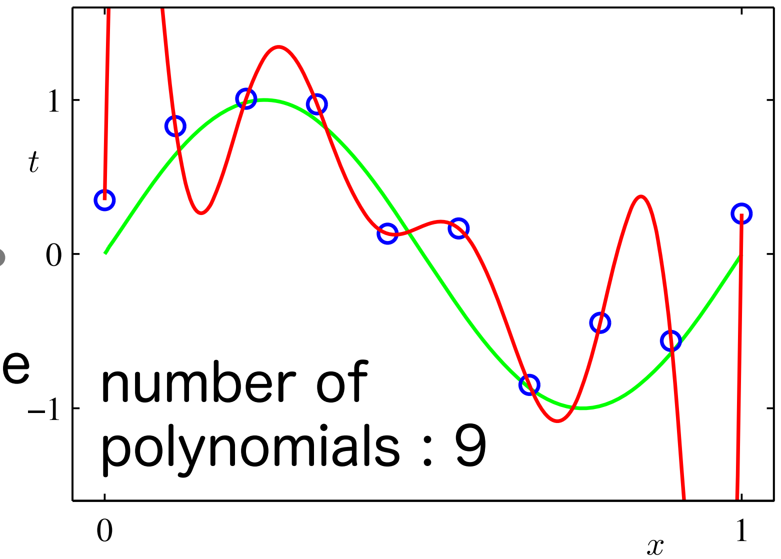
$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right)$$

Large $\lambda$ : # of model parameters becomes small.

Small $\lambda$ : # of model parameters becomes large.

For small $\lambda$ case,
the observed magnetization curve
is well fitted.

Correspondence




*Overfitting will be observed…*

# *How to determine hyperparameter*

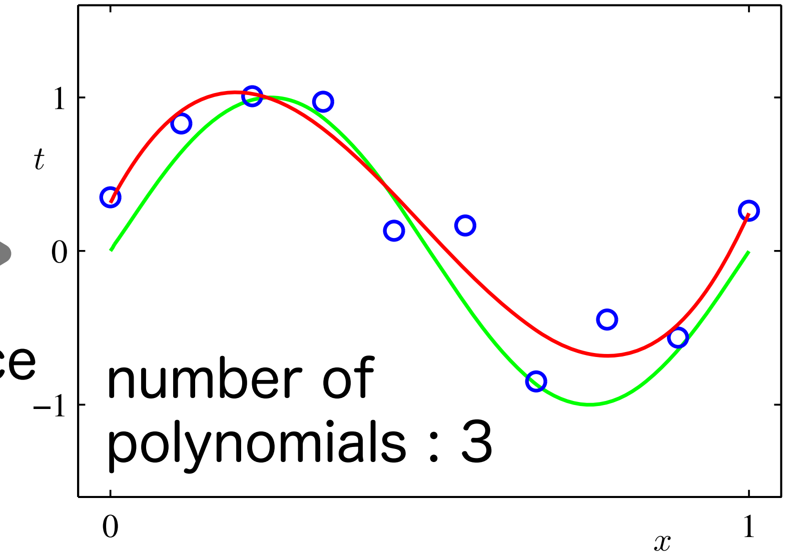
**L1 regularization**

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right)$$

→ Large $\lambda$ : # of model parameters becomes small.
Small $\lambda$ : # of model parameters becomes large.

For plausible $\lambda$ case,
the observed magnetization curve
is correctly fitted.

⟷ Correspondence



# *How to determine hyperparameter*

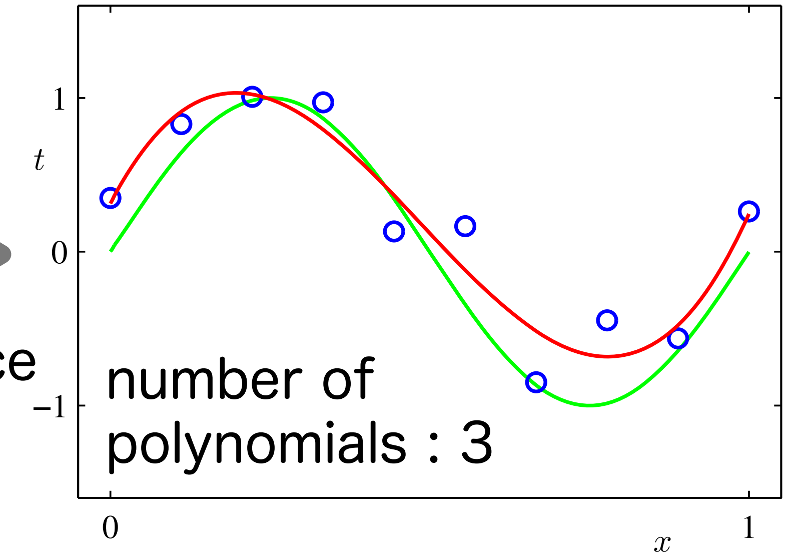
**L1 regularization**

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right)$$

→ Large $\lambda$ : # of model parameters becomes small.
Small $\lambda$ : # of model parameters becomes large.

For plausible $\lambda$ case,
the observed magnetization curve
is correctly fitted.

