Analytic Continuation of Quantum Monte Carlo Data

Erik Koch
Computational Materials Science
Jülich Supercomputer Centre, Forschungszentrum Jülich

$$G(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\omega\tau}}{1+\mathrm{e}^{-\beta\tau}}\, A(\omega)\, d\omega$$

JÜLICH
FORSCHUNGSZENTRUM

analytic continuation

- why analytic continuation?
- analytic properties of $G(\tau)$
- preparing data and discretization
- least-squares
- MaxEnt
- average spectrum

analytic continuation

why analytic continuation?

expectation values

$$\langle A \rangle = \frac{\sum_n e^{-\beta E_n} \langle n | A | n \rangle}{\sum_n e^{-\beta E_n}} = \frac{1}{Z} \operatorname{Tr}\left(e^{-\beta H} A \right)$$

sample trace using Monte Carlo (sign problem)

correlation functions

$$\big\langle A(t) B(0) \big\rangle = \frac{1}{Z} \operatorname{Tr} \mathrm{e}^{-\beta H}\, \mathrm{e}^{\mathrm{i}Ht} A\, \mathrm{e}^{-\mathrm{i}Ht}\, B$$

complex time evolution: phase problem

could be avoided in imaginary time

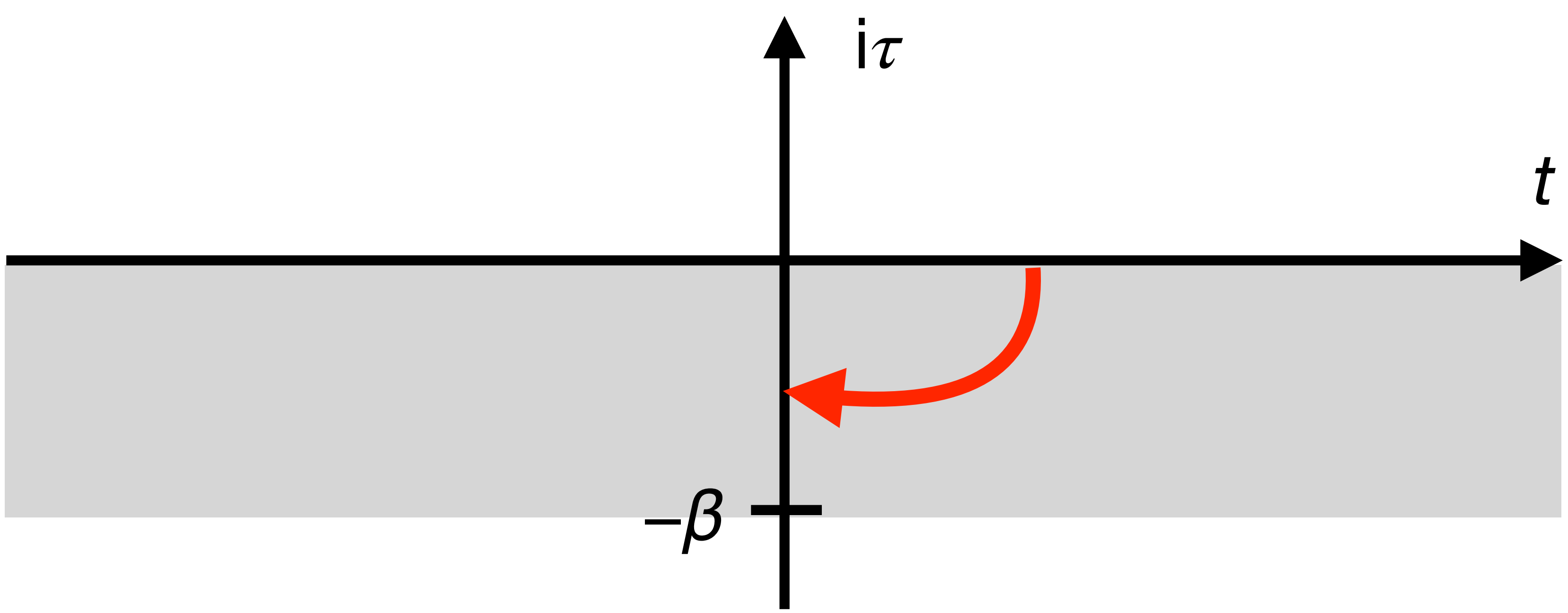
analytic continuation

complex time: $\zeta = t - i\tau$ $A(-i\tau) = e^{\tau H} A\, e^{-\tau H}$

$$\langle A(t - i\tau) B \rangle = \frac{1}{Z} \operatorname{Tr} e^{(it+\tau-\beta)H} A\, e^{-(it+\tau)H} B$$

$$= \frac{1}{Z} \sum_{n,m} e^{(it+\tau-\beta)E_n} e^{-(it+\tau)E_m} \langle n|A|m\rangle \langle m|B|n\rangle$$

well defined for $\tau \in [0, \beta]$



calculate $C_{AB}(\tau) := \langle A(-i\tau)B \rangle$ for $\tau \in [0, \beta]$ by Monte Carlo

how to go back to real axis?

QMC gives only function values $C_{AB}(\tau) := \langle A(-i\tau)B \rangle$ for $\tau \in [0, \beta]$

need functional form for analytic continuation

$$C_{AB}(\tau) = \frac{1}{Z} \sum_{n,m} e^{(\tau-\beta)E_n} \, e^{-\tau E_m} \langle n|A|m\rangle\langle m|B|n\rangle$$

involves all eigenvalues & matrix elements…

more compact: **spectral function**

$$\int_{-\infty}^{\infty} dt \, e^{i\omega t} \langle A(t)B \rangle = \frac{2\pi}{Z} \sum_{n,m} e^{-\beta E_n} \langle n|A|m\rangle\langle m|B|n\rangle \, \delta\big(\omega - (E_m - E_n)\big) =: \rho_{AB}(\omega)$$

$$\boxed{C_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, e^{-\omega\tau} \rho_{AB}(\omega)} = \langle A(-i\tau)B \rangle$$

solve for spectral function

then we can analytically continue back to real axis

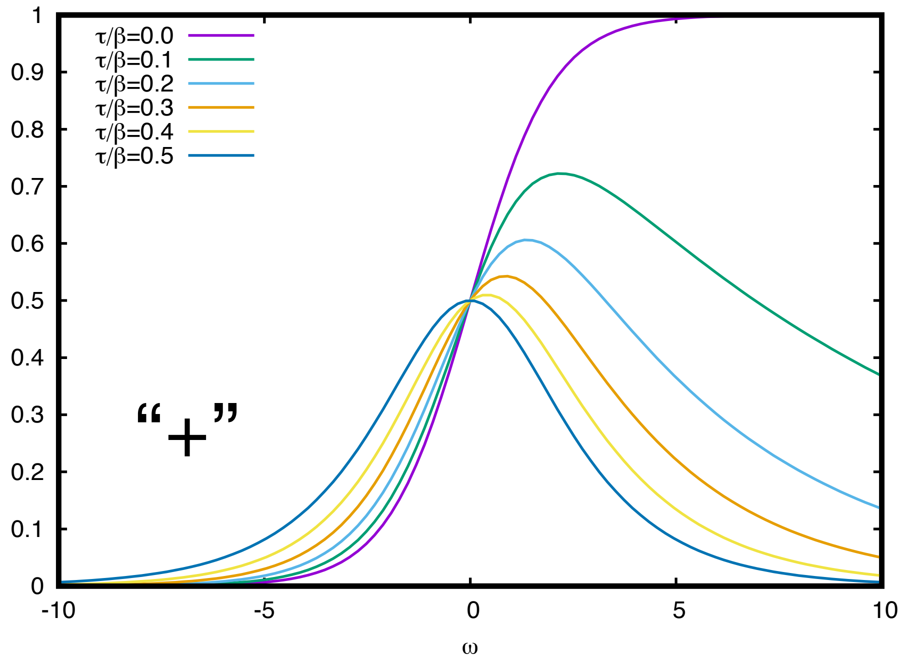
Fredholm equation of 1st kind

$$C_{AB}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, e^{-\omega\tau} \rho_{AB}(\omega)$$

diverging **integral kernel** for $\omega \to -\infty$

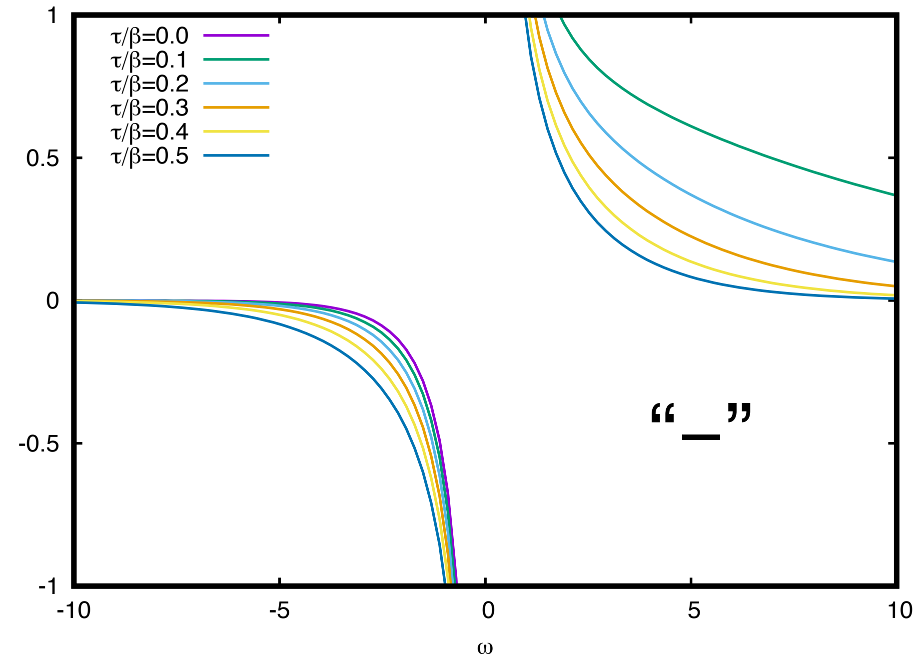
modify kernel

$$C_{AB}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \frac{e^{-\omega\tau}}{\mu(\omega)} \underbrace{\mu(\omega)\rho_{AB}(\omega)}_{=:\tilde{\rho}(\omega)}$$



convenient choice

$$\mu(\omega) = 1 \pm e^{-\beta\omega}$$



meaning of modified spectral function

$$\tilde{\rho}^{\pm}_{AB}(\omega) = \left(1 \pm \mathrm{e}^{-\beta\omega}\right) \rho_{AB}(\omega) = \rho_{AB}(\omega) \pm \rho_{BA}(-\omega)$$

is spectral function of

$$\mathrm{i}G^{\pm}_{AB}(t) := \left\langle A(t)B \right\rangle \pm \left\langle B(-t)A \right\rangle = \left\langle A(t)B \right\rangle \pm \left\langle BA(t) \right\rangle = \left\langle [A(t),\, B]_{\pm} \right\rangle$$

contains retarded/advanced Green function:

$$G^{R\pm}_{AB}(t) = \Theta(t)\, G^{\pm}_{AB}(t)$$
$$G^{A\pm}_{AB}(t) = -\Theta(-t)\, G^{\pm}_{AB}(t)$$

$\langle A(t)B \rangle$ can be analytically continued to stripe below real axis, $\langle BA(t) \rangle$ above

define **Matsubara Green function**

$$-G^{M\pm}_{AB}(\tau) := \left\langle \mathcal{T}^{\pm}_{\tau}\, A(-\mathrm{i}\tau)B(0) \right\rangle$$

where imaginary-time ordering selects the term that is analytic for the given τ

$$\mathcal{T}^{\pm}_{\tau}\, A(-\mathrm{i}\tau)B(0) = \Theta(\tau)A(-\mathrm{i}\tau)B(0) \mp \Theta(-\tau)B(0)A(-\mathrm{i}\tau)$$

introducing discontinuity at τ=0

$$G^{M\pm}_{AB}(0^{+}) - G^{M\pm}_{AB}(0^{-}) = -\left\langle [A,\, B]_{\pm} \right\rangle$$

Fredholm equation for Matsubara GF

$$G^{M\pm}_{AB}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \frac{\mathrm{e}^{-\omega\tau}}{1 \pm \mathrm{e}^{-\beta\omega}} \, \tilde{\rho}^{\pm}_{AB}(\omega) \quad \text{for } \tau \in [0,\beta]$$

sum rule $\quad \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \tilde{\rho}^{\pm}_{AB}(\omega) = \langle [A,\ B]_{\pm} \rangle$

special case B=$A^\dagger$

$$\tilde{\rho}^{\pm}_{AA^\dagger}(\omega) = \frac{2\pi}{Z}\sum_{n,m}\left(\mathrm{e}^{-\beta E_n} \pm \mathrm{e}^{-\beta E_m}\right) |\langle n|A|m\rangle|^2 \, \delta\big(\omega - (E_m - E_n)\big)$$

$$G^{M+}_{AA^\dagger}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \frac{\mathrm{e}^{-\omega\tau}}{1 + \mathrm{e}^{-\beta\omega}} \, \tilde{\rho}^{+}_{AA^\dagger}(\omega)$$

$$G^{M-}_{AA^\dagger}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \frac{\omega \, \mathrm{e}^{-\omega\tau}}{1 - \mathrm{e}^{-\beta\omega}} \, \frac{\tilde{\rho}^{-}_{AA^\dagger}(\omega)}{\omega}$$

non-negative

properties of $G(\tau)$

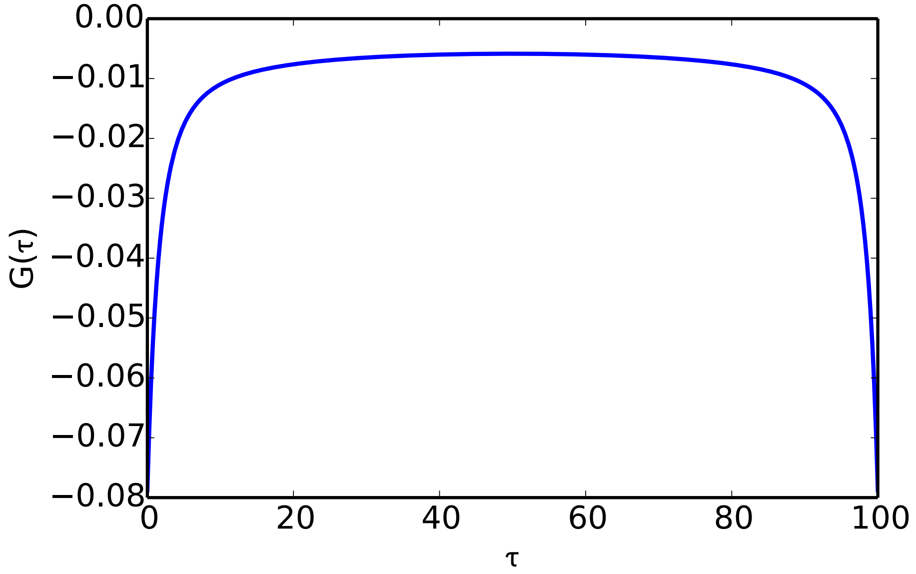
where is the information in $G^M(\tau)$?

$$G(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, \frac{e^{-\omega\tau}}{1+e^{-\beta\omega}} \, \rho(\omega) \quad \text{for } \tau \in [0,\beta]$$

model spectral function: $\rho(\omega) = \sum_i w_i \, \delta(\omega - \varepsilon_i)$

$$G(\tau) = -\frac{1}{2\pi}\sum_i w_i \big(1 - n_{\mathrm{FD}}(\varepsilon_i)\big) \, e^{-\varepsilon_i \tau}$$

$$= -\frac{1}{2\pi}\sum_i w_i \qquad n_{\mathrm{FD}}(\varepsilon_i) \; e^{+\varepsilon_i(\beta-\tau)}$$



linear combination of decaying exponentials

information on spectral features at large $|\omega|$ concentrated at 0 and β

Euler polynomials and moments

fermionic kernel is generating function of Euler polynomials

$$\frac{2e^{xt}}{e^t+1} = \sum_{n=0}^{\infty} E_n(x)\,\frac{t^n}{n!} \quad \text{for } x \in [0,1]$$

bosonic: Bernoulli poly

$$\frac{t\,e^{xt}}{e^t-1} = \sum_{n=0}^{\infty} B_n(x)\,\frac{t^n}{n!}$$

$$G^{M+}(\tau) = -\frac{1}{4\pi}\sum_{n=0}^{\infty} E_n(\tau/\beta)\,\frac{(-\beta)^n}{n!} \int_{-\infty}^{\infty} d\omega\,\omega^n\,\tilde{\rho}^+(\omega)$$

n^{th} moment

to determine moments from Matsubara GF
we need dual functions of Euler polynomials

$$\int_0^1 dx\,E^n(x)\,E_m(x) = \delta_{n,m}$$

$$\int_0^1 dx\,E^n(x)\,\frac{2e^{xt}}{e^t+1} \overset{!}{=} \frac{t^n}{n!} \quad\Rightarrow\quad E^n(x) = \frac{(-1)^n}{2\,n!}\Big(\delta^{(n)}(x-1) + \delta^{(n)}(x)\Big)$$

moments and derivative discontinuities

$$\frac{d^n G^{M+}(0^+)}{d\,\tau^n} - \frac{d^n G^{M+}(0^-)}{d\,\tau^n} = -\frac{(-1)^n}{2\pi}\int_{-\infty}^{\infty} d\omega\;\omega^n\,\tilde{\rho}^+(\omega)$$

discontinuities also describe algebraic decay of FT on Matsubara freq.

