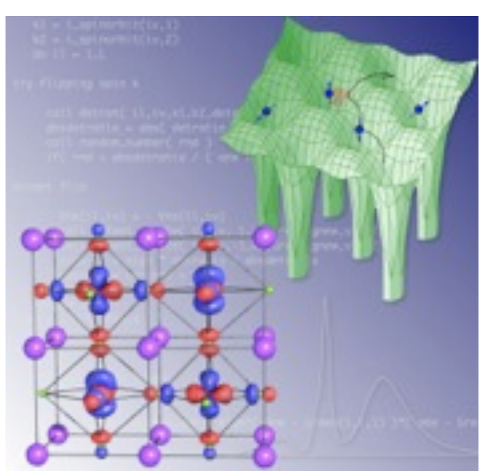
Stochastic sampling for the analytic continuation of imaginary-time data

Khaldoon Ghanem and Erik Koch
Computational Materials Science
German Research School for Simulation Sciences, Jülich

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



DFG FOR 1346

German Research School
for Simulation Sciences

outline

- analytic continuation
- singular value decomposition
- ill-posedness and role of non-negativity constraint
- average spectrum approach
- blocked mode sampling
- results depend on discretization
- separating numerics from default model
- a practical approach

Green/correlation function

spectral function $A(\omega)$ defines analytic function
in upper/lower half of complex frequency plane

$$G(z) = \int_{-\infty}^{\infty} d\omega \, \frac{A(\omega)}{z - \omega}$$

reconstruct $A(\omega)$ from $G(i\omega_n)$

$$\omega_n = \begin{cases} (2n+1)\pi/\beta & \text{fermions} \\ 2n\pi/\beta & \text{bosons} \end{cases}$$

$$G(i\omega_n) = \int_{-\infty}^{\infty} d\omega \, \frac{A(\omega)}{i\omega_n - \omega}$$

reconstruct $A(\omega)$ from $G(\tau)$

$$G(\tau) = \frac{1}{\beta} \sum_n G(i\omega_n) \, e^{-i\omega_n \tau} = \pm \int_{-\infty}^{\infty} d\omega \, \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}} \, A(\omega)$$

$\tau \in [0, \beta]$

analytic continuation

$A(\omega) \geq 0$ for fermions

$A(-\omega) = -A(\omega)$ for bosons: consider $A(\omega)/\omega \geq 0$ *for* $\omega \geq 0$

$$G(i\omega_n) = \int_0^\infty d\omega \, \frac{2\omega^2}{\omega_n^2 + \omega^2} \, \frac{A(\omega)}{\omega}$$

e.g. optical conductivity

$$\Pi(i\omega_n) = \frac{1}{\pi} \int_0^\infty d\omega \, \frac{\omega^2}{\omega_n^2 + \omega^2} \quad \sigma(\omega)$$

$$\Pi(\tau) = \frac{1}{\pi} \int_0^\infty d\omega \, \frac{\omega e^{-\tau\omega}}{1 - e^{-\beta\omega}} \; \sigma(\omega)$$

Fredholm equation

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

data kernel model

$$|g\rangle = K|f\rangle$$

inverse problem: given *g(y)* find *f(x)*

ill posed

singular value decomposition

$$|g\rangle = K|f\rangle$$

SVD of kernel

$$K = USV^\dagger$$

$V = (|v_1\rangle, |v_2\rangle, \ldots)$ orthonormal basis in model space

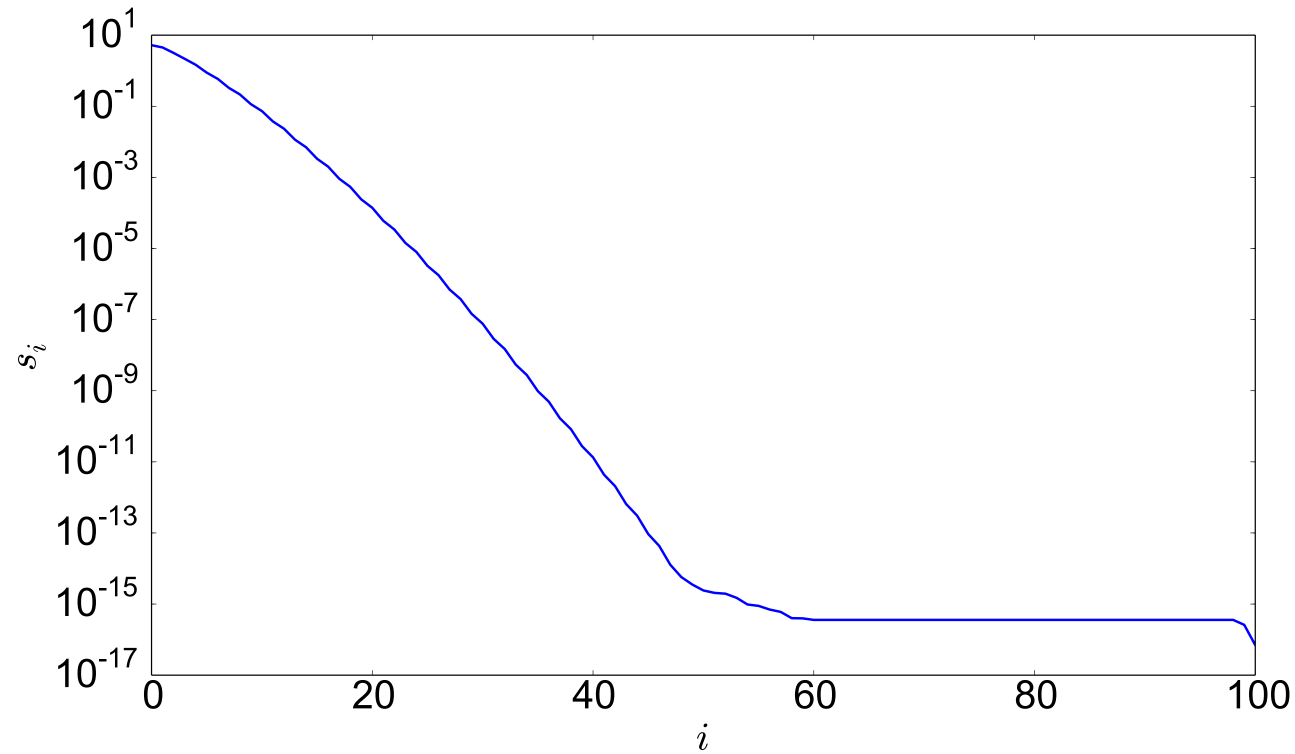
$U = (|u_1\rangle, |u_2\rangle, \ldots)$ orthonormal basis in data space

S diagonal rectangular matrix with $s_i \geq s_{i+1}$ (singular values)

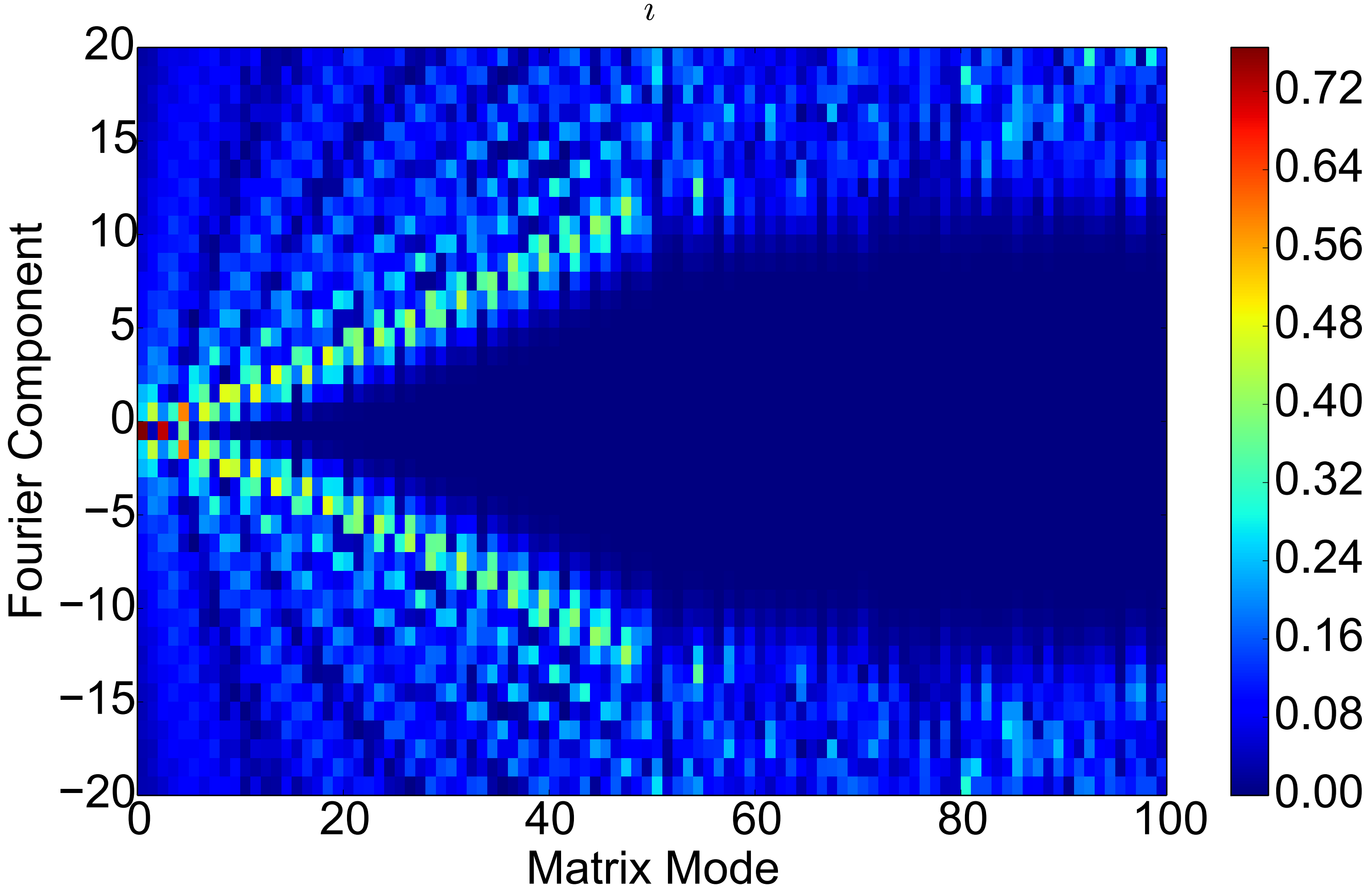
spectral representation of kernel

$$K|v_i\rangle = s_i|u_i\rangle$$

singular values and singular modes



singular values
decay exponentially



singular modes v_i
oscillate more

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

kernel smoothes model

least-squares solution

$$|g\rangle = K|f\rangle$$

to find best f, minimize $\left\| K|f\rangle - |g\rangle \right\|^2$

$$|f\rangle = \sum |v_i\rangle\langle v_i|f\rangle = \sum f_i\,|v_i\rangle \quad \rightsquigarrow \quad K|f\rangle = \sum f_i s_i |u_i\rangle$$

$$|g\rangle = \sum |u_i\rangle\langle u_i|g\rangle = \sum g_i\,|u_i\rangle$$

$$\frac{\partial}{\partial \bar{f}_i} \left\| K|f\rangle - |g\rangle \right\|^2 = s_i\left(f_i s_i - g_i\right) \stackrel{!}{=} 0$$

$$f_i = \frac{g_i}{s_i}$$

ill posedness

$$|f_{LS}\rangle = \sum \frac{g_i}{s_i} |v_i\rangle$$

modes with small singular value hardly affect data
small errors in data are catastrophically amplified

the affected modes tend to be highly oscillating
sawtooth noise in least-squares fit

Picard condition

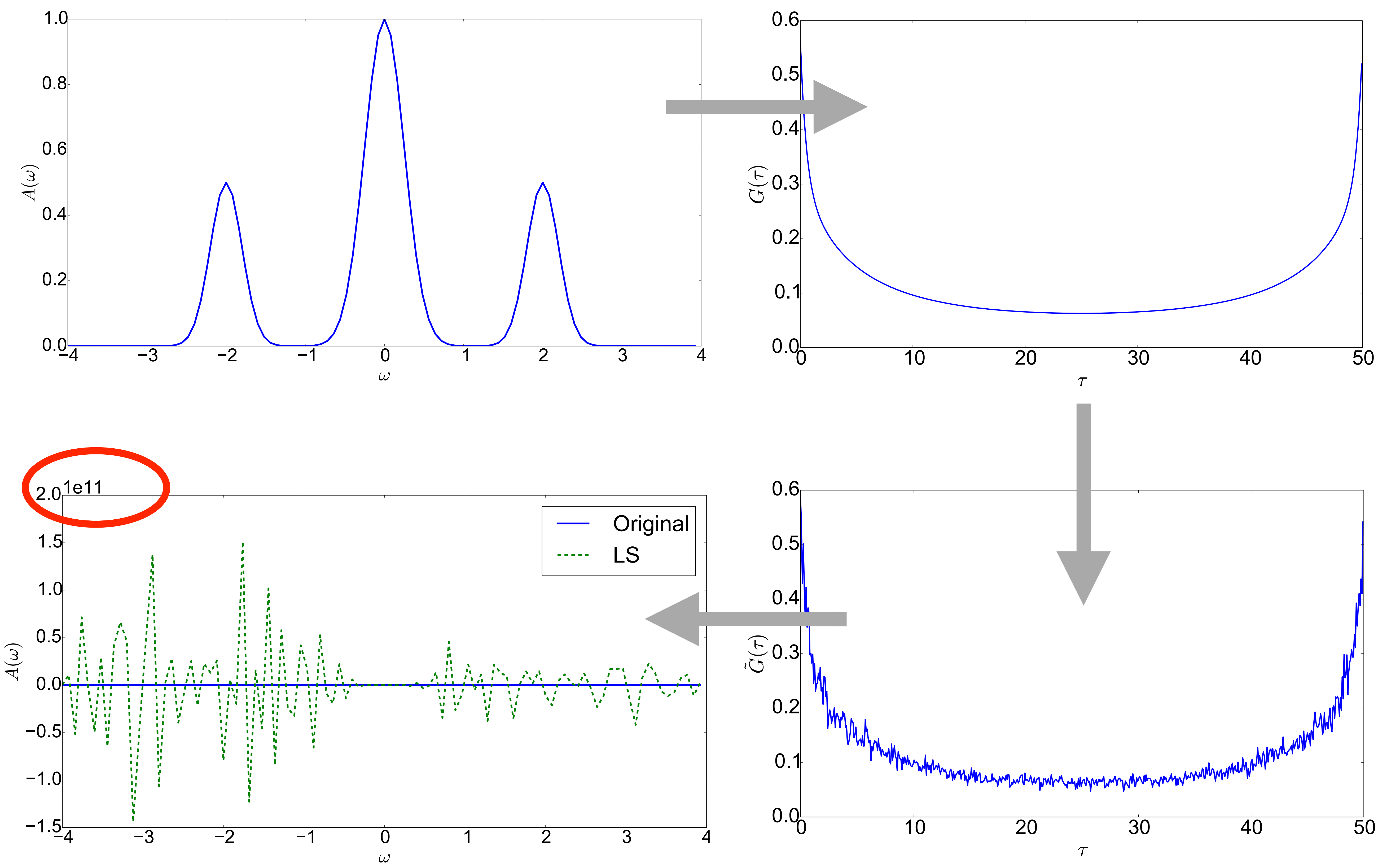
$$\big\| |f\rangle \big\|^2 = \sum |\langle v_i|f\rangle|^2 = \sum \frac{|\langle u_i|g\rangle|^2}{s_i^2} < \infty$$

implies smoothness of data

noise in data leads to violation of Picard condition

$f_i = g_i/s_i$ variance in data $\sigma_i \Rightarrow$ variance in model: σ_i/s_i