↔ Correspondence




**Materials scientist want to know the minimum model.**

↔ Correspondence

**To prevent the overfitting**

We prevent the overfitting. → **We determine $\lambda$ so that the prediction error is minimized.**

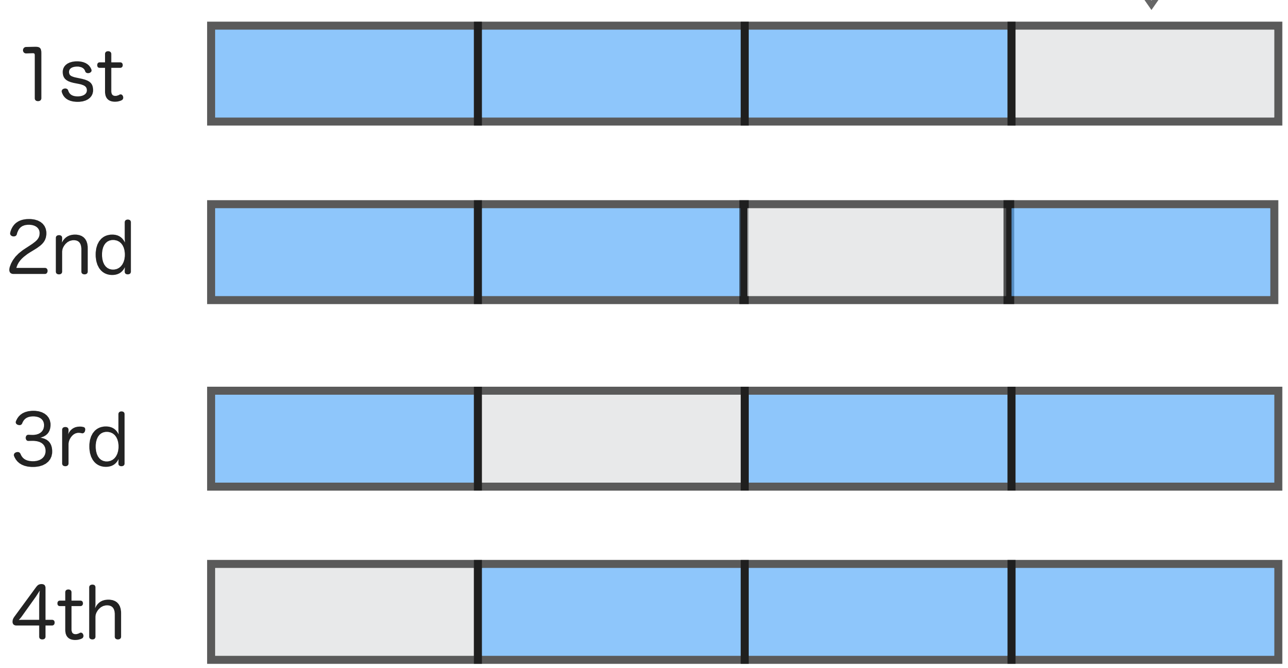
# *Cross validation*

To calculate the prediction error → We divide data into training data and test data.

e.g. We divide the data into 4 groups.

Data used in estimation of model parameters $\mathbf{x}^*$ (Training data)

Data used in validation for predicted performance (Test data)




Error between test data and estimated magnetization

$$\Delta(\lambda) := \frac{4}{L}\sum_{l'=1}^{L/4}\left(m_{l'}^{\mathrm{ex}} - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle \mathbf{s}_i\rangle_{H_{l'},\mathbf{x}^*}\right|\right)^2$$

