2010.11.4-6　第13回情報論的学習理論ワークショップ（IBIS2010）

1分子計測データに対するパラメータ推定

Parameter estimation on single-molecule time series

京都大学理学研究科 物理学専攻

宮崎　牧人

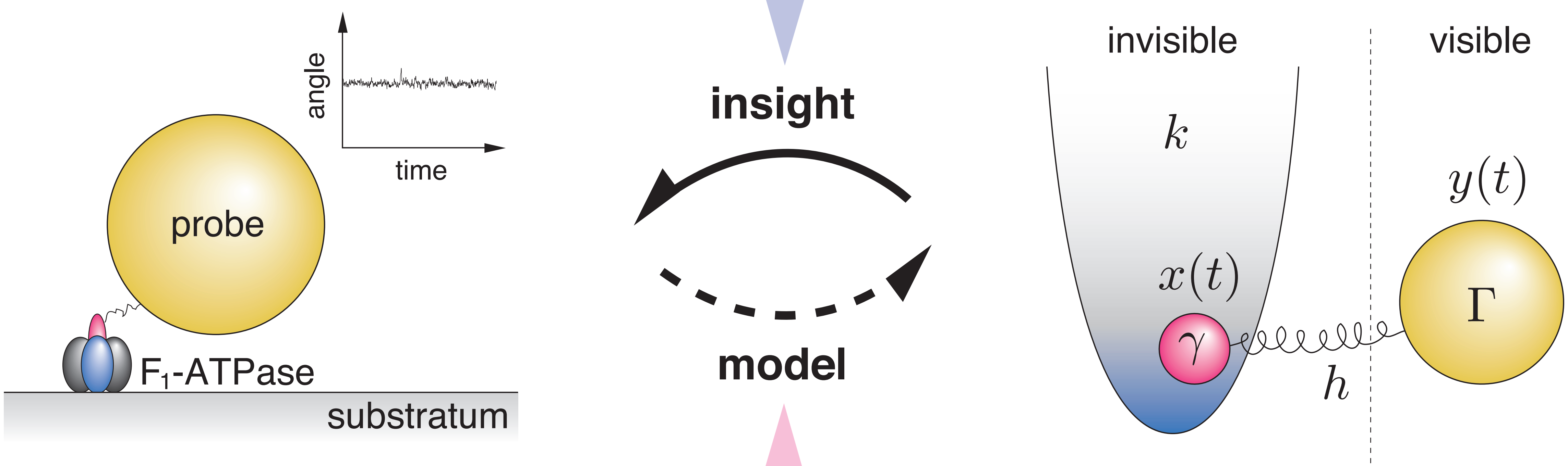
Dept of Physics, Kyoto Univ
Makito Miyazaki

Introduction

progress in statistical physics of non-equilibrium small systems

ex.) Jarzynski equality, fluctuation theorem, Hatano-Sasa equality, Harada-Sasa equality, ...

but ... the model parameters should be determined.



development in a single-molecule observation & manipulation technique

ex.) optical microscope, FRET, TIRF, optical tweezer, AFM, ...

but ... observable degrees of freedom are limited.

problem: there are no general frameworks to derive system information from restricted measurements.

General framework

invisible degrees of freedom : $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$

visible degrees of freedom : $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$

system parameter : $\boldsymbol{\Pi} = (\Pi_1, \Pi_2, \ldots, \Pi_p)$

trajectory of $\boldsymbol{x}$ and $\boldsymbol{y}$ are denoted as $[\boldsymbol{x}]$ and $[\boldsymbol{y}]$

$*$: true value, $\wedge$: estimated value

problem : How can we estimate the values of $\boldsymbol{\Pi}$ on the bases of $[\boldsymbol{y}]$ WITHOUT referring to $[\boldsymbol{x}]$?

simple idea : maximize $P(\boldsymbol{\Pi}|[\boldsymbol{y}])$ with respect to $\boldsymbol{\Pi}$

Are there any general evidence that the idea actually works ?

General framework

simple idea : maximize $P(\boldsymbol{\Pi}|[\boldsymbol{y}])$ with respect to $\boldsymbol{\Pi}$

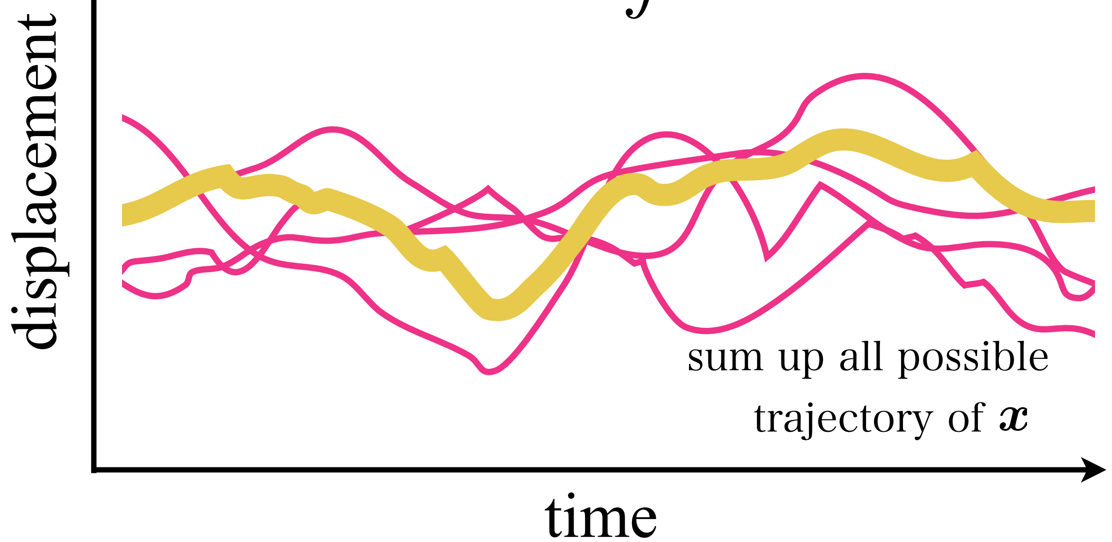
$$P(\boldsymbol{\Pi}|[\boldsymbol{y}]) = \frac{P([\boldsymbol{y}]|\boldsymbol{\Pi})P(\boldsymbol{\Pi})}{P([\boldsymbol{y}])}$$ **(Bayes theorem)**

$\downarrow$ $P(\boldsymbol{\Pi}) = const.$ (noninformative prior)

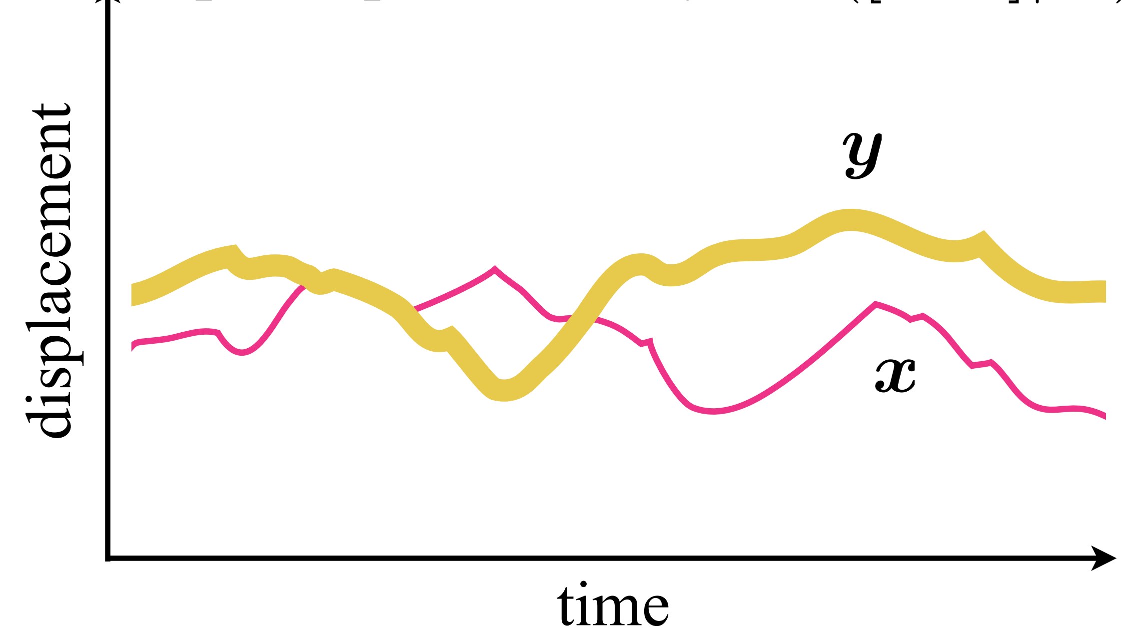
$$P(\boldsymbol{\Pi}|[\boldsymbol{y}]) \propto P([\boldsymbol{y}]|\boldsymbol{\Pi})$$ **(marginal likelihood)**

marginal likelihood :

$$P([\boldsymbol{y}]|\boldsymbol{\Pi}) \equiv \int \mathcal{D}x \; P([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Pi})$$



path probability : $P([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Pi})$



General framework

The maximization of $P(\boldsymbol{\Pi}|[\boldsymbol{y}])$ is replaced with the minimization of "Hamiltonian" defined as $\mathcal{H}_\tau(\boldsymbol{\Pi})[\boldsymbol{y}] \equiv -\tau^{-1} \ln P([\boldsymbol{y}]|\boldsymbol{\Pi})$

general proof :

In the limit of $\tau \longrightarrow \infty$, "Hamiltonian" has a minimal at $\boldsymbol{\Pi} = \boldsymbol{\Pi}^*$.

$$\left.\frac{\partial \mathcal{H}_\infty(\boldsymbol{\Pi})}{\partial \boldsymbol{\Pi}}\right|_{\boldsymbol{\Pi}=\boldsymbol{\Pi}^*} = \mathbf{0}.$$

$$(H_{ij} \equiv) \quad \left.\frac{\partial^2 \mathcal{H}_\infty(\boldsymbol{\Pi})}{\partial \Pi_i \partial \Pi_j}\right|_{\boldsymbol{\Pi}=\boldsymbol{\Pi}^*} = \left\langle \frac{\partial \ln P([y]|\boldsymbol{\Pi}^*)}{\partial \Pi_i^*} \frac{\partial \ln P([y]|\boldsymbol{\Pi}^*)}{\partial \Pi_j^*} \right\rangle_{\boldsymbol{\Pi}^*}$$

(non-negative definite)

... but in practice, the idea should work with FINITE observation data.

General consequence in the case of finite data

general proof :

The error of the estimate decreases proportional to the observation time length.

$$\hat{\mathbf{\Pi}}[y] = \mathbf{\Pi}^* - \varepsilon \left[H^{-1} \cdot \left. \frac{\partial \psi(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\mathbf{\Pi}^*} \right] + O(\varepsilon^2) \qquad (\varepsilon \equiv \tau_0/\tau)$$

$$\psi(\mathbf{\Pi}) \equiv \lim_{\varepsilon \to 0} [\mathcal{H}_\tau(\mathbf{\Pi})[y] - \mathcal{H}_\infty(\mathbf{\Pi})]/\varepsilon$$

General framework

Estimation of the system parameter: $\mathbf{\Pi}$

Minimize "Hamiltonian", $\mathcal{H}_\tau(\mathbf{\Pi})[\boldsymbol{y}] \equiv -\tau^{-1} \ln P([\boldsymbol{y}]|\mathbf{\Pi})$ with respect to $\mathbf{\Pi}$.

We obtained the general consequence that the framework works.

Estimation of the motion of hidden part: $[\boldsymbol{x}]$

Maximize the path probability, $P([\boldsymbol{x}, \boldsymbol{y}]|\mathbf{\Pi})$ with respect to $[\boldsymbol{x}]$.