ill posedness

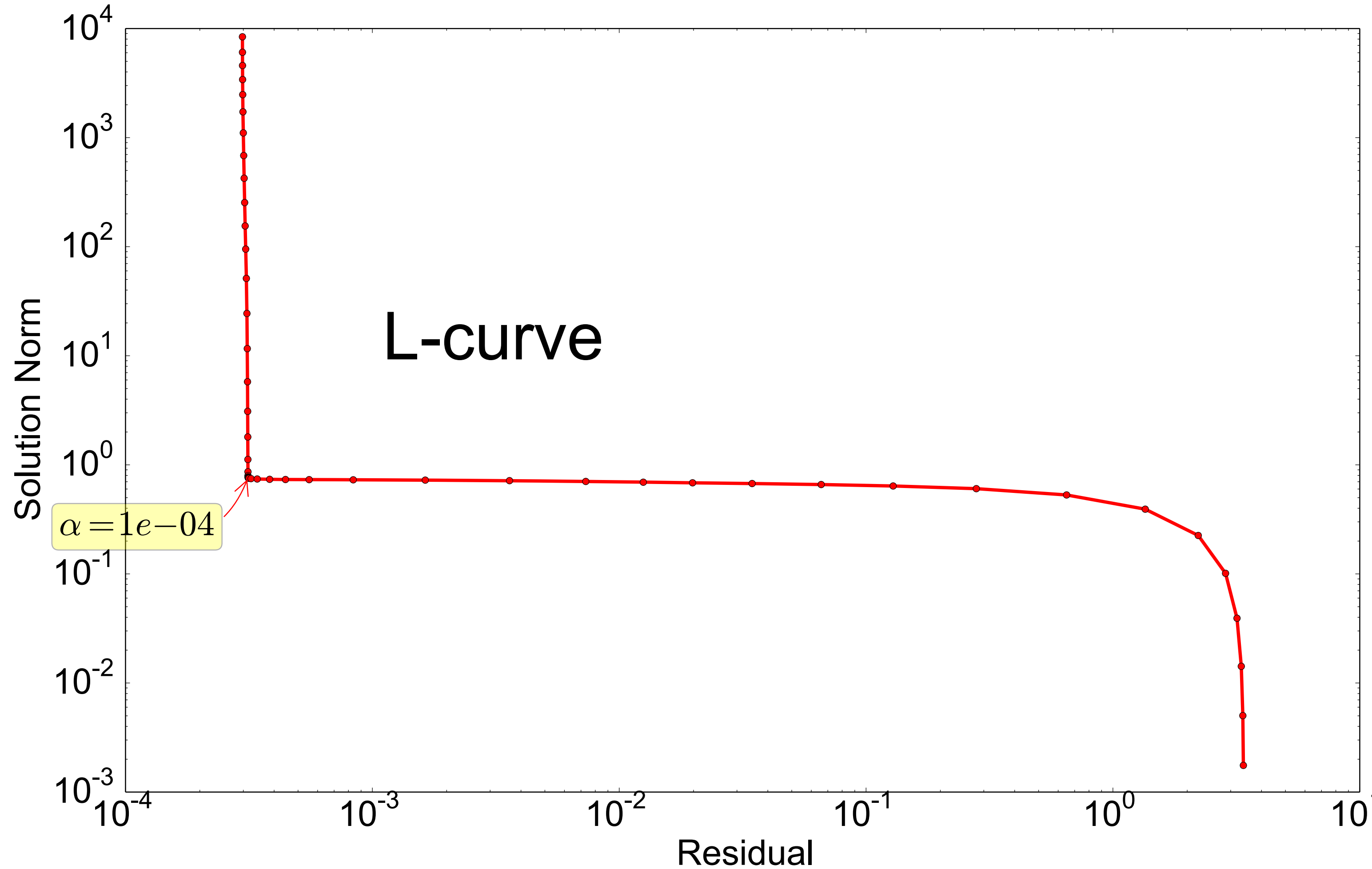


Tikhonov regularization

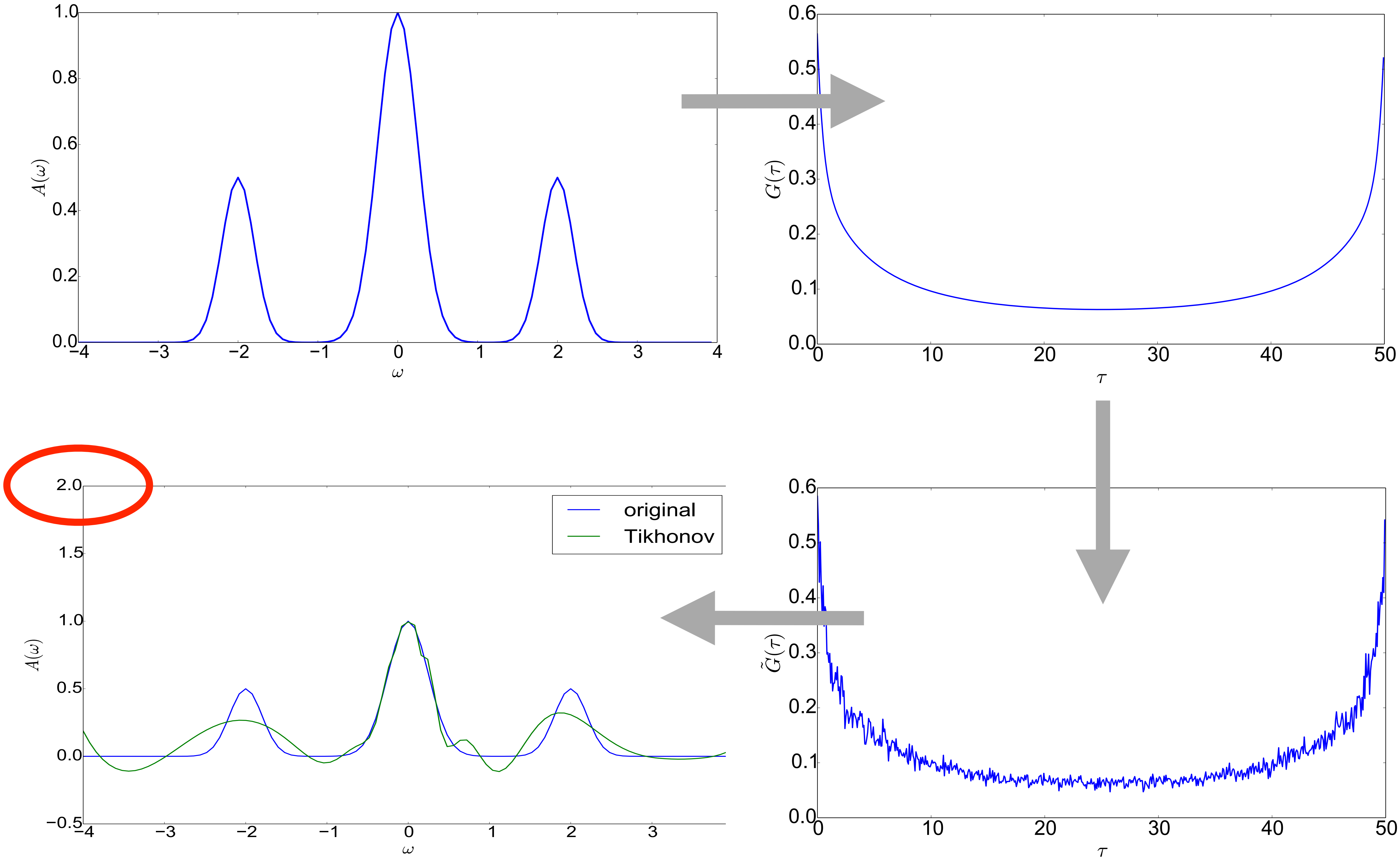
minimize $$\| K|f\rangle - |g\rangle \|^2 + \alpha \| |f\rangle \|^2$$

$$f_i = \frac{s_i^2}{s_i^2 + \alpha^2} \frac{g_i}{s_i}$$

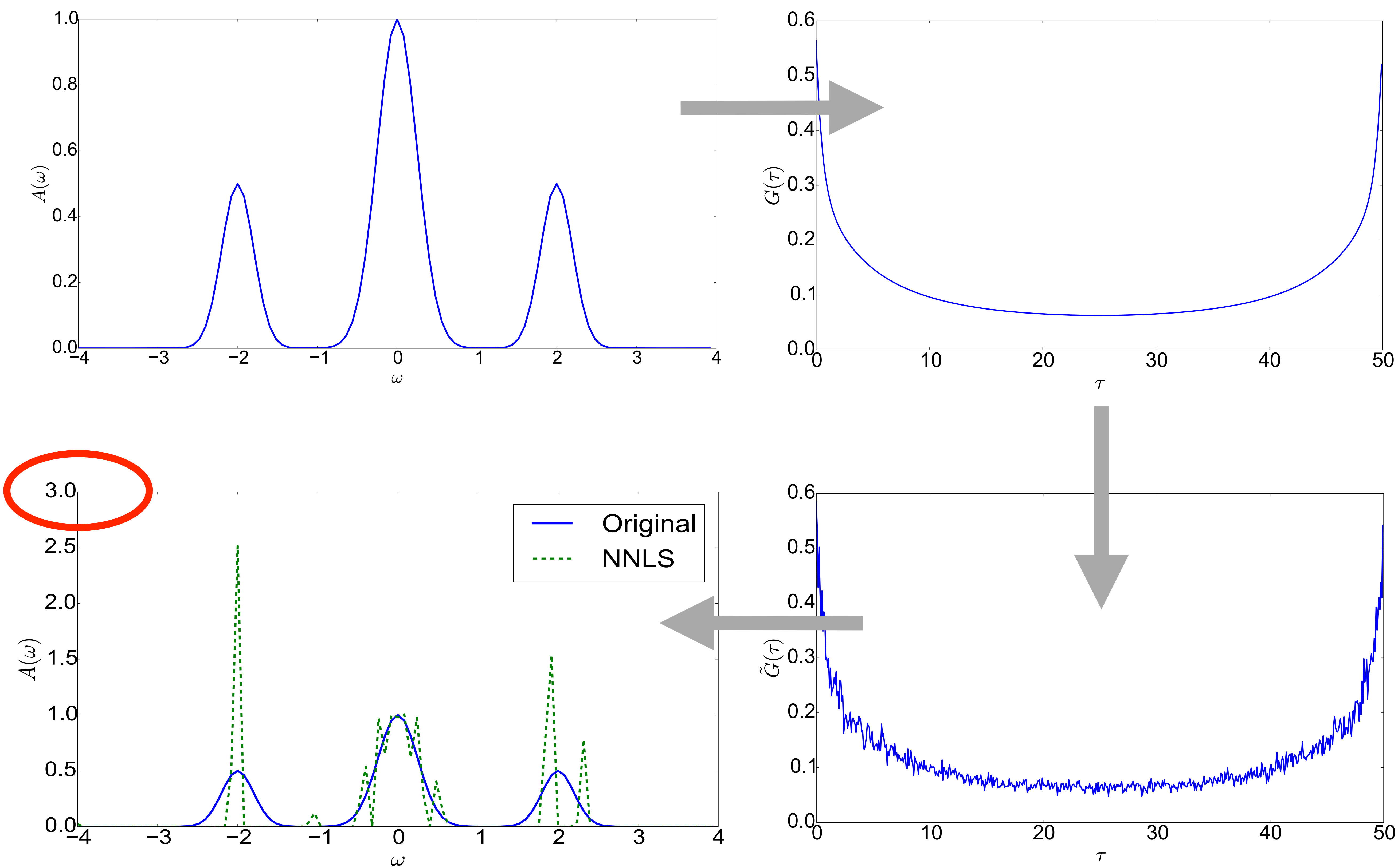
determine regularization parameter α



effect of regularization

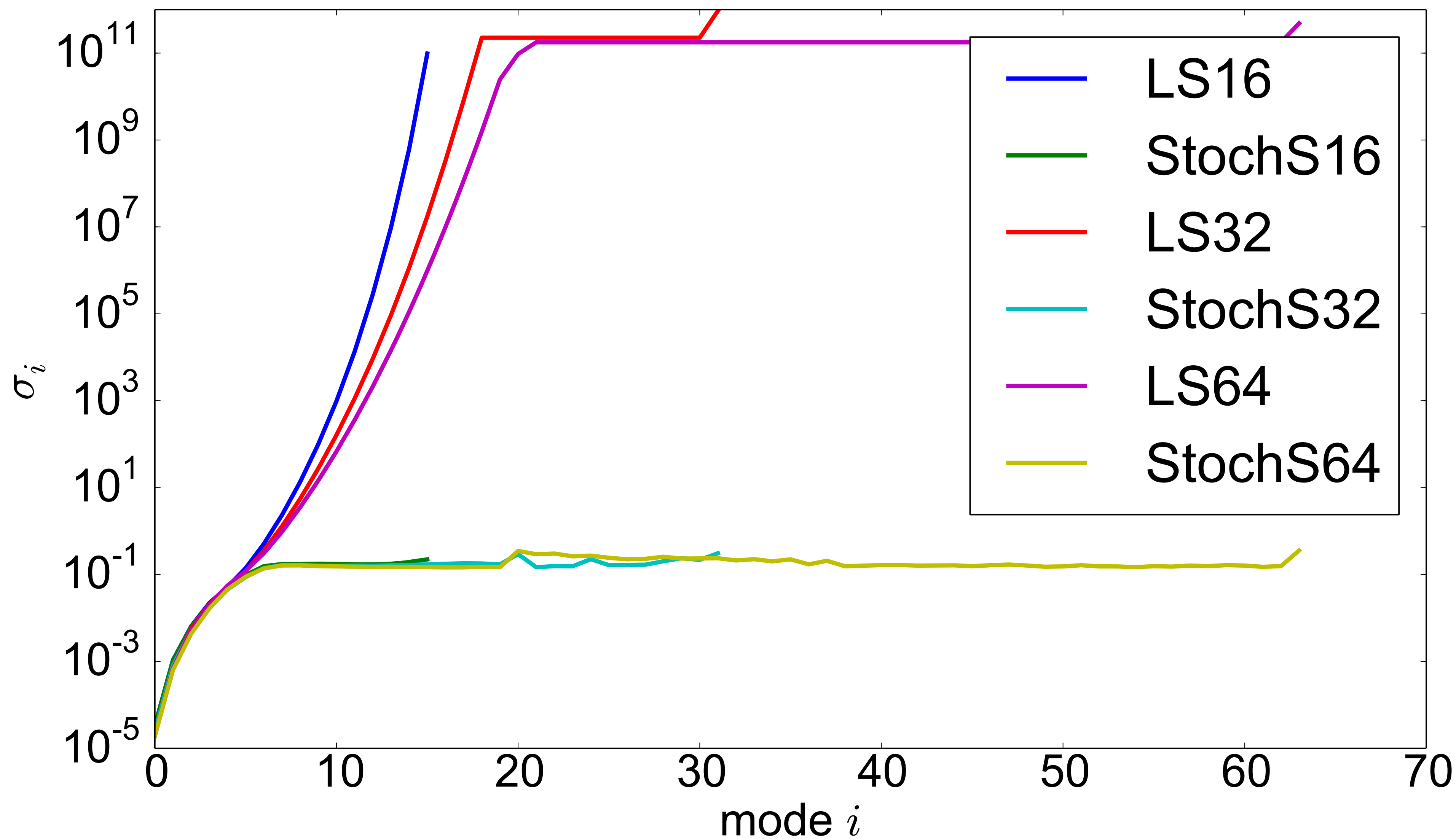


effect of non-negativity constraint



regularization by constraint

non-negativity constraint regularizes singular modes:
modes with small s_i adjust such that leading modes can be optimized



Bayesian inference

take quality of data and *a priori* knowledge about model into account

$$|g\rangle = K|f\rangle$$

data given by mean $\bar{g}$ and covariance Σ

$$P(g|\bar{g},\Sigma) \propto e^{-\frac{1}{2}\langle g-\bar{g}|\Sigma^{-1}|g-\bar{g}\rangle}$$

probability of model *f* given measured data $\bar{g}$, Σ

$$P(f|\bar{g},\Sigma) = \frac{P(\bar{g},\Sigma|f)\,P(f)}{P(\bar{g},\Sigma)}$$

likelihood $$P(\bar{g},\Sigma|f) = \int \mathcal{D}g\; P(\bar{g}|g)\,P(g|f) = P(\bar{g},\Sigma|Kf)$$

$$\propto e^{-\frac{1}{2}\langle Kf-\bar{g}|\Sigma^{-1}|Kf-\bar{g}\rangle}$$

assume $P(g) = P(\bar{g},\Sigma)$

prior probability *P(f)*: *a priori* information about model

likelihood function

$$P(\bar{g}, \Sigma | f) \propto e^{-\frac{1}{2}\langle Kf - \bar{g} | \Sigma^{-1} | Kf - \bar{g}\rangle}$$

absorb covariance in data/kernel

Cholesky decomposition $\Sigma^{-1} = W^{\dagger} W$

$$\begin{aligned}
\langle Kf - \bar{g} | \Sigma^{-1} | Kf - \bar{g}\rangle &= \langle \quad Kf - \quad \bar{g} | W^{\dagger} W | \quad Kf - \quad \bar{g}\rangle \\
&= \langle WKf - W\bar{g} \quad | \quad WKf - W\bar{g}\rangle \\
&= \langle \quad \tilde{K}f - \quad \tilde{g} \quad | \quad \tilde{K}f - \quad \tilde{g}\rangle
\end{aligned}$$

prior probabilities

- Tikhonov $$P(f) \propto \exp\left(-\frac{\alpha}{2}\langle f|f\rangle\right)$$

  favor solutions with small norm

- MaxEnt $$P(f) \propto \exp\left(-\alpha \int f(x)\,\ln\left(f(x)/d(x)\right)\,dx\right)$$

  favor solutions close to default model $d(x)$

- non-informative $$P(f) \propto \begin{cases} 0 & \text{if } \exists x : f(x) < 0 \\ 1 & \text{otherwise} \end{cases}$$

  all *a priori* admissible solutions equally probably

how to pick model given $P(f\,|\bar{g},\Sigma)$?

minimize cost of being wrong

best estimate minimizes expected loss

$$\int \mathcal{D}f \, L(\bar{f}, f) \, P(f|\bar{g}, \Sigma)$$

loss function $L(\bar{f},f)$

$$L(\bar{f}, f) = \begin{cases} 0 & \text{when } \bar{f} = f \\ 1 & \text{otherwise} \end{cases} \quad \leadsto \quad \bar{f} \text{ maximizes } P(f|\bar{g}, \Sigma)$$

$$L(\bar{f}, f) = \left\| |\bar{f} - f\rangle \right\|^2 \quad \leadsto \quad \bar{f} = \int \mathcal{D}f \, f \, P(f|\bar{g}, \Sigma)$$

average over posterior

respects constraint
if set of admissible functions
is convex

Stochastic Sampling

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

S.R. White

Department of Physics, University of California, Irvine, CA 92717, USA

In this paper we present the average spectrum method, a new method for obtaining real-frequency information from imaginary-time quantum Monte Carlo data. This technique does not require the adjustable parameters, smoothness constraints, or model forms of some previous techniques, yet produces smooth, consistent spectra from noisy data. Various tests of the method on mock data are presented, as well as realistic applications to the two-dimensional Hubbard model.