**We use an average of $\Delta(\lambda)$ of 4 times as the prediction error.**

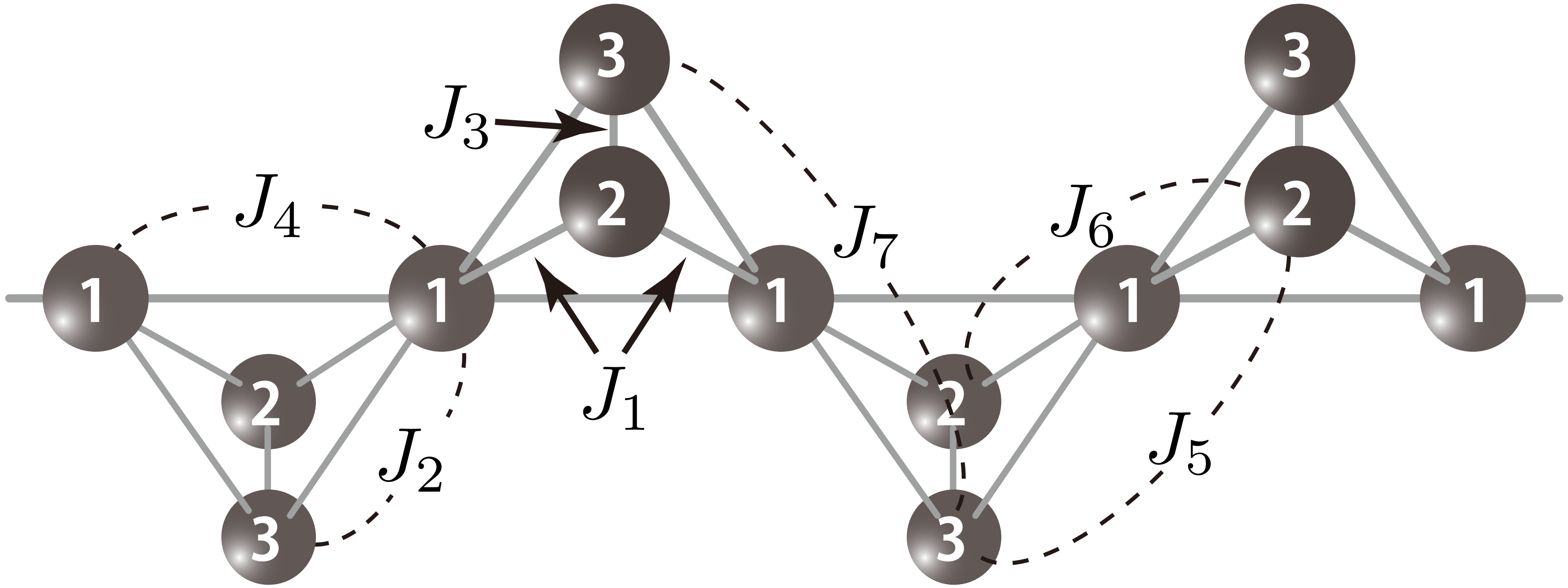
# *Validation by theoretical model*

Classical Heisenberg model with biquadratic interactions
(magnetization plateau is appeared)

$$\mathcal{H} = \sum_{\langle i,j \rangle} J_{ij} \left[ \mathbf{s}_i \cdot \mathbf{s}_j - b_{ij} (\mathbf{s}_i \cdot \mathbf{s}_j)^2 \right] - H \sum_i s_i^z \qquad b_{ij} = bJ_{ij}$$

$\mathbf{s}_i$ : Classical Heisenberg spin (S=1/2)

Crystal structure




Type of interactions

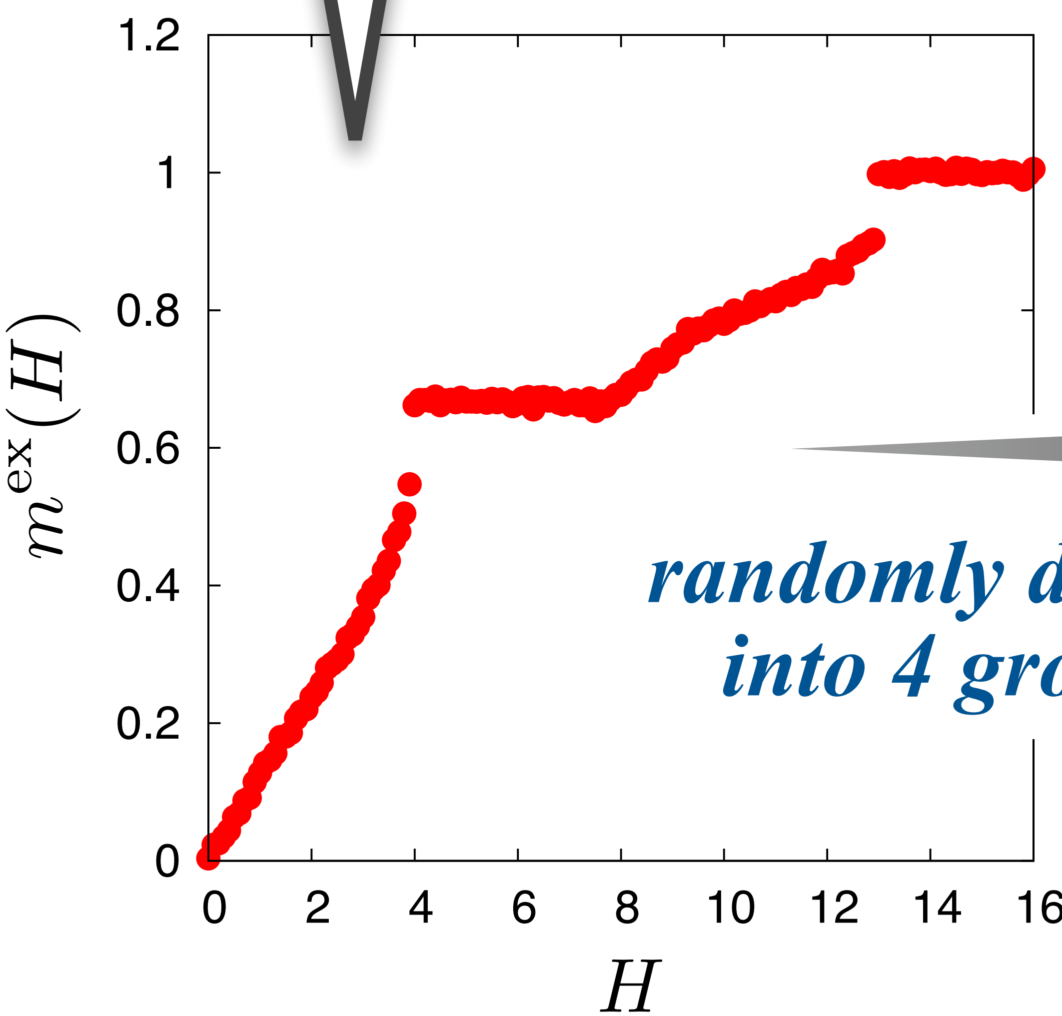
| | number | distance |
|---|---|---|
| $J_1$ | $n_1 = 2$ | $r_1 = 1$ |
| $J_2$ | $n_2 = 2$ | $r_2 = 1$ |
| $J_3$ | $n_3 = 1$ | $r_3 = 1$ |
| $J_4$ | $n_4 = 1$ | $r_4 = 1$ |
| $J_5$ | $n_5 = 2$ | $r_5 = \sqrt{3}$ |
| $J_6$ | $n_6 = 1$ | $r_6 = 2$ |
| $J_7$ | $n_7 = 1$ | $r_7 = 2$ |