higher derivatives very hard to get from $G(\tau)$ directly …

$$-G^{M+}(\tau) = \frac{1}{Z}\,\mathrm{Tr}\,\mathrm{e}^{-\beta H}\mathrm{e}^{\tau H}A\mathrm{e}^{-\tau H}B$$

$$-\frac{dG^{M+}(\tau)}{d\tau} = \frac{1}{Z}\,\mathrm{Tr}\,\mathrm{e}^{-\beta H}\mathrm{e}^{\tau H}[H,\,A]\,\mathrm{e}^{-\tau H}B$$

repeated derivatives produce $[H;A]_n := [H;[H;A]_{n\text{-}1}]$ and $[H;A]_0 := A$

$$\big\langle \big[[H;\,A]_n,\,B\big]\big\rangle = -\frac{(-1)^n}{2\pi}\int_{-\infty}^{\infty} d\omega\;\omega^n\,\tilde{\rho}^+(\omega)$$

preparing data
and discretization

preparing QMC data

Fredholm equation

$$g(y) = \int K(y,x)\, f(x)\, dx$$

QMC data: discrete (mesh or expansion coeff.) $\quad g(y) \to \boldsymbol{g} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{pmatrix}$

statistically independent samples

mean $\quad \bar{\boldsymbol{g}} = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{g}_k$

covariance $\quad \boldsymbol{C} = \frac{1}{K(K-1)} \sum_{k=1}^{K} (\boldsymbol{g}_k - \bar{\boldsymbol{g}})(\boldsymbol{g}_k - \bar{\boldsymbol{g}})^\dagger$

$$= \frac{1}{K(K-1)} \sum_{k} \boldsymbol{g}_k \boldsymbol{g}_k^\dagger - \frac{1}{K-1} \bar{\boldsymbol{g}} \bar{\boldsymbol{g}}^\dagger$$

$$p(\bar{\boldsymbol{g}}, \boldsymbol{C} | \boldsymbol{g}_{\text{exact}}) = \frac{1}{(2\pi)^{M/2} \det \boldsymbol{C}} \mathrm{e}^{-(\bar{\boldsymbol{g}} - \boldsymbol{g}_{\text{exact}})^\dagger \boldsymbol{C}^{-1} (\bar{\boldsymbol{g}} - \boldsymbol{g}_{\text{exact}})/2}$$

blocking of correlated data

sequence of MC data $m_1, \ldots, m_K$ $\qquad \bar{m} = \frac{1}{K}\sum_k m_k$

$$\sigma^2(\bar{m}) = \frac{1}{K(K-1)}\sum (m_k - \bar{m})^2$$

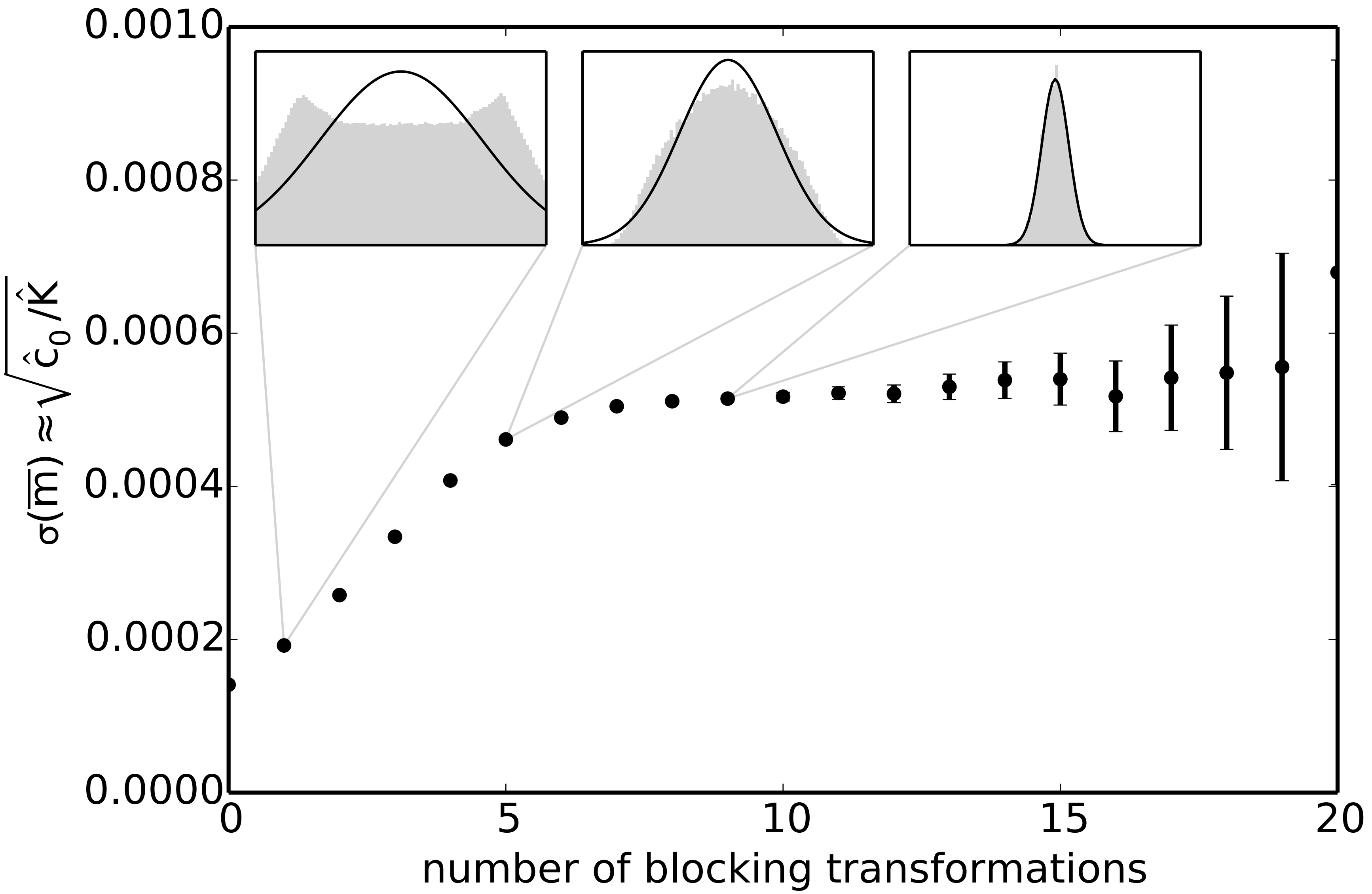
underestimates variance when data correlated

renormalization step

$$\hat{m}_{\hat{k}} := \frac{m_{2\hat{k}-1} + m_{2\hat{k}}}{2}$$

$$\hat{K} := K/2$$

removes correlations
approaches normal distr.



matrix representation

$$g(y) = \int K(y,x)\, f(x)\, dx$$

discretize also model function $f(x) \to \boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}$

$$g(y_m) = \sum_n K(y_m, x_n)\, w_n\, f(x_n) \quad \text{or} \quad g_m = \sum_n \langle \psi_m | K | \varphi_n \rangle\, f_n$$

matrix equation ***g*** = ***Kf***

absorb covariance in kernel matrix
factorize ***C***$^{-1}$ = ***T***†***T***

$$\tilde{\boldsymbol{g}} := \boldsymbol{T}\bar{\boldsymbol{g}}$$

$$\tilde{\boldsymbol{K}} := \boldsymbol{T}\boldsymbol{K}$$

$$p(\bar{\boldsymbol{g}}, \boldsymbol{C} | \boldsymbol{f}) \propto \mathrm{e}^{-(\bar{\boldsymbol{g}} - \boldsymbol{K}\boldsymbol{f})^\dagger \boldsymbol{C}^{-1} (\bar{\boldsymbol{g}} - \boldsymbol{K}\boldsymbol{f})/2} \to p(\bar{\boldsymbol{g}}, \mathbb{1} | \boldsymbol{f}) \propto \mathrm{e}^{-\frac{1}{2}||\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}||^2}$$

solving $g = Kf$

Bayesian inversion

probability for measuring $\bar{\boldsymbol{g}}$ with covariance $\boldsymbol{C}$ when true model is $\boldsymbol{f}$

$$p(\tilde{\boldsymbol{g}}|\,\boldsymbol{f}) \propto \mathrm{e}^{-\frac{1}{2}||\tilde{\boldsymbol{g}}-\tilde{\boldsymbol{K}}\boldsymbol{f}||^2}$$

want probability of $\boldsymbol{f}$ given data $\bar{\boldsymbol{g}}$ and $\boldsymbol{C}$

$$p(\boldsymbol{f}|\,\tilde{\boldsymbol{g}}) = \frac{p(\tilde{\boldsymbol{g}}|\,\boldsymbol{f})\,p(\boldsymbol{f})}{p(\tilde{\boldsymbol{g}})}$$

posterior ∝ likelihood × prior

Bayes' theorem: $p(B|\,A)\,p(A) = p(A, B) = p(A|\,B)\,p(B)$

maximize posterior probability

assume $p(\boldsymbol{f})$ = const. (uninformative prior)

maximum likelihood estimator

minimize χ^2

least-squares fit

singular value decomposition

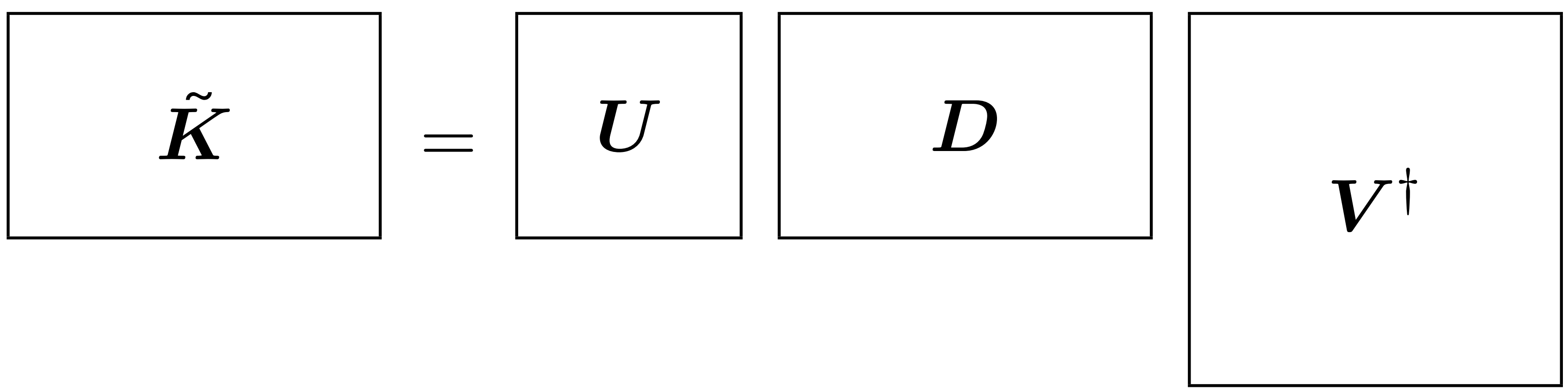
singular value decomposition of $M{\times}N$ kernel matrix

$$\tilde{\boldsymbol{K}} = \sum_i |\boldsymbol{u}_i\rangle d_i \langle \boldsymbol{v}_i|$$

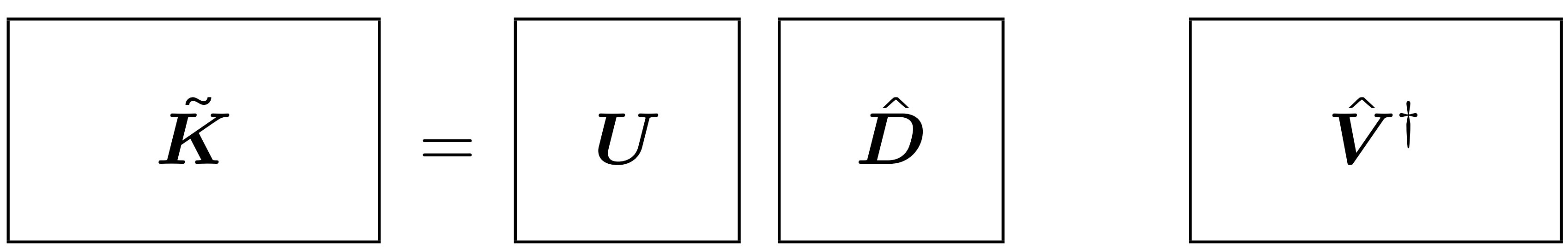
$d_1 \geq d_2 \geq \ldots \geq d_{\min(M,N)} \geq 0$
small singular value d_i
small contribution to result

eigenvalue decomposition: $\boldsymbol{H} = \sum_i |\boldsymbol{v}\rangle e_i \langle \boldsymbol{v}_i|$

full SVD (for $N > M$)



reduced SVD (drop zero modes)



least-squares fit

$$\text{minimize } \chi^2(\boldsymbol{f}) = ||\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}||^2$$

$$|\tilde{\boldsymbol{g}}\rangle - \tilde{\boldsymbol{K}}|\boldsymbol{f}\rangle = |\tilde{\boldsymbol{g}}\rangle - \sum_i |\boldsymbol{u}_i\rangle\, d_i\, \langle \boldsymbol{v}_i|\boldsymbol{f}\rangle$$

$$= \sum_i |\boldsymbol{u}_i\rangle \Big(\langle \boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle - d_i \langle \boldsymbol{v}_i|\boldsymbol{f}\rangle \Big)$$

least squares solution

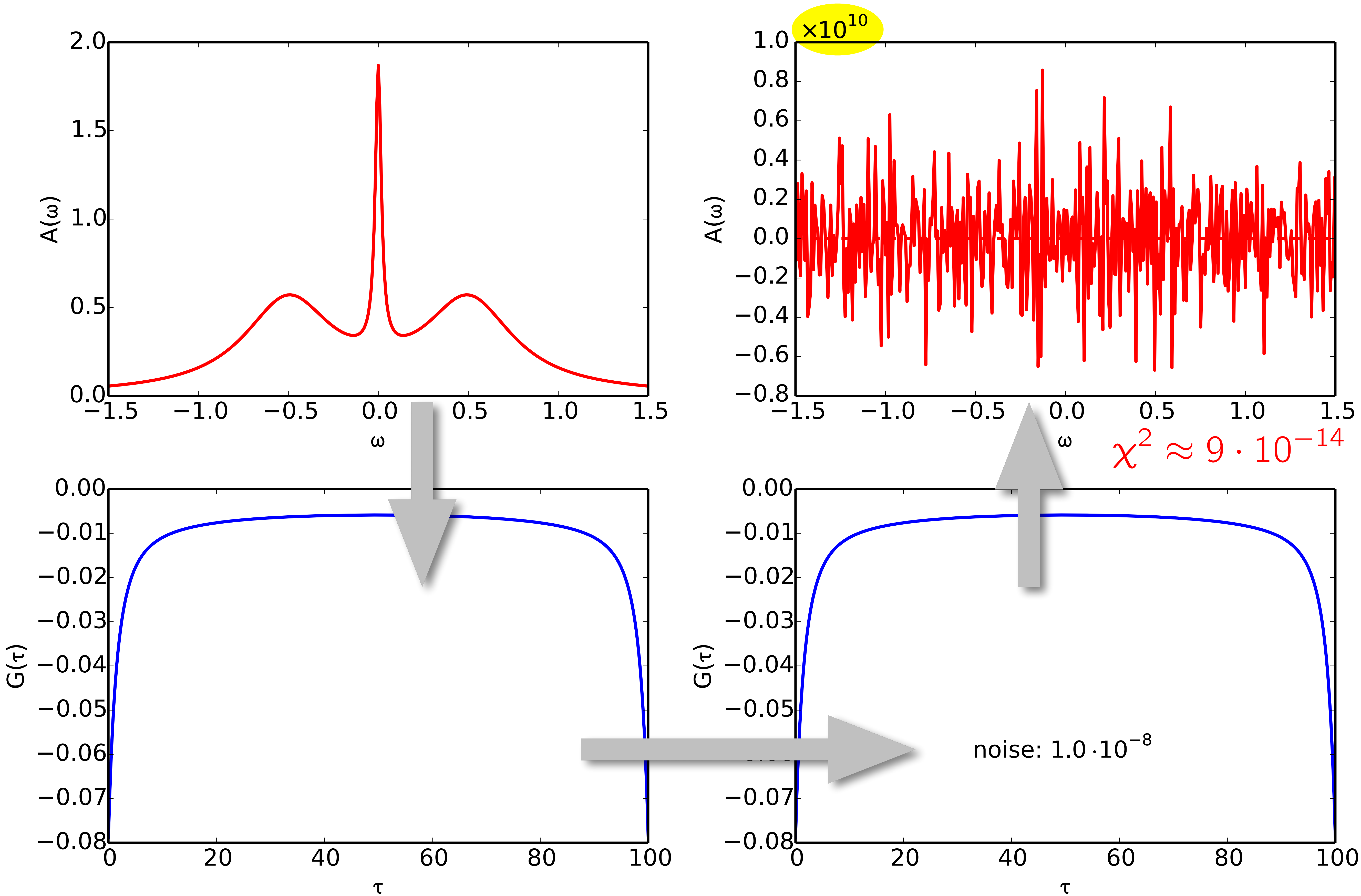
$$|\boldsymbol{f}_{\mathrm{LS}}\rangle = \sum_i \frac{\langle u_i|\tilde{\boldsymbol{g}}\rangle}{d_i}\, |\boldsymbol{v}_i\rangle \qquad \boldsymbol{f}_{\mathrm{LS}} = \hat{\boldsymbol{V}}\hat{\boldsymbol{D}}^{-1}\boldsymbol{U}^{\dagger}\tilde{\boldsymbol{g}}$$