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

Stochastic Sampling

impose *a priori* via constraints
variance in data smoothes model

- S.R. White:
  The Average Spectrum Method for the Analytic Continuation of Imaginary-Time Data
  Springer Proceedings in Physics **53**, 145 (1991)
- A.W. Sandvik:
  Stochastic method for analytic continuation of quantum Monte Carlo data
  Phys. Rev. B **57**, 10287 (1998)
- K. Vafayi and O. Gunnarsson:
  Analytical continuation of spectral data from imaginary time axis to real frequency axis using statistical sampling
  Phys. Rev. B **76**, 035115 (2007)
- K.S.D. Beach:
  Identifying the maximum entropy method as a special limit of stochastic analytic continuation
  cond-mat/0403055
- O.F. Syljuåsen:
  Using the average spectrum method to extract dynamics from quantum Monte Carlo simulations
  Phys. Rev. B **78**, 174429 (2008)

Gaussian distribution in *f*

$$|\bar{f}\rangle \propto \int_{\mathcal{F}} \mathcal{D}f \; e^{-\frac{1}{2}\langle \tilde{K}f-\tilde{g}|\tilde{K}f-\tilde{g}\rangle} \, |f\rangle$$

rewrite as Gaussian in *f*

pseudo-inverse of kernel *K*

$$K = \sum_i |u_i\rangle s_i \langle v_i| \qquad K^- = \sum_{s_i>0} \frac{|u_i\rangle\langle v_i|}{s_i}$$

$$\langle \tilde{K}f-\tilde{g}|\tilde{K}f-\tilde{g}\rangle = \langle f-\tilde{K}^-\tilde{g}|\tilde{K}^\dagger\tilde{K}|f-\tilde{K}^-\tilde{g}\rangle - \langle g_\perp|g_\perp\rangle$$

$$\sum_i \tilde{s}_i^2|f_i|^2 - \tilde{s}_i(f_i^*\tilde{g}_i + \tilde{g}_i^* f_i) + |\tilde{g}_i|^2 = \sum_{\tilde{s}_i>0} \tilde{s}_i^2 \left| f_i - \frac{\tilde{g}_i}{\tilde{s}_i} \right| + \sum_{\tilde{s}_i=0} |\tilde{g}_i|^2$$

$g_\perp$: component of data that cannot be described by model

Gibbs sampling

write model in (finite) basis $$|f\rangle = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \end{pmatrix}$$

sample component $f_i \rightarrow f_i'$ keeping all other components fixed

$$|f\rangle = \begin{pmatrix} f_i \\ f_\perp \end{pmatrix} \qquad \tilde{K} = (\tilde{K}_i, \tilde{K}_\perp) \qquad \tilde{g}_i = \tilde{g} - \tilde{K}_\perp f_\perp$$

$$P(f_i \rightarrow f_i' | f_\perp) = \exp\left(-\frac{1}{2} \frac{\langle \tilde{K}_i f_i' - \tilde{g}_i | \tilde{K}_i f_i' - \tilde{g}_i \rangle}{\langle \tilde{K}_i f_i - \tilde{g}_i | \tilde{K}_i f_i - \tilde{g}_i \rangle} \right)$$

Gaussian with $$\sigma^2 = 1 / \tilde{K}_i^\dagger \tilde{K}_i \qquad \mu = \sigma^2 \, \tilde{K}_i^\dagger \tilde{g}_i$$

non-negativity constraint limits allowed values → truncated Gaussian

can be efficiently sampled

mode sampling

in SVD basis the singular modes have independent Gaussians

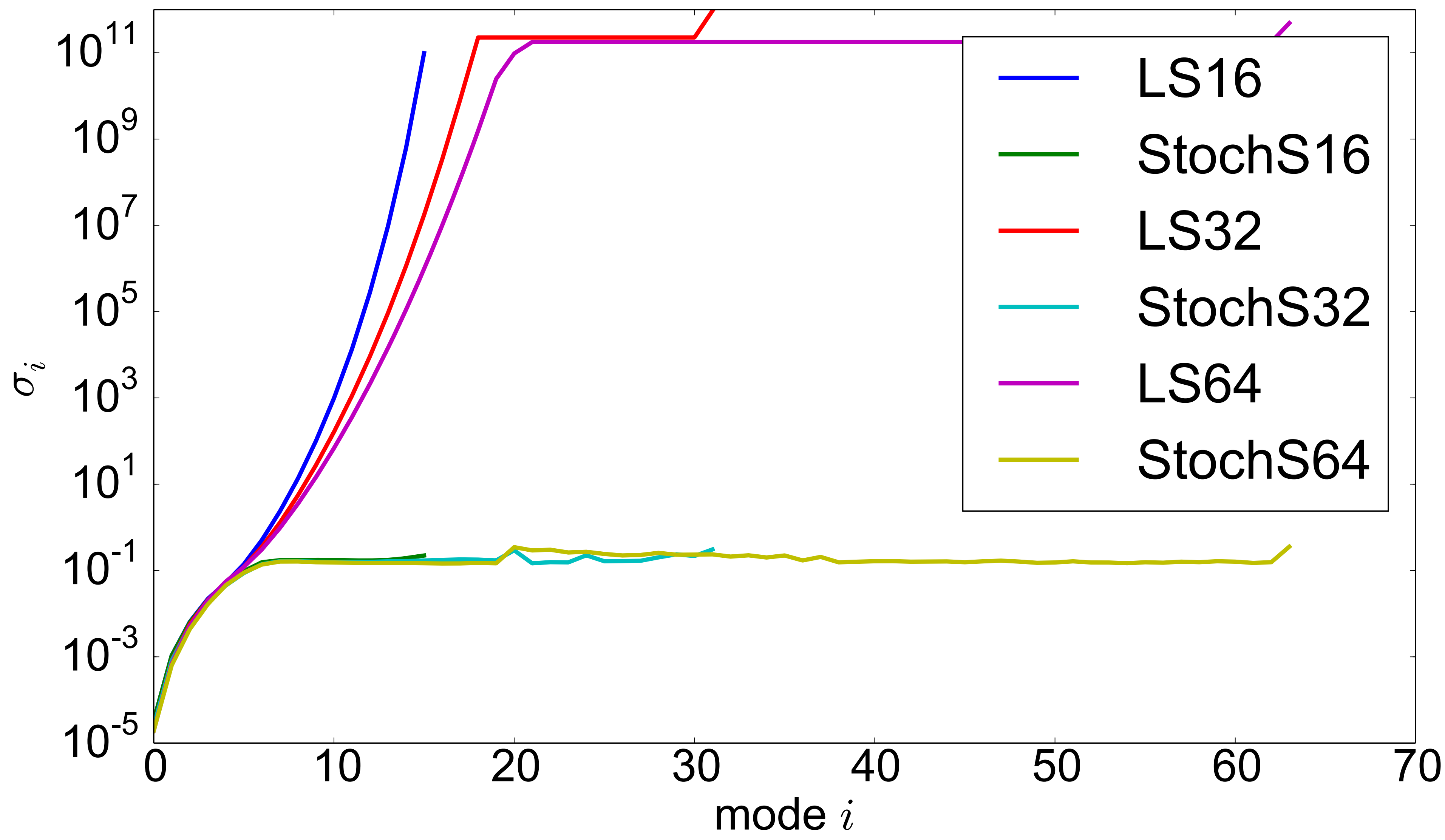
$$P(f_i \to f_i' | \mathbf{f}) = \exp\left(-\frac{1}{2} \frac{|f_i' - \tilde{g}_i / \tilde{s}_i|^2}{|f_i - \tilde{g}_i / \tilde{s}_i|^2} \right)$$

but are coupled by constraint $f(\omega) = \sum f_i \, v_i(\omega) \geq 0$

coupling of modes reduces variance of modes with small s_i

regularization by constraint

non-negativity constraint regularizes singular modes:
modes with small s_i adjust such that leading modes can be optimized



component sampling

in frequency basis $f(\omega_i)$

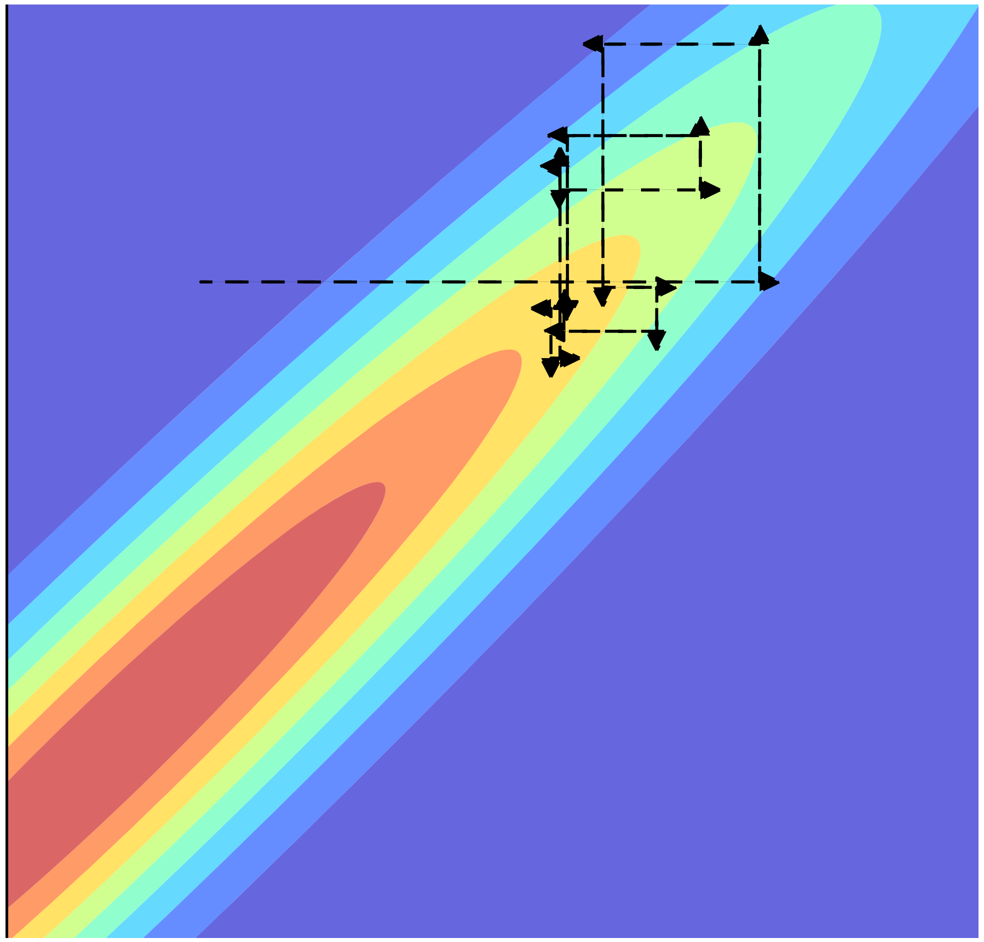
non-negativity constraint: $f(\omega_i) \geq 0$

$$P(f(\omega_i) \to f'(\omega_i)|f_\perp) = \text{ Gaussian with}$$

$$\sigma^{-2} = \int d\tau \, |\tilde{K}(\tau, \omega_i)|^2 \qquad \mu = \sigma^2 \int d\tau \, g_i(\tau) \, K^*(\tau, \omega_i)$$