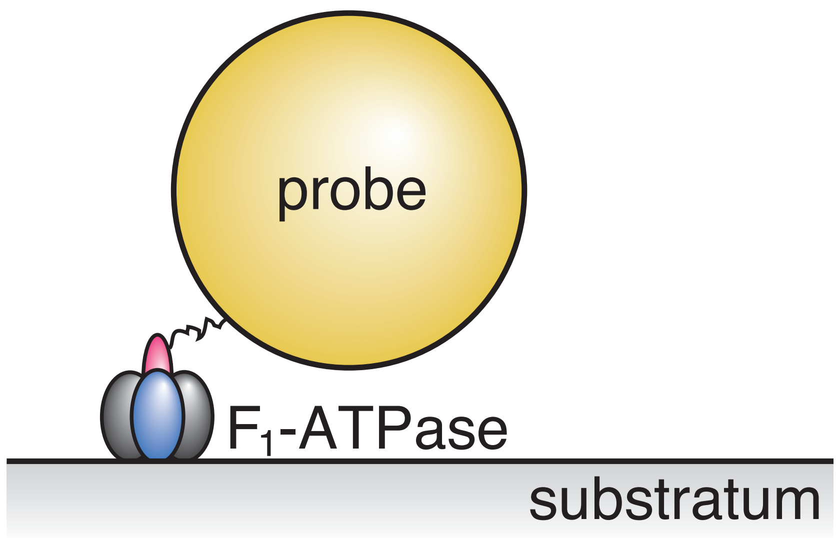
$$\frac{\delta}{\delta x(t)} \ln P([x]|[y]; \hat{\mathbf{\Pi}}) = 0$$ (Euler-Lagrange eq.)

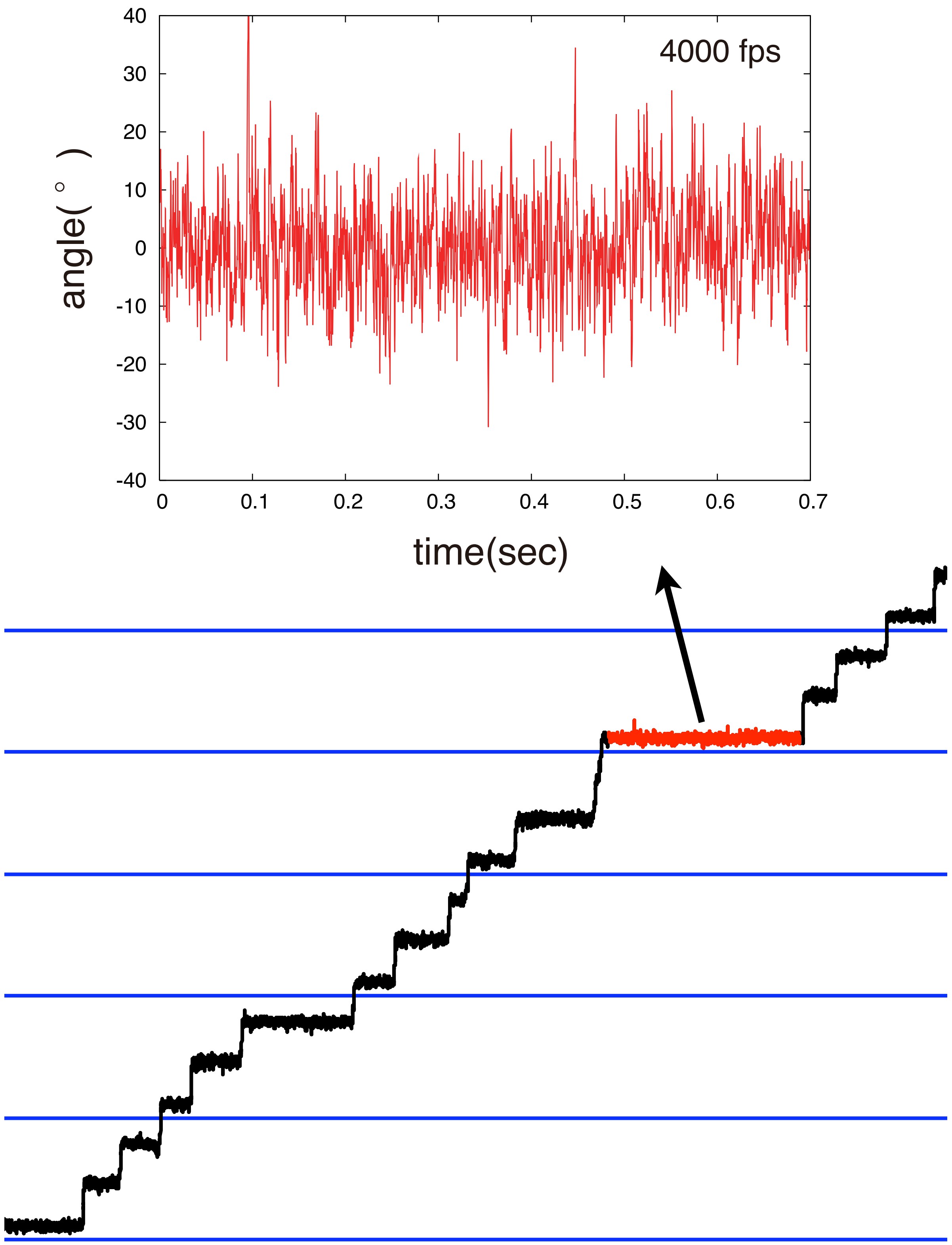
$$P([x]|[y]; \hat{\mathbf{\Pi}}) = \frac{P([x, y]|\hat{\mathbf{\Pi}})}{P([y]|\hat{\mathbf{\Pi}})}$$ (Bayes theorem)

General framework

Bayesian inference	Techniques in physics
calculation of the path probability : $P([\boldsymbol{x}, \boldsymbol{y}] \vert \boldsymbol{\Pi})$	path-integral approach $\cdots \times P(\boldsymbol{x}_t, \boldsymbol{y}_t \vert \boldsymbol{x}_{t-\Delta t}, \boldsymbol{y}_{t-\Delta t}) \times P(\boldsymbol{x}_{t+\Delta t}, \boldsymbol{y}_{t+\Delta t} \vert \boldsymbol{x}_t, \boldsymbol{y}_t) \times \cdots$
calculation of the marginal likelihood : $P([\boldsymbol{y}] \vert \boldsymbol{\Pi}) \equiv \int \mathcal{D}x \; P([\boldsymbol{x}, \boldsymbol{y}] \vert \boldsymbol{\Pi})$	minimization of the action functional $S([\boldsymbol{x}, \boldsymbol{y}] \vert \boldsymbol{\Pi})$ with respect to $[\boldsymbol{x}]$ (solving the Euler-Lagrange eq.) $\frac{\delta}{\delta[\boldsymbol{x}]} S([\boldsymbol{x}, \boldsymbol{y}] \vert \boldsymbol{\Pi}) = 0$ $\longrightarrow$ “classical trajectory” $[\hat{\boldsymbol{x}}]$ WKB (Wentzel-Kramers-Brillouin) approximation around $[\hat{\boldsymbol{x}}]$

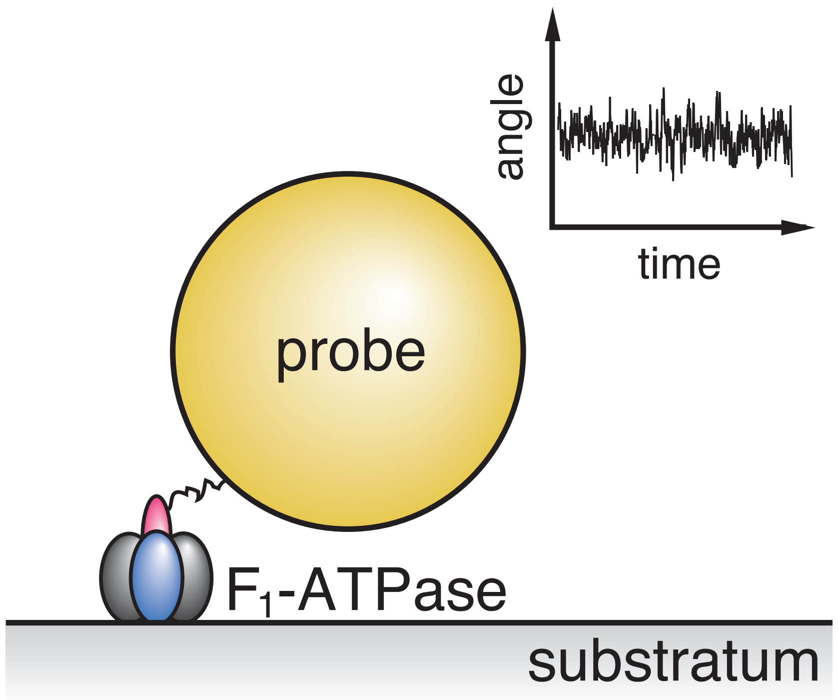
Experimental data of F1-ATPase

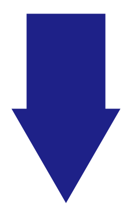




the data given by T. Nishizaka

Model





Estimating parameters: $\mathbf{\Pi}^* \equiv \{k^*, h^*, \gamma^*, \Gamma^*\}$

a) interaction potential between the central subunit (magenta) and the surrounding subunits (blue and black), k^*

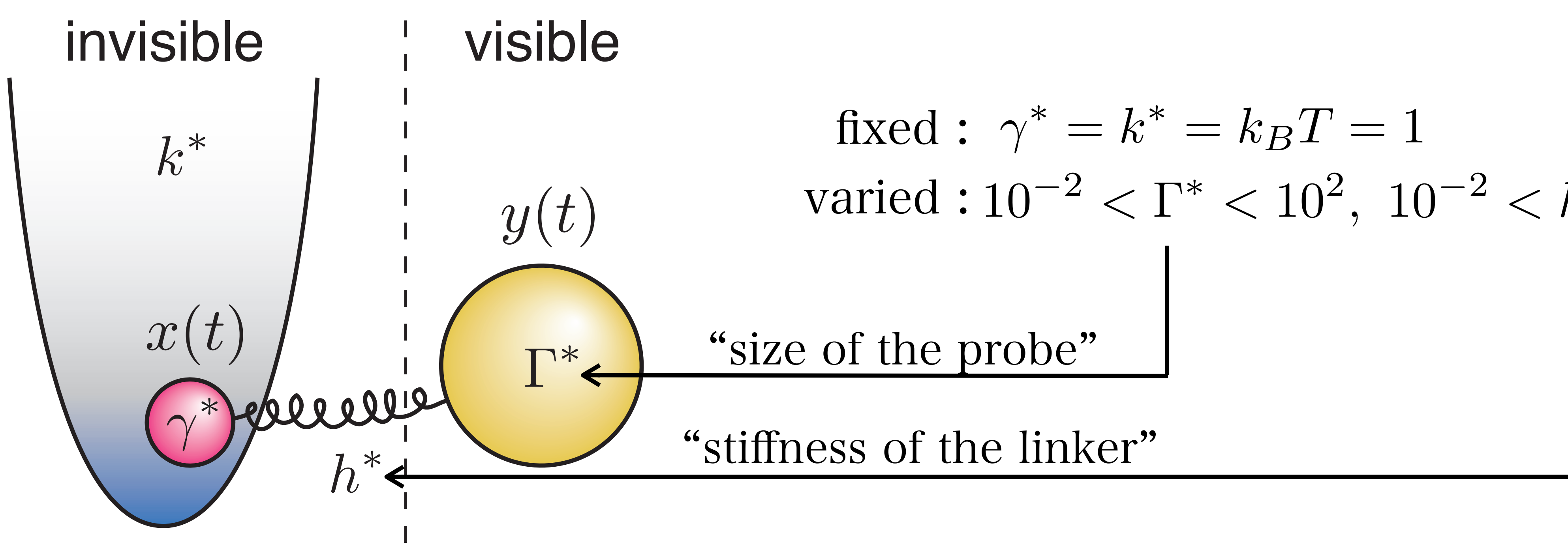
b) friction coefficient for the rotation of the central subunit, γ^*