modes with d_i=0 do not contribute to χ^2 — can be chosen at will [**ill-posed**]

perfect fit (χ^2=0) if d_M>0

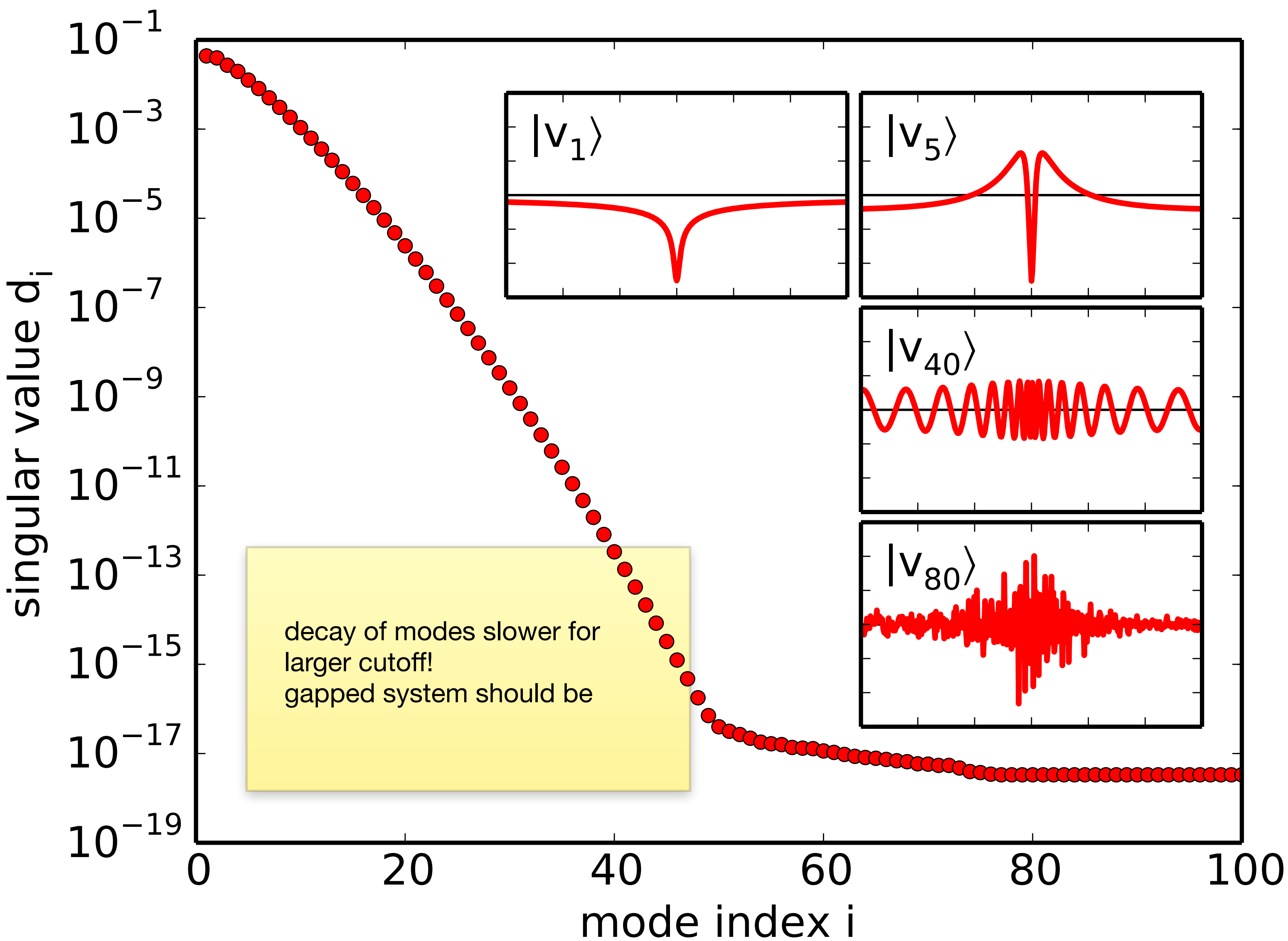
problem solved!

least-squares fit



why ill-conditioned?

$$\tilde{g} = \tilde{g}_{\text{exact}} + \tilde{g}_{\text{noise}} \quad \Rightarrow \quad |f_{\text{LS}}\rangle = |f_{\text{exact}}\rangle + \sum_i \frac{\langle u_i|\tilde{g}_{\text{noise}}\rangle}{d_i} |v_i\rangle$$



modes with small singular value

- responsible for noise in model
- hardly represented in data

$$\int K(y,x)\, v_i(x)\, dx = d_i\, u_i(y)$$

oscillate rapidly
⇒ necessary to resolve sharp features

limit contribution of these modes by imposing non-negativity: $A(\omega) \geq 0$

non-negative least-squares

$$p(\boldsymbol{f}|\,\tilde{\boldsymbol{g}}) = \frac{p(\tilde{\boldsymbol{g}}|\,\boldsymbol{f})\,p(\boldsymbol{f})}{p(\tilde{\boldsymbol{g}})}$$

$$\text{prior:}\;\; p(\boldsymbol{f}) = \begin{cases} \text{const.} & \text{for } \boldsymbol{f} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

constrained minimization:

true extremum

$$f_n > 0 \text{ and } \frac{\partial \chi^2(\boldsymbol{f})}{\partial f_n} = 0$$

boundary

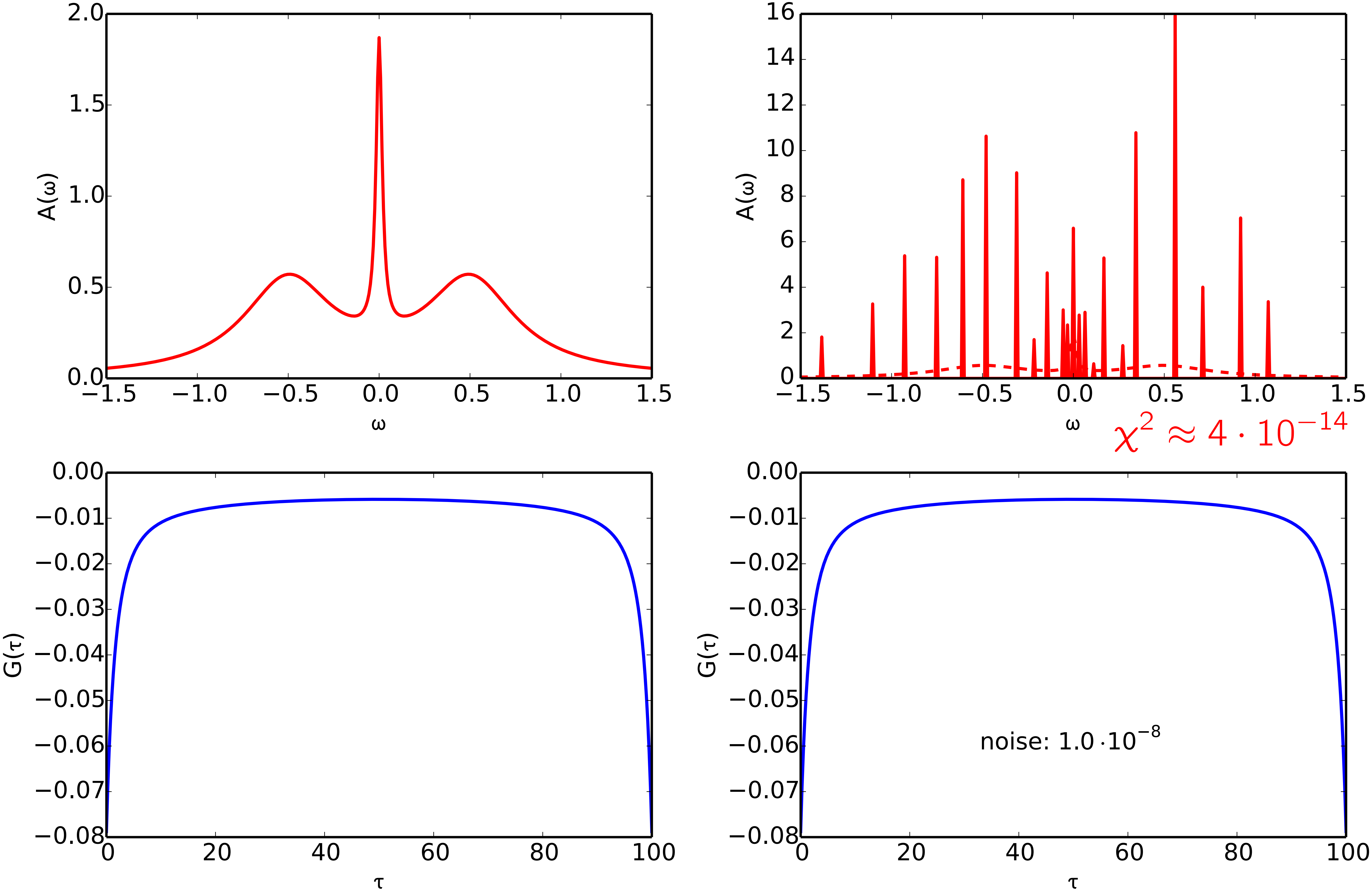
$$f_n = 0 \text{ and } \frac{\partial \chi^2(\boldsymbol{f})}{\partial f_n} \geq 0$$

Karush-Kuhn-Tucker conditions

naive algorithm: check all partitions; better: Appendix 2

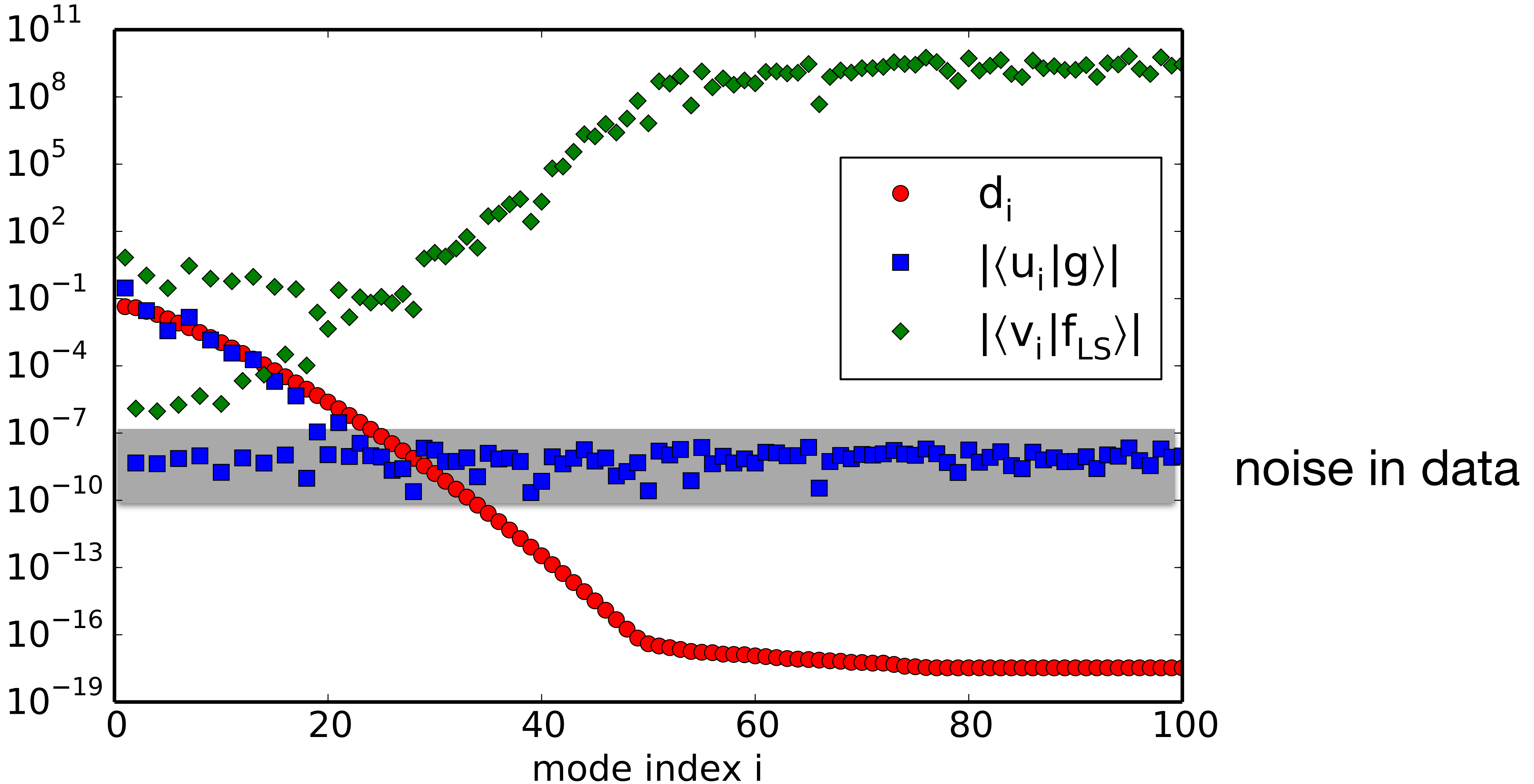
now modes with d_i=0 are well defined! [**well posed**]
help d_i>0 modes to move as close to optimum as possible
without violating constraint

non-negative least-squares fit



Picard plot

problem with LS/NNLS: **overfitting** noise in data



Picard condition: $|\langle \boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle| \lesssim d_i$

modes with small singular value contribute little

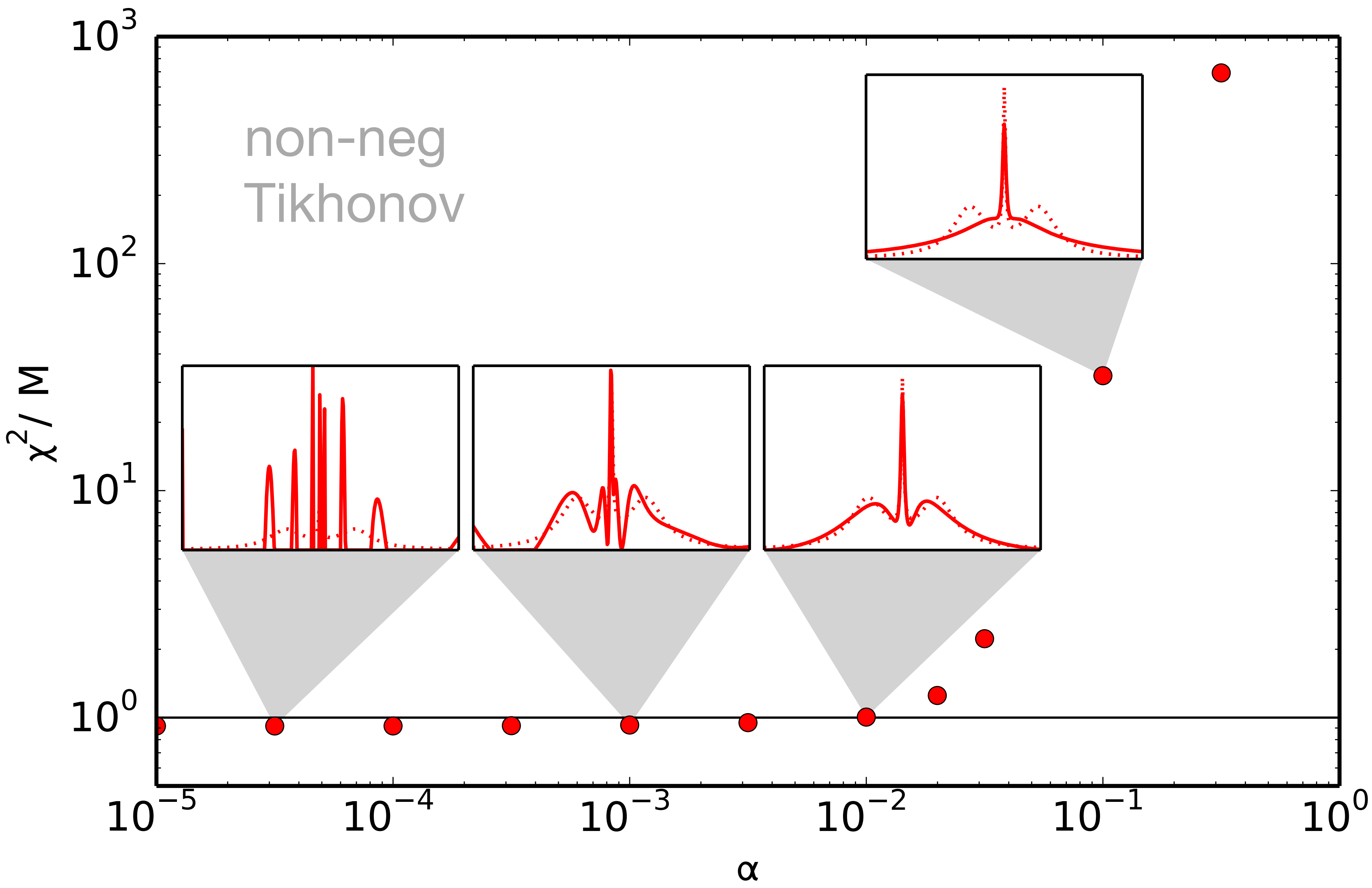
otherwise data contains little information about $\boldsymbol{f}$

regularization

avoid fitting noise in data

Tikhonov regularization $$\boldsymbol{f}_{\mathrm{T}}(\alpha) = \sum_{i=1}^{M} \frac{d_i}{d_i^2 + \alpha^2} \langle \boldsymbol{u}_i | \tilde{\boldsymbol{g}} \rangle$$

regularization parameter α scales with noise in data



discrepancy principle

regularization methods

Tikhonov: minimize $||\tilde{\boldsymbol{K}}\boldsymbol{f} - \tilde{\boldsymbol{g}}||^2 + \alpha^2||\boldsymbol{f}||^2$