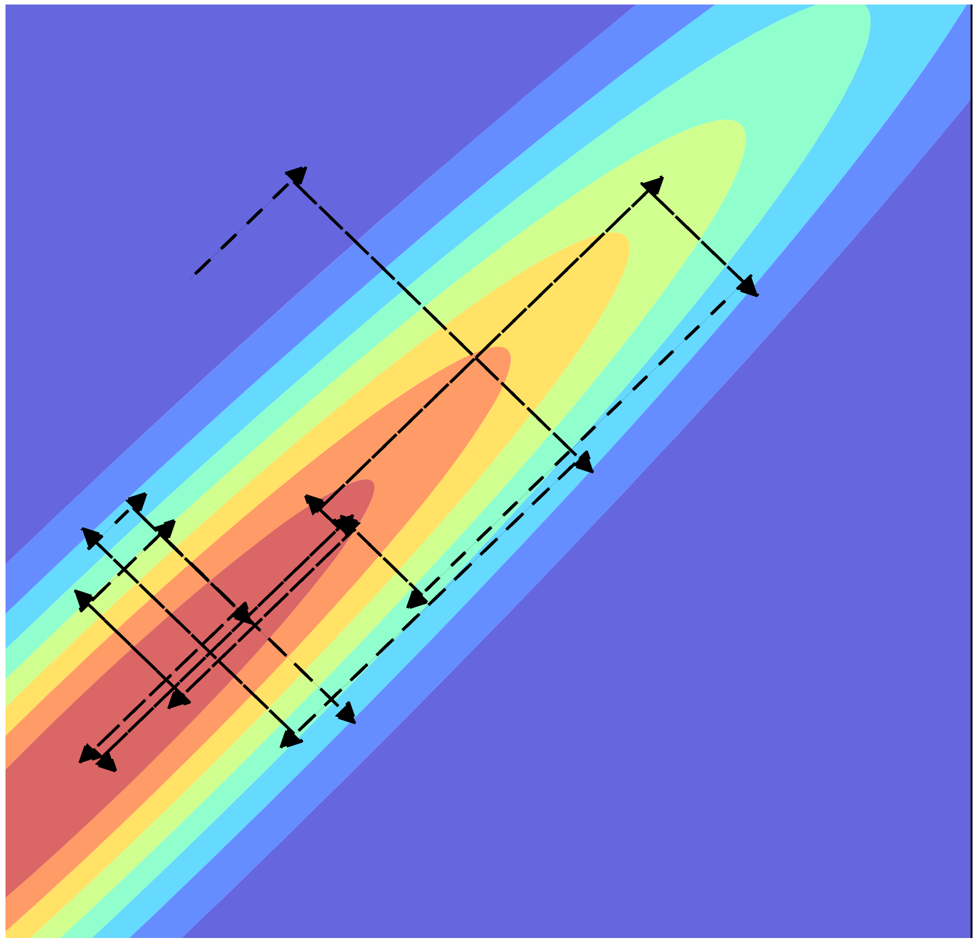
choice of basis

component sampling



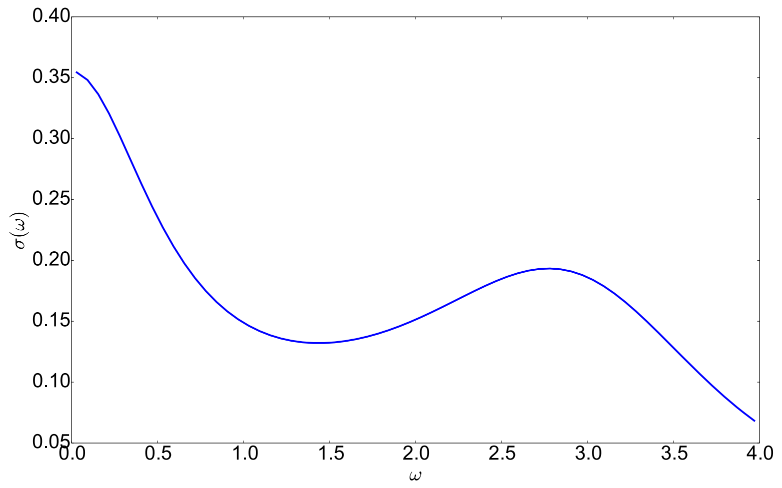
best at constraint

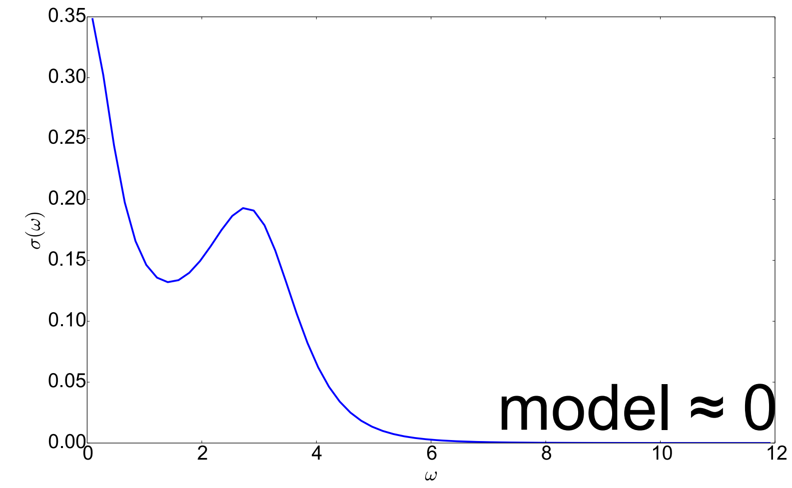
mode sampling

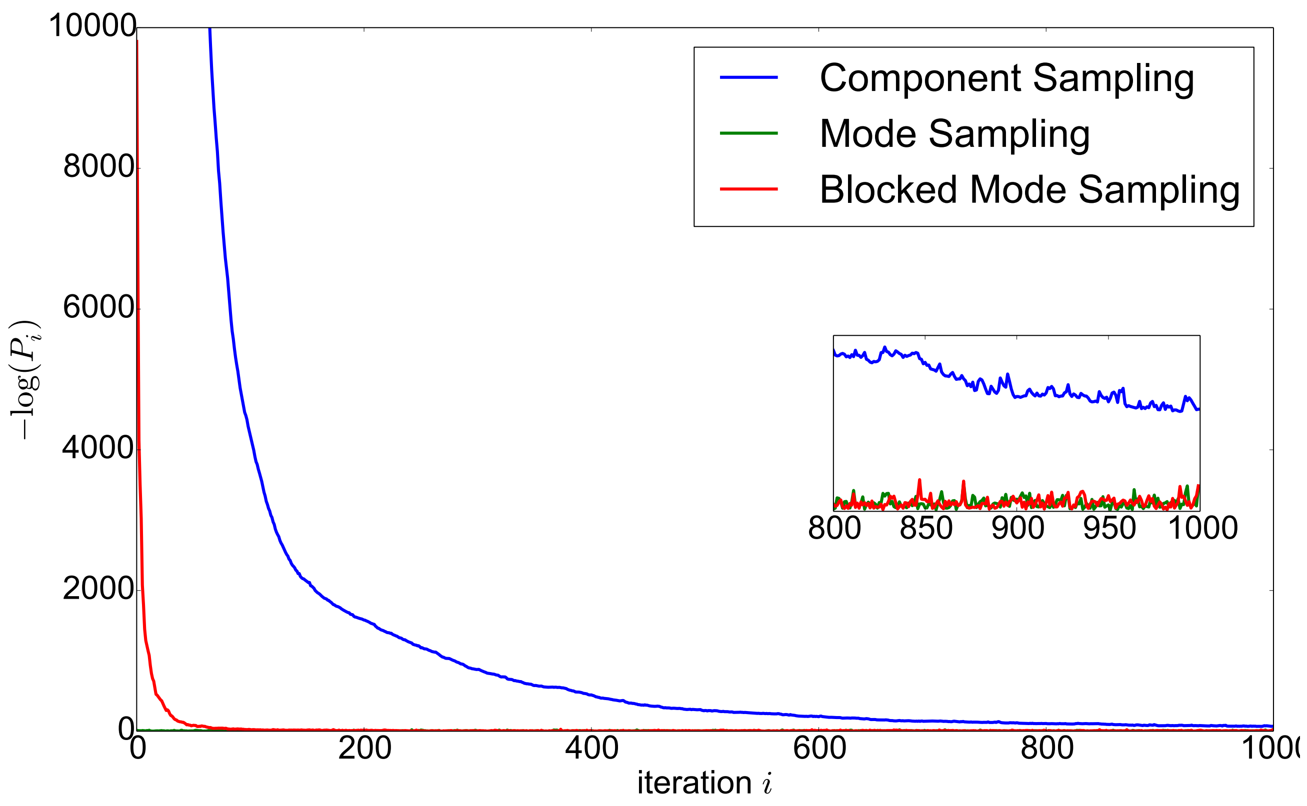


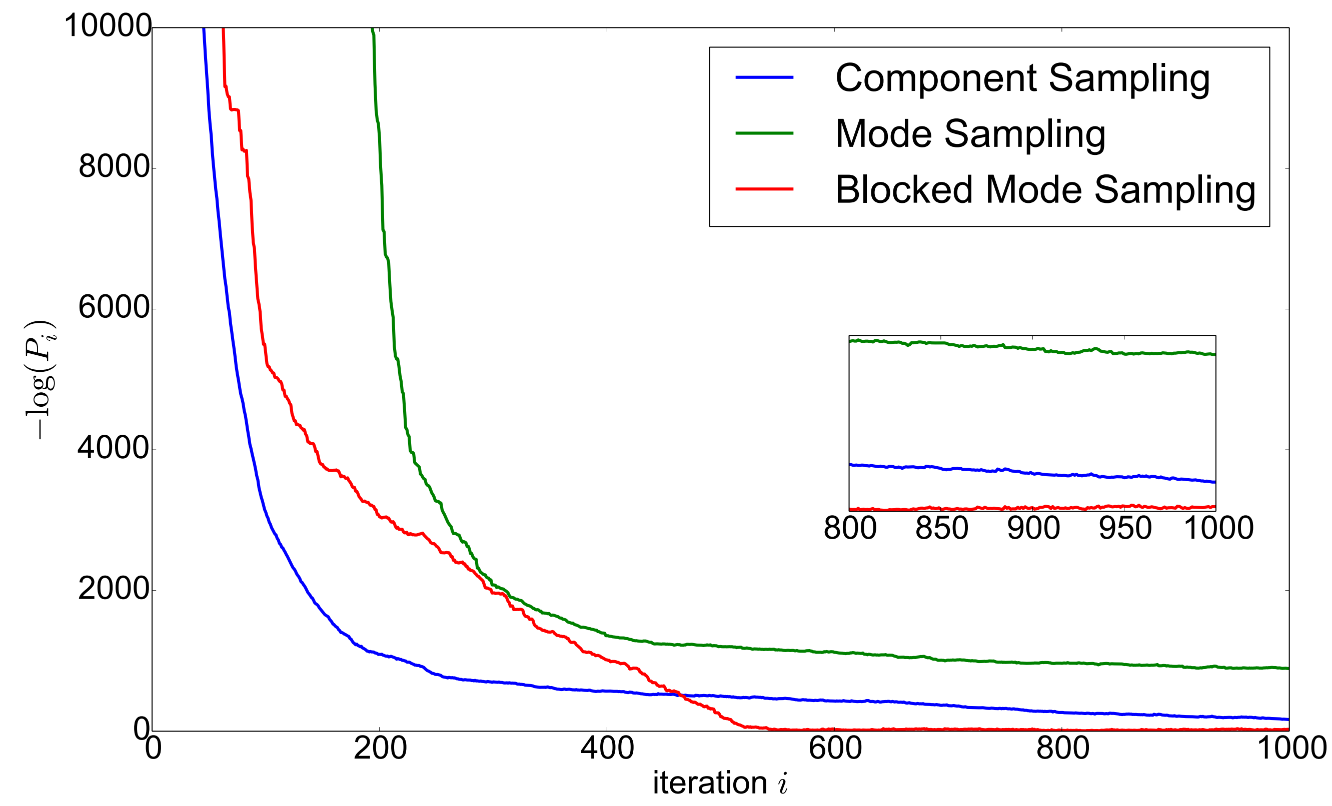
best at maximum

efficiency: thermalization

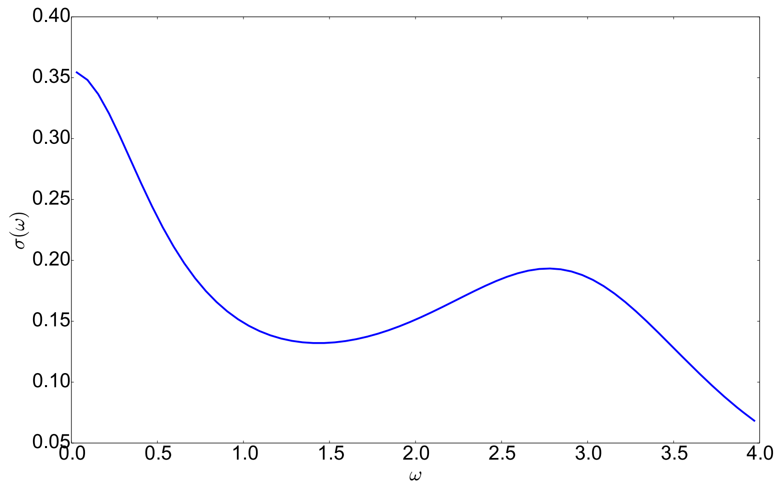


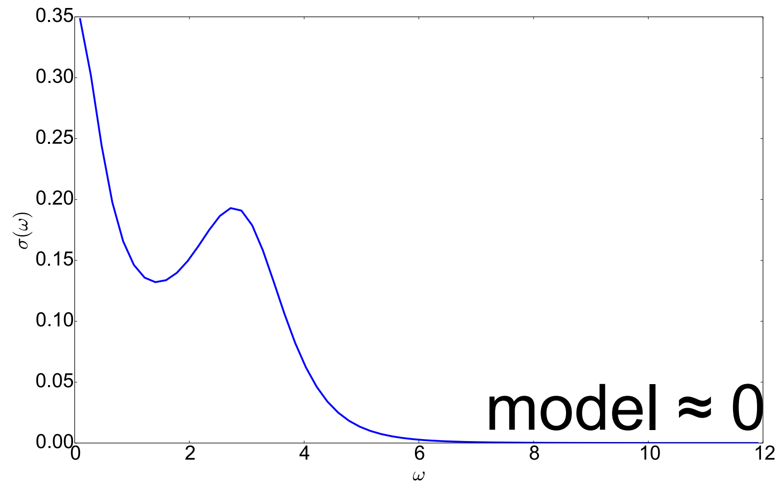


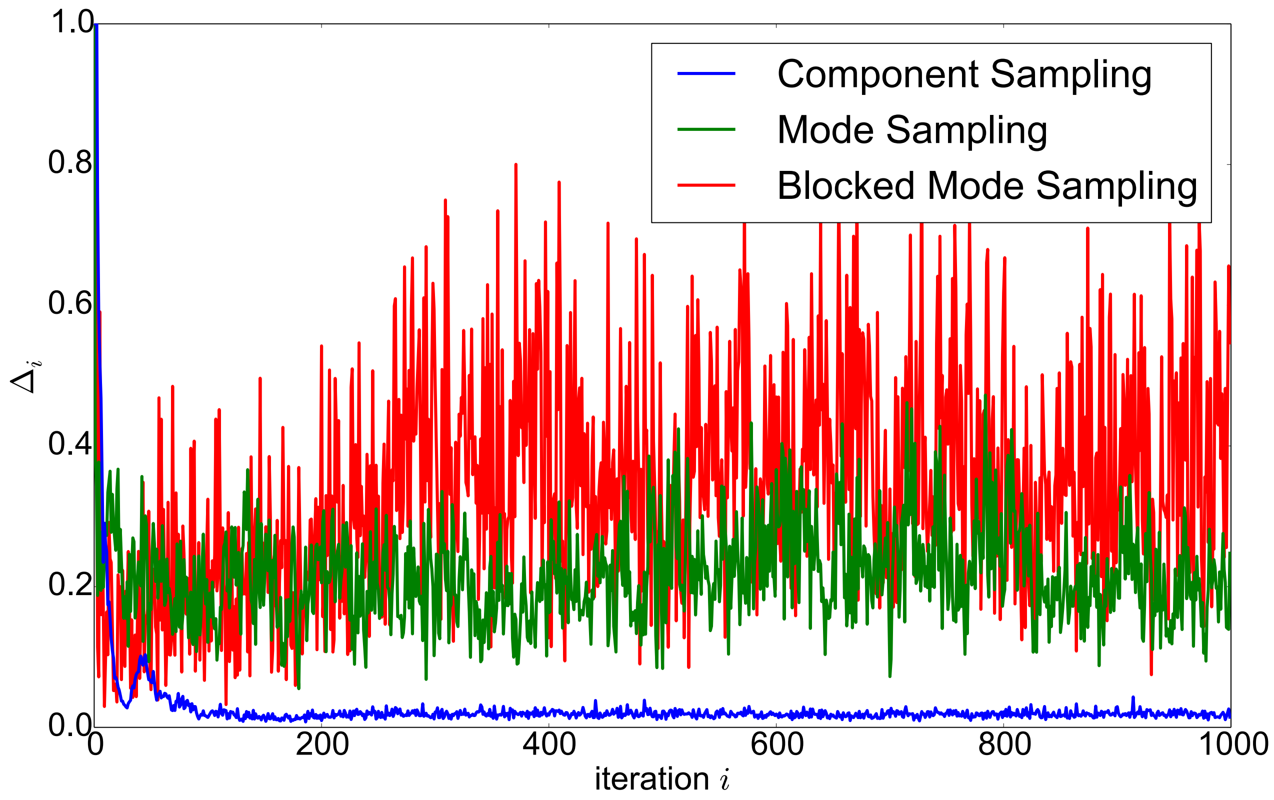


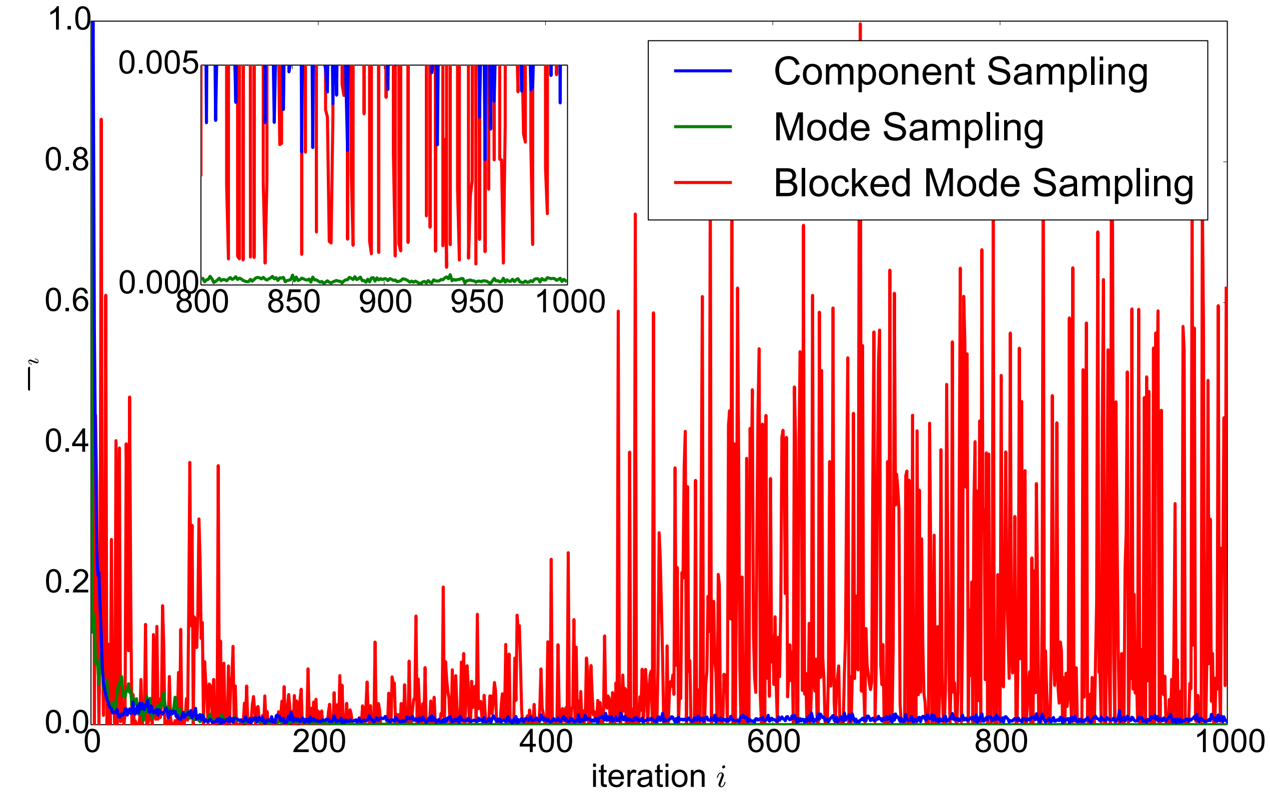


efficiency: autocorrelation



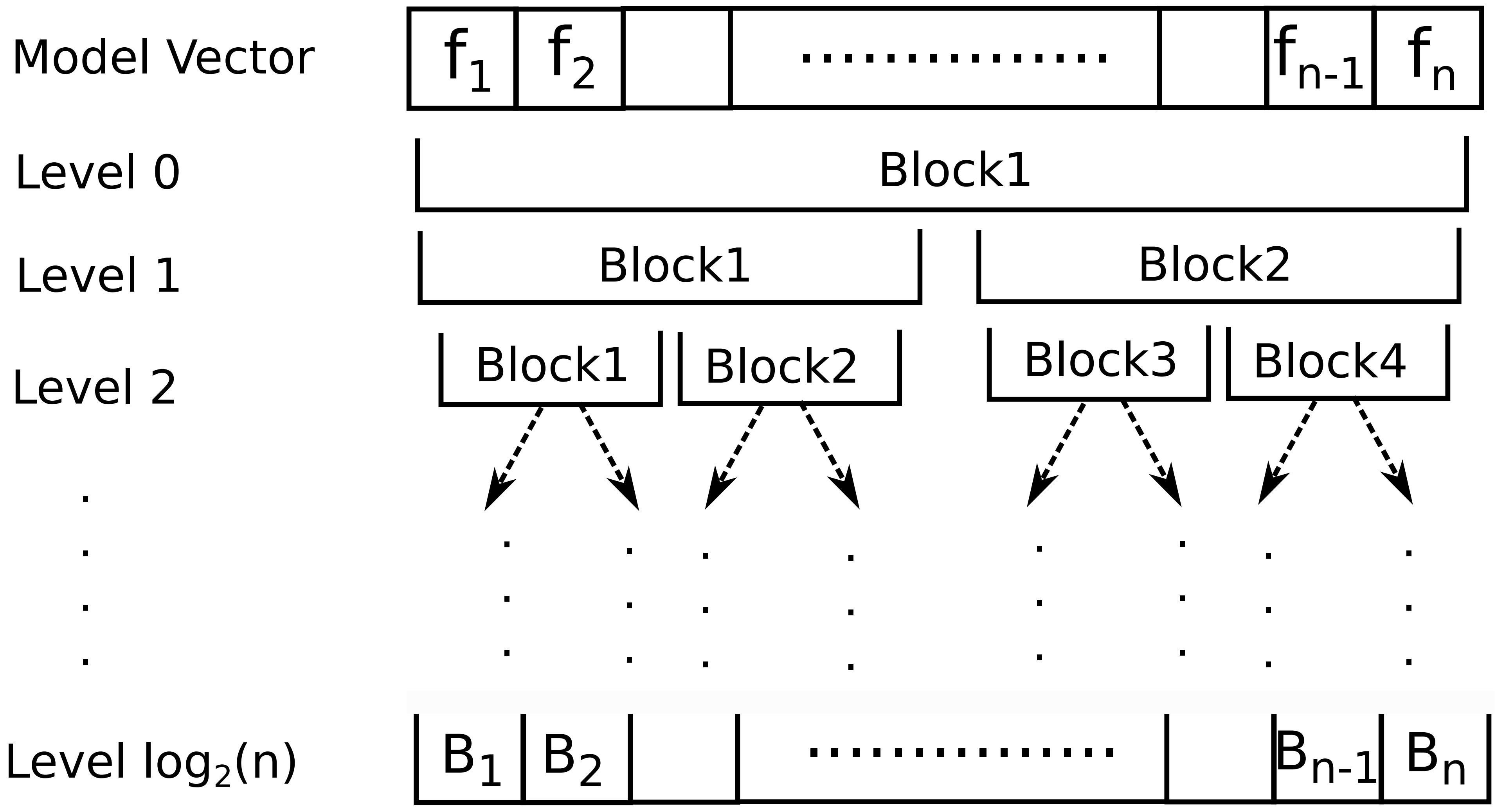






blocked mode sampling

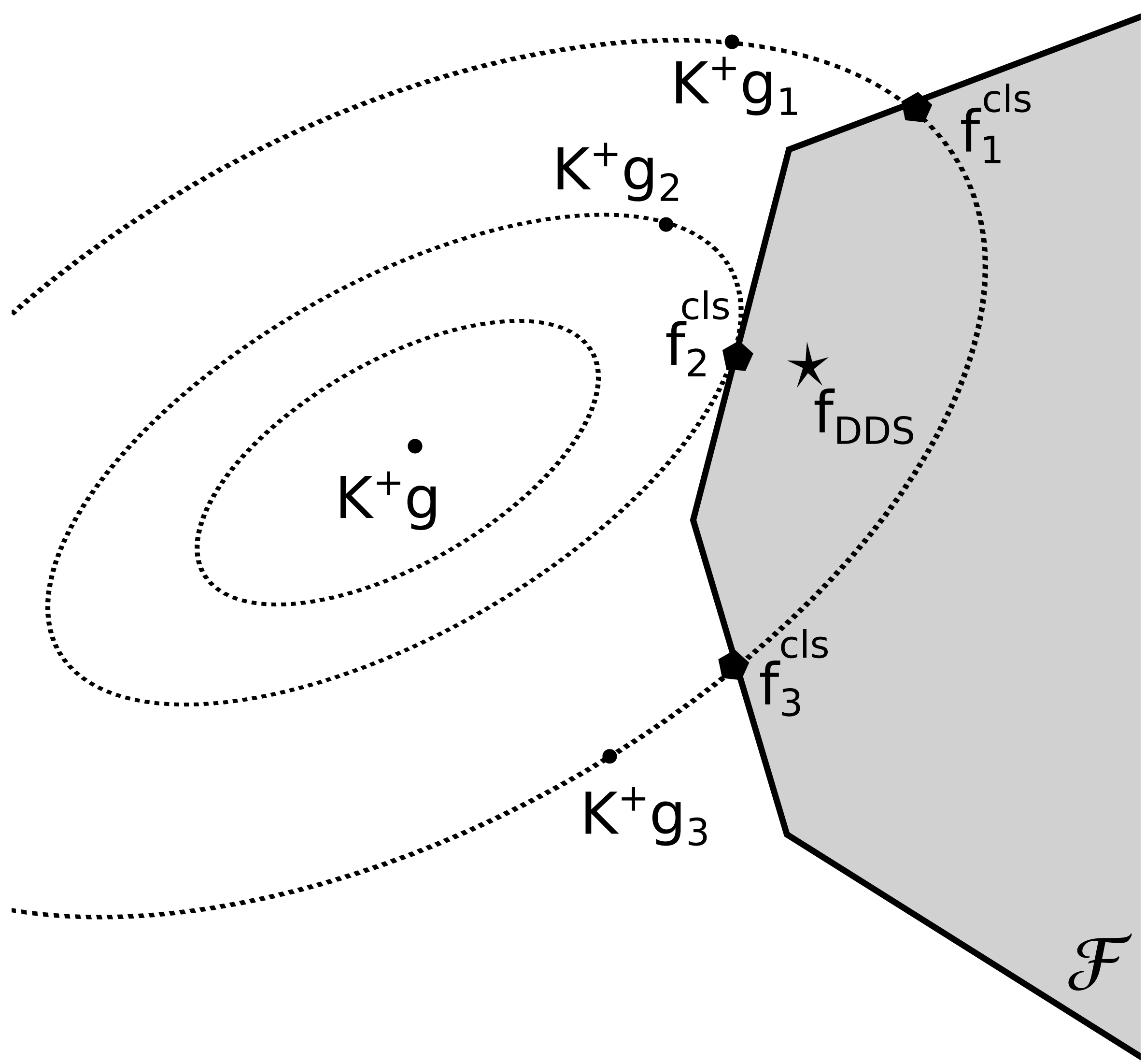