c) spring constant of the linker, h^*

d) friction coefficient of the probe particle, Γ^*

fixed : $\gamma^* = k^* = k_B T = 1$

varied : $10^{-2} < \Gamma^* < 10^2, \ 10^{-2} < h^* < 10^2$

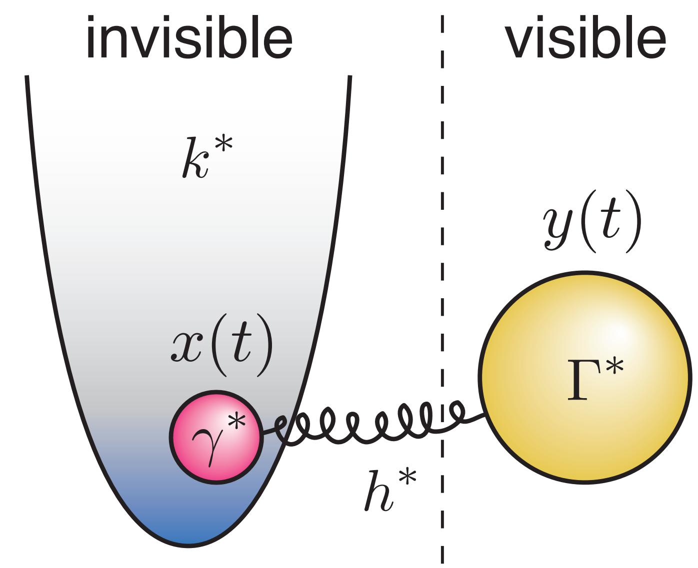


Outline of the numerical experiment

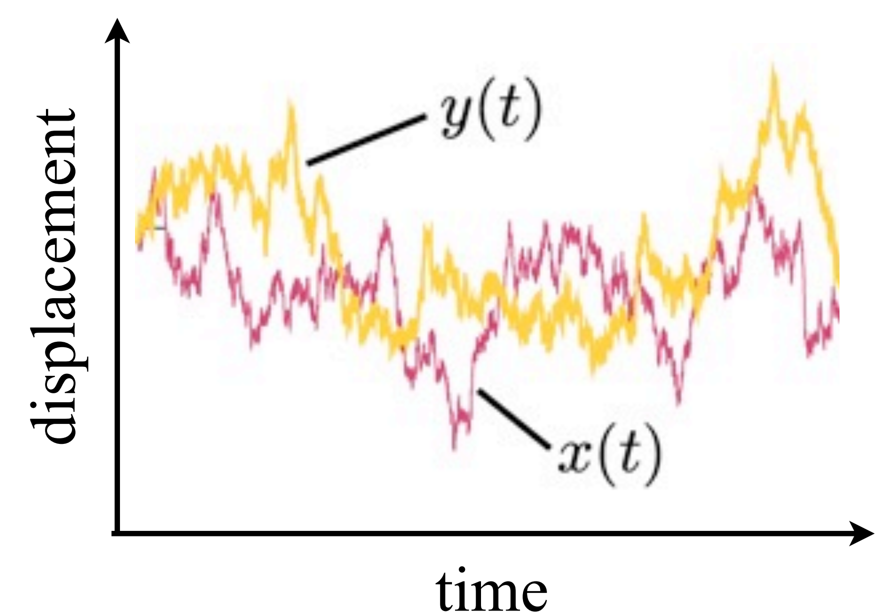
Langevin equations : $\gamma\dot{x} = -kx - h(x-y) + \xi(t)$
$\Gamma\dot{y} = -h(y-x) + \eta(t)$

$$\langle \xi(t)\xi(0)\rangle = 2\gamma k_{\mathrm{B}} T\delta(t)\ ,\ \langle \eta(t)\eta(0)\rangle = 2\Gamma k_{\mathrm{B}} T\delta(t)$$

estimating parameters : $\mathbf{\Pi} = \{k, h, \gamma, \Gamma\}$

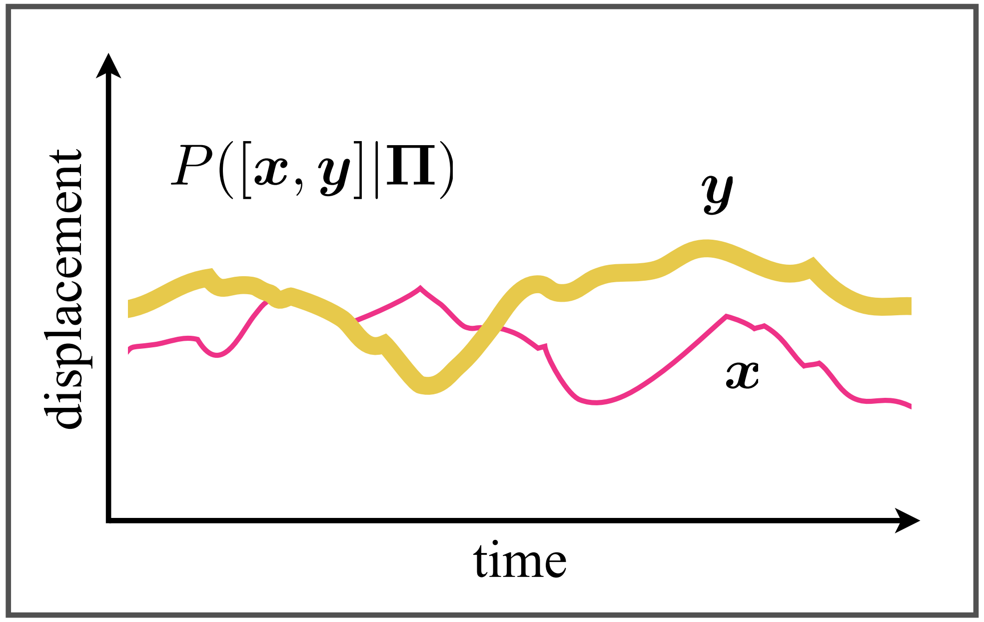


Under the *true* parameter values, trajectories were simulated.



On the basis of $[y]$ alone, $\mathbf{\Pi} = \{k, h, \gamma, \Gamma\}$ were estimated by the minimization of "Hamiltonian", $\mathcal{H}_\tau(\mathbf{\Pi})[y]$

Path probability



$$P([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Pi}) = P_{init}(x_0, y_0|\boldsymbol{\Pi})\, P_{tr}((x_0, y_0) \to [x, y]|\boldsymbol{\Pi})$$

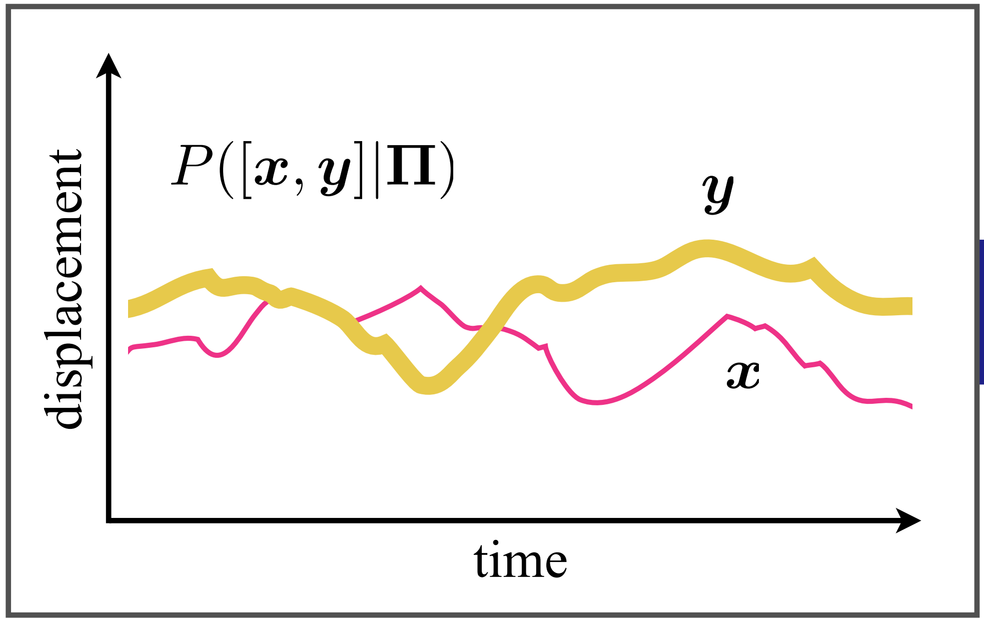
initial distribution : $P_{init}((x_0, y_0)|\boldsymbol{\Pi}) = \dfrac{\sqrt{kh}}{2\pi T} \exp[-\beta H((x_0, y_0); \boldsymbol{\Pi})]$

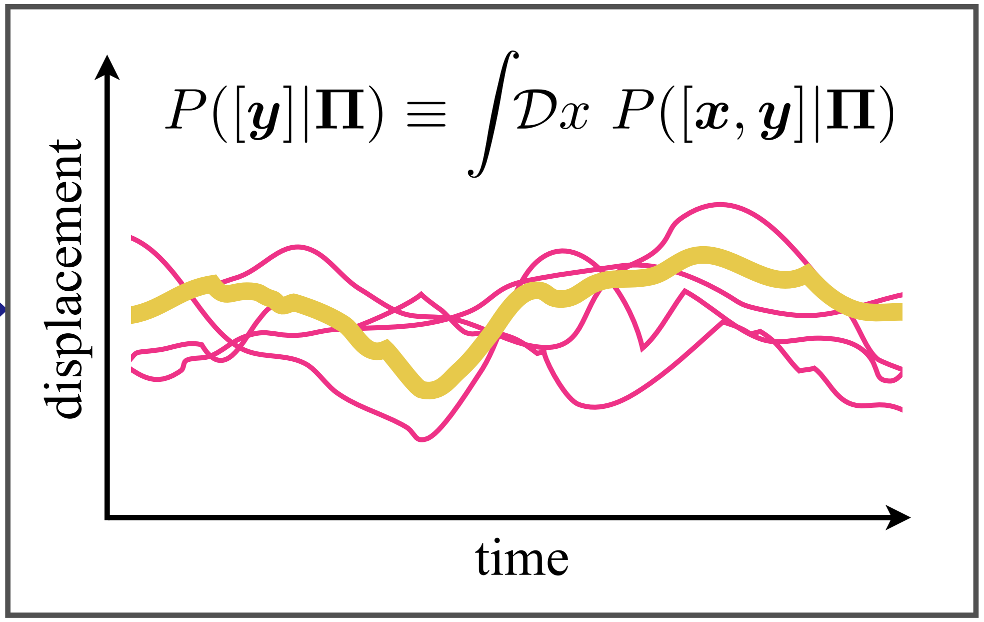
$$H(x, y; \boldsymbol{\Pi}) = \frac{k}{2}x^2 + \frac{h}{2}(x-y)^2$$

transition probability : $P_{tr}((x_0, y_0) \to [x, y]|\boldsymbol{\Pi}) = \left(\dfrac{\sqrt{\gamma\Gamma}}{4\pi T \Delta t}\right)^N \exp[-\beta S([x, y]; \boldsymbol{\Pi})]$

$$S([x, y]; \boldsymbol{\Pi}) \equiv \frac{1}{4\gamma}\int_0^\tau dt\, [\gamma\dot{x} + kx + h(x-y)]^2 + \frac{1}{4\Gamma}\int_0^\tau dt\, [\Gamma\dot{y} + h(y-x)]^2 - \frac{T(k+h)\tau}{2\gamma} - \frac{Th\tau}{2\Gamma}$$

Marginal likelihood





1.) calculate the MAP (Maximum *A Posteriori*) estimator $[\hat{x}]$

Euler-Lagrange equation: $$\frac{\partial}{\partial x_i} \ln P_{init}(x_0, y_0|\boldsymbol{\Pi}) \, P_{tr}((x_0, y_0) \to [x, y]|\boldsymbol{\Pi}) = 0$$