$\Rightarrow$ prior

use smarter regularizers?

suppress modes based on first derivative

$$\sum_{n=1}^{N-1} |f_n - f_{n+1}|^2 = \langle \boldsymbol{f}| \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix} |\boldsymbol{f}\rangle$$

would also affect stable modes with large d_i

Tikhonov is more adaptive: suppresses highly oscillating modes with small d_i

reasonable modification: $||\boldsymbol{f}||^2 \Rightarrow \sum_n f_n^2/\rho_n^2$

default model (result in absence of data $\boldsymbol{f} \propto \boldsymbol{\rho}$)

maximum entropy

maximum entropy

$$\min_{\boldsymbol{f}} \left(||\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}||^2 - \alpha H(\boldsymbol{f};\boldsymbol{\rho}) \right)$$

entropy: non-linear regularizer

$$H(\boldsymbol{f};\boldsymbol{\rho}) = -\sum_n \left(f_n \ln \frac{f_n}{\rho_n} - f_n + \rho_n \right)$$

default model $\boldsymbol{\rho}$ and non-negativity
(linearization about default model gives Tikhonov)

how to choose regularization parameter?

- discrepancy principle: historic MaxEnt
- most probable value: classic MaxEnt
- average value: Bryan's MaxEnt

Bryan's method

determine regularization parameter from Bayes' theorem

$$p(B|\,A)\,p(A) = p(A,B) = p(A|\,B)\,p(B)$$

MaxEnt prior: $p(\boldsymbol{f}\,|\boldsymbol{\rho},\alpha) \propto \mathrm{e}^{+\alpha H(\boldsymbol{f};\boldsymbol{\rho})}$

$$\boldsymbol{f}(\alpha): \quad \min_{\boldsymbol{f}} \left(\chi^2(\boldsymbol{f}) - \alpha H(\boldsymbol{f};\boldsymbol{\rho})\right)$$

$$\boldsymbol{f}_{\mathrm{Bryan}} = \int_0^\infty d\alpha\, \boldsymbol{f}(\alpha)\, p(\alpha\,|\tilde{\boldsymbol{g}},\rho)$$

probability for regularizer:

$$p(\alpha\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho}) = \int \prod_n \frac{df_n}{\sqrt{f_n}}\; p(\boldsymbol{f},\alpha\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho})$$

functional integral replaced by Gaussian approximation…

average spectrum

The Average Spectrum Method for the Analytic Continuation of Imaginary-Time Data

S.R. White

Department of Physics, University of California, Irvine, CA92717, USA

Springer Proceedings in Physics, Vol. 53 145
Computer Simulation Studies in Condensed Matter Physics III
Editors: D.P. Landau · K.K. Mon · H.-B. Schüttler © Springer-Verlag Berlin, Heidelberg 1991

1991

In the average spectrum method, we avoid the need to introduce biases in $\Pr[A]$ and avoid any adjustable parameters by abandoning the maximum likelihood method and instead *average* over $\Pr[A|G^{\rm data}]$. We calculate the average spectrum via

unbiased

$$\langle A(p,\omega)\rangle = \int \mathcal{D}A(p,\omega)\Pr[A|G^{\rm data}]A(p,\omega) \qquad (9)$$

$$f_{\rm ASM} \propto \int_{f(x)\geq 0} \mathcal{D}f\; f(x)\, e^{-\chi^2[f]/2}$$

with

$$\Pr[A] = \begin{cases} \text{const.}, & \text{if } A(p,\omega) \geq 0 \text{ for all } \omega; \\ 0, & \text{otherwise.} \end{cases} \qquad (10)$$

The integral here is a path integral over all positive definite spectra. Each path in the integral is a possible spectrum weighted by the probability of that spectrum, given the data. The averaging over spectra automatically smears out statistically insignificant features. Because of this we need only include in $\Pr[A]$ what we really know: that the result is positive definite and in some cases that it is an even function. The average spectrum method makes fewer assumptions and is conceptually simpler than any of the maximum likelihood methods; the only disadvantage is that the path integral may be somewhat more time-consuming to compute. In the next section we discuss numerical

noise regularizes

The Average Spectrum Method for the Analytic Continuation of Imaginary-Time Data

S.R. White

Department of Physics, University of California, Irvine, CA92717, USA

Springer Proceedings in Physics, Vol. 53
Computer Simulation Studies in Condensed Matter Physics III
Editors: D.P. Landau · K.K. Mon · H.-B. Schüttler

The average spectrum method makes fewer assumptions and is conceptually simpler than any of the maximum likelihood methods; the only disadvantage is that the path integral may be somewhat more time-consuming to compute.

PHYSICAL REVIEW B VOLUME 57, NUMBER 17 1 MAY 1998-I

Stochastic method for analytic continuation of quantum Monte Carlo data

Anders W. Sandvik
Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801
(Received 28 October 1997)

A method for analytic continuation of quantum Monte Carlo data is presented. The spectrum $A(\omega)$ is parametrized as a sum of δ functions, the weights A_i of which are sampled according to a distribution $p(A) \sim \exp(-\chi^2/\Theta)$. It is argued that the calculated entropy S provides a criterion for determining the Θ corresponding to the ''best'' averaged spectrum. The appearance of spurious structure is signaled by a sharp drop in $\langle S(\Theta)\rangle$, which in test cases is preceded by a local maximum. Results for the dynamic spin structure factor of a 16-site Heisenberg chain obtained at this maximum are in better agreement with exact results than ''classic'' maximum-entropy results. [S0163-1829(98)02717-9]

before the method can be applied to more complicated spectra than the single-maximum case considered here. A problem for practical use of the method is that the sampling needed for an accurate determination of Θ^* as well as the averaging needed to obtain a final result are quite time consuming. The good agreement with the exact results obtained here should motivate further work along these lines.

PHYSICAL REVIEW B **76**, 035115 (2007)

Analytical continuation of spectral data from imaginary time axis to real frequency axis using statistical sampling

K. Vafayi and O. Gunnarsson
Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany
(Received 27 February 2007; revised manuscript received 2 May 2007; published 19 July 2007)

We present a method for performing analytical continuation of spectral data from imaginary time to real frequencies based on a statistical sampling method. Compared with the maximum entropy method (MEM), an advantage is that no default model needs to be introduced. For the problems studied here, the statistical sampling method gives comparable or slightly better results than MEM using quite accurate default models.

DOI: 10.1103/PhysRevB.76.035115 PACS number(s): 72.15.Eb, 02.70.Ss

We have presented a statistical sampling method, where the optical conductivity $\sigma(\omega)$ is averaged using the probability in Eq. (7) as a weight function. Comparing with the maximum entropy method (MEM), an advantage is that there is no need to provide a default model, which influences the MEM results if the method is close to its limit of applicability. For the problems considered here, the statistical sampling method gives comparable or slightly better results than MEM using default models close to the exact result. The price we have to pay is that the method is many orders of magnitude slower than MEM. With present day computers this is not a serious drawback, in particular, since the time spent for the analytical continuation is typically small compared with the time needed for the QMC calculation of the $J(\tau)$ data.

PHYSICAL REVIEW E **81**, 056701 (2010)

Analytic continuation of quantum Monte Carlo data by stochastic analytical inference

Sebastian Fuchs* and Thomas Pruschke
Institut für Theoretische Physik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Mark Jarrell
Louisiana State University, Baton Rouge, Louisiana 70803, USA

(Received 3 January 2010; revised manuscript received 11 March 2010; published 4 May 2010)

We present an algorithm for the analytic continuation of imaginary-time quantum Monte Carlo data which is strictly based on principles of Bayesian statistical inference. Within this framework we are able to obtain an explicit expression for the calculation of a weighted average over possible energy spectra, which can be evaluated by standard Monte Carlo simulations, yielding as by-product also the distribution function as function of the regularization parameter. Our algorithm thus avoids the usual *ad hoc* assumptions introduced in similar algorithms to fix the regularization parameter. We apply the algorithm to imaginary-time quantum Monte Carlo data and compare the resulting energy spectra with those from a standard maximum-entropy calculation.

One apparent drawback of the method is the necessity to perform simulations for a broad range of values for α, independent of whether one chooses the Wang-Landau approach or $\chi^2 \sim N$, respectively, $\alpha = 1$ [23] to fix α. Although this can be performed efficiently with parallel tempering techniques, the required computer resources for one single spectrum can sum up to about 20 processor hours and are hence orders of magnitude larger than for standard MEM approaches. Especially for QMC data at higher temperatures, more computer time may be needed for the analytic continuation than for the simulation of the Monte Carlo data itself. As the resulting spectra tend to be less regularized one has to ponder the gain in details in the structures against the significant increase in computer time.

sampling

$$f_{\mathrm{ASM}}(\tilde{\boldsymbol{g}}; x) \propto \int_{f(x)\geq 0} \mathcal{D}f \; f(x)\, e^{-\frac{1}{2}\chi^2[f]}$$

discretize $\longrightarrow \prod_{n=1}^{N} \int_0^{\infty} df_n \, e^{-\frac{1}{2}||\tilde{\boldsymbol{g}}-\tilde{\boldsymbol{K}}\boldsymbol{f}||^2} \; \boldsymbol{f}$

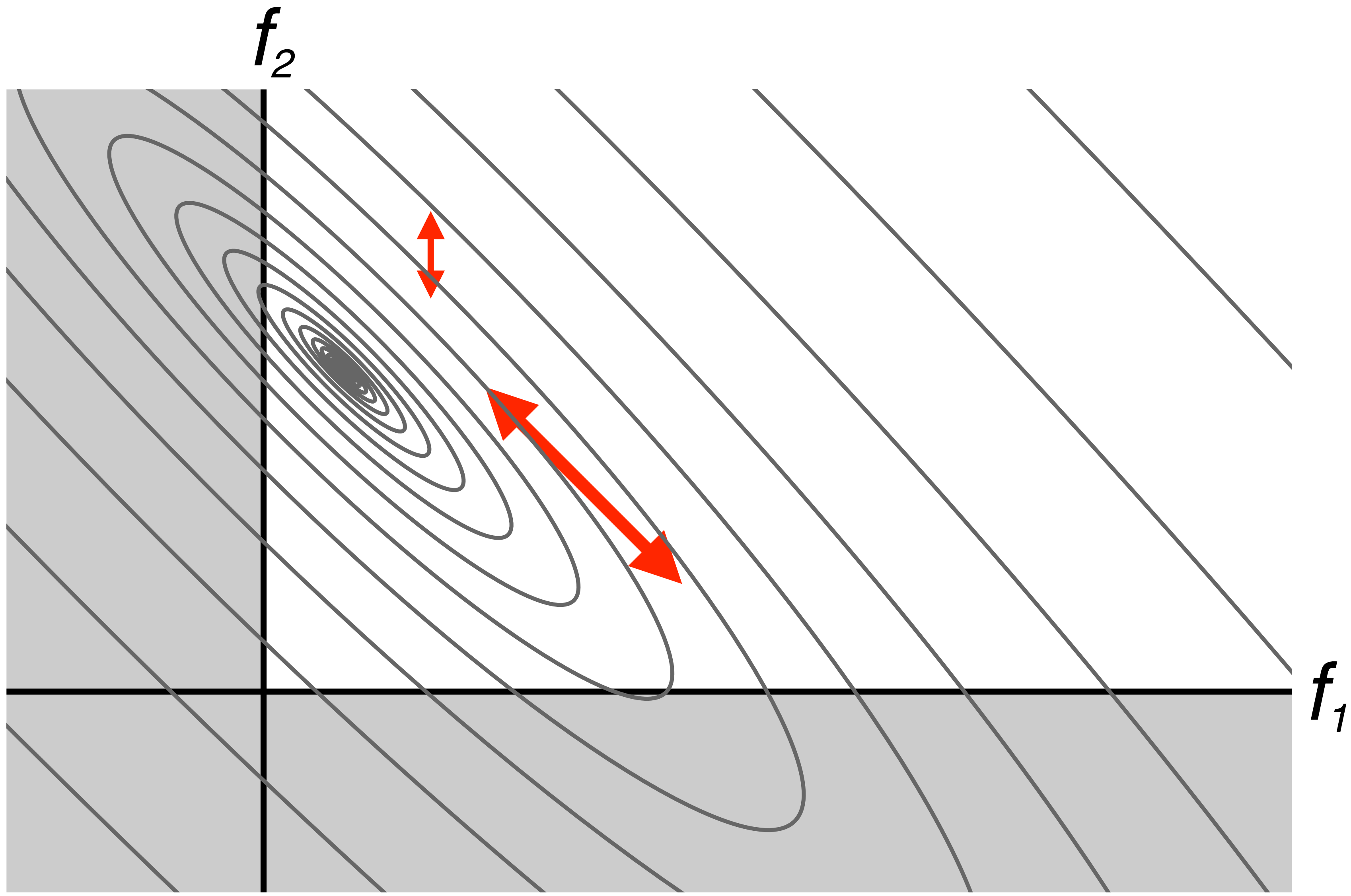
Gibbs sampling: sample component $f_n \rightarrow f_n' \in [0,\infty)$ keeping all others fixed

$$\chi^2(\boldsymbol{f}; f_n') = ||\underbrace{\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f} + \tilde{\boldsymbol{K}}_n f_n}_{=:\tilde{\boldsymbol{g}}_n} - \tilde{\boldsymbol{K}}_n f_n'||^2 = \underbrace{\tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{K}}_n}_{1/\sigma^2} \left(f_n' - \underbrace{\frac{\tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{g}}_n}{\tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{K}}_n}}_{=\mu} \right)^2$$

sample truncated Gaussian, quite slow since σ very narrow

sampling

sample truncated Gaussian, still quite slow since σ very narrow



better to sample along principal axes of $\boldsymbol{\chi}^2$

modes sampling

unitary transformations to singular modes

$$\boldsymbol{h} := \tilde{\boldsymbol{U}}^\dagger \tilde{\boldsymbol{g}} \qquad \chi^2(\boldsymbol{f}) = ||\tilde{\boldsymbol{U}}^\dagger \tilde{\boldsymbol{g}} - \tilde{\boldsymbol{S}}\tilde{\boldsymbol{V}}^\mathsf{T} \boldsymbol{f}||^2 = \sum_{i=1}^{M} (h_i - s_i\, e_i)^2$$

$$\boldsymbol{e} := \tilde{\boldsymbol{V}}^\mathsf{T} \boldsymbol{f}$$

$$\prod_{n=1}^{N} \int_0^\infty df_n\, e^{-\frac{1}{2}||\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}||^2}\, \boldsymbol{f} \;\; \rightarrow \;\; \prod_{i=1}^{N} \int_{\boldsymbol{f} \geq 0} de_i\, e^{-\frac{(s_i\, e_i - h_i)^2}{2}}\, e_i$$