model parameters : $\mathbf{x} = \{J_1, J_2, J_3, J_4, J_5, J_6, J_7, b\}$

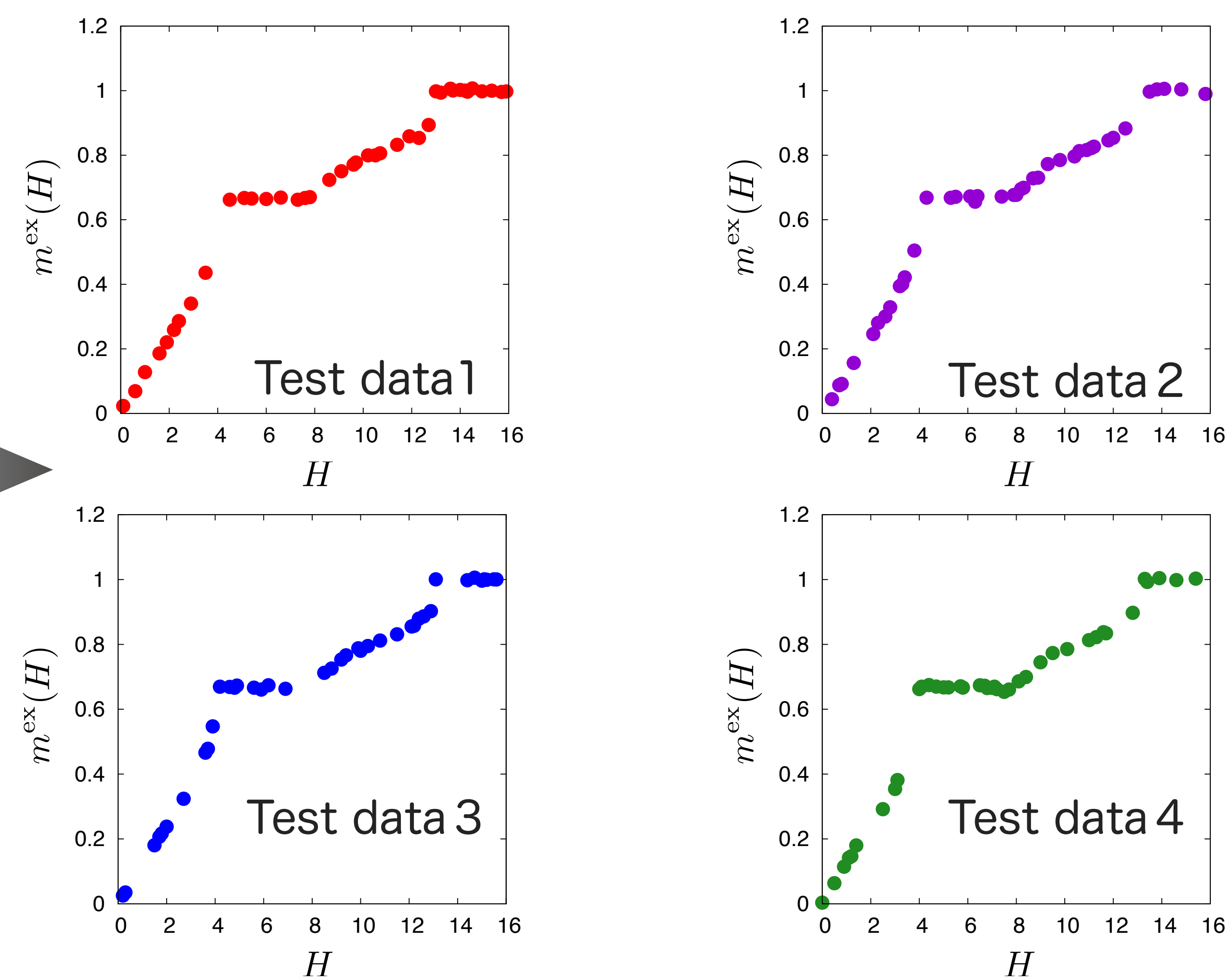
# *Inputted observed magnetization*

Zero temperature magnetization curve for $J_1 = 1, J_2 = 4, J_3 = 5, J_4 = 6, b = 0.1$, $J_5 = J_6 = J_7 = 0$ + Gaussian noise

**Magnetization is calculated by the steepest descent method.**




***randomly divided into 4 groups***



# *Simulation methods*

We search the maximizer of the posterior distribution
by Markov chain Monte Carlo method and exchange method.

Energy function for MCMC

$$E(\mathbf{x}|\lambda,\sigma,K) = \frac{1}{2\sigma^2}\sum_{l=1}^{L}\left(m^{\mathrm{ex}}(H_l) - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle \mathbf{s}_i\rangle_{H_l,\mathbf{x}}\right|\right)^2 + \lambda\sum_{k=1}^{K}|x_k|$$

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l=1,\cdots,L}) \propto \exp\left[-E(\mathbf{x}|\lambda,\sigma,K)\right]$$

→ ***Boltzmann distribution !***

Transition probability for Markov chain $(\mathbf{x} \to \mathbf{x}')$

$$\min\left\{1, \exp[-(E(\mathbf{x}'|\lambda,\sigma,K) - E(\mathbf{x}|\lambda,\sigma,K))]\right\}$$

Dynamical variables in this MC simulation are the model parameters.