2.) apply WKB method around $[\hat{x}]$

$$\mathcal{H}_\tau(\boldsymbol{\Pi})[y] = \frac{s([\hat{x}, y]; \boldsymbol{\Pi})}{k_\mathrm{B}T} + \frac{\sqrt{\phi^2 + \chi^2}}{2} - \frac{\ln(\Gamma/4\pi k_\mathrm{B}T\Delta t)}{2\Delta t} + \frac{c(\hat{x}_0, y_0; \boldsymbol{\Pi})}{\tau}$$

$$s([\hat{x}, y]; \boldsymbol{\Pi}) \equiv \tau^{-1} S([\hat{x}, y]; \boldsymbol{\Pi})$$

$$\phi^2 \equiv (k+h)^2/\gamma^2, \;\; \chi^2 \equiv h^2/\gamma\Gamma$$

Error & sample averaged variance

optimized solution set $\{\mathbf{\Pi}^{\ddagger}_{1,j}\ , \cdots , \mathbf{\Pi}^{\ddagger}_{i,j}\ , \cdots \mathbf{\Pi}^{\ddagger}_{r,j}\} \ \rightarrow$ the best $\hat{\mathbf{\Pi}}_j$

i : index number of the different initial settings of the optimization process **total → r**
j : index number of the data of $[y]$ **total → s**

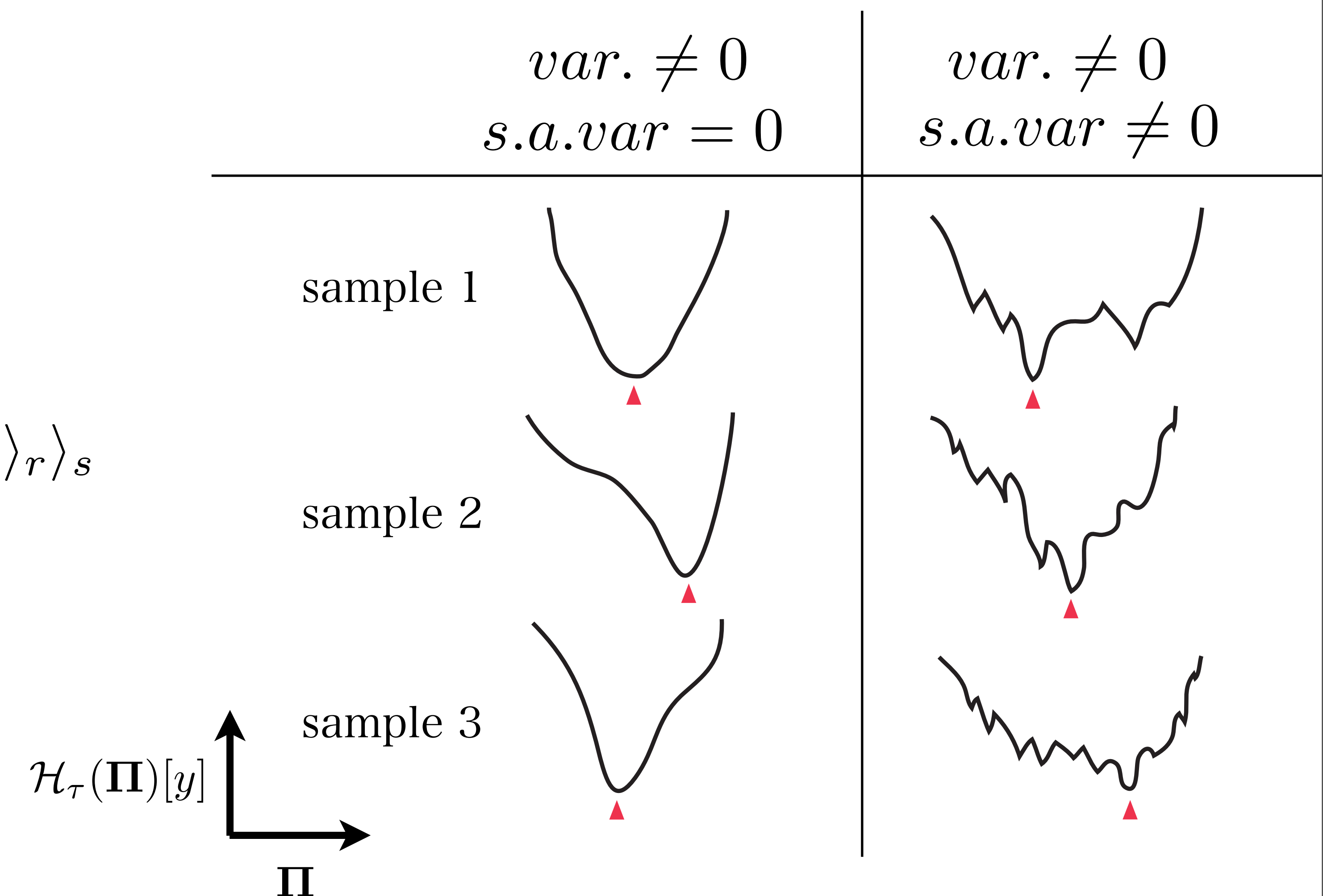
error :

$$err. \ \equiv \ \langle \| \log \hat{\mathbf{\Pi}}_j - \log \mathbf{\Pi}^* \| \rangle_s$$

sample averaged variance :

$$s.a.var. \ \equiv \ \langle\langle \ \| \log \mathbf{\Pi}^{\ddagger}_{i,j} - \langle \log \mathbf{\Pi}^{\ddagger}_{i,j} \rangle_r \|^2 \ \rangle_r \rangle_s$$

→ characteristic parameter of
the roughness of $\mathcal{H}_\tau(\mathbf{\Pi})[y]$

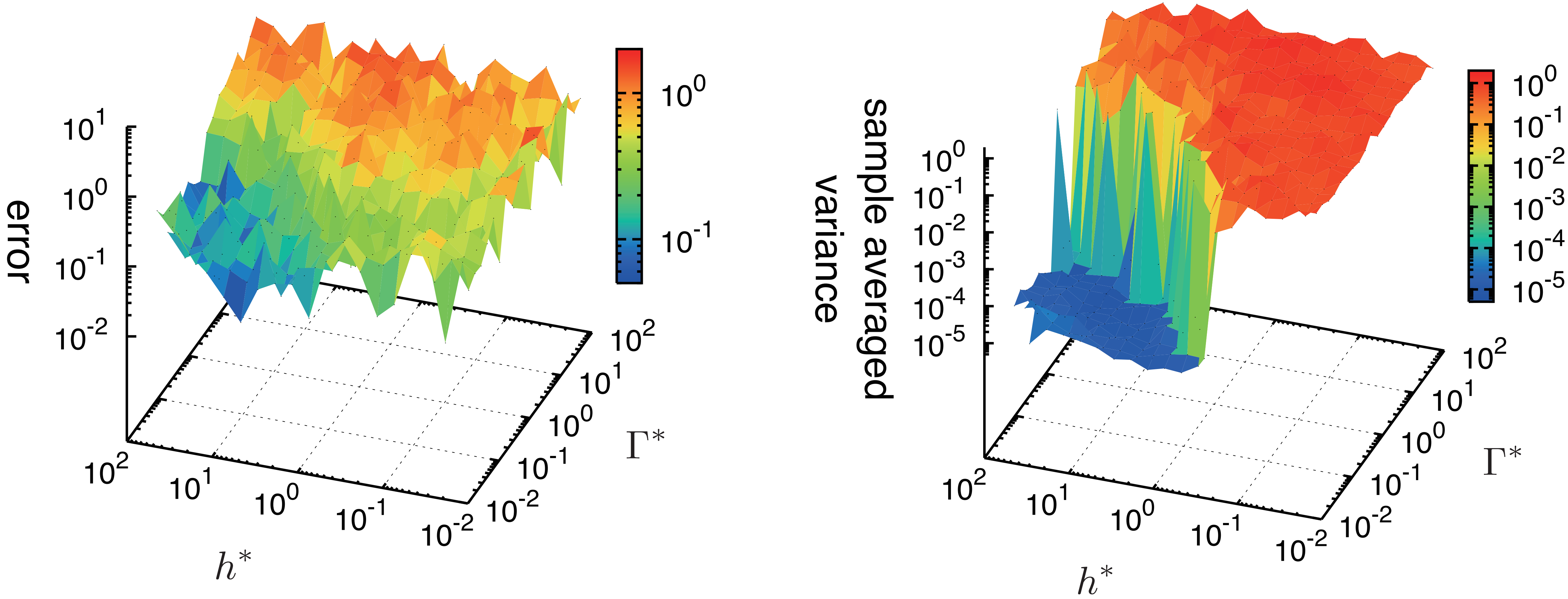


$*$: true value, $\hat{}$: estimated value

$$\langle \cdots \rangle_s \ \equiv \ \frac{1}{s} \Sigma^s_{j=1} \ \cdots$$

$$\langle \cdots \rangle_r \ \equiv \ \frac{1}{r} \Sigma^r_{i=1} \ \cdots$$

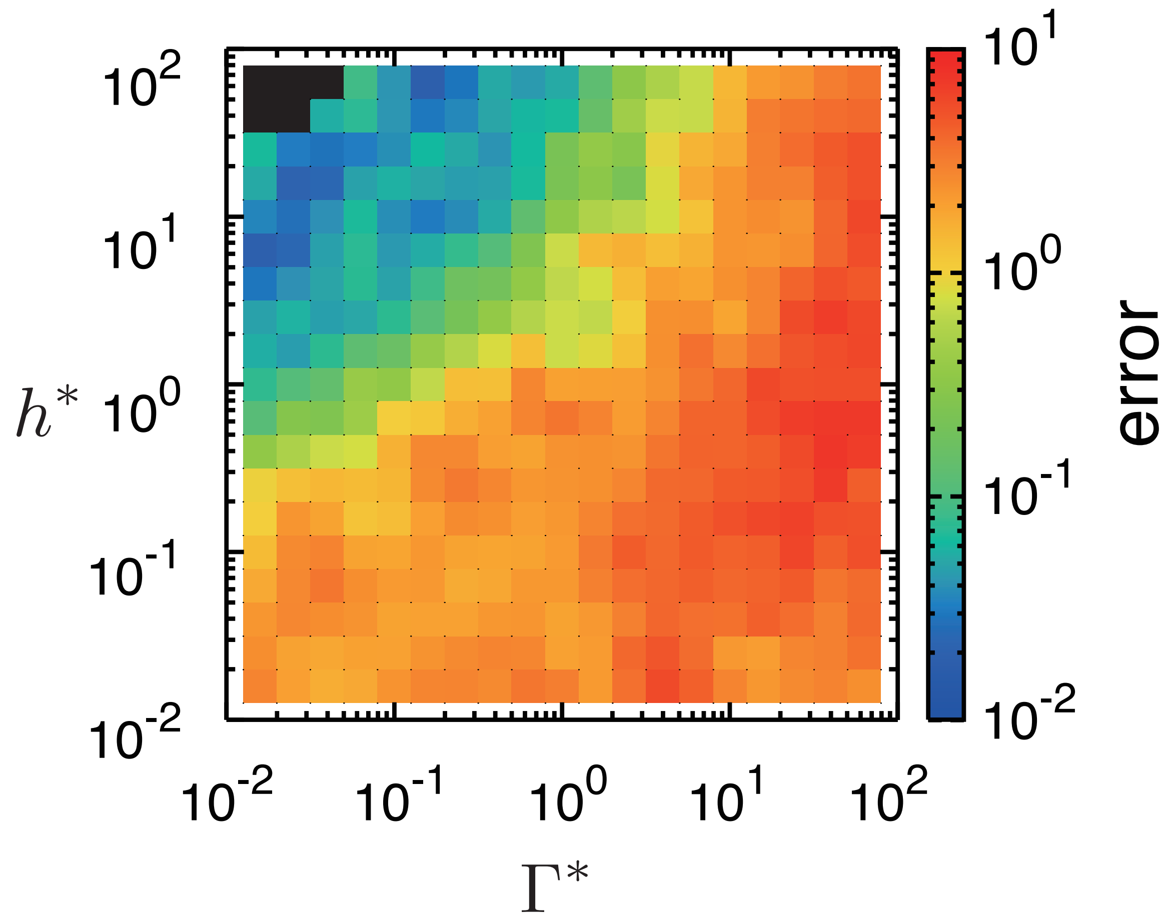
Estimation results

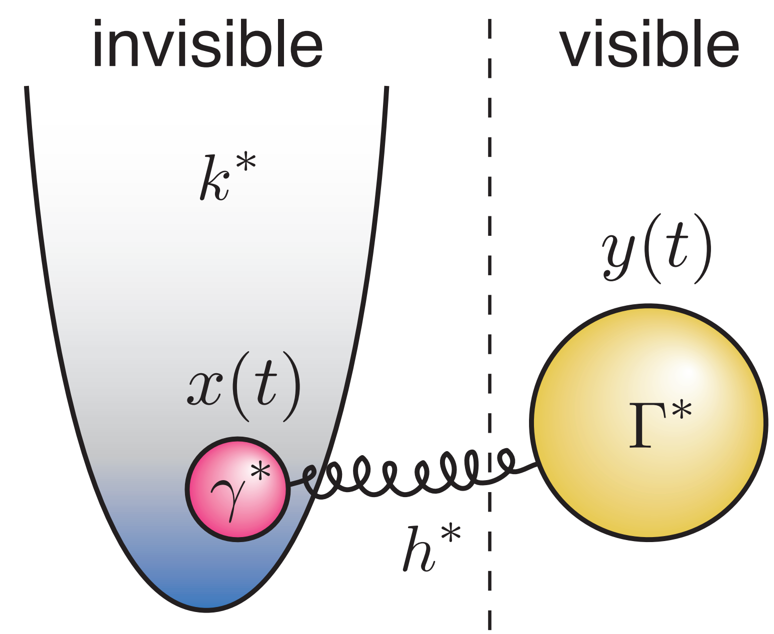


The sample averaged variance is abruptly increase at a certain curve.