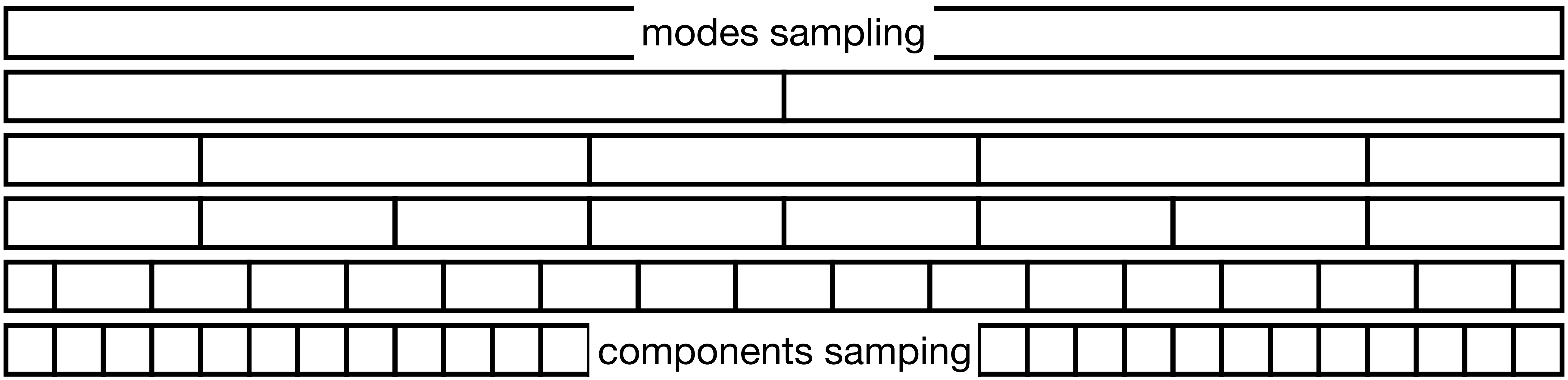
non-negativity constraint $f_n \geq 0$

$$\boldsymbol{f}' = \boldsymbol{f} + (e_i' - e_i)\boldsymbol{V}_i \geq 0 \;\; \Rightarrow \;\; \max\left\{ \frac{f_n}{V_{ni}} \middle| V_{ni} < 0 \right\} \leq e_i - e_i' \leq \min\left\{ \frac{f_n}{V_{ni}} \middle| V_{ni} > 0 \right\}$$

optimal sampling, except when f_n becomes small (e.g. in tail)

blocked modes sampling

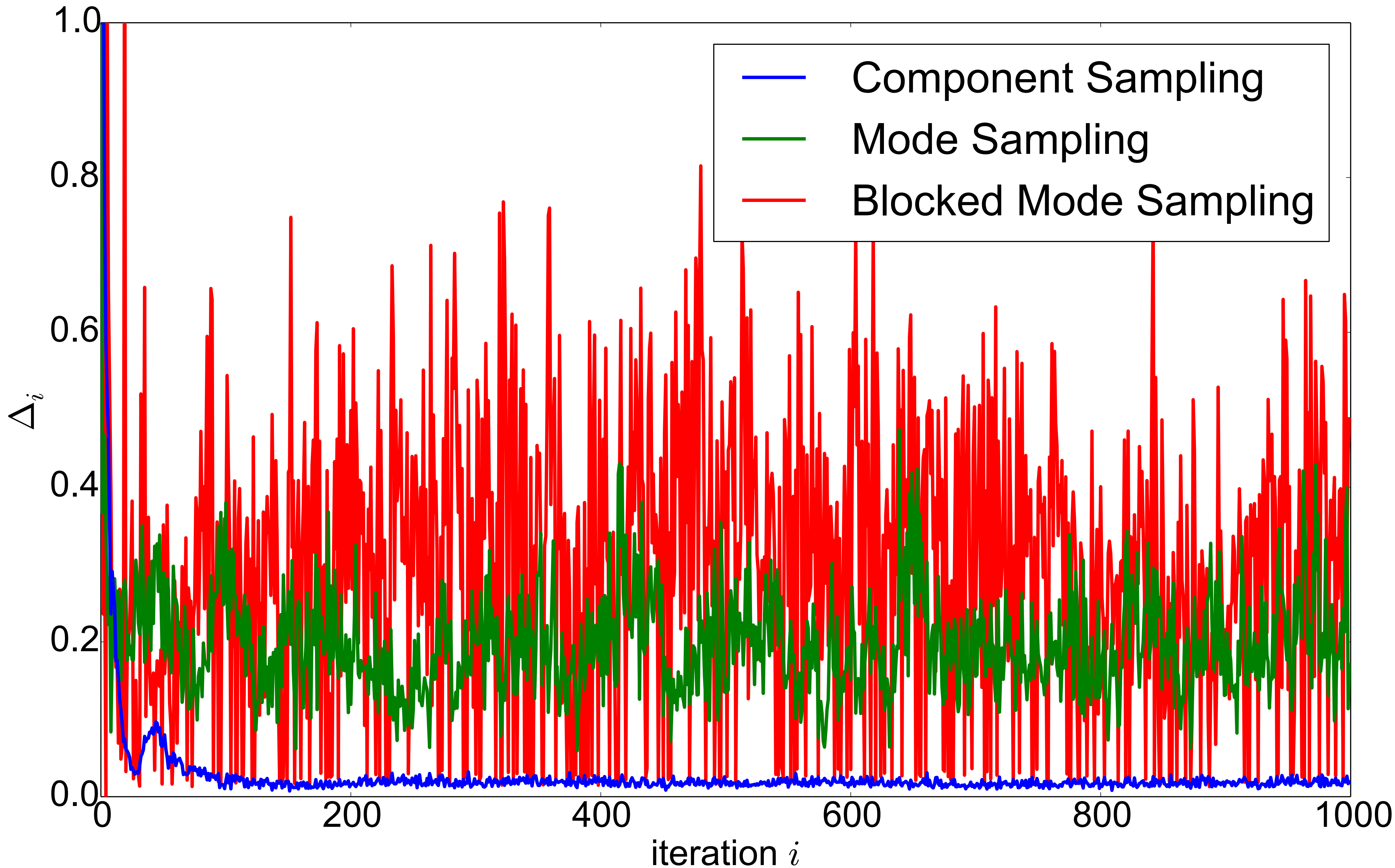
introduce hierarchy of grid partitionings in blocks
and sample modes on blocks



in each update choose random level in hierarchy
and sample modes on blocks

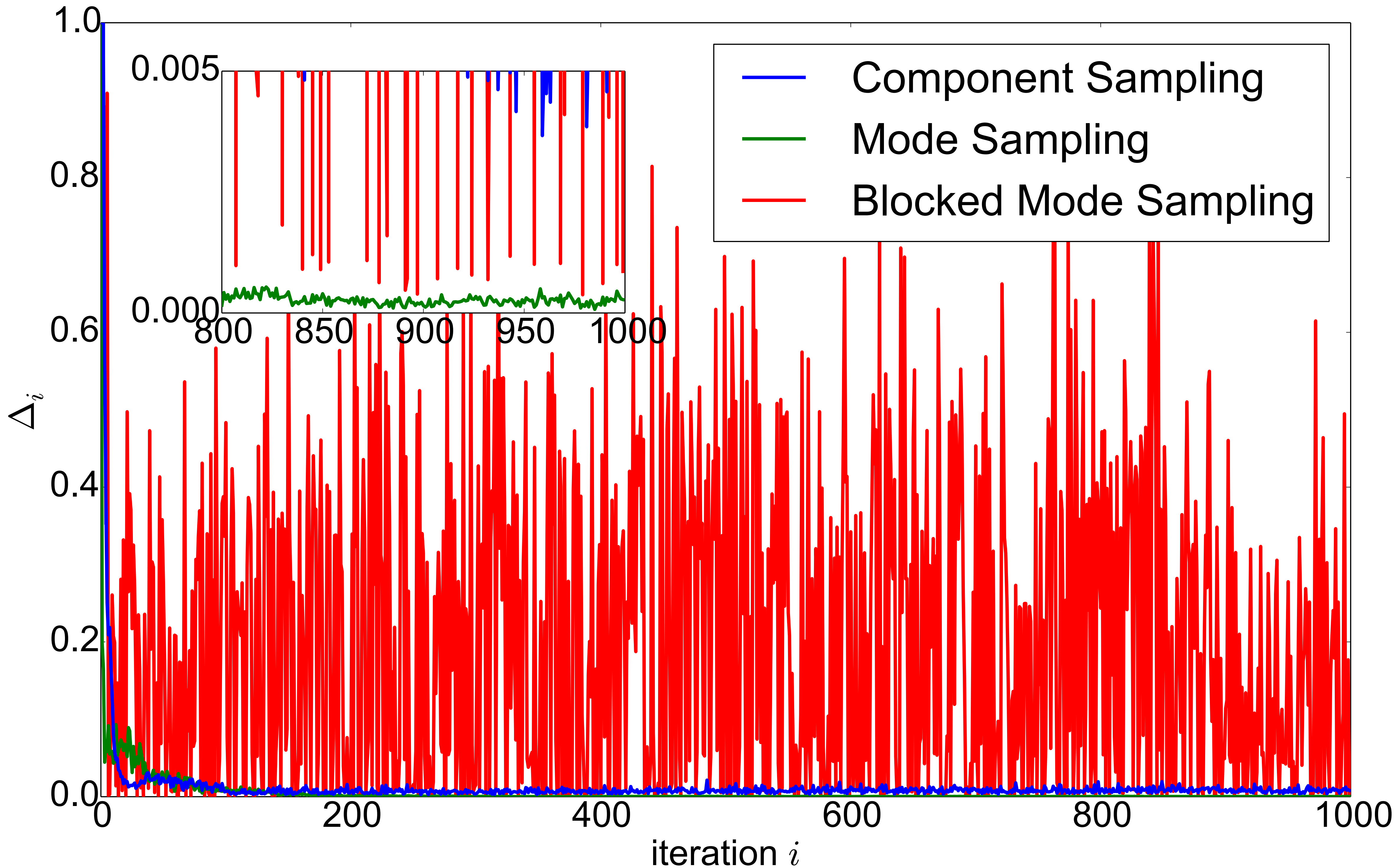
efficiency: small cutoff

step size: $\Delta_i = ||\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i-1}||/||\boldsymbol{f}^{(i)}||$



efficiency: large cutoff

step size: $\Delta_i = ||\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i-1)}||/||\boldsymbol{f}^{(i)}||$



test cases

PHYSICAL REVIEW B **82**, 165125 (2010)

Analytical continuation of imaginary axis data for optical conductivity

O. Gunnarsson,[1] M. W. Haverkort,[1] and G. Sangiovanni[2]
[1]*Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany*
[2]*Institut für Festkörperphysik, Technische Universität Wien, Vienna, Austria*
(Received 24 June 2010; revised manuscript received 13 August 2010; published 28 October 2010)

We compare different methods for performing analytical continuation of spectral data from the imaginary time or frequency axis to the real frequency axis for the optical conductivity $\sigma(\omega)$. We compare the maximum entropy (MaxEnt), singular value decomposition (SVD), sampling, and Padé methods for analytical continuation. We also study two direct methods for obtaining $\sigma(0)$. For the MaxEnt approach we focus on a recent modification. The data are split up in batches, a separate MaxEnt calculation is done for each batch and the results are averaged. For the problems studied here, we find that typically the SVD, sampling, and modified MaxEnt methods give comparable accuracy while the Padé approximation is usually less reliable.

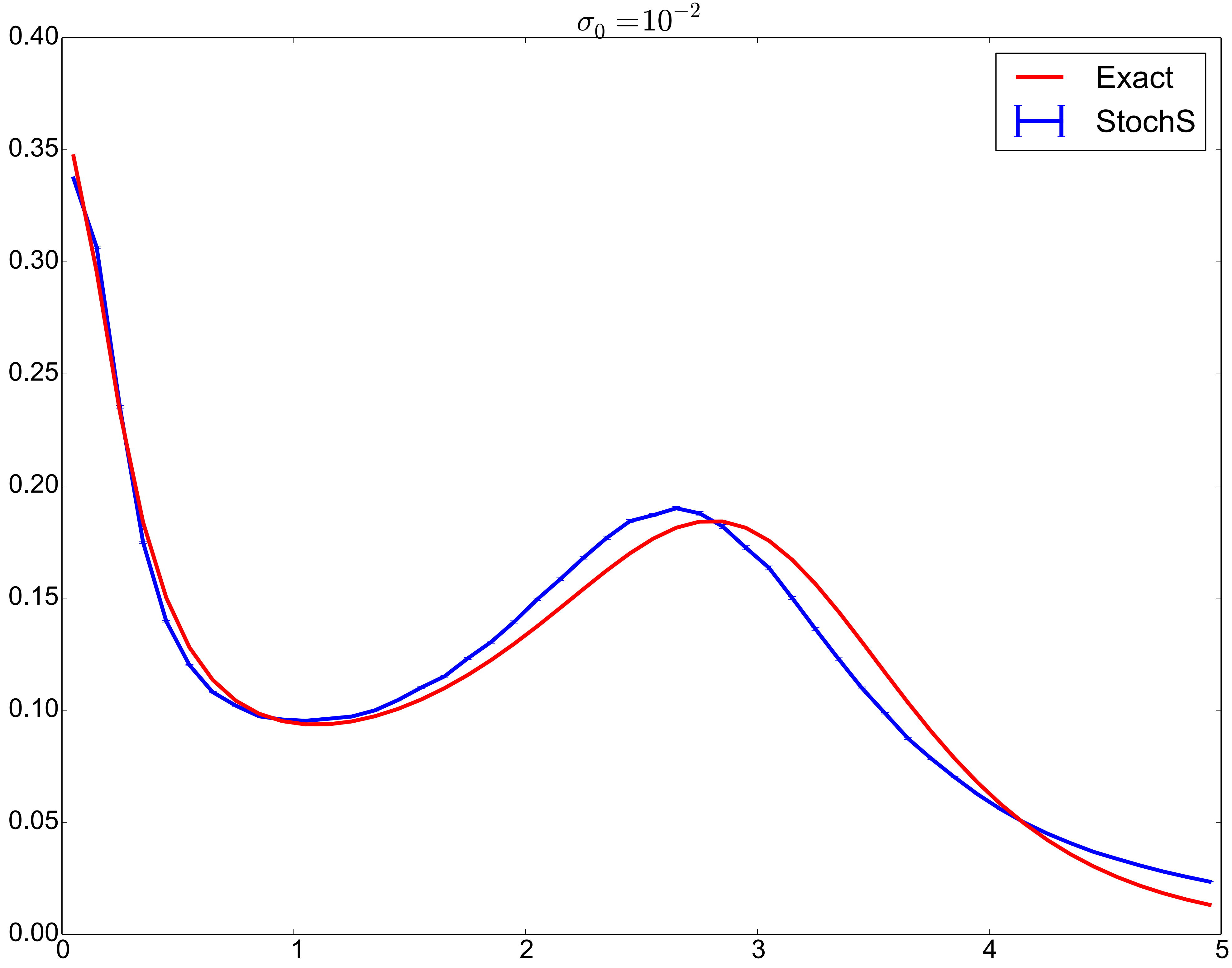
DOI: 10.1103/PhysRevB.82.165125 PACS number(s): 72.15.Eb, 02.70.Ss

optical conductivity $$\Pi(\nu_n) = \frac{2}{\pi}\int_0^\infty \frac{\omega^2}{\nu_n^2+\omega^2}\,\sigma(\omega)\,d\omega$$