# Simulation methods

We search the maximizer of the posterior distribution by Markov chain Monte Carlo method and exchange method.

Introduction of virtual temperature

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l=1,\cdots,L}) \propto \exp\left[-\frac{1}{T}E(\mathbf{x}|\lambda,\sigma,K)\right]$$

Exchange probability between replicas

$$\min\left\{1, \exp\left[(E(\mathbf{x}_i|\lambda,1,K) - E(\mathbf{x}_j|\lambda,1,K))\left(\frac{1}{T_i}-\frac{1}{T_j}\right)\right]\right\}$$

- Monte Carlo steps to update the model parameters was $10^4$.
- 20 replicas with virtual temperatures were prepared between 0.001 and 10.

# *Estimated model parameters*

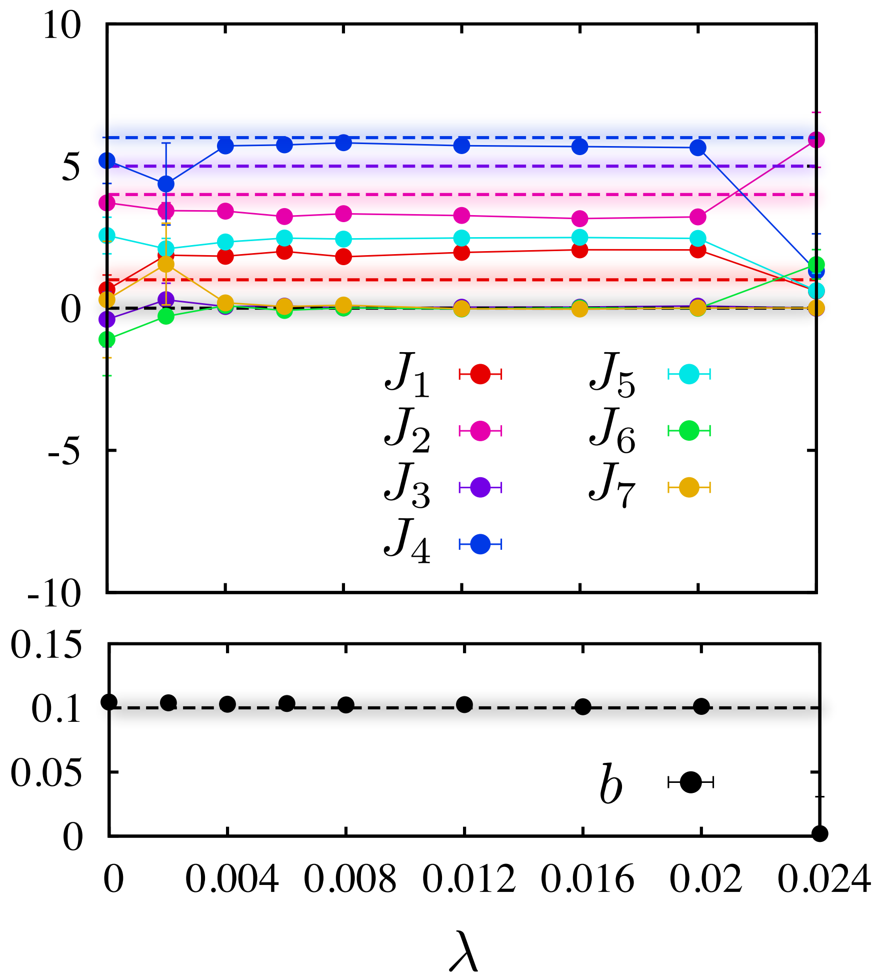
Type of prior distribution

**Searching by Monte Carlo simulation**

Type I

$$P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7}|J_k| + |b|\right)\right]$$