Estimation results

The error is small when h^* is large and Γ^* is small.



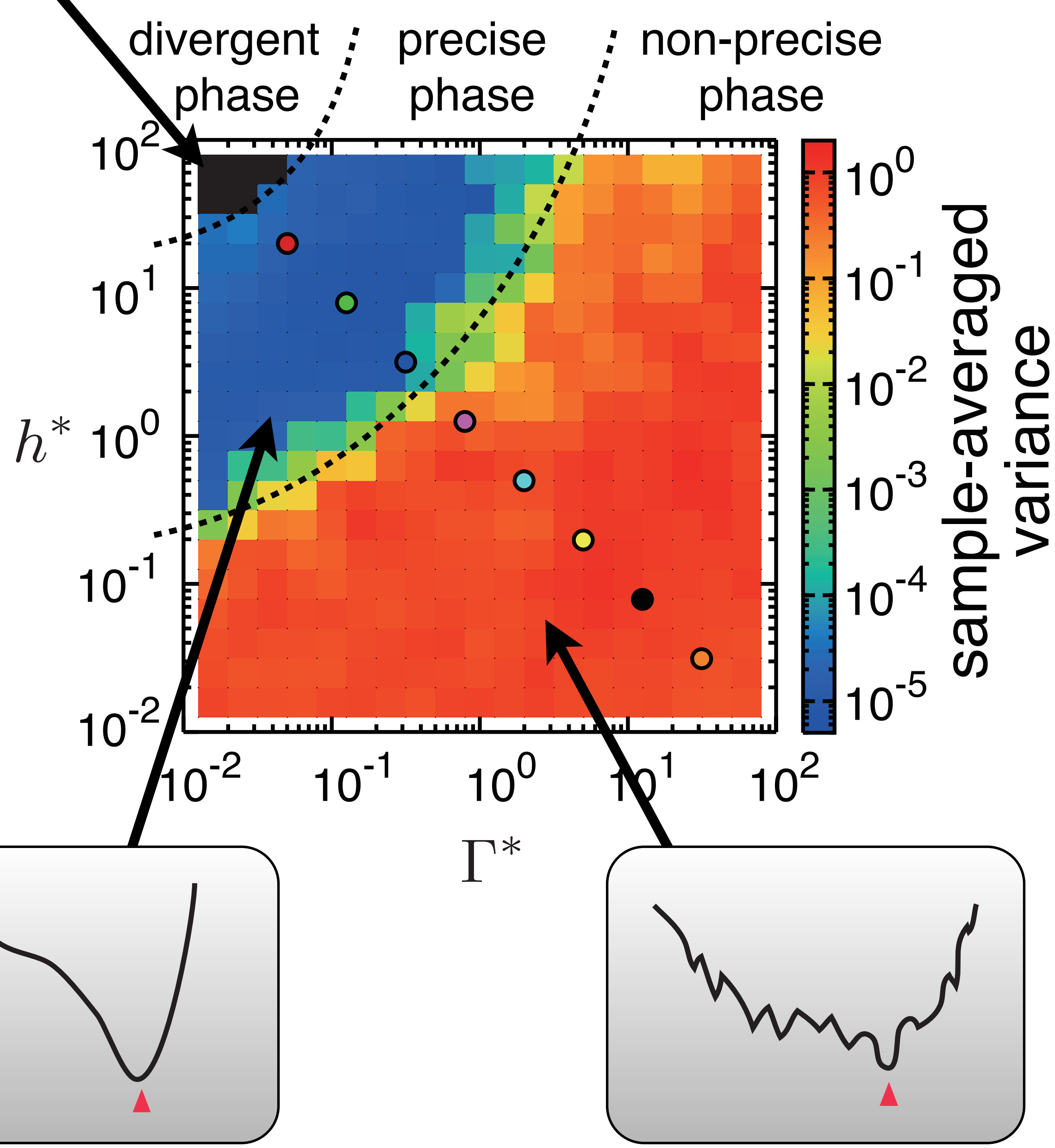


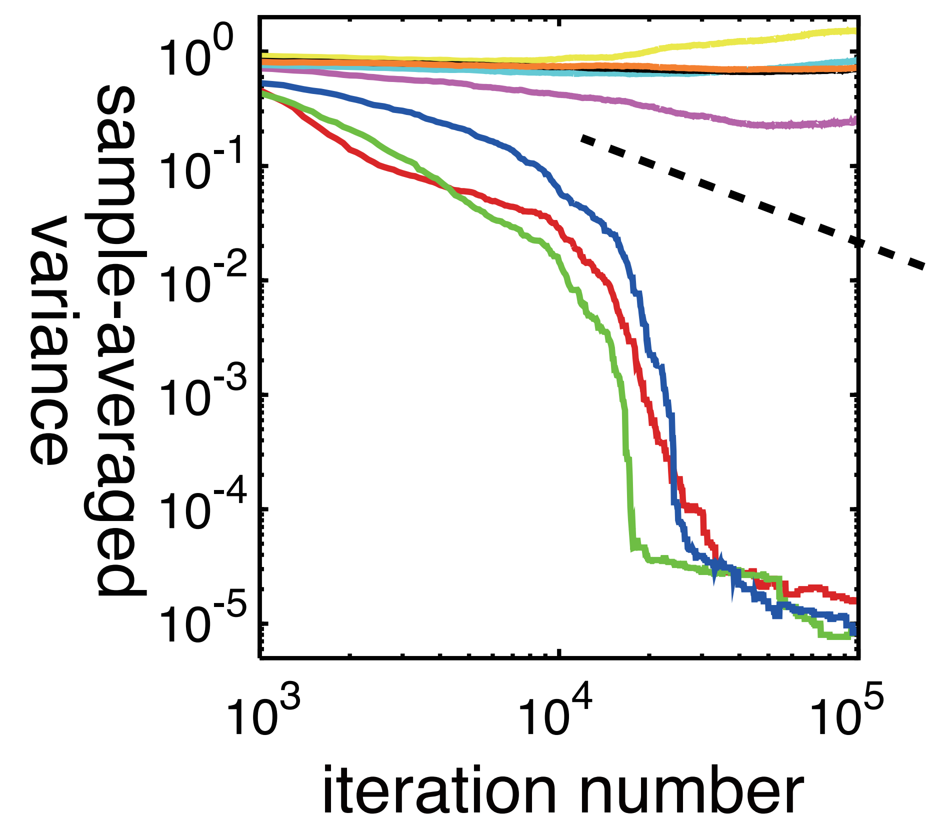
$$\gamma^* = k^* = k_B T = 1,\ \Delta t = 10^{-4}, \tau = 10^2$$

Estimation results

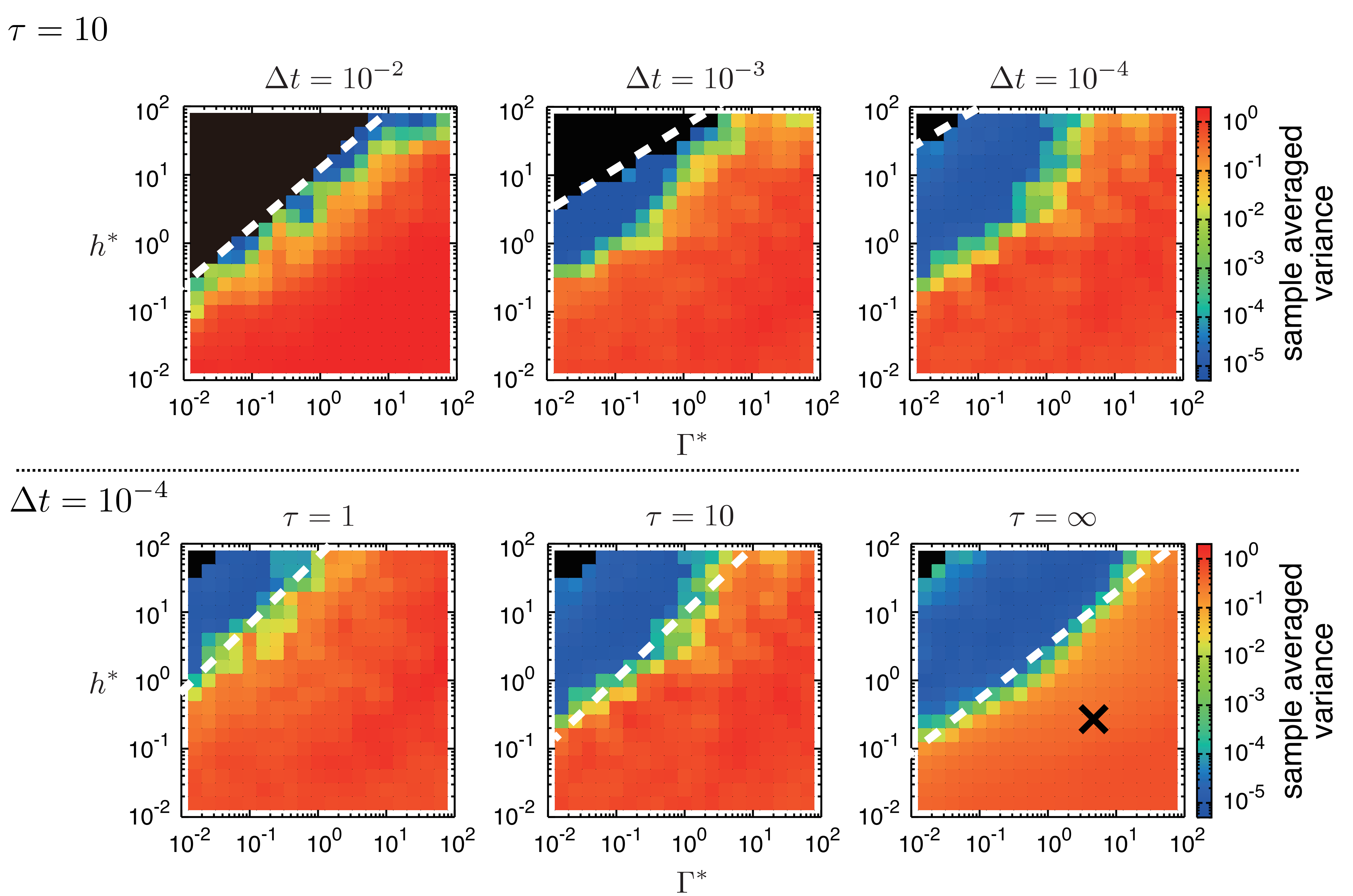
Δt is not enough to follow the probe motion.

The parameter space can be classified into 3 distinct regions.