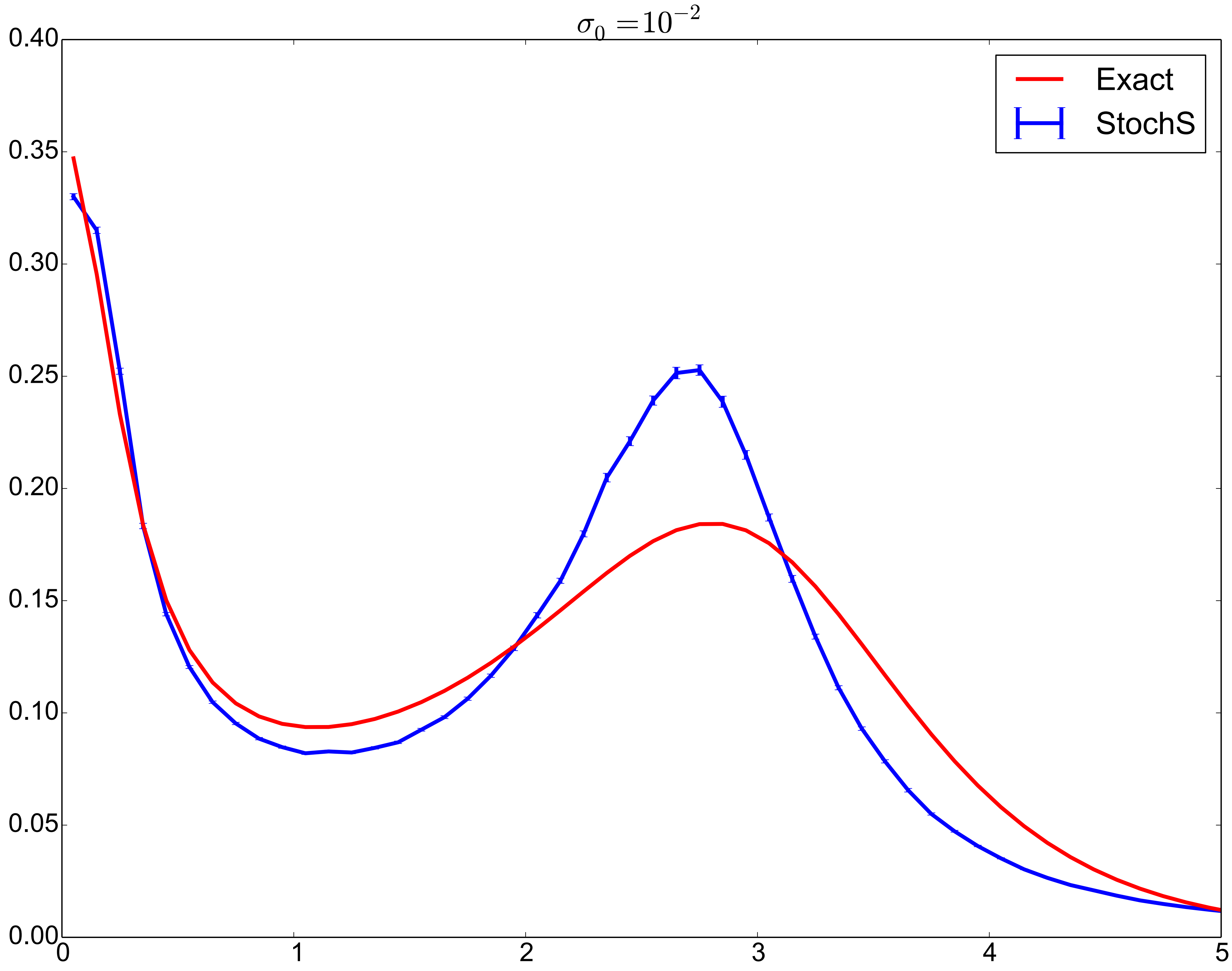
$$\sigma(\omega) = \left\{\frac{W_1}{1+(\omega/\Gamma_1)^2} + \frac{W_2}{1+[(\omega-\epsilon)/\Gamma_2]^2} + \frac{W_2}{1+[(\omega+\epsilon)/\Gamma_2]^2}\right\}\frac{1}{1+(\omega/\Gamma_3)^6}$$

$$\Gamma_1 = 0.3 \text{ or } 0.6, \Gamma_2 = 1.2, \Gamma_3 = 4, \epsilon = 3, W_1 = 0.3, W_2 = 0.2$$