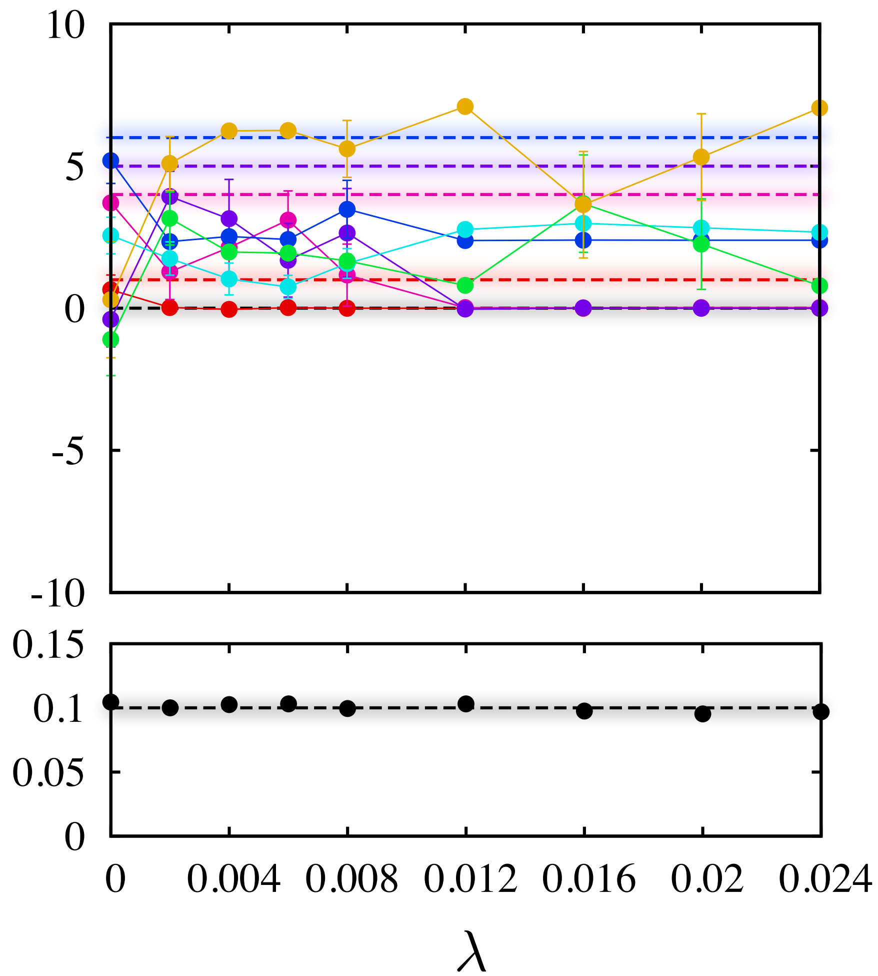
L1 regularization




Type II

$$P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7}|n_k J_k| + |b|\right)\right]$$