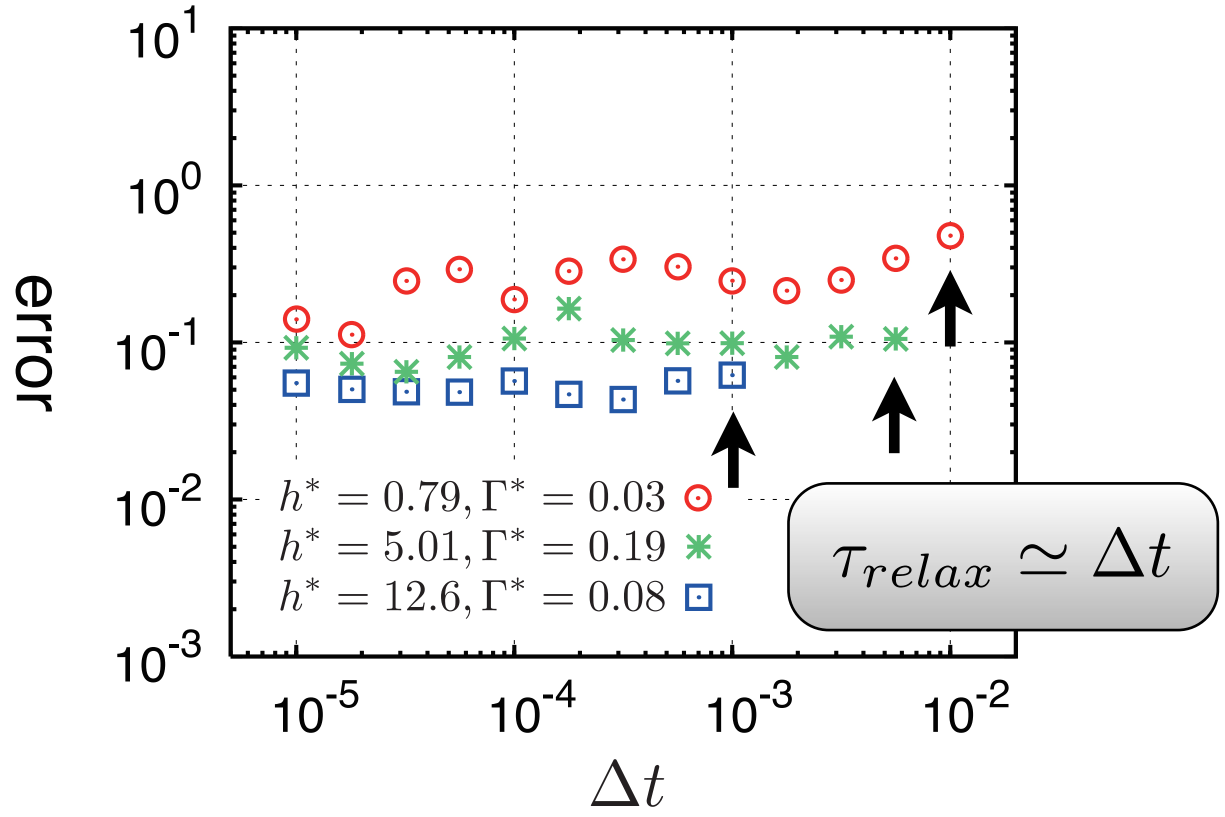


Dependency on τ and Δt



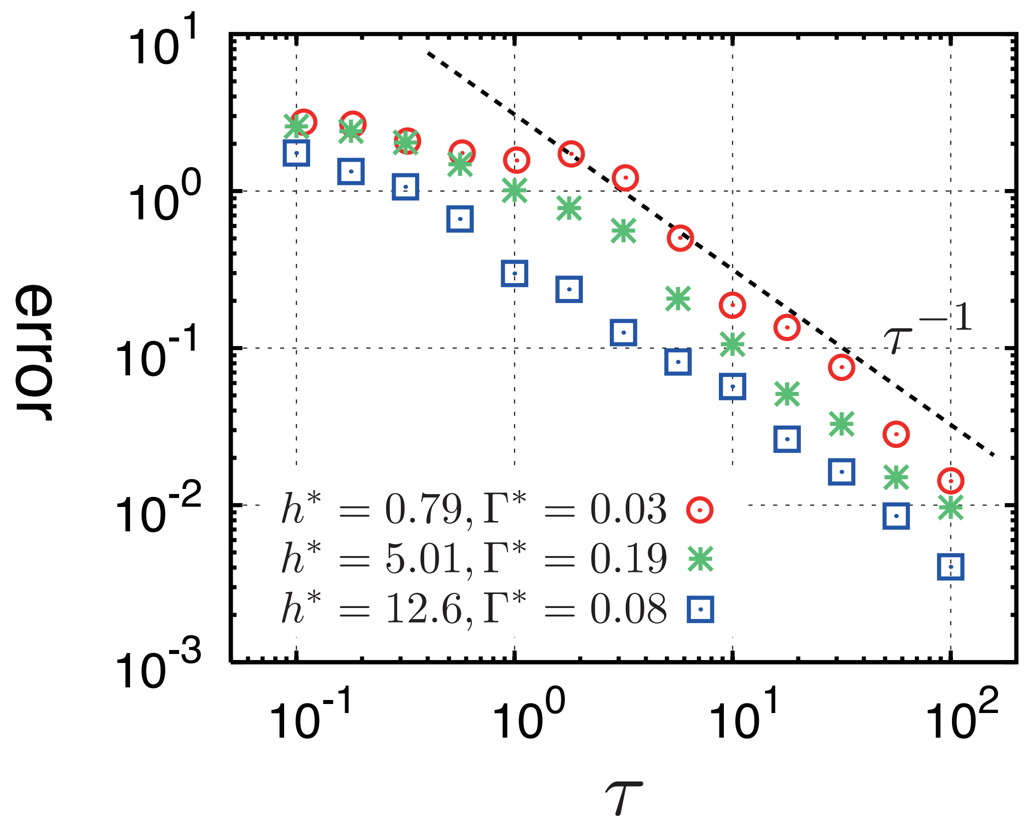
Dependency on τ and Δt

$\tau = 10$



The error is almost independent of Δt .

$\Delta t = 10^{-4}$

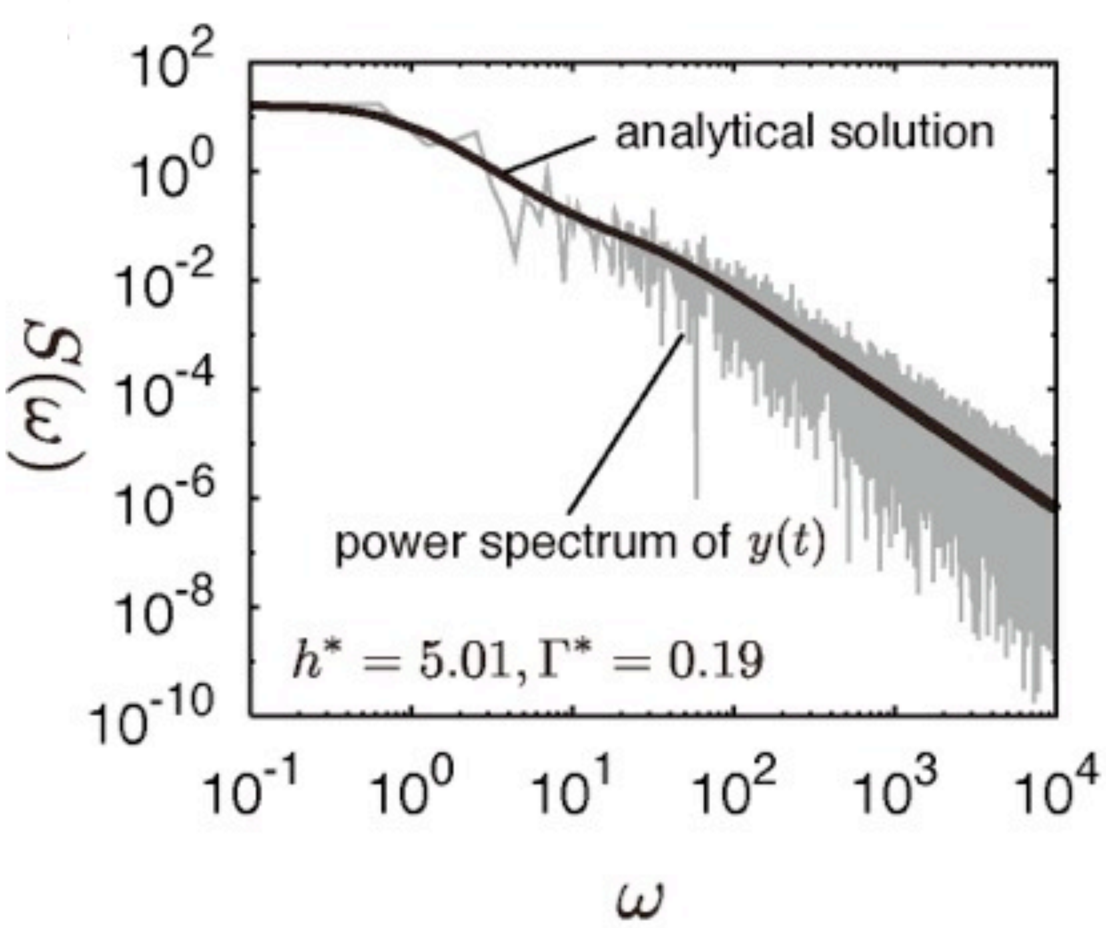


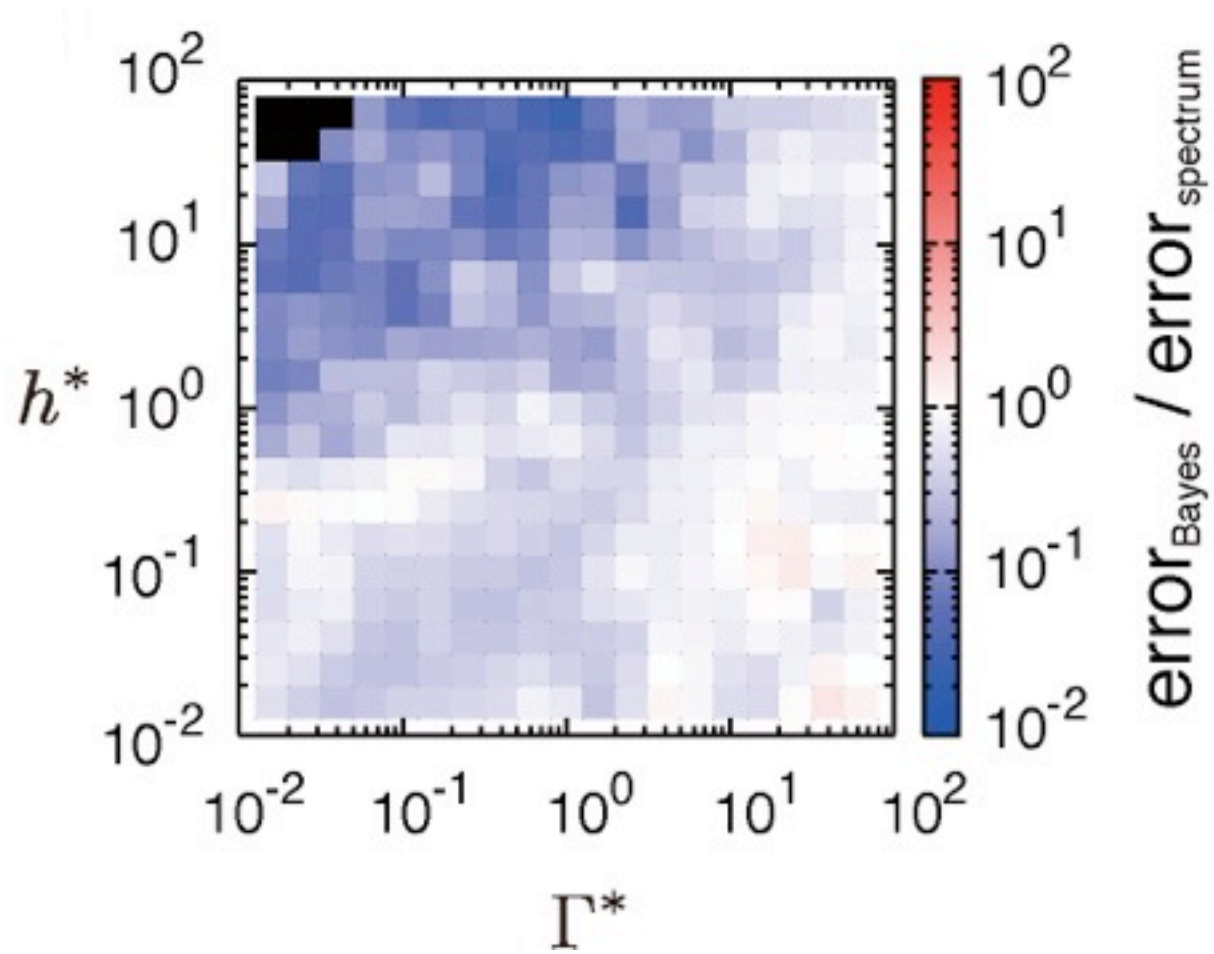
The error decreases proportional to τ .
--> consistent with the theory

Comparison with the spectrum fitting

the power spectrum of $[y]$ was fitted by the analytical solution :

$$S(\omega) = \frac{2k_BT}{\Gamma} \frac{\omega^2 + \frac{h^2}{\gamma\Gamma} + \left(\frac{k+h}{\gamma}\right)^2}{\left\{\omega^2 + i\omega\left(\frac{k+h}{\gamma} + \frac{h}{\Gamma}\right) - \frac{kh}{\gamma\Gamma}\right\}\left\{\omega^2 - i\omega\left(\frac{k+h}{\gamma} + \frac{h}{\Gamma}\right) - \frac{kh}{\gamma\Gamma}\right\}}$$





Our method is more precise than the spectrum fitting up to 100 times.