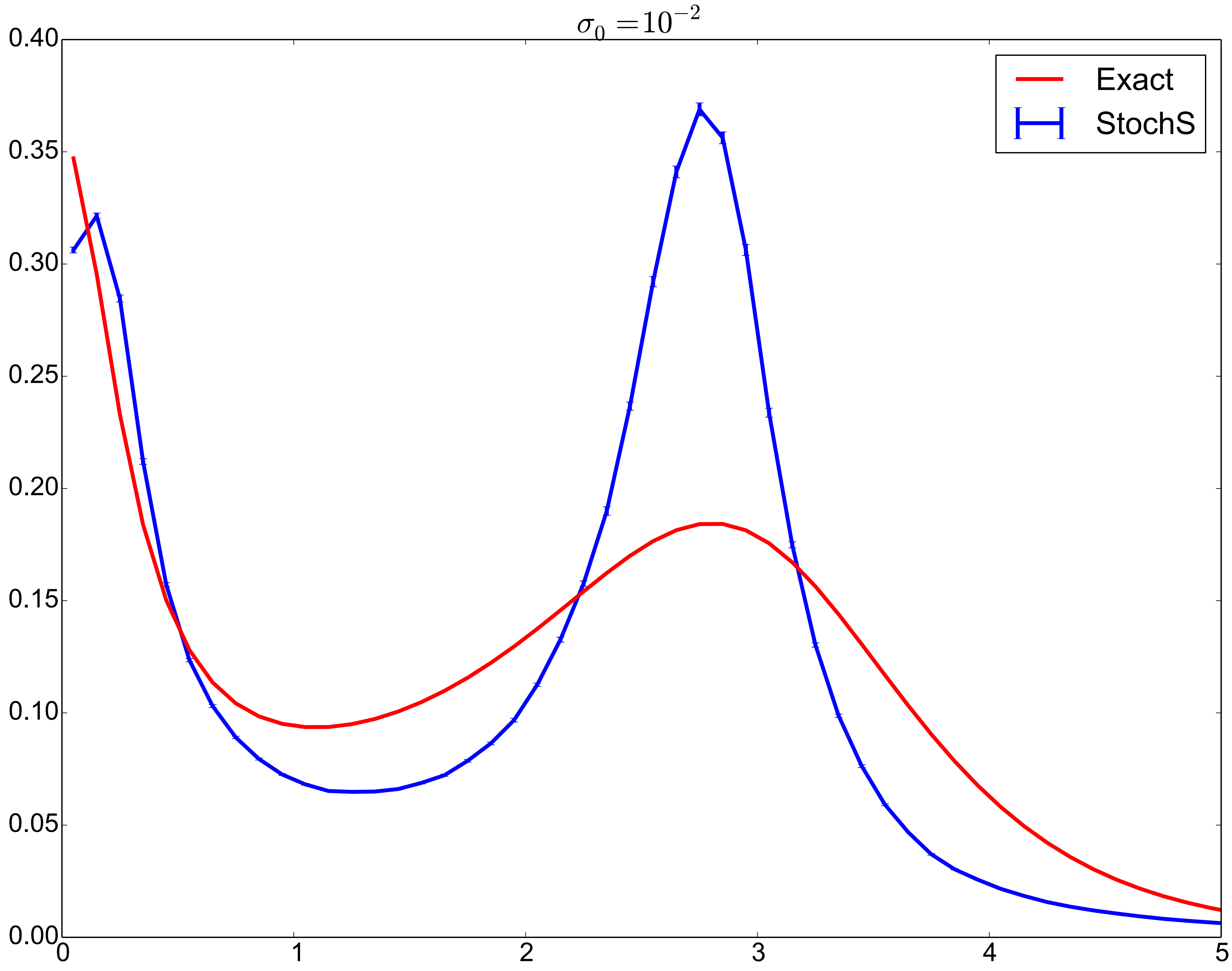
average spectrum



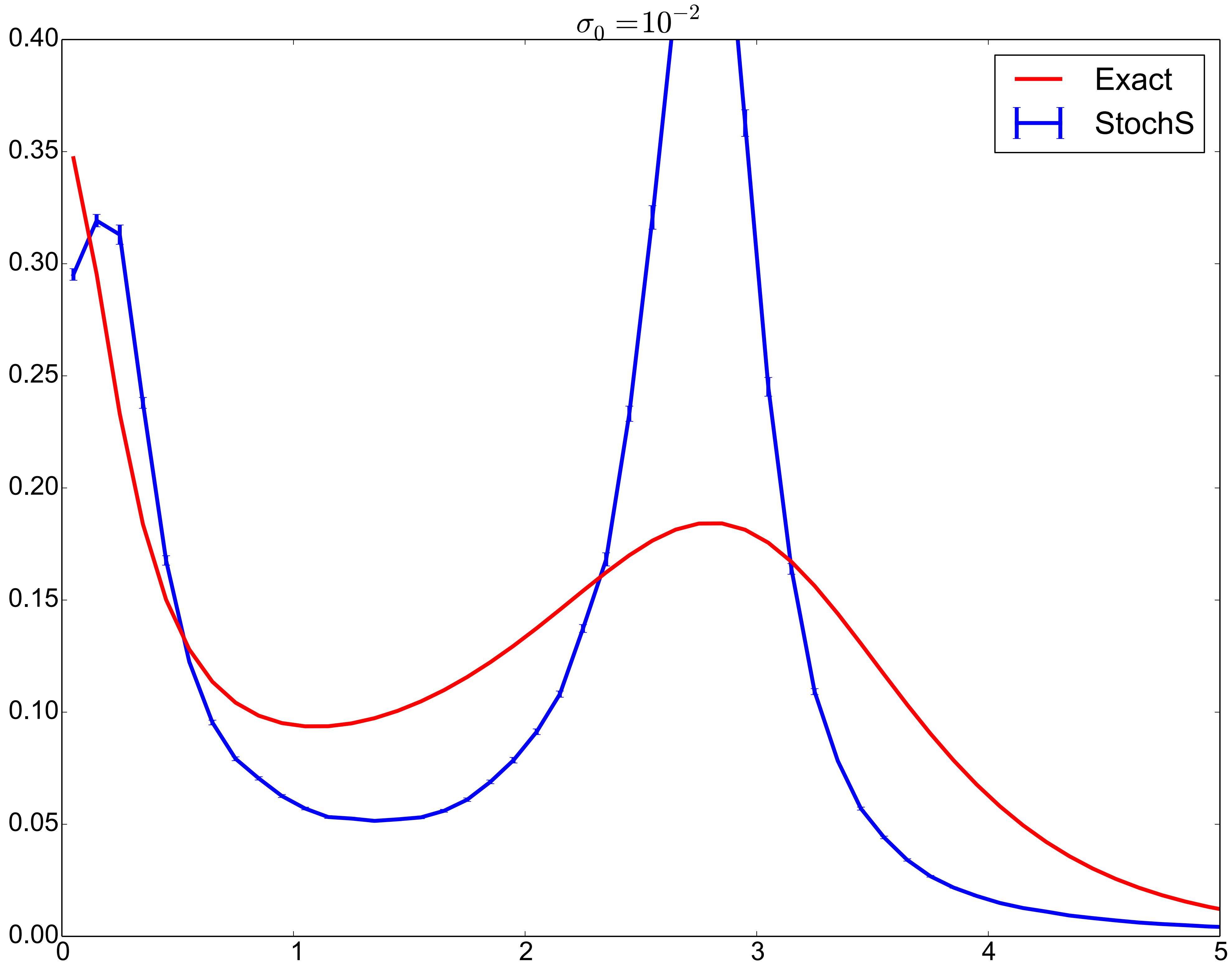
improve cut-off: $\omega_{max} = 8$



improve cut-off: $\omega_{max} = 16$



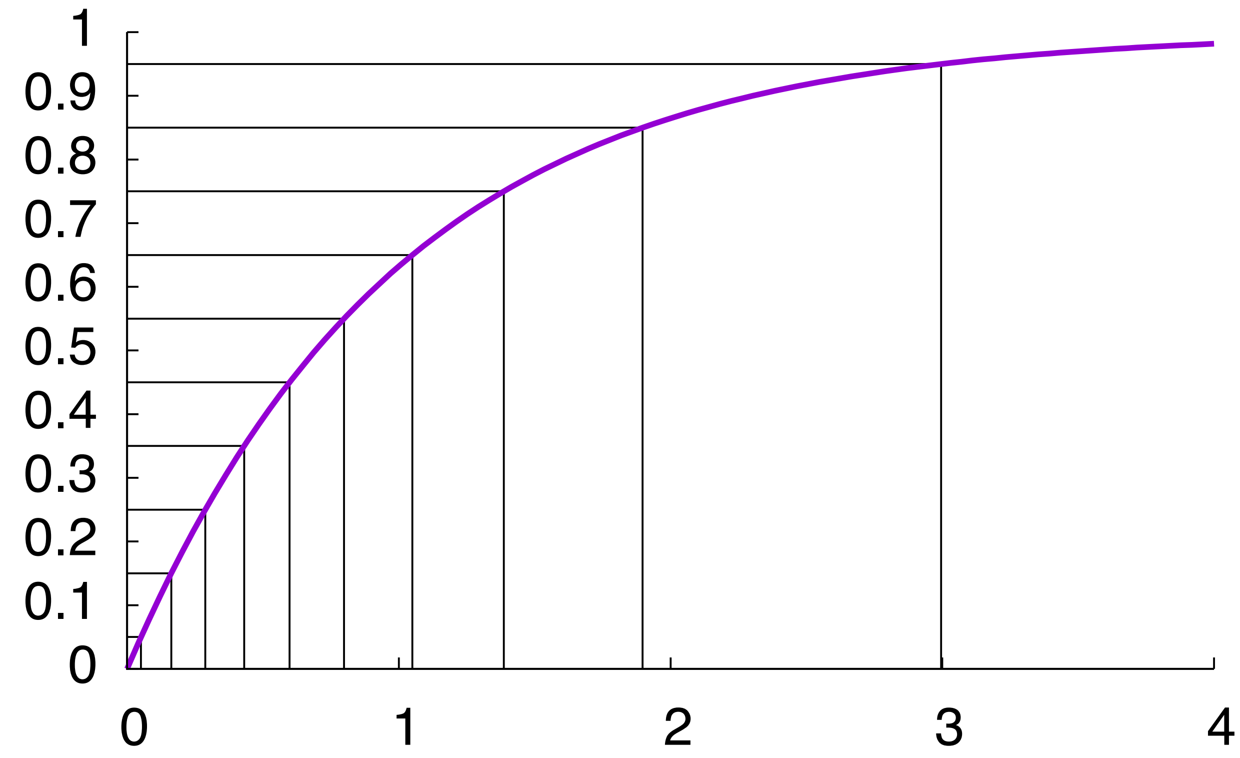
improve cut-off: $\omega_{max} = 32$



remove cutoff

variable transform

$$\omega \to z = \int_0^{\omega} \rho(\omega')\, d\omega'$$



exponential grid

$$\rho(\omega) = \lambda\, e^{-\lambda\omega}$$

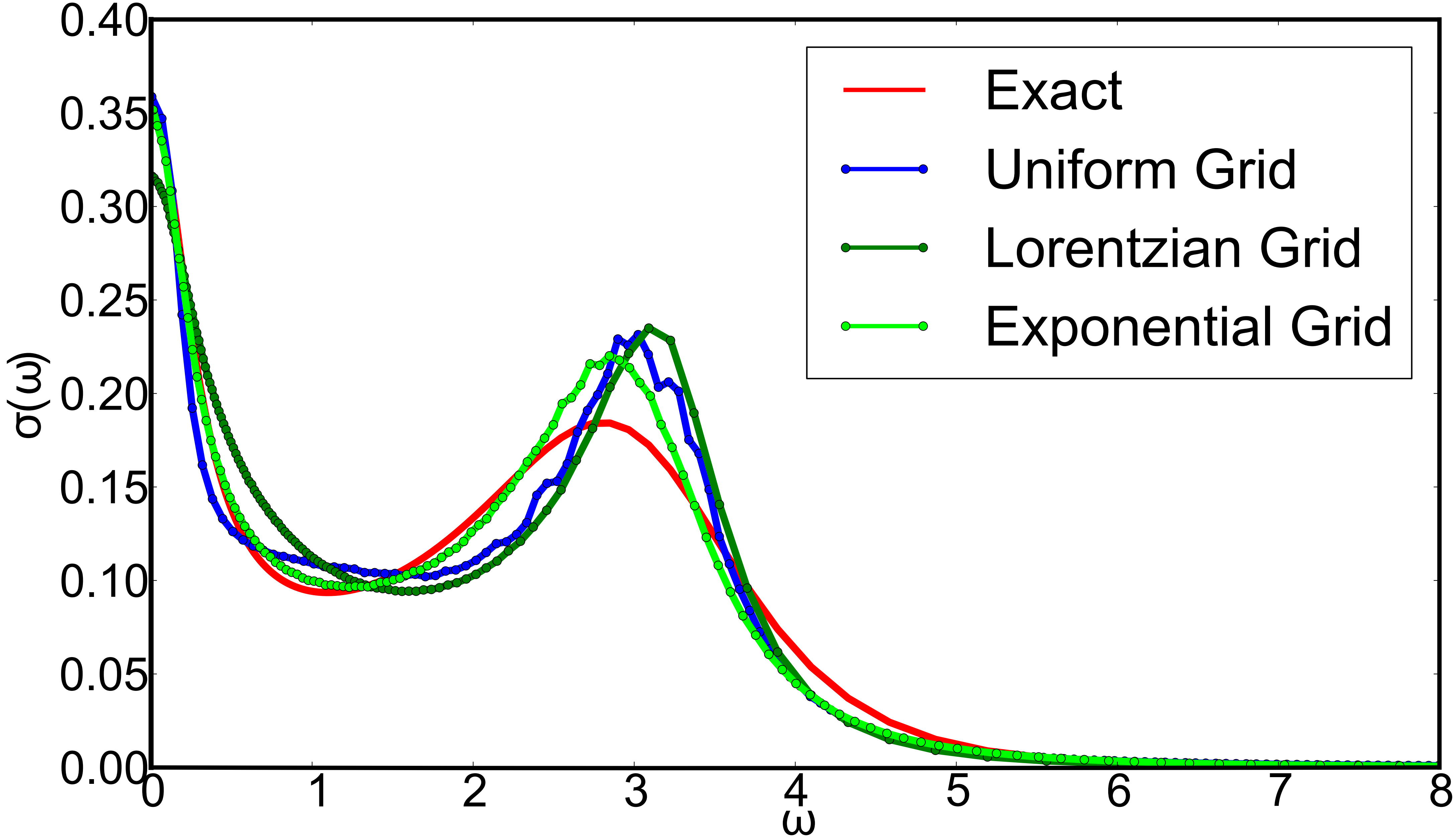
Lorentzian grid

$$\rho(\omega) = \frac{\gamma}{\pi} \frac{1}{\omega^2 + \gamma^2}$$

Gaussian grid

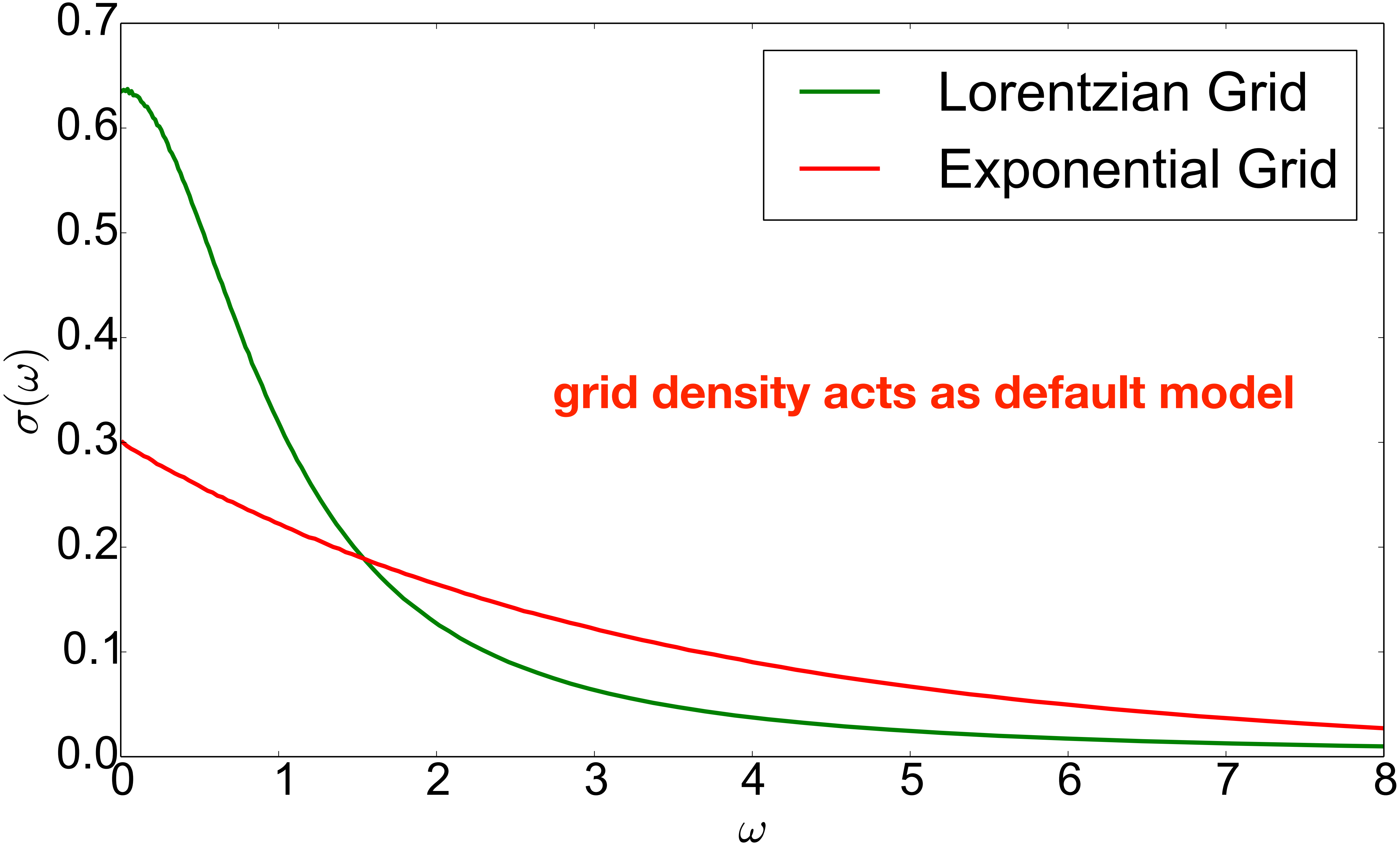
$$\rho(\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\omega^2/2\sigma^2}$$

grid dependence !?



default model

average spectral **without data** (except normalization)



consider different grids

$$\int_{f(x)\geq 0} \mathcal{D}f \, e^{-\frac{1}{2}\chi^2[f]} \, f(x) \to \prod_{n=1}^{N} \int_0^\infty df_n \, e^{-\frac{1}{2}||\tilde{\boldsymbol{g}}-\tilde{\boldsymbol{K}}\boldsymbol{f}||^2} \, \boldsymbol{f}$$

sample component $f_n \geq 0$ uniformly on grid interval n

inconsistent with choice of different grids

example: coarsening of grid interval $I = I_1 \cup I_2$

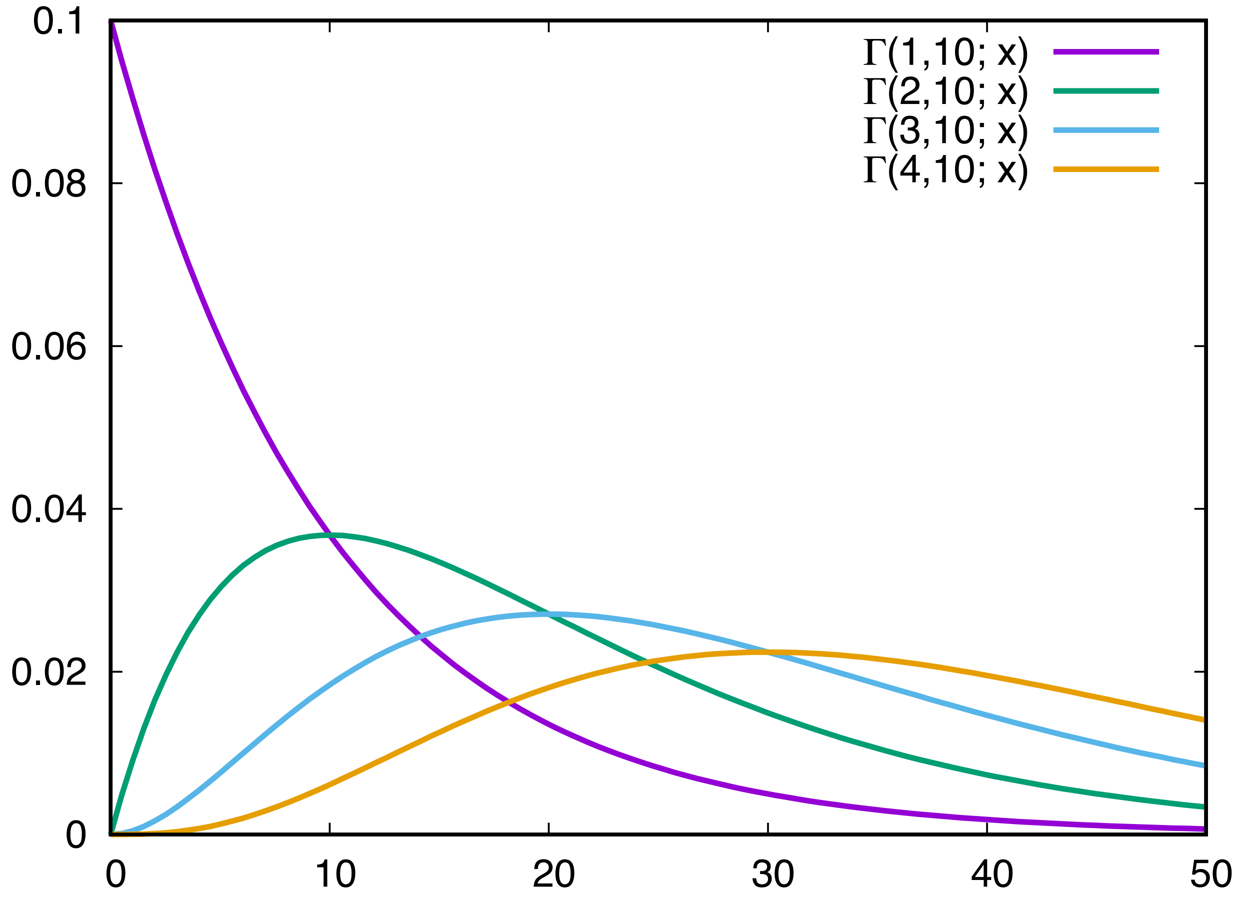
$$f = f_1 + f_2 \quad \Rightarrow \quad p_I(f) = \int_0^f df_1 \; p_1(f_1) \, p_2(f - f_1)$$

for uniform $p_n(f)=1$ we get $p_I(f)=f$

Gamma distribution

$$\Gamma(k,\lambda;x) = \frac{(x/\lambda)^{k-1} e^{-x/\lambda}}{\lambda \Gamma(k)}$$

$$\Gamma(1,\lambda;x) = \frac{1}{\lambda} e^{-x/\lambda}$$



$$\lim_{\lambda\to\infty} \Gamma(1,\lambda;x) = \text{uniform}$$

$$\int_0^\infty dx\, x \Gamma(k,\lambda;x) = \lambda k$$

$$\int_0^\infty dx\, x^n \Gamma(k,\lambda;x') = \lambda^n \frac{\Gamma(k+n)}{\Gamma(n)}$$

$$\int_0^x dx'\, \Gamma(k_1,\lambda;x')\, \Gamma(k_2,\lambda;x-x') = \Gamma(k_1+k_2,\lambda;x)$$

functional measure

sampling on grid with k-fold resolution: reweight with

$$\frac{p_{\mathrm{coarse}}(f_1, f_2, \ldots, f_{kN})}{p_{\mathrm{fine}}(f_1, f_2, \ldots, f_{kN})} \sim \lim_{\lambda\to\infty} \frac{\prod_n e^{-f_n/\lambda}}{\prod_n f_n^{k-1} e^{-f_n/\lambda}} = \prod_n f_n^{1-k}$$

$$\prod_{n=1}^{N} \int_0^\infty df_n \, e^{-\frac{1}{2}||\tilde{\boldsymbol{g}}-\tilde{\boldsymbol{K}}\boldsymbol{f}||^2} \boldsymbol{f} \to \prod_{n=1}^{kN}$$

check reweighing…

$$\tilde{\boldsymbol{K}}\boldsymbol{f}||^2 \boldsymbol{f} \to \int_{f(x)\geq 0} \mathcal{D}f \, e^{-\frac{1}{2}\chi^2[f]} f(x)$$

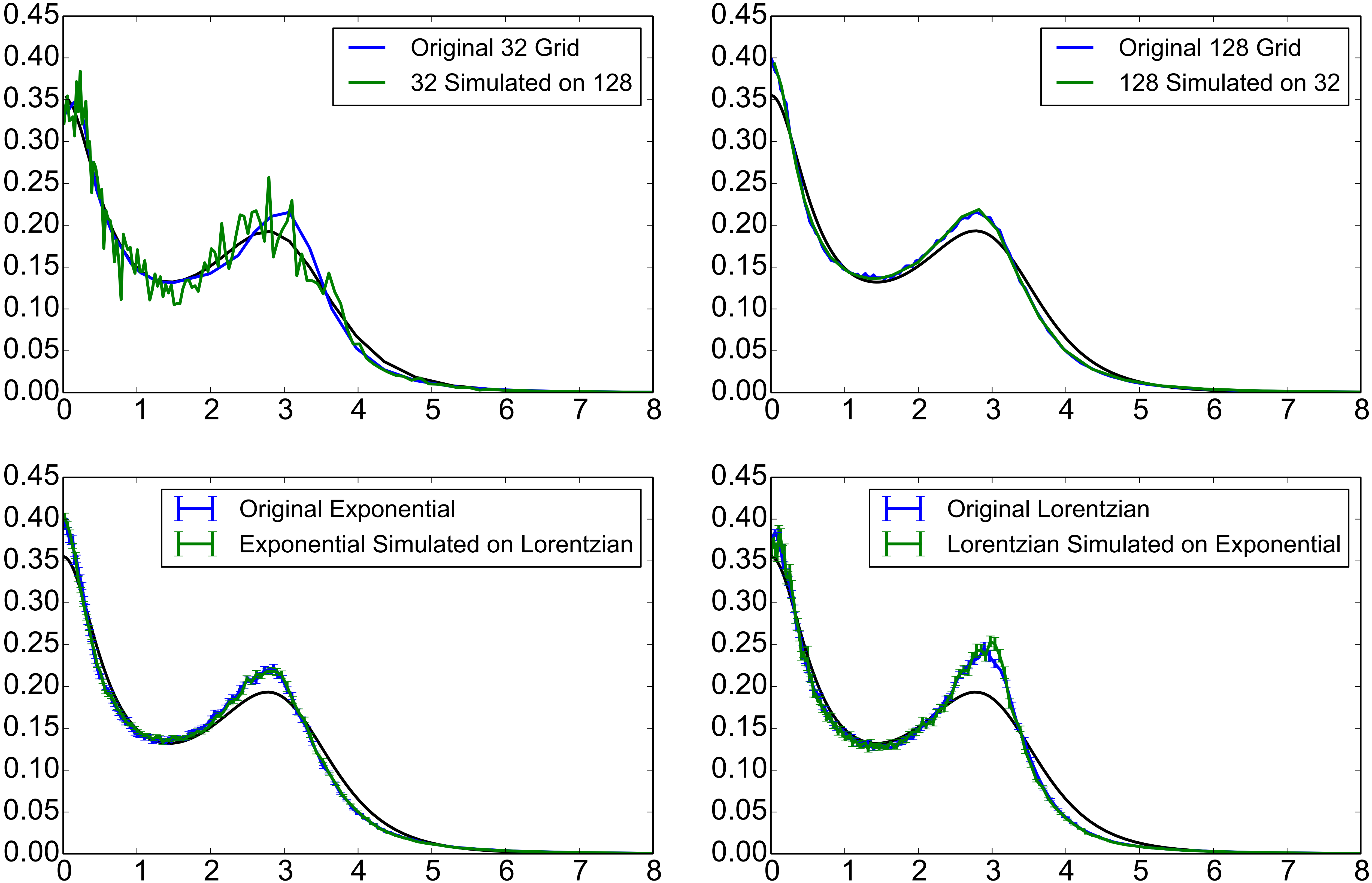
sampling uniformly on a grid implies a certain functional measure

reweighting for grids of different density

$$\frac{\tilde{p}(f_1, f_2, \ldots, f_N)}{p(f_1, f_2, \ldots, f_N)} = \prod_n f_n^{1-N\rho(x_n)/\tilde{N}\tilde{\rho}(x_n)}$$

separate grid from default model

simulating different grids



how to choose grid?

estimate width of model from NNLS

$$\mu = \int f(\omega)\, d\omega \quad \sigma^2 = \int (\omega - \mu)^2 f(\omega)\, d\omega$$