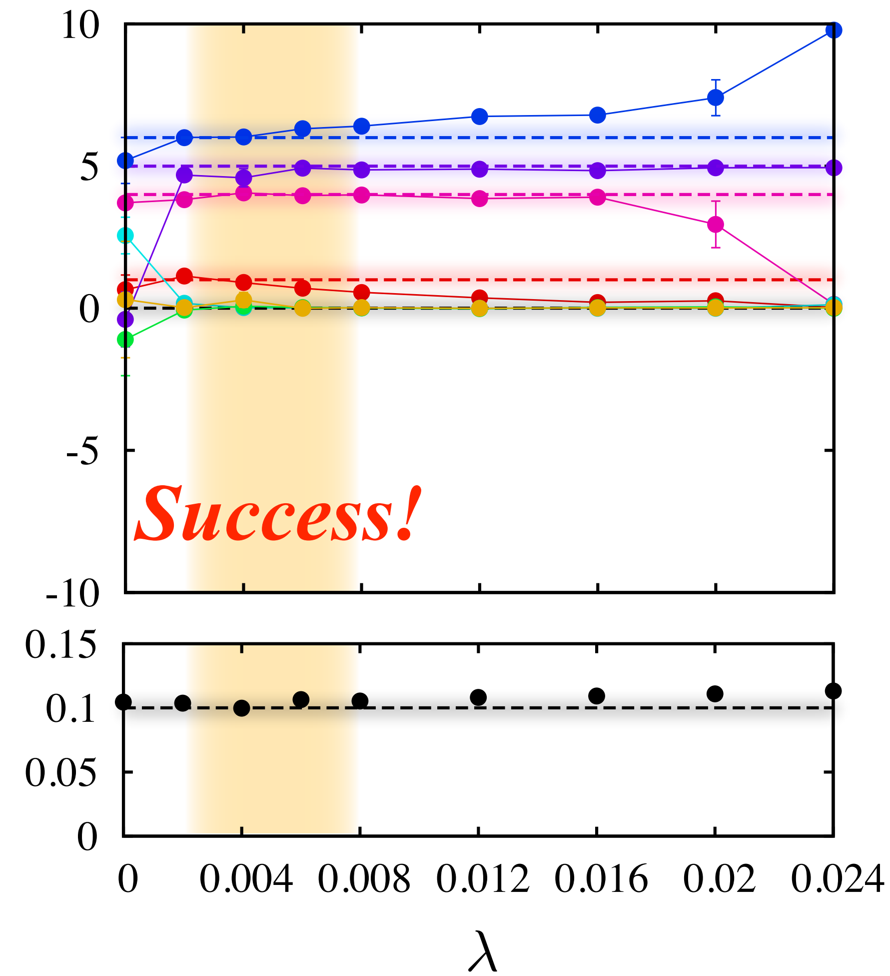
Including # of interactions




Type III

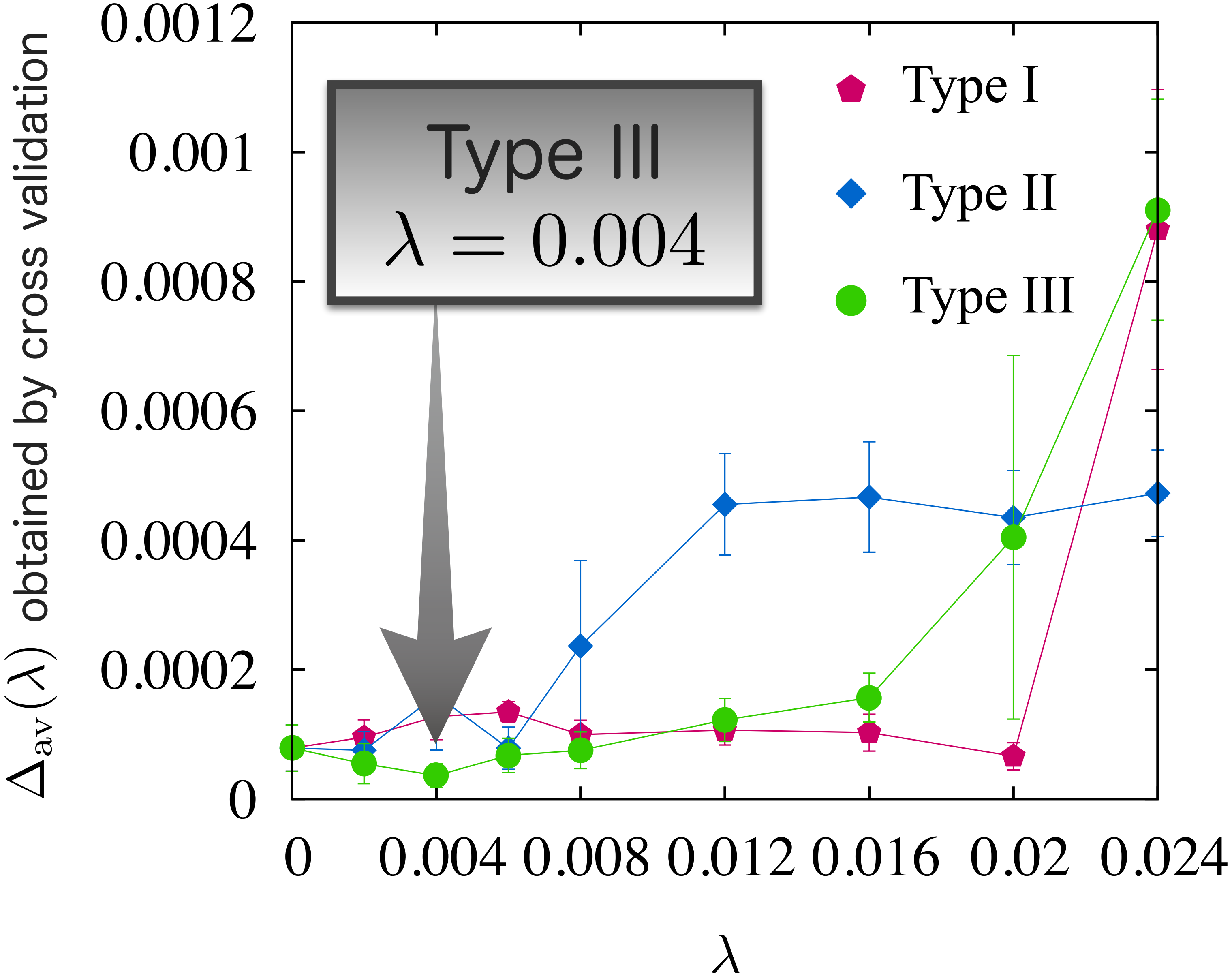
$$P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7}|n_k r_k J_k| + |b|\right)\right]$$