subdivide ω-axis into subintervals on which to do mode samping



use different primes for subdivision to avoid seams

starting point

NNLS solutions tend to be on boundary of admissible set
take linear combination of several NNLS solutions
obtained from adding noise to data



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

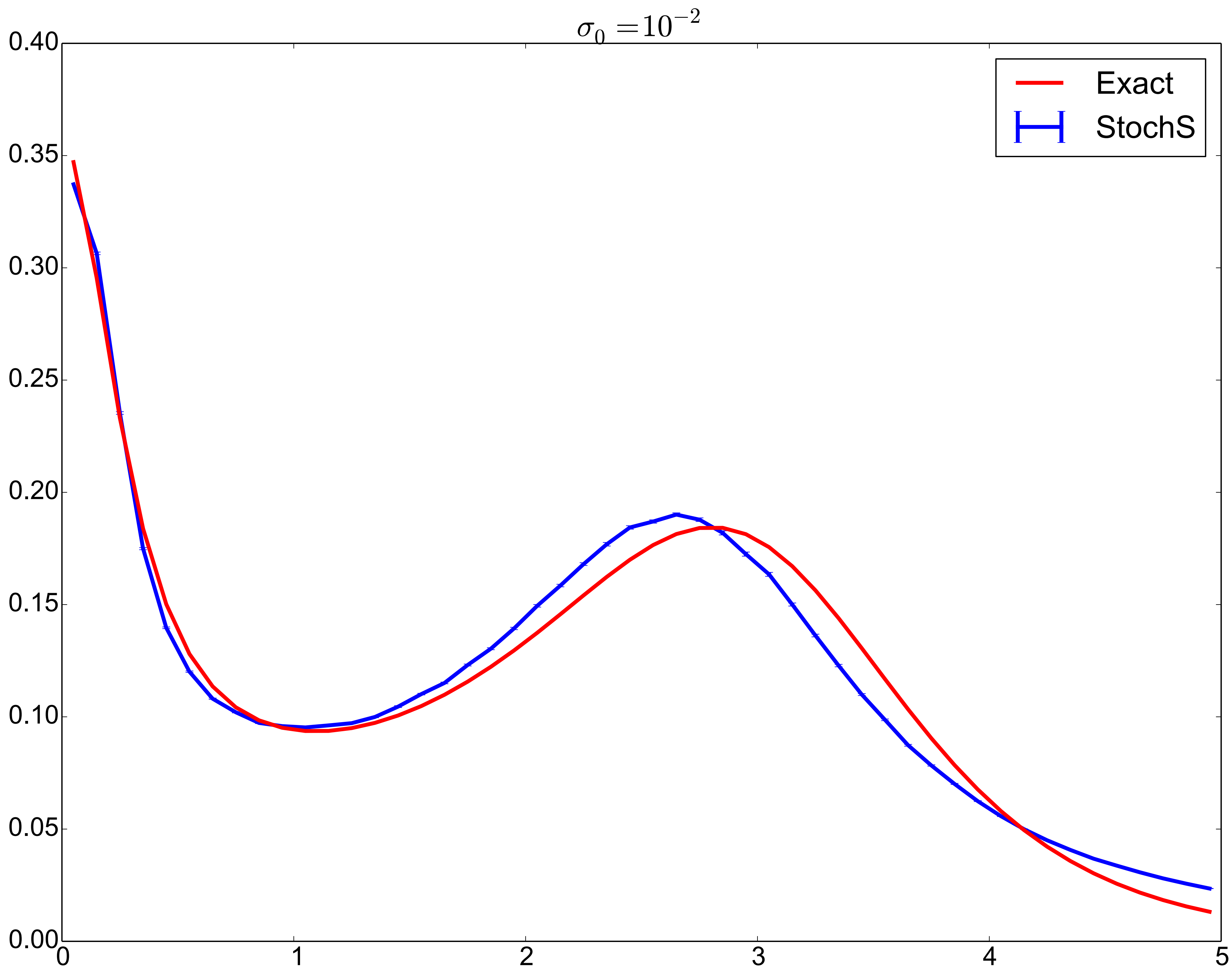
optical conductivity $$\Pi(\nu_n) = \frac{2}{\pi}\int_0^\infty \frac{\omega^2}{\nu_n^2+\omega^2}\,\sigma(\omega)\,d\omega$$

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

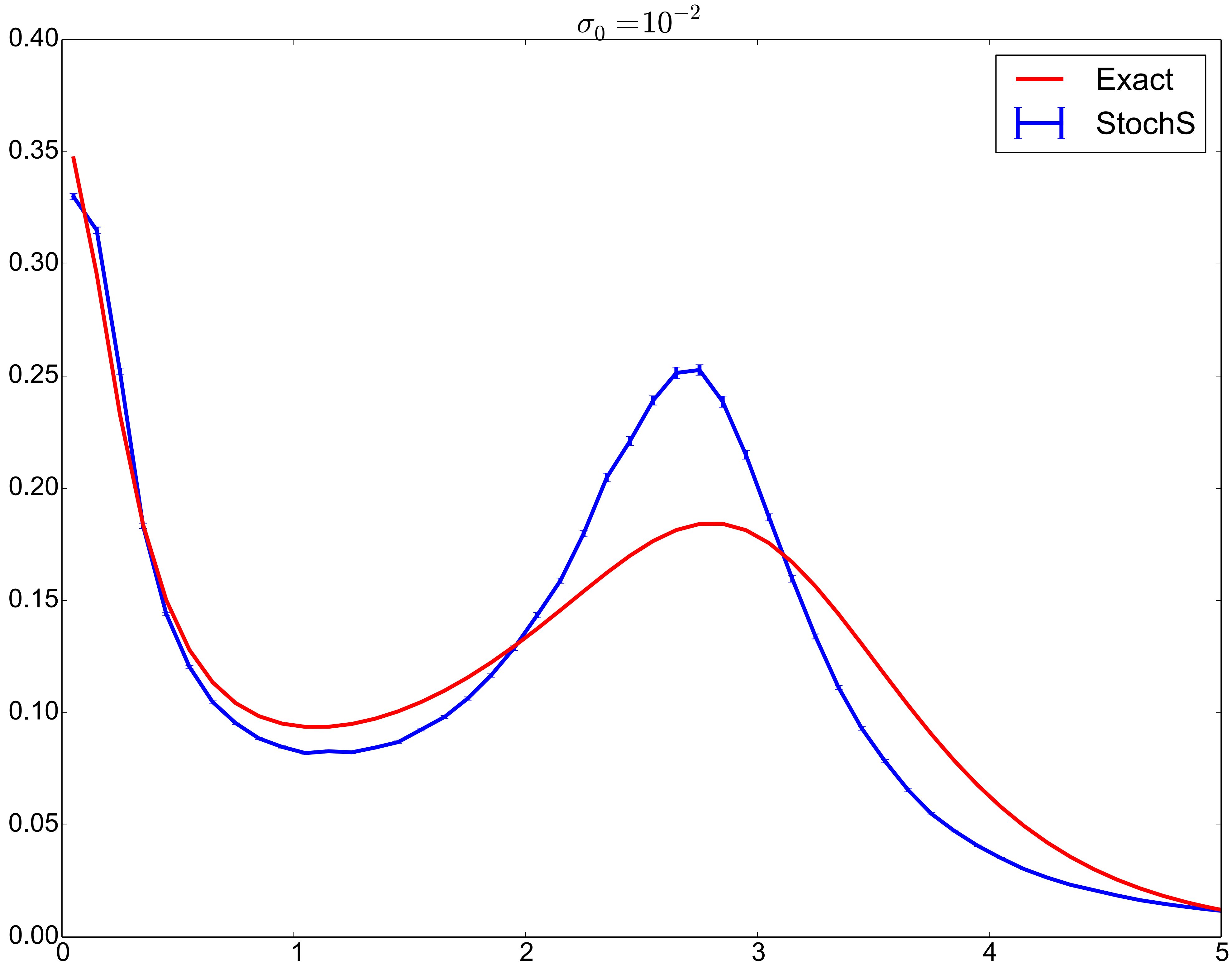
$$\Gamma_1 = 0.3 \text{ or } 0.6, \Gamma_2 = 1.2, \Gamma_3 = 4, \epsilon = 3, W_1 = 0.3, W_2 = 0.2$$