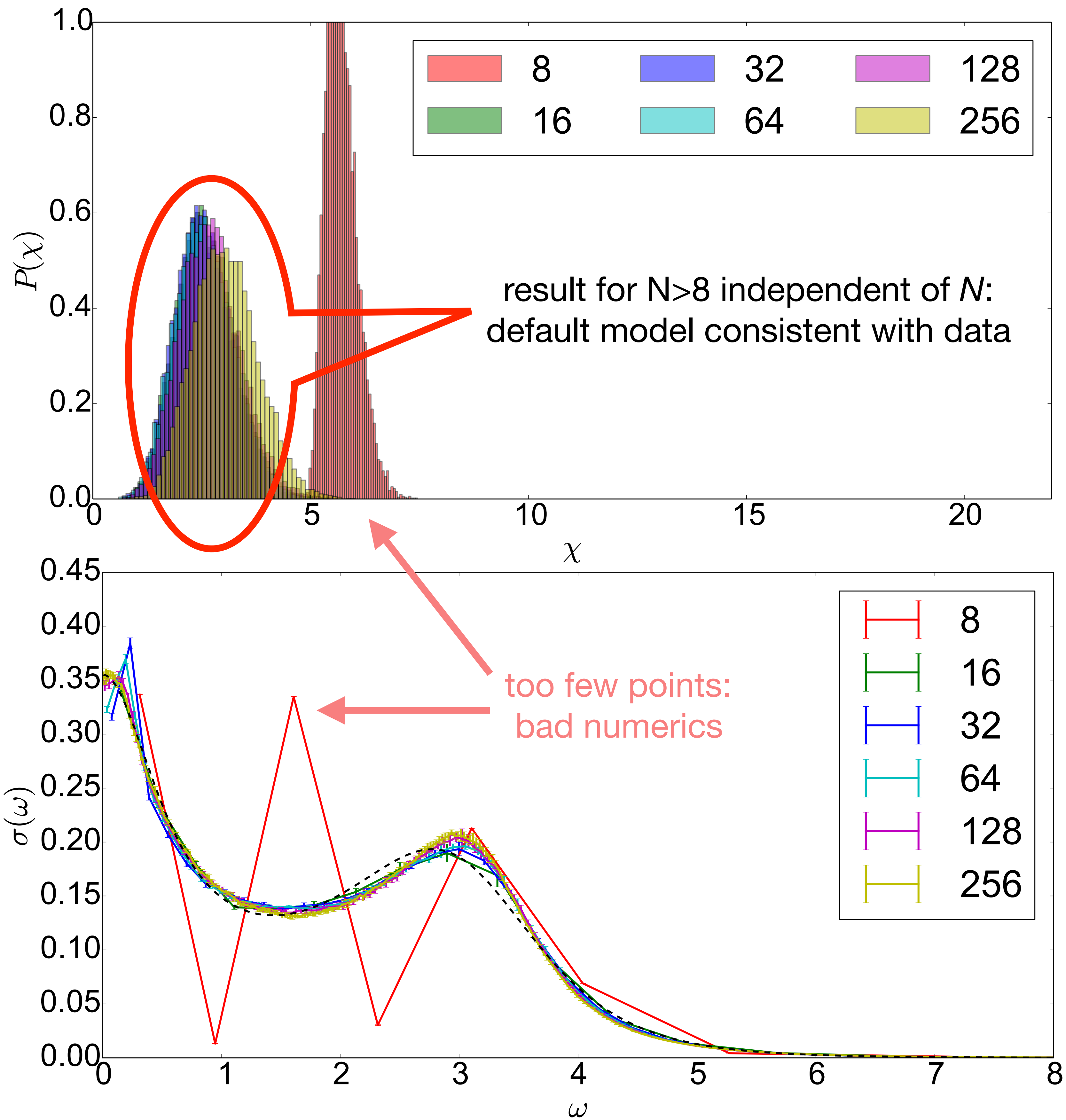
default model: least informative function that reproduces lowest moment

$\omega \geq 0$: $$\rho(\omega) = \frac{1}{\mu}\, e^{-\omega/\mu}$$

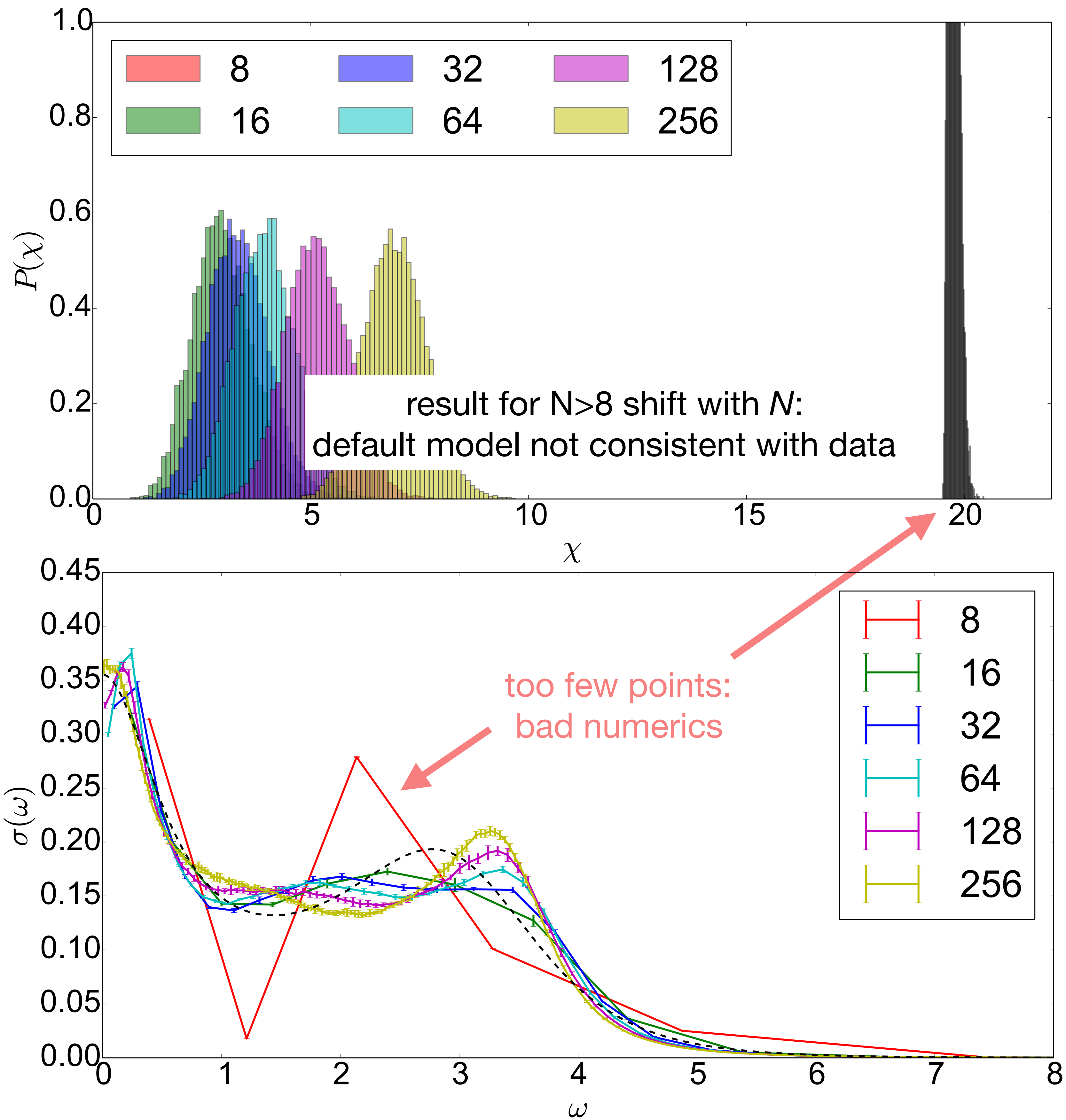
otherwise: $$\rho(\omega) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(\omega-\mu)^2/2\sigma^2}$$

grid: make numerical error in evaluating kernel smaller than noise in data

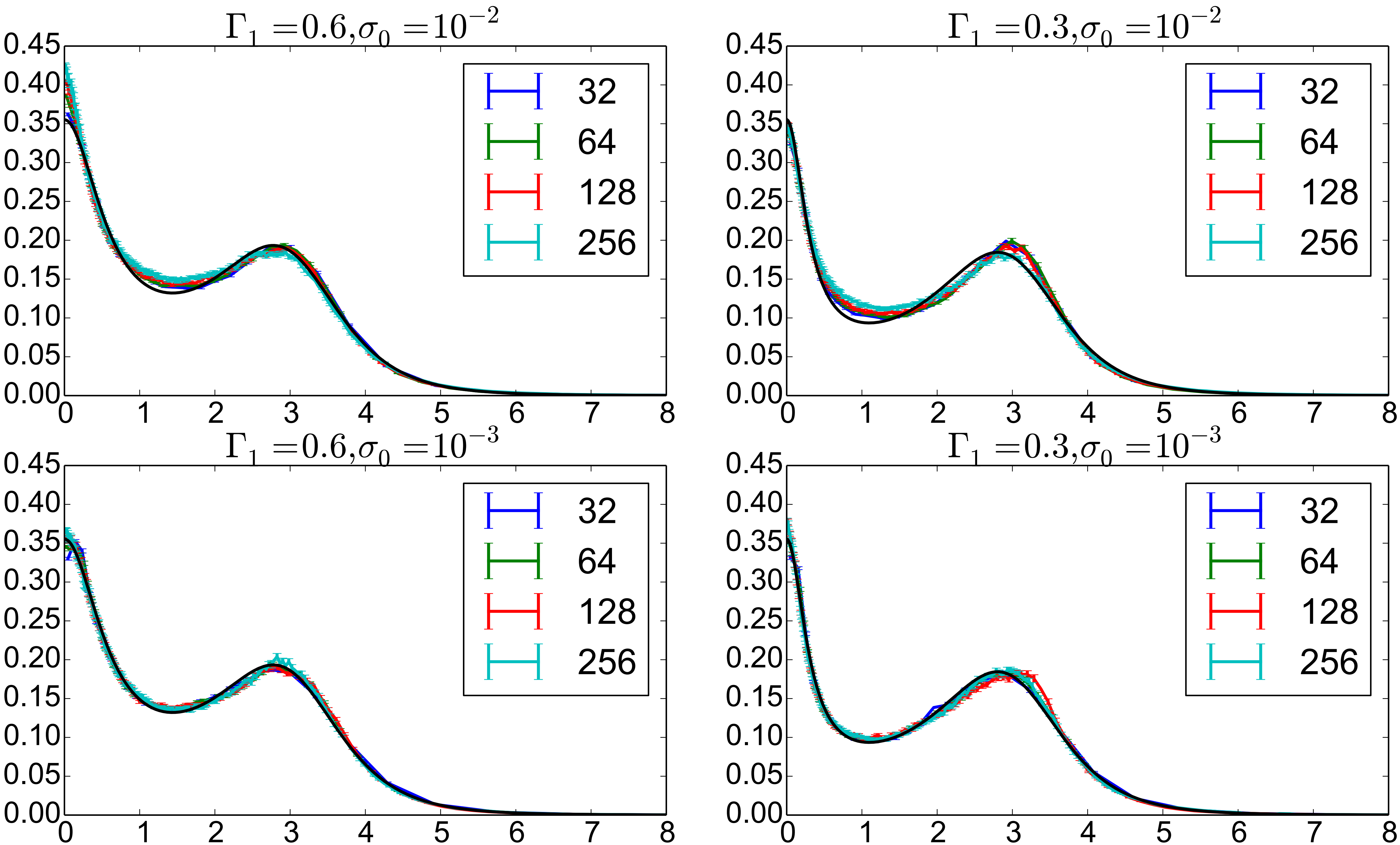
fit histogram: reliable results



fit histogram: unreliable results



test case

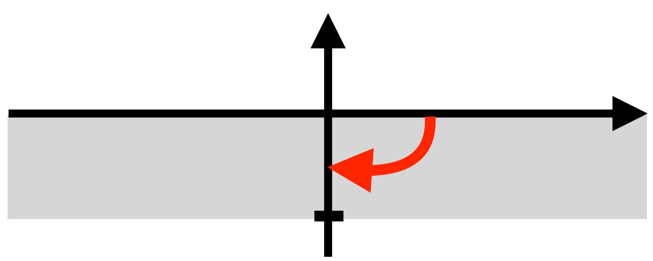


$\Gamma_1 = 0.6, \sigma_0 = 10^{-2}$
$\Gamma_1 = 0.3, \sigma_0 = 10^{-2}$
$\Gamma_1 = 0.6, \sigma_0 = 10^{-3}$
$\Gamma_1 = 0.3, \sigma_0 = 10^{-3}$
32
64
128
256

summary

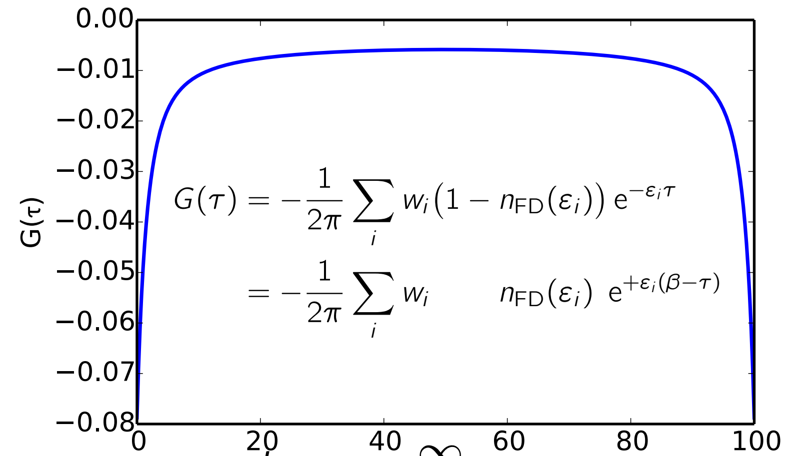
analytic continuation

$$\langle A(t)B(0)\rangle = \frac{1}{Z}\,\mathrm{Tr}\, e^{-\beta H}\, e^{iHt} A\, e^{-iHt} B$$



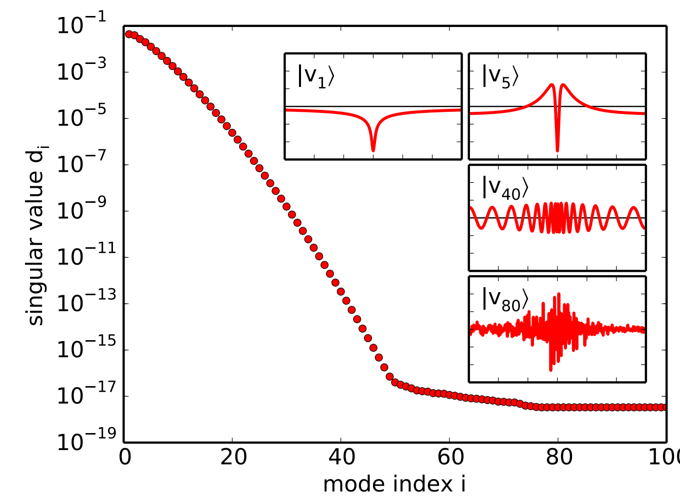
$$-G^{M\pm}_{AB}(\tau) := \langle\, \mathcal{T}^{\pm}_{\tau}\, A(-i\tau)B(0)\,\rangle$$

where is information?

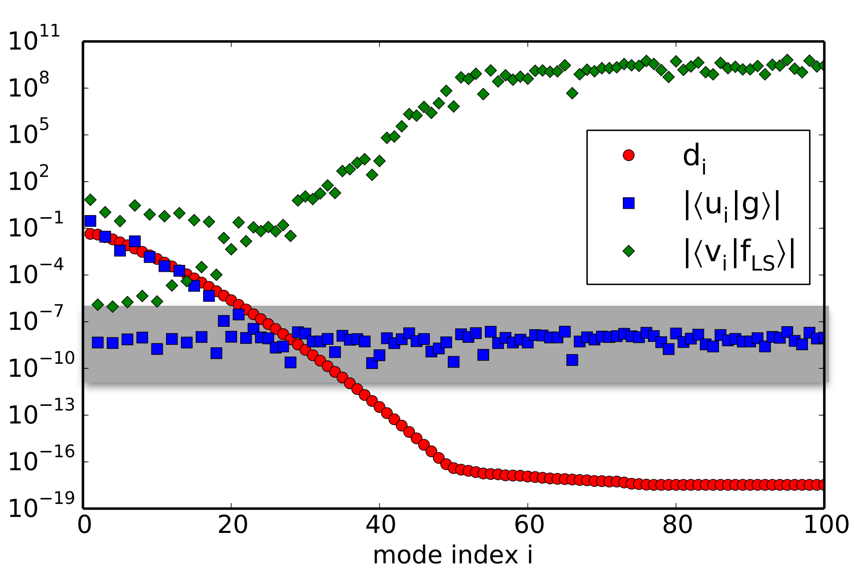


$$\frac{2e^{xt}}{e^t + 1} = \sum_{n=0}^{\infty} E_n(x)\,\frac{t^n}{n!}$$

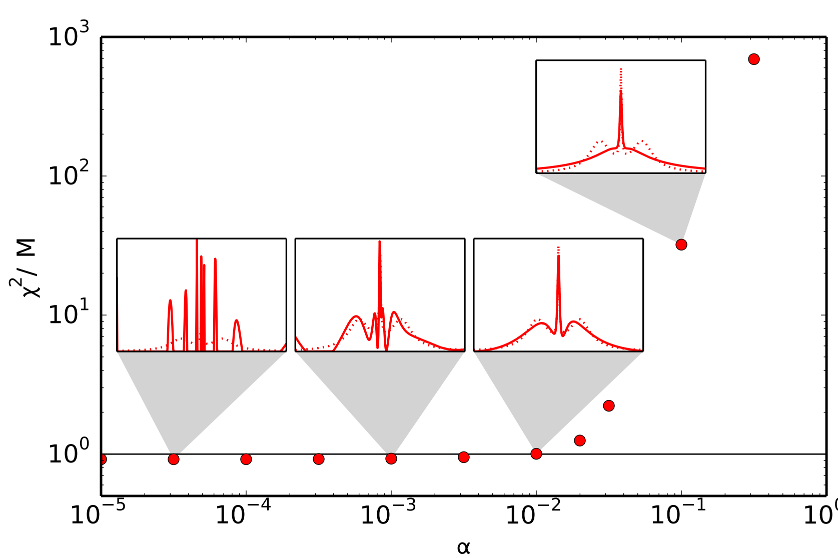
why ill conditioned?



LS/NNLS & Picard condition



regularization Tikhonov/MaxEnt



average spectrum

$$\int_{f(x)\geq 0} \mathcal{D}f\; e^{-\frac{1}{2}\chi^2[f]}\; f(x)$$

(sum-rule)