Estimation of the motion of hidden part

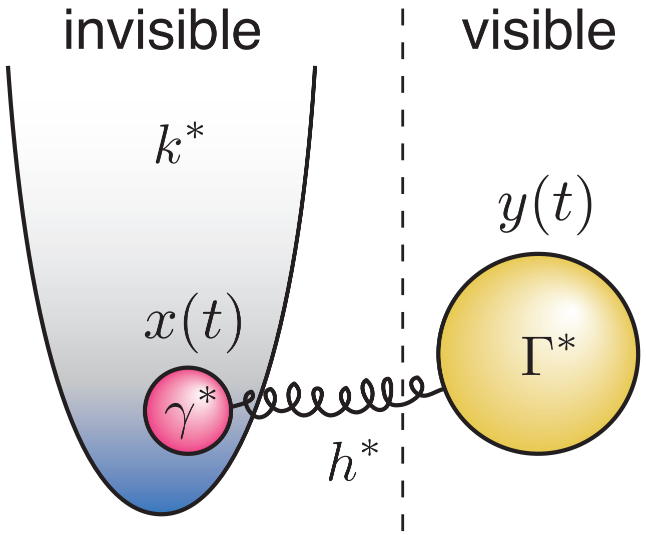
Is the MAP estimator $\hat{x}(t)$ a good approximation of $x(t)$?

⟶ only in the case ; $h^*/k^* \gg 1$

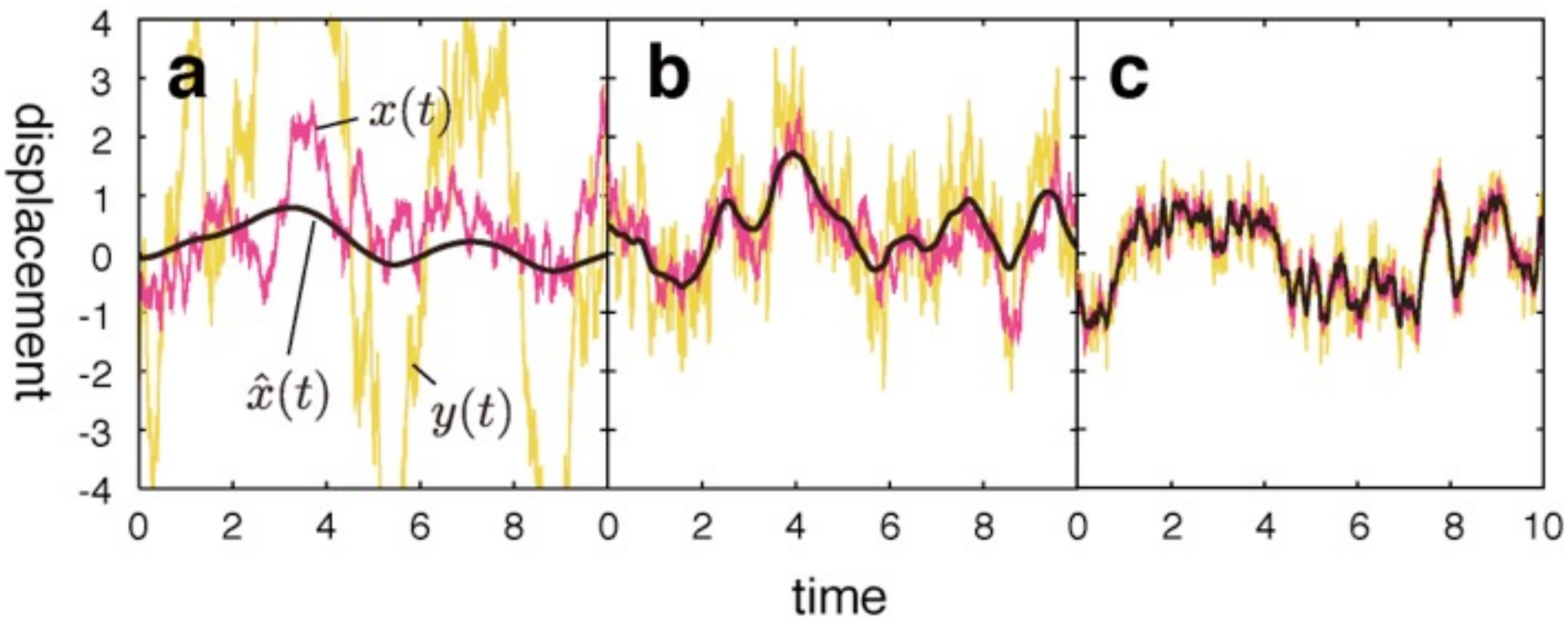
Analytical solution:

$$\langle(\hat{x}(t)-x(t))^2\rangle = \frac{1}{\beta h^*}\left[\frac{2}{\kappa^*} - \frac{1}{\sqrt{\kappa^{*2}+g^*}}\right]$$

$\kappa^* \equiv 1+k^*/h^*$, $g^* \equiv \gamma^*/\Gamma^*$



$\Gamma^* = 0.12$:fixed



Precise estimation of the hidden motion does not required to estimate the system parameters.

Conclusion

We propose the general framework of the parameter estimation method with hidden degrees of freedom.

We obtained the general consequence that the error of the estimate decreases as τ^{-1}.

The practical utility of our method is confirmed by the simple model.

- We found "loss-of-precision transition" and "loss-of-estimate transition".

For precise estimate, the "probe" should be small and tightly coupled to the hidden part.

In "precise phase", the error of the estimate decreases as τ^{-1}, which is consistent with the theoretical prediction.

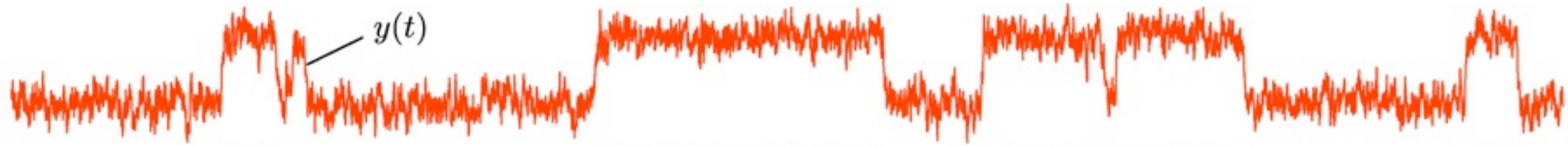
If we mistake the choice of the "probe" and the "linker", we fail to obtain the precise estimates even if we take the data over long times with super-high time resolution.

- The proposed method is more accurate than the conventional method.

The spectrum method cannot be extended to non-linear systems.

Estimation of the motion of hidden part

observed data



⇩

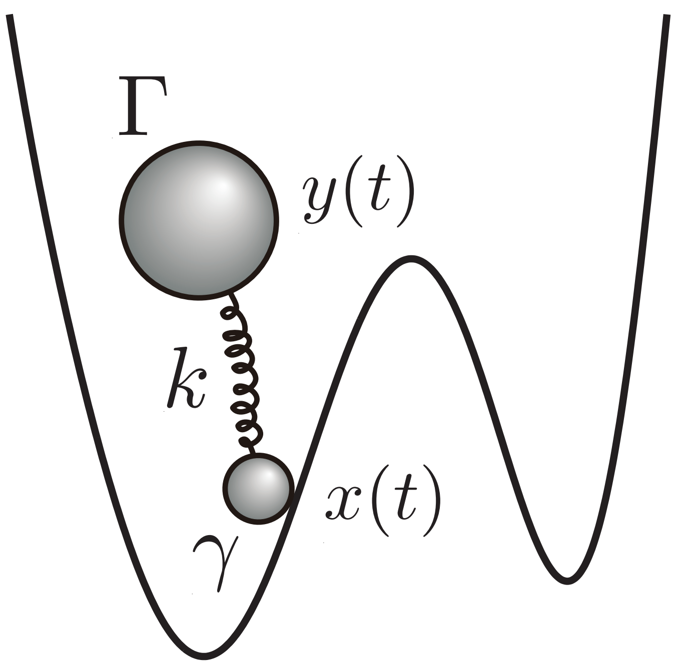
the motion of hidden part: $x(t)$?

DRP (Dominant Reaction Pathway)

Model (Langevin eqs.):

$$\gamma \dot{x} = F(x) - \frac{\partial U(x,y)}{\partial x} + \xi(t)$$

$$\Gamma \dot{y} = -\frac{\partial U(x,y)}{\partial y} + \eta(t)$$



$$F(x) \equiv f(x) - \partial_x V(x)$$

$$\langle \xi(t)\xi(0)\rangle = 2\gamma k_{\mathrm{B}} T \delta(t) \;,\; \langle \eta(t)\eta(0)\rangle = 2\Gamma k_{\mathrm{B}} T \delta(t)$$

MAP estimator in a multiwell potential

Task: estimate the MAP estimator $[\hat{x}]$ from $[y]$.

Maximize the path probability $C\exp[-\beta S([x,y];\mathbf{\Pi})]$ with respect to $[x]$.

→ Solve Euler-Lagrange eq.: $\dfrac{\delta S([x,y];\mathbf{\Pi})}{\delta x} = 0$

× **relaxation method :** too much computational cost
cf. TDGL (time-dependent Ginzburg-Landau) eq.