vary noise on data

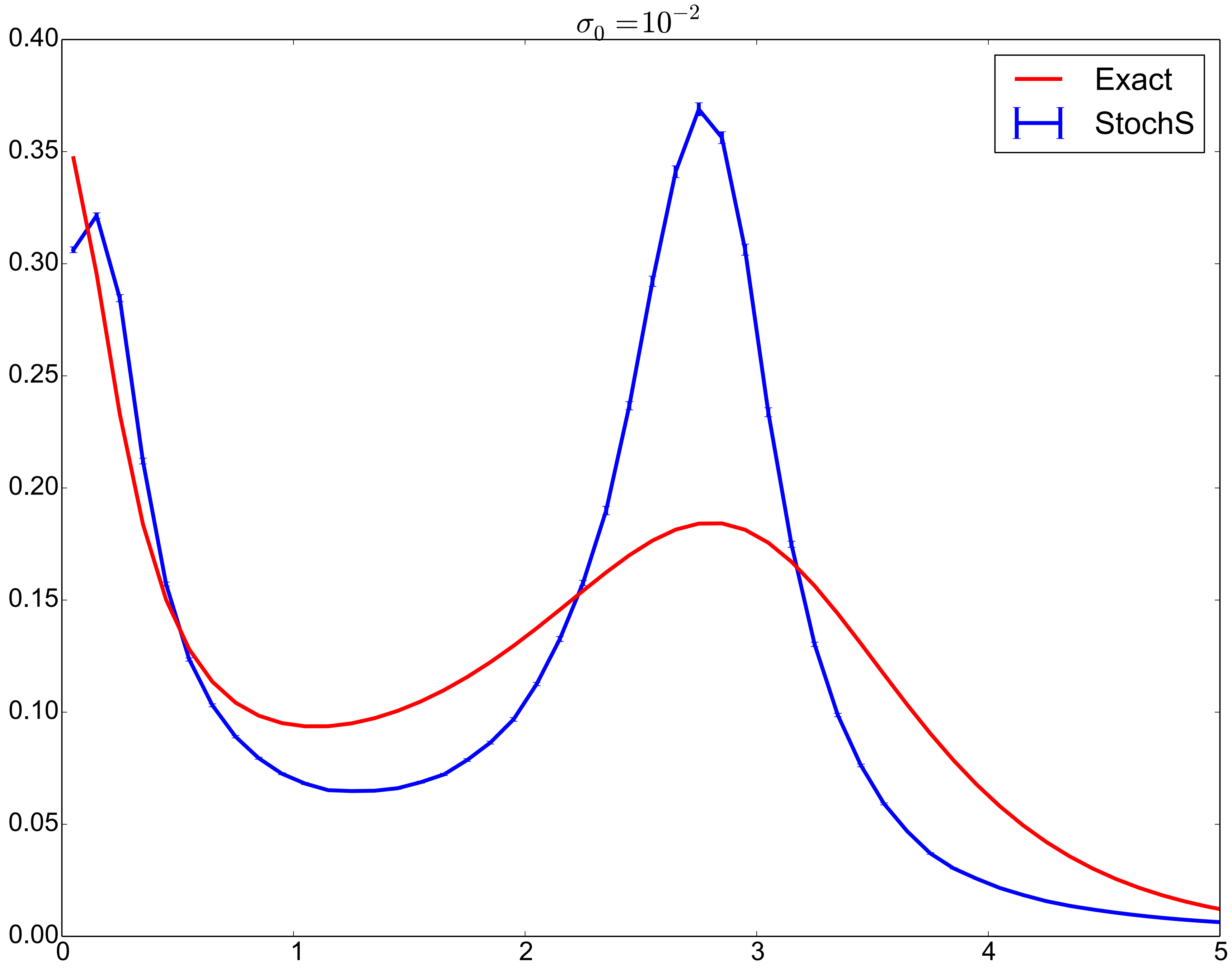
stochastic sampling



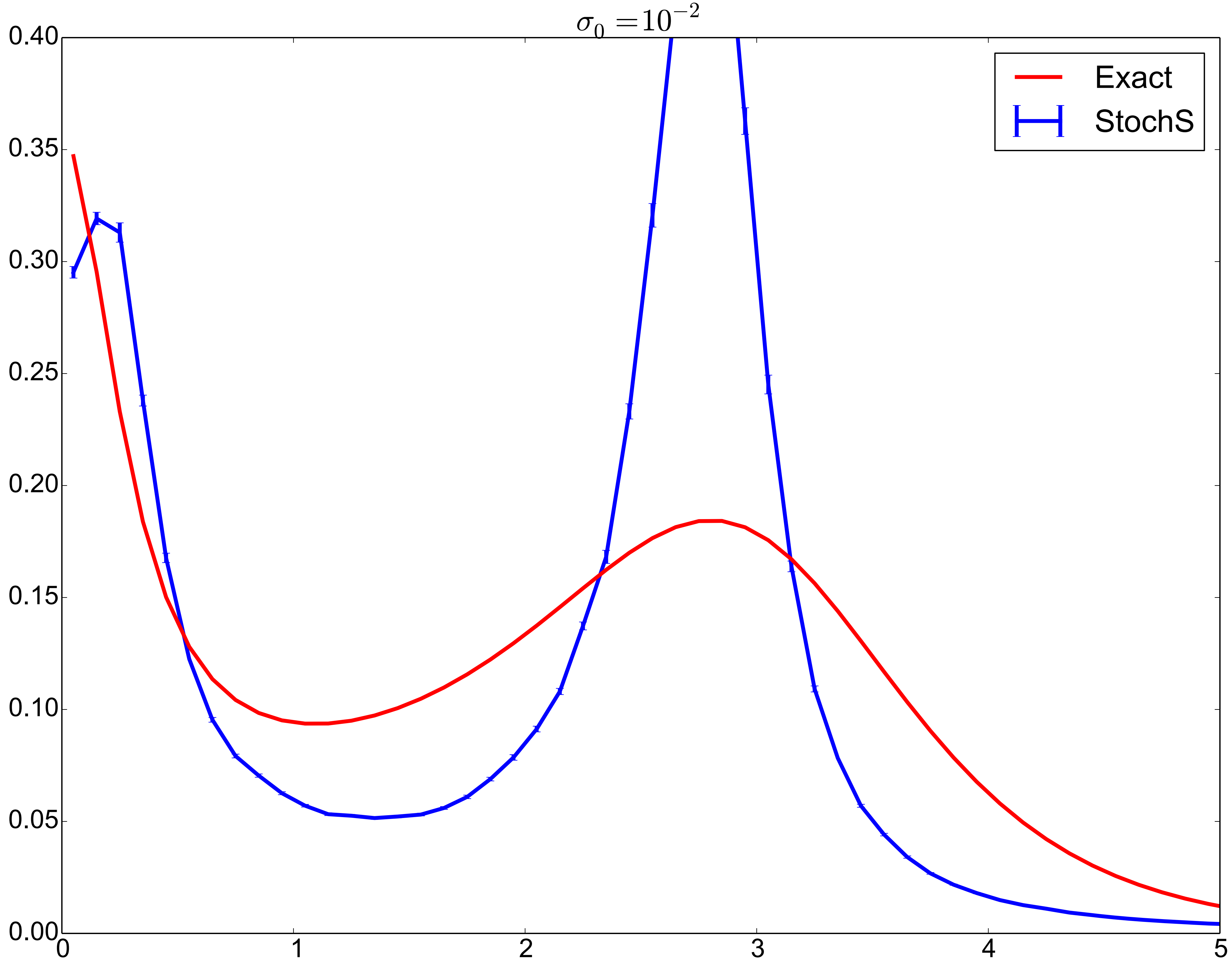
improve cutoff: $\omega_{\text{cut}} = 8$



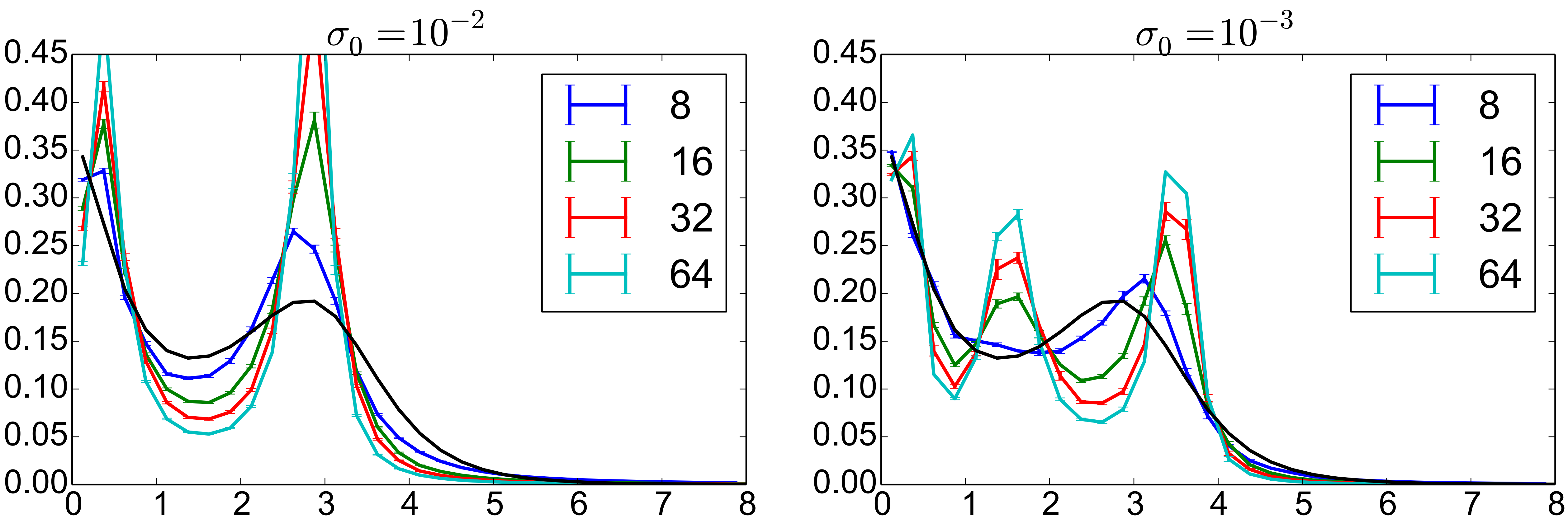
improve cutoff: $\omega_{\text{cut}} = 16$



improve cutoff: $\omega_{cut} = 32$



cutoff dependence

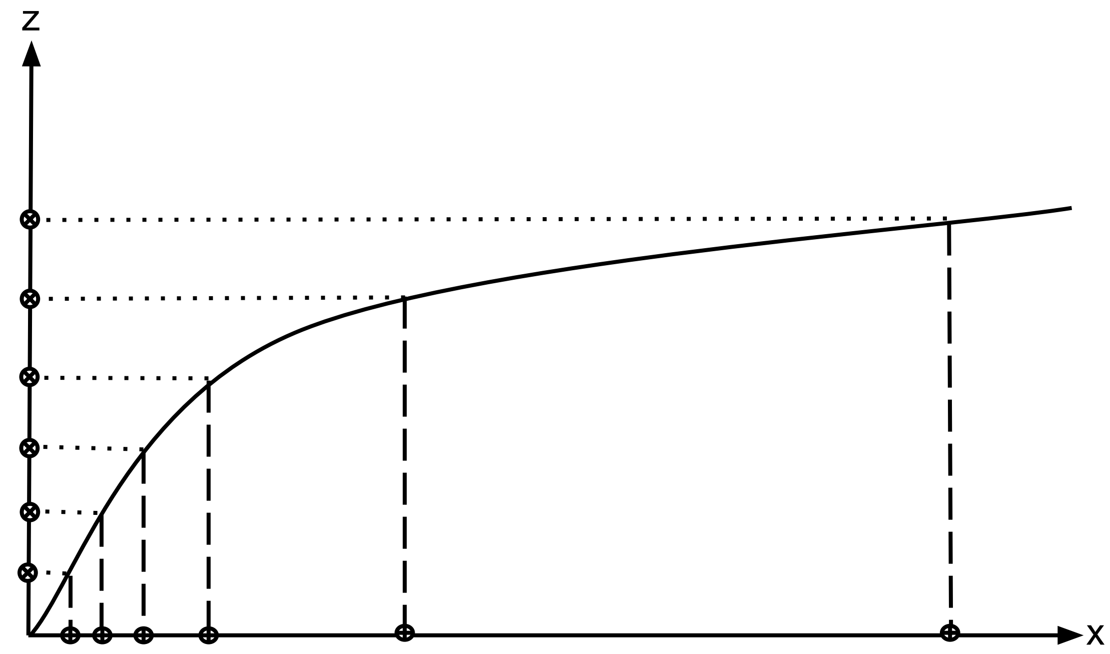


same model, more accurate data

effect of discretization

variable transform

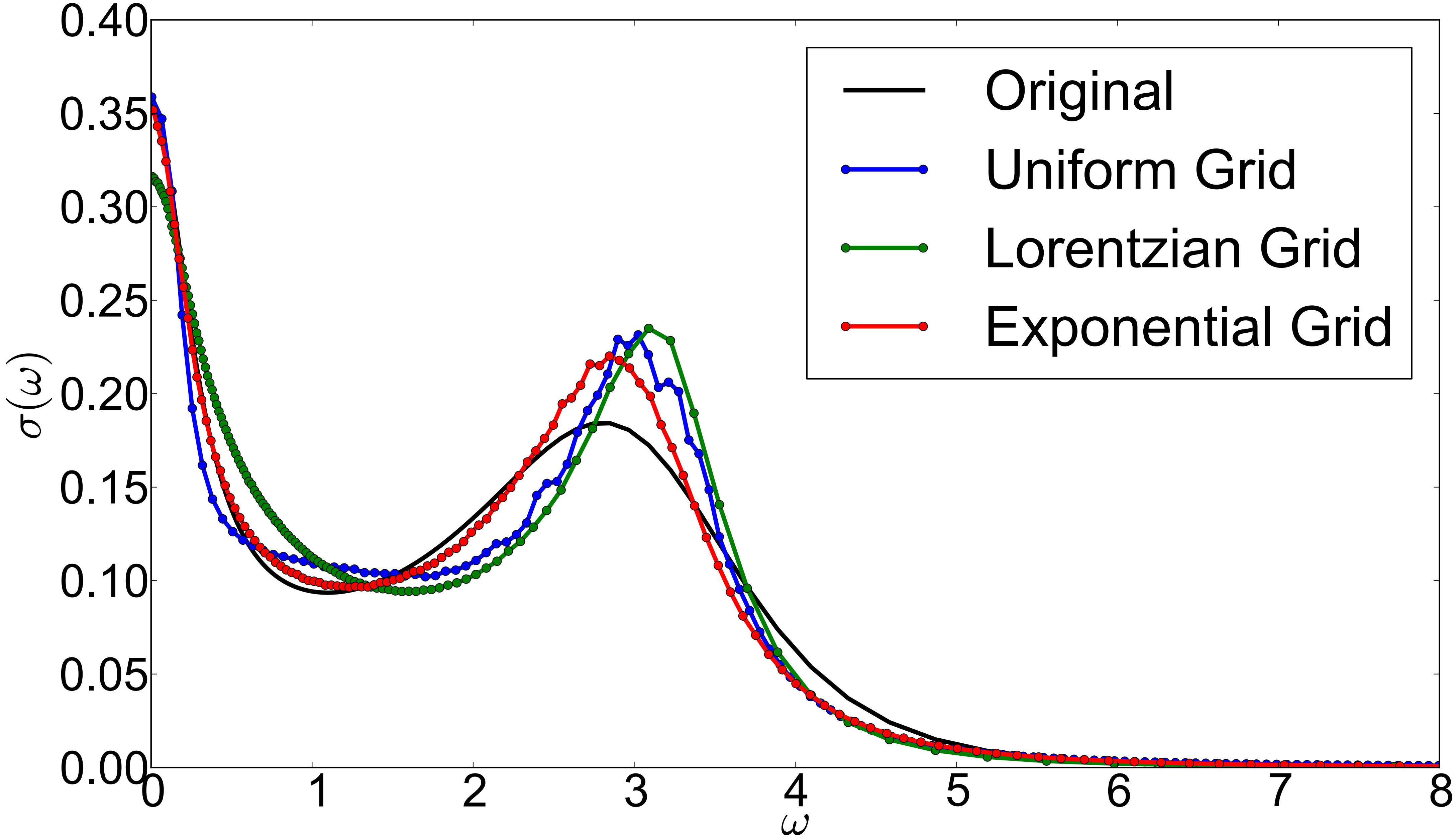
$$\omega \to z = \int_0^{\omega} \rho(\omega')\,d\omega'$$