Including # and distance of interactions



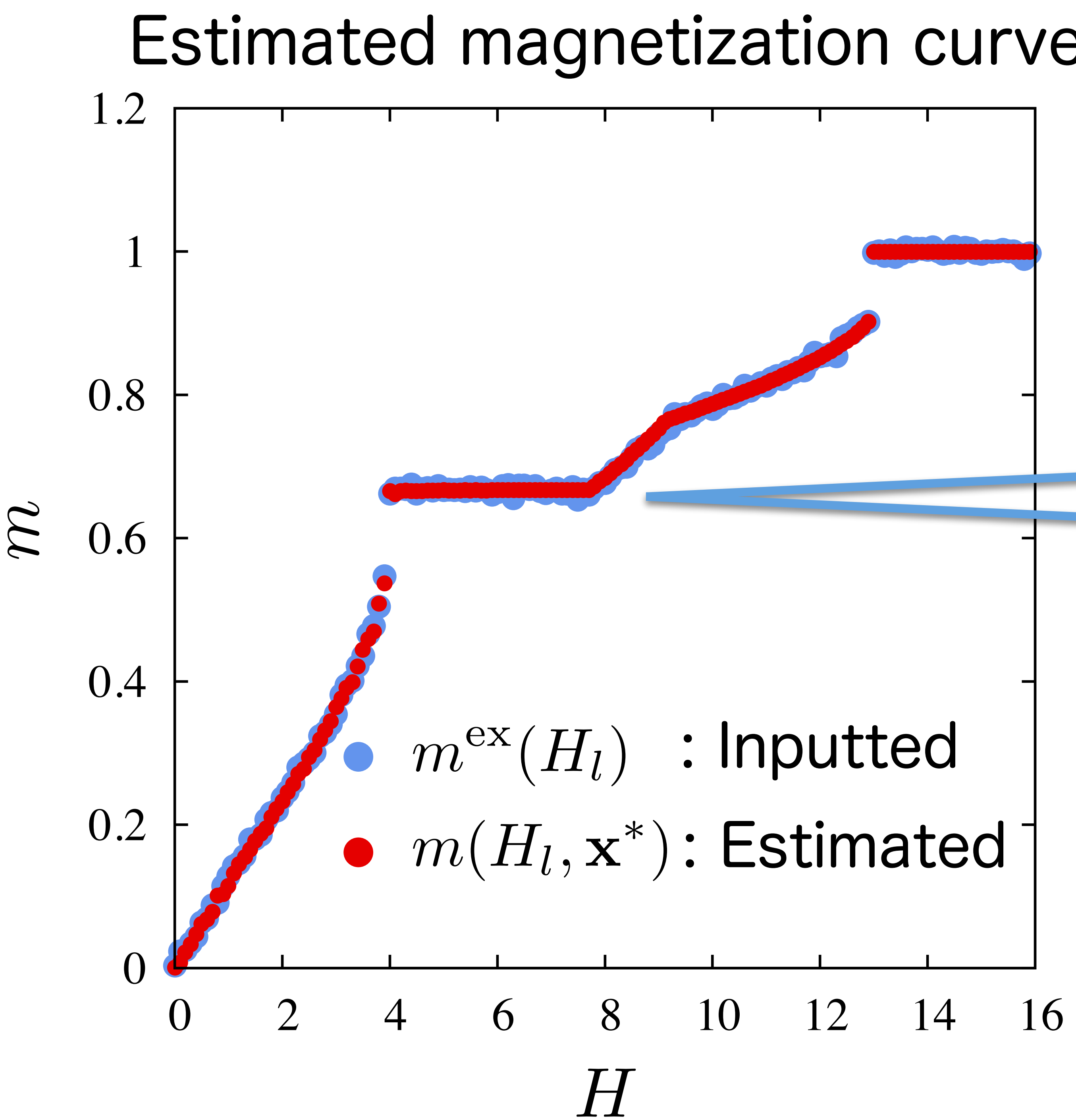
# *Prediction errors*

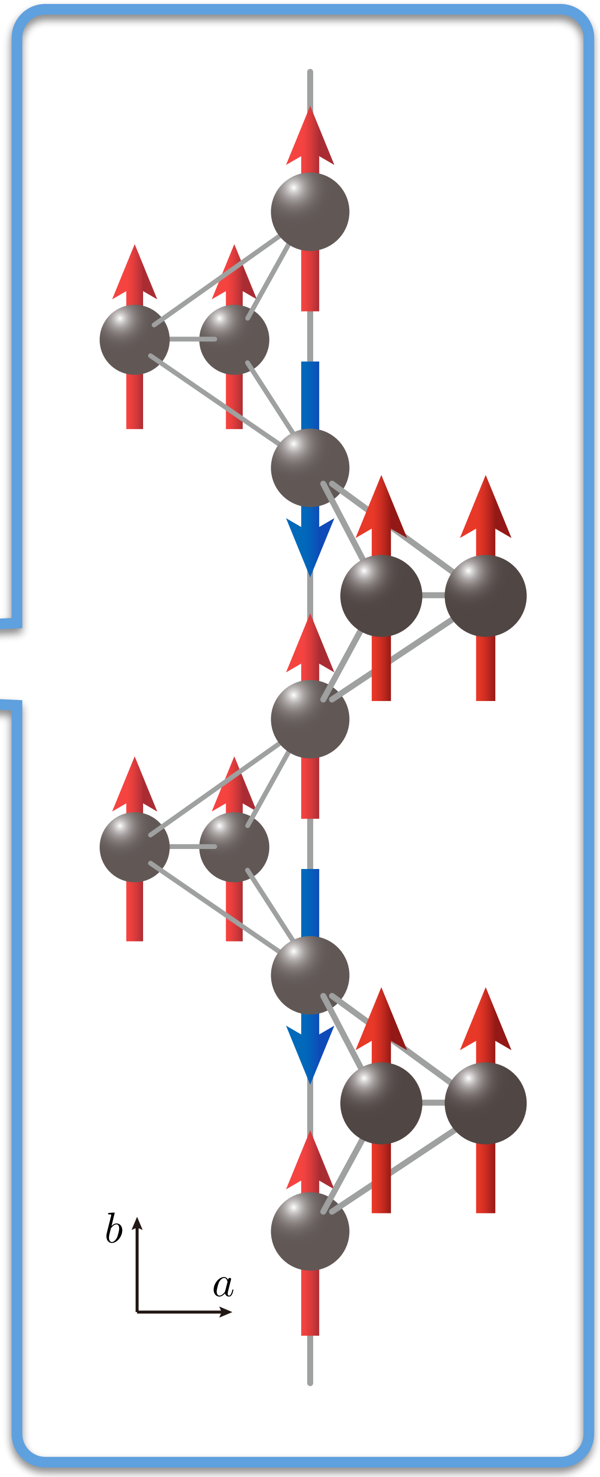
Prediction error depending on $\lambda$




| | Estimated interactions | Inputted interactions |
|---|---|---|
| $J_1$ | 1.074 | 1.000 |
| $J_2$ | 3.850 | 4.000 |
| $J_3$ | 5.012 | 5.000 |
| $J_4$ | 6.051 | 6.000 |
| $J_5$ | 0.011 | 0.000 |
| $J_6$ | -0.051 | 0.000 |
| $J_7$ | 0.002 | 0.000 |
| $b$ | 0.102 | 0.100 |

# Prediction errors



# *Effective model estimation method*

Posterior distribution by Bayesian statistics

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l=1,\cdots,L}) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{l=1}^{L}\left(m^{\mathrm{ex}}(H_l) - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle\mathbf{s}_i\rangle_{H_l,\mathbf{x}}\right|\right)^2 - \lambda\sum_{k=1}^{K}|x_k|\right]$$

Least square mean
between calculated data and inputted data

**Regularization**
**&**
**Prediction error by cross validation**

**We get plausible effective model for experimental results.**
**(selection of important model parameters)**

R. Tamura and K. Hukushima, Phys. Rev. B **95**, 064407 (2017).

*Thank you !!*

Nano Revolution for the Future

World Premier International Research Center Initiative

R. Tamura and K. Hukushima, Phys. Rev. B **95**, 064407 (2017).