○ **Go-and-Back method :** $\Gamma \gg \gamma$ & perturbation expansion

solve $x_{(i)}$ from $t=0$ to $t=\tau$. **(Go)**

$$\gamma \dot{x}_{(i)} = G(x_{(i)}, y) + v_{(i)} + O(\varepsilon^{i+1})$$

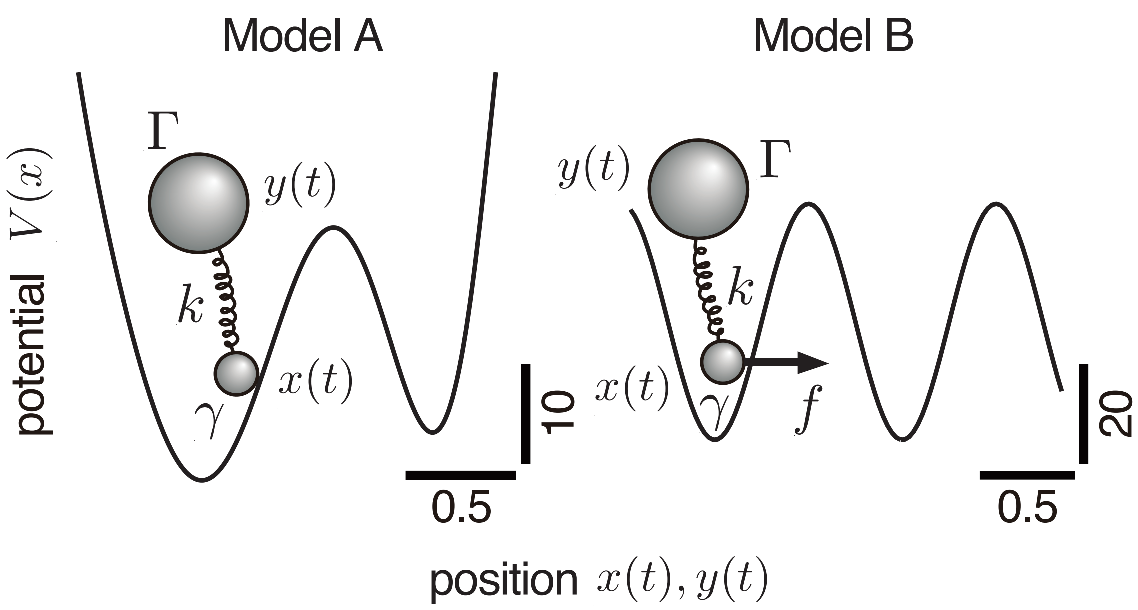
$i+1 \to i$

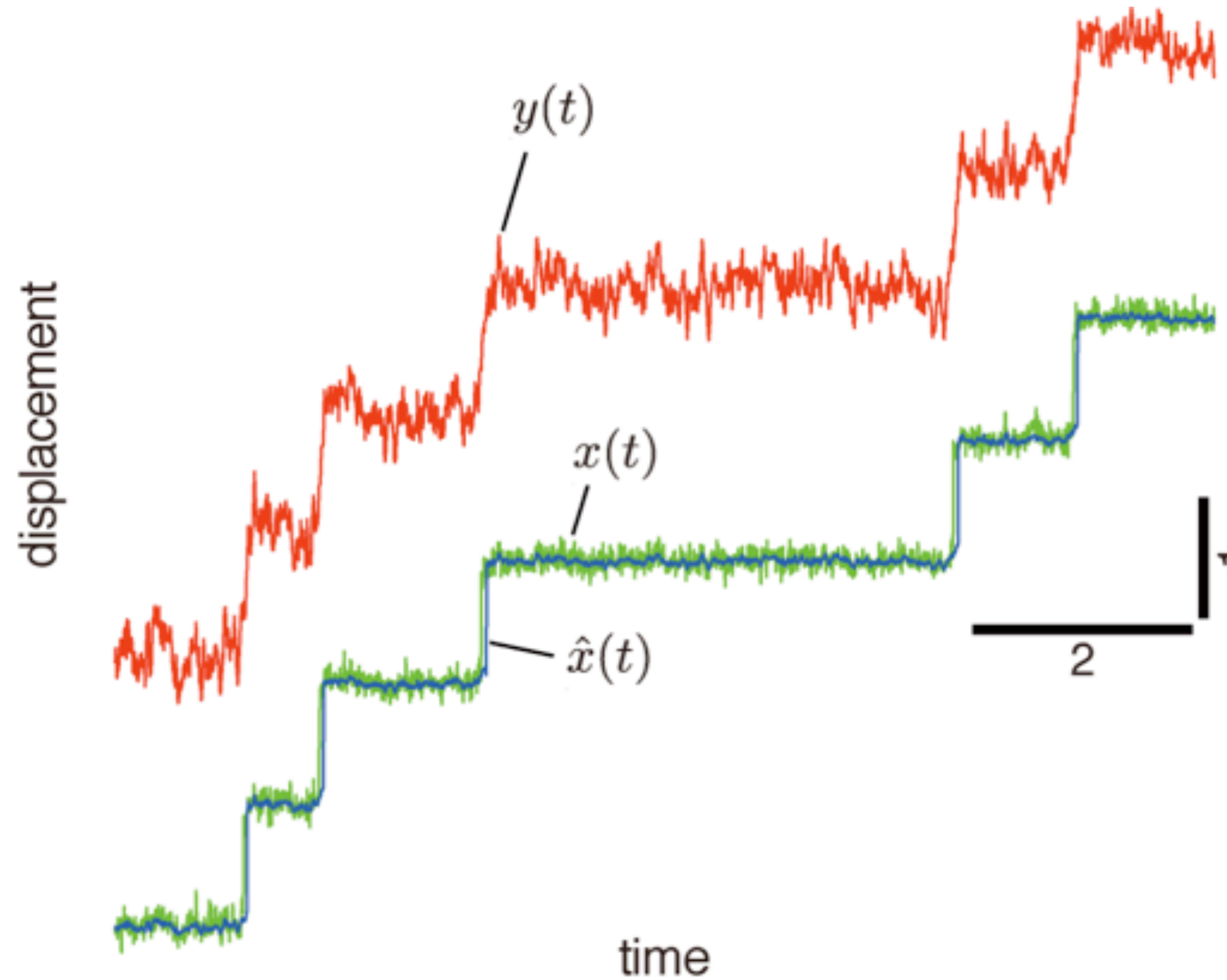
solve $v_{(i+1)}$ from $t=\tau$ to $t=0$. **(Back)**

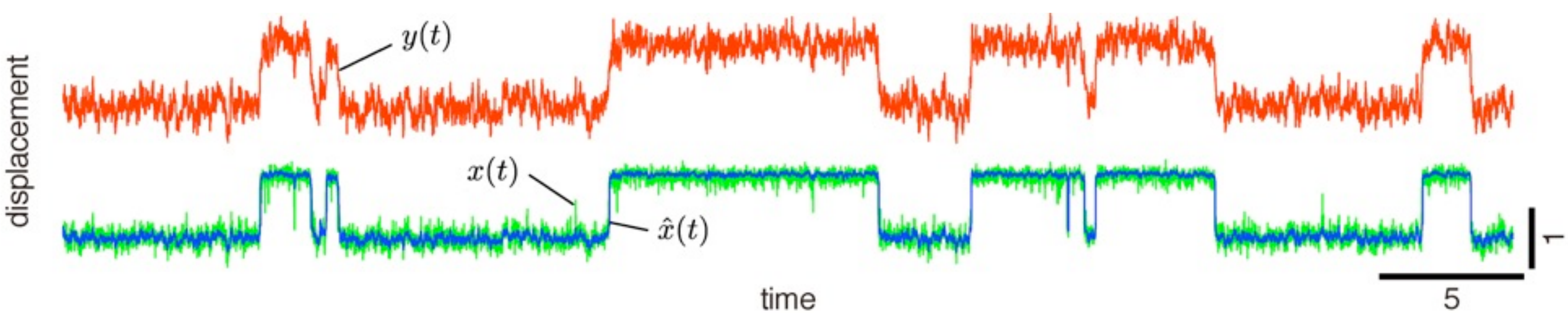
$$\gamma \dot{v}_{(i+1)} = -G_x(x_{(i)}, y) v_{(i+1)} + \varepsilon\, k_{\mathrm{B}} T G_{xx}(x_{(i)}, y) + O(\varepsilon^{i+2})$$

Result

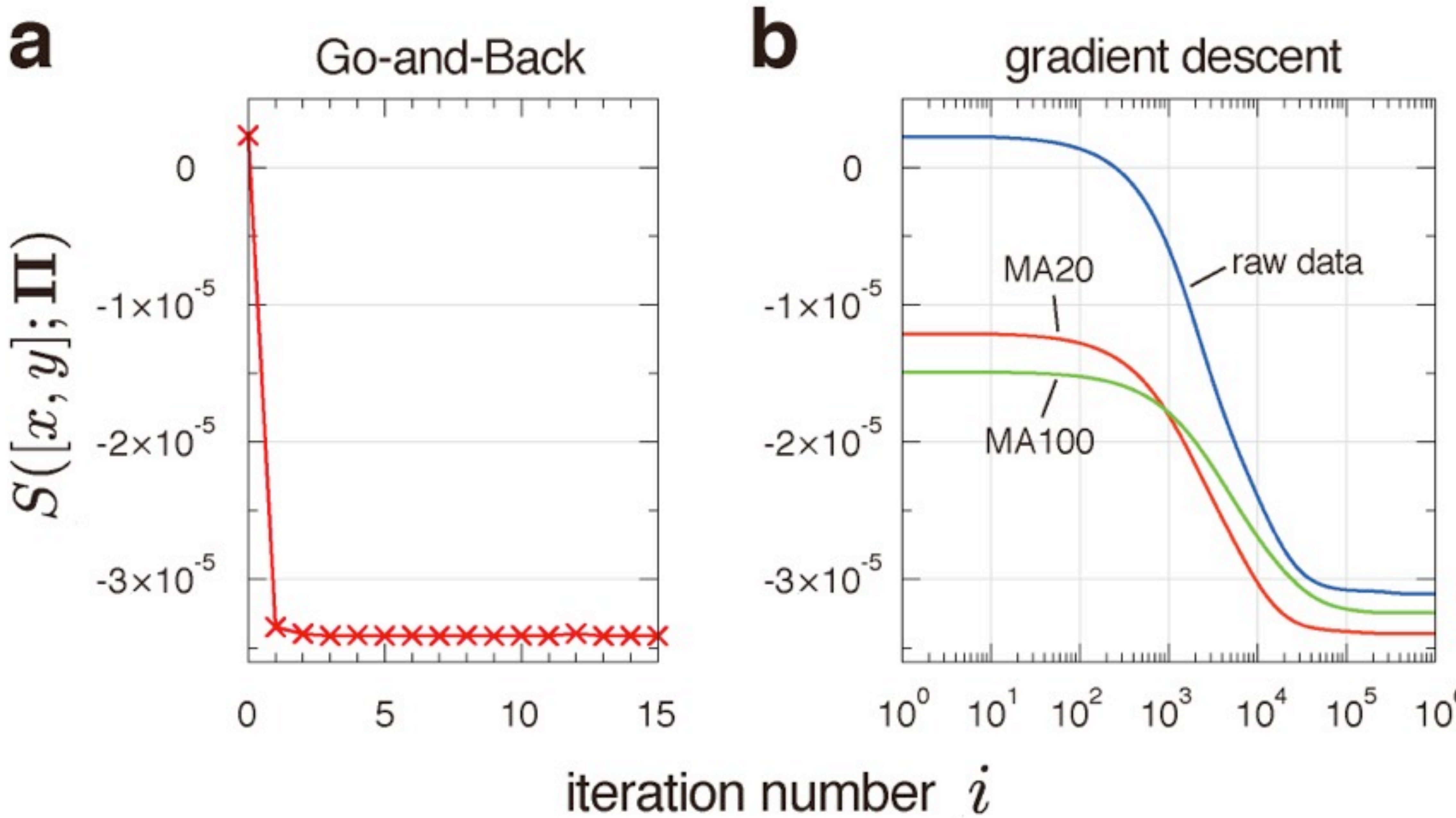
Go-and-Back method yields reasonable estimates.







Result



Go-and-Back method is faster than the gradient descent ~10^4 times.

Reference & Acknowledgement

Miyazaki, M., Harada, T. (2010) “Bayesian estimation of the internal structure of proteins from single-molecule measurements”, *J. Chem. Phys.*, **133**, to appear, November 21, 2010.

Miyazaki, M., Harada, T. (2010) “Go-and-Back method: Effective estimation of the hidden motion of proteins from single-molecule time series”, manuscript in preparation.

Single-molecule techniques

Takayuki Nishizaka (Gakushuin Univ)
Sugawa Mitsuhiro (Gakushuin Univ)
Tetsuaki Okamoto (Kanazawa Univ)

Physics

Shoichi Toyabe (Chuo Univ)
Miki Matsuo-Yamamoto (The Univ of Tokyo)

Bayesian statistics

Tetsuya J. Kobayashi (The Univ of Tokyo)
Ryota Tomioka (The Univ of Tokyo)

特定領域研究「膜超分子モーターの革新的ナノサイエンス」
京都大学GCOEプログラム「普遍性と創発性から紡ぐ次世代物理学」