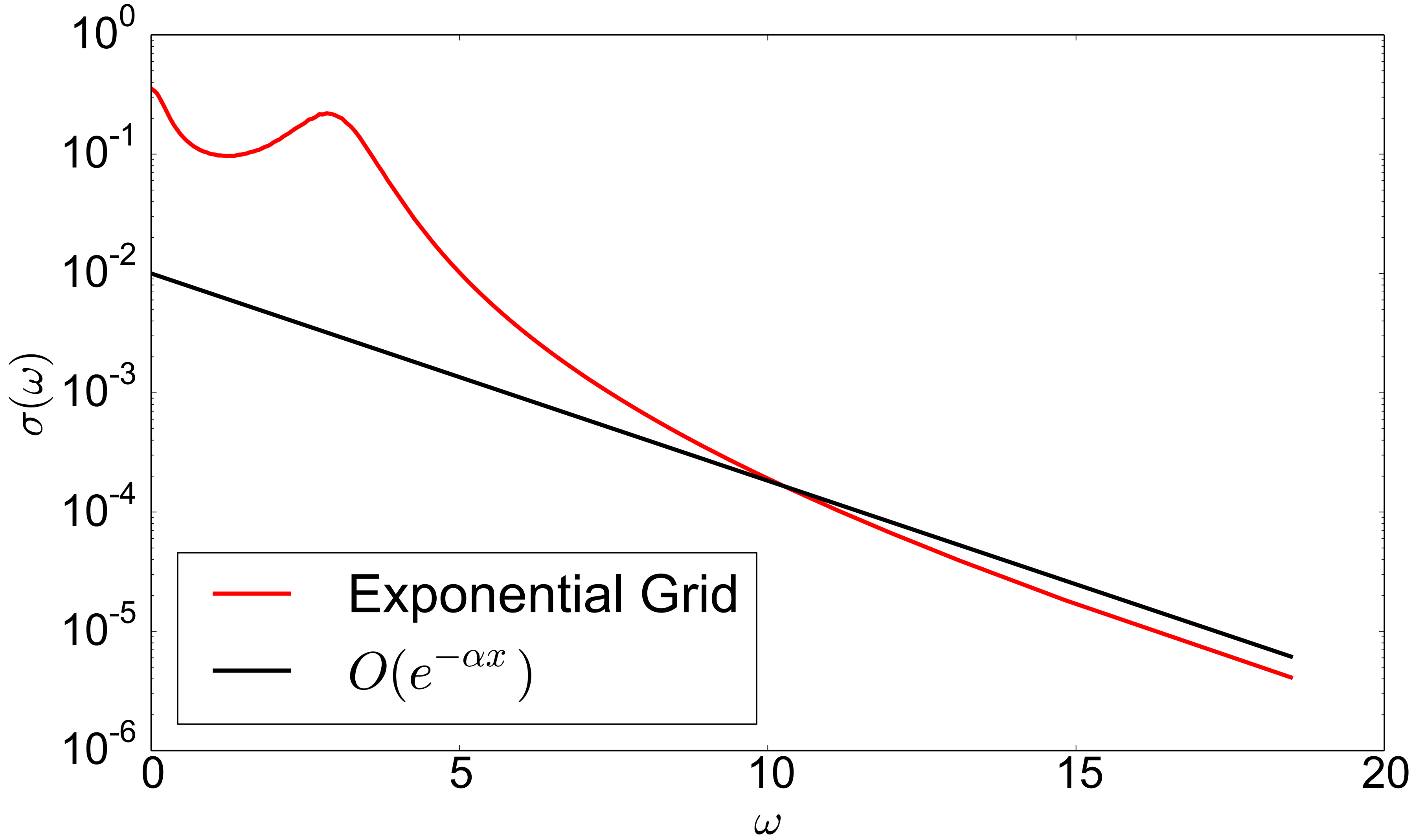
exponential grid $\quad \rho(\omega) = \lambda\, e^{-\lambda\omega)}$

Lorentzian grid $\quad \rho(\omega) = \dfrac{\gamma}{\pi}\dfrac{1}{\omega^2 + \gamma^2}$

grid dependence

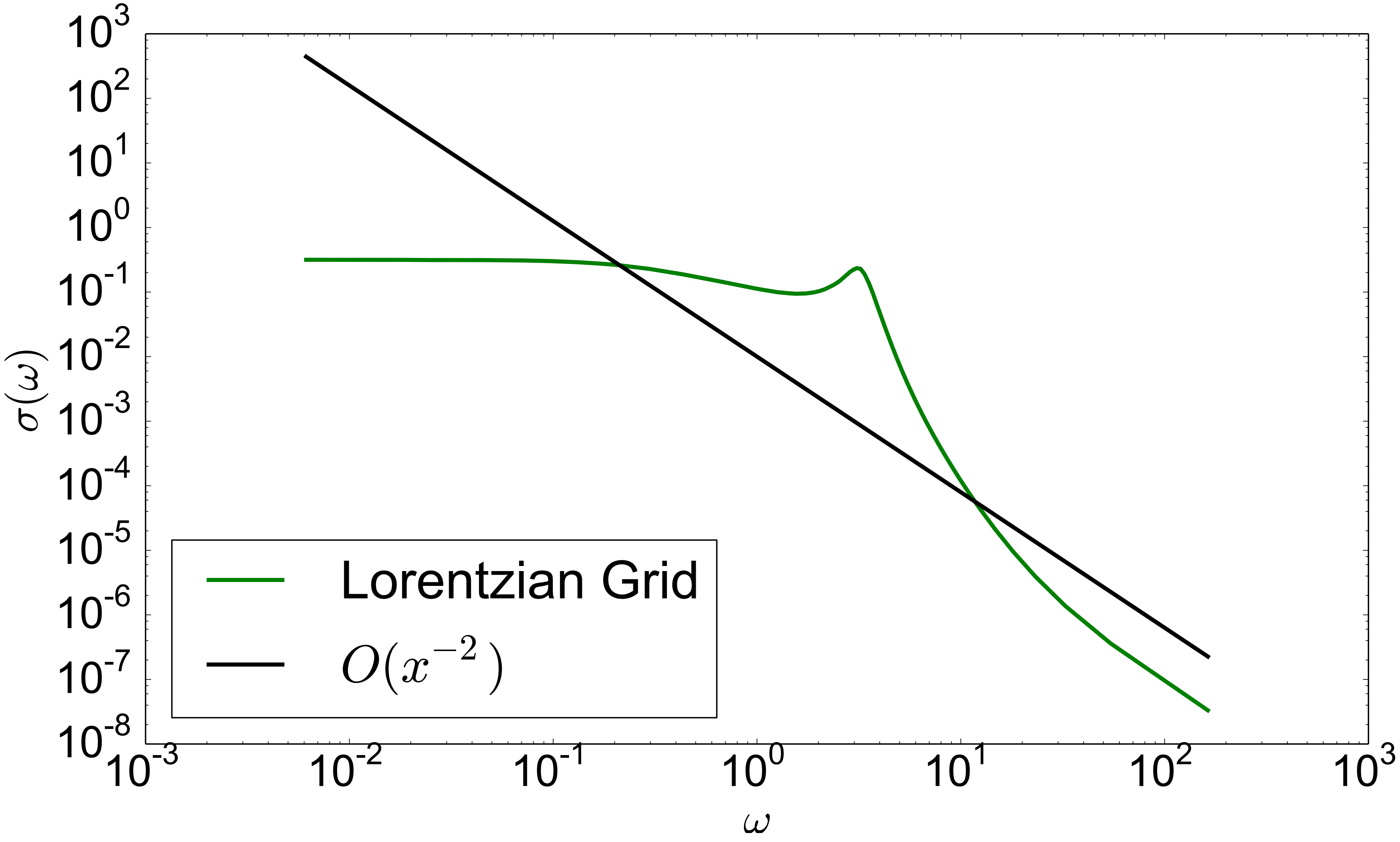


behavior where data is weak



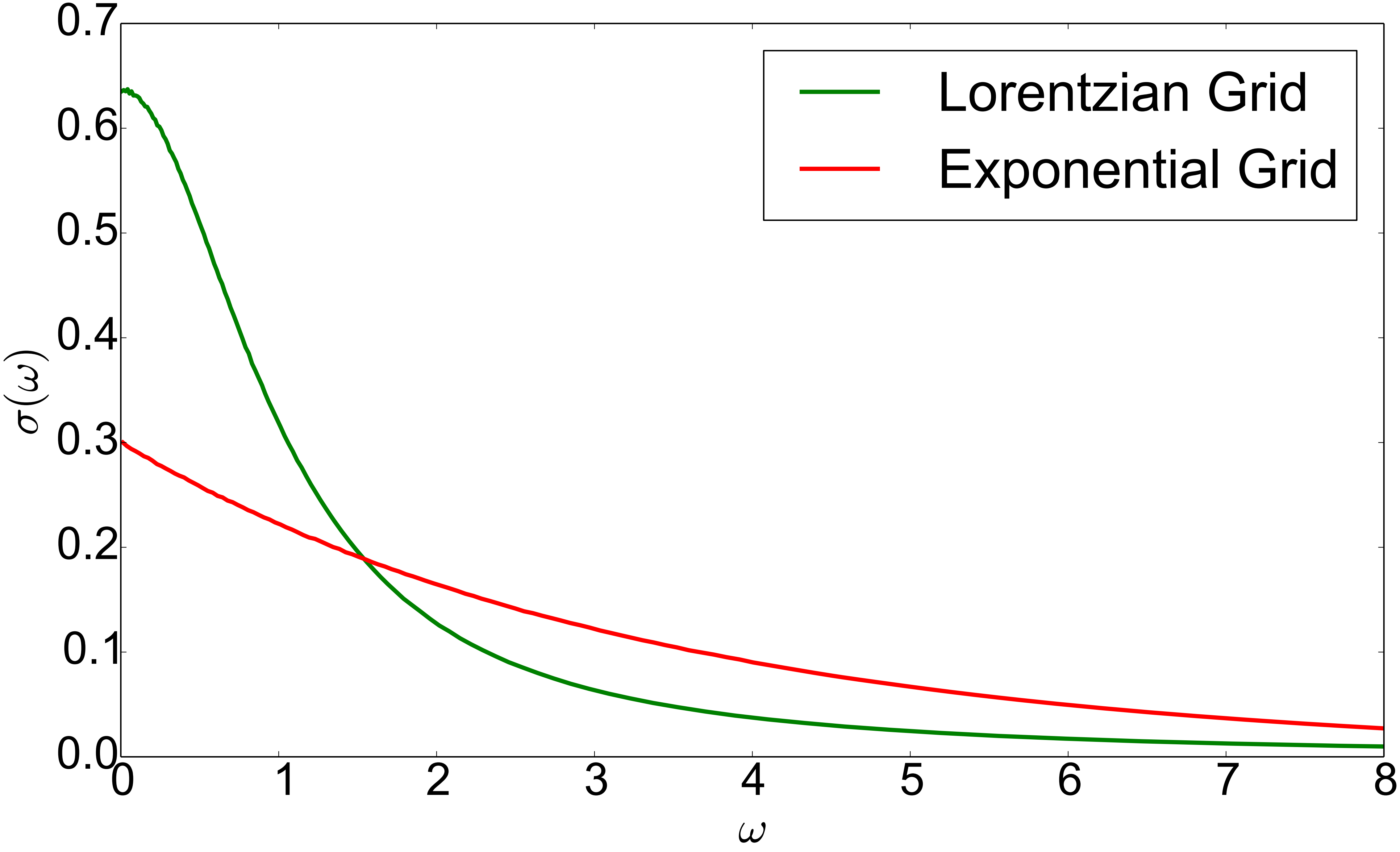
for large ω stochastic sampling result
behaves like density function $\varrho(\omega)$

behavior where data is weak



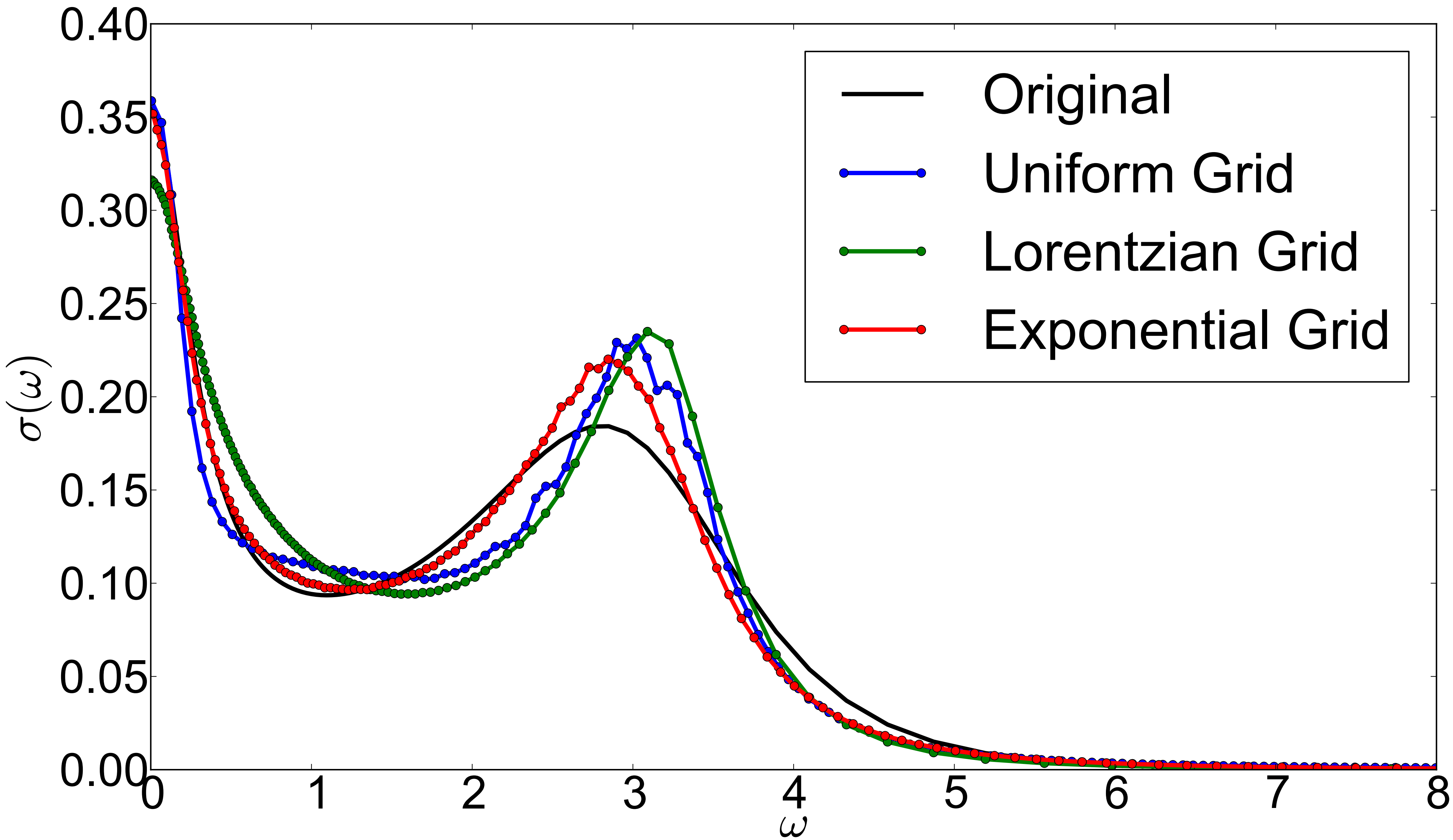
for large ω stochastic sampling result
behaves like density function $\varrho(\omega)$

limit of no data



in absence of data stochastic sampling recovers grid density
default model

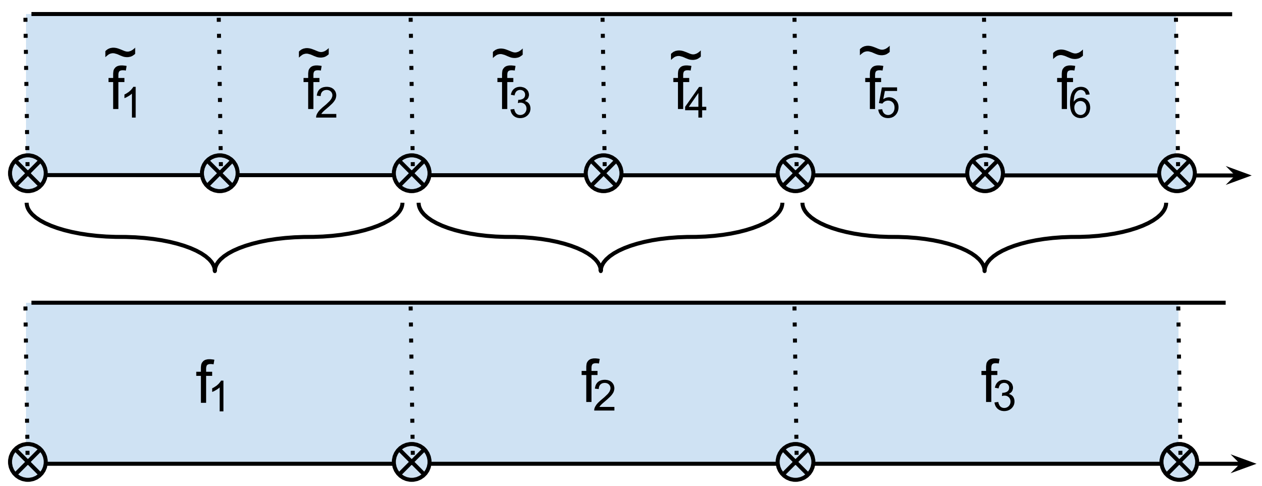
understand grid dependence?



grid resolution

sample uniformly on fine grid

$$U(x) = \lim_{\lambda \to \infty} \mathrm{Exp}(\lambda; x) = \lim_{\lambda \to \infty} \frac{1}{\lambda} e^{-x/\lambda}$$



$\tilde{f}_i \sim \mathrm{Exp}(\lambda) = \mathrm{Gamma}(1, \lambda)$

$\Downarrow$

$f_i \sim \mathrm{Gamma}(2, \lambda)$

sum of *k* exponentially distributed variables

$$p_{\sum}(x) = \int_0^x p(x')\, p(x - x')\, dx$$

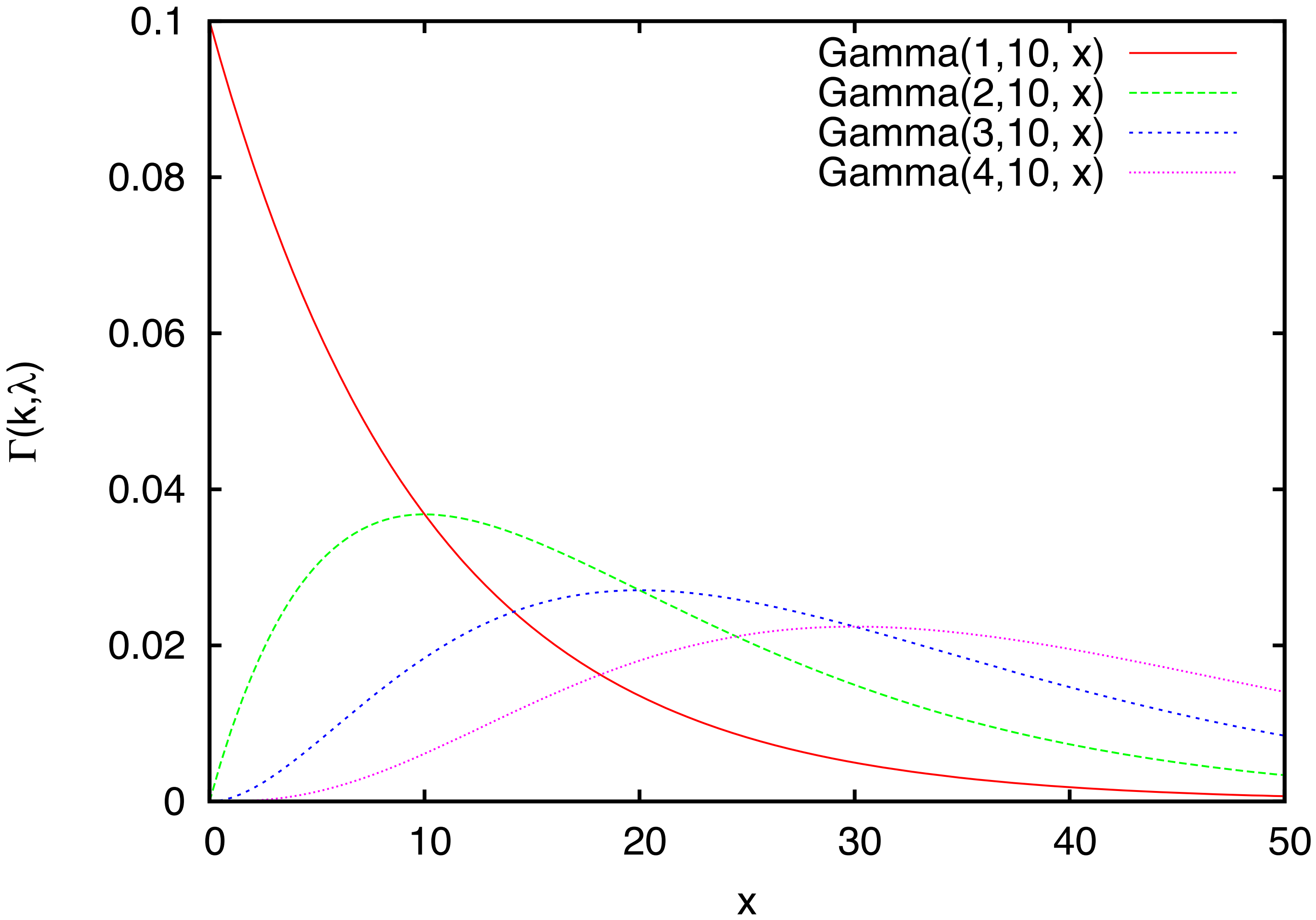
$$\mathrm{Gamma}(k, \lambda; x) = \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-x/\lambda}$$

to simulate finer grid, reweight on coarser grid

$$\frac{p_{\mathrm{fine}}(f_1, f_2, \ldots, f_n)}{p_{\mathrm{coarse}}(f_1, f_2, \ldots, f_n)} \propto \lim_{\lambda \to \infty} \frac{\prod_{i=1}^n f_i^{k-1} e^{-f_i/\lambda}}{\prod_{i=1}^n e^{-f_i/\lambda}} = \prod_{i=1}^n f_i^{k-1}$$

Gamma distribution

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

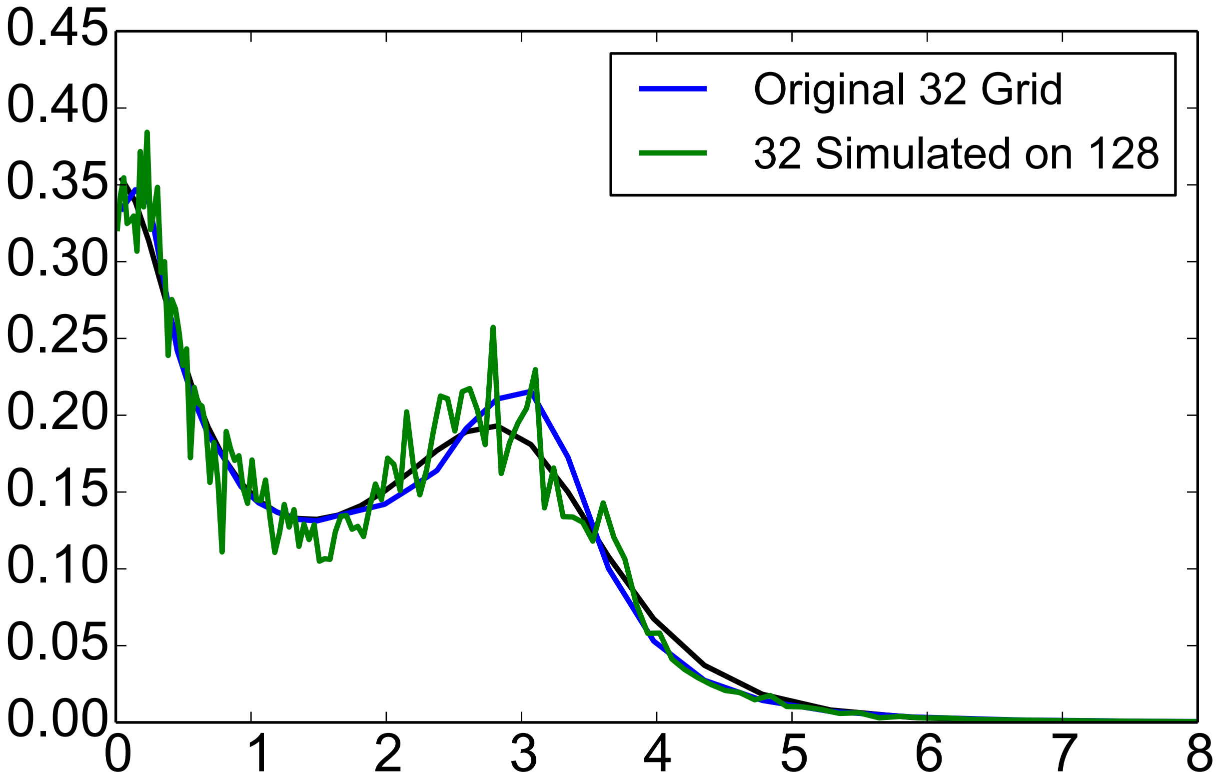


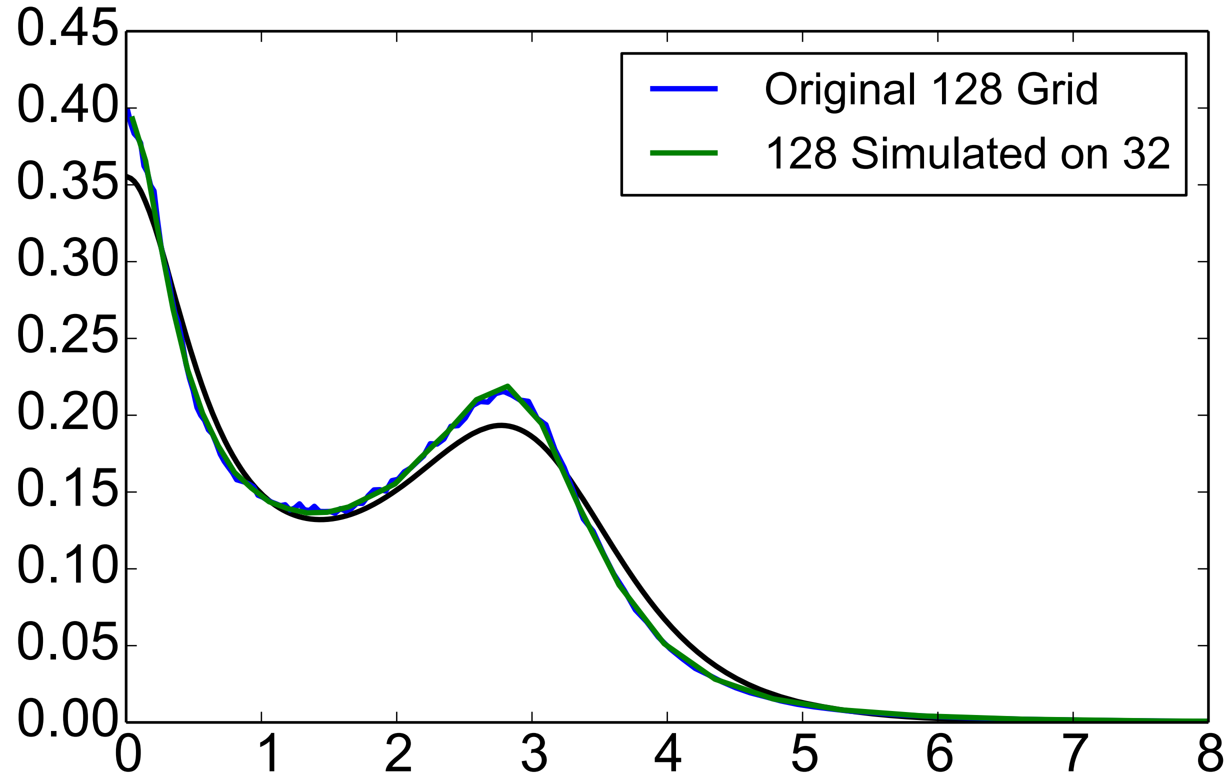
$$\int_0^\infty x\,\mathrm{Gamma}(k,\lambda;x')\;dx = \lambda\,k$$

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

simulating grids of different resolution

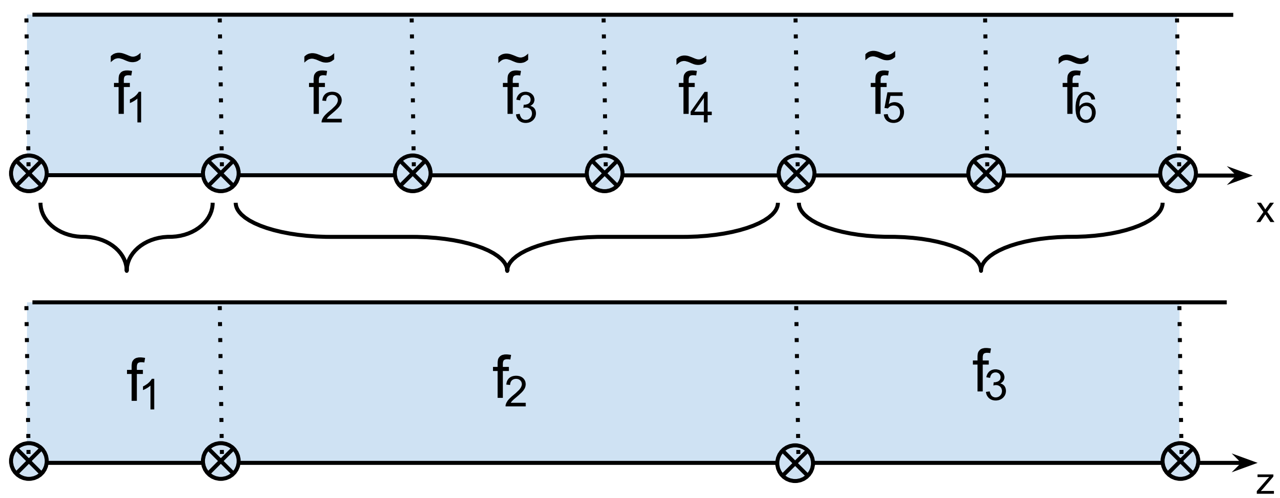
exponential grid with 32 and 128 points





$$\frac{p_{\text{fine}}(f_1, f_2, \ldots, f_n)}{p_{\text{coarse}}(f_1, f_2, \ldots, f_n)} \propto \lim_{\lambda \to \infty} \frac{\prod_{i=1}^{n} f_i^{k-1} e^{-f_i/\lambda}}{\prod_{i=1}^{n} e^{-f_i/\lambda}} = \prod_{i=1}^{n} f_i^{k-1}$$

grid type



$\tilde{f}_i \sim \mathrm{Exp}(\lambda) = \mathrm{Gamma}(1, \lambda)$

$\Downarrow$

$f_1 \sim \mathrm{Gamma}(1, \lambda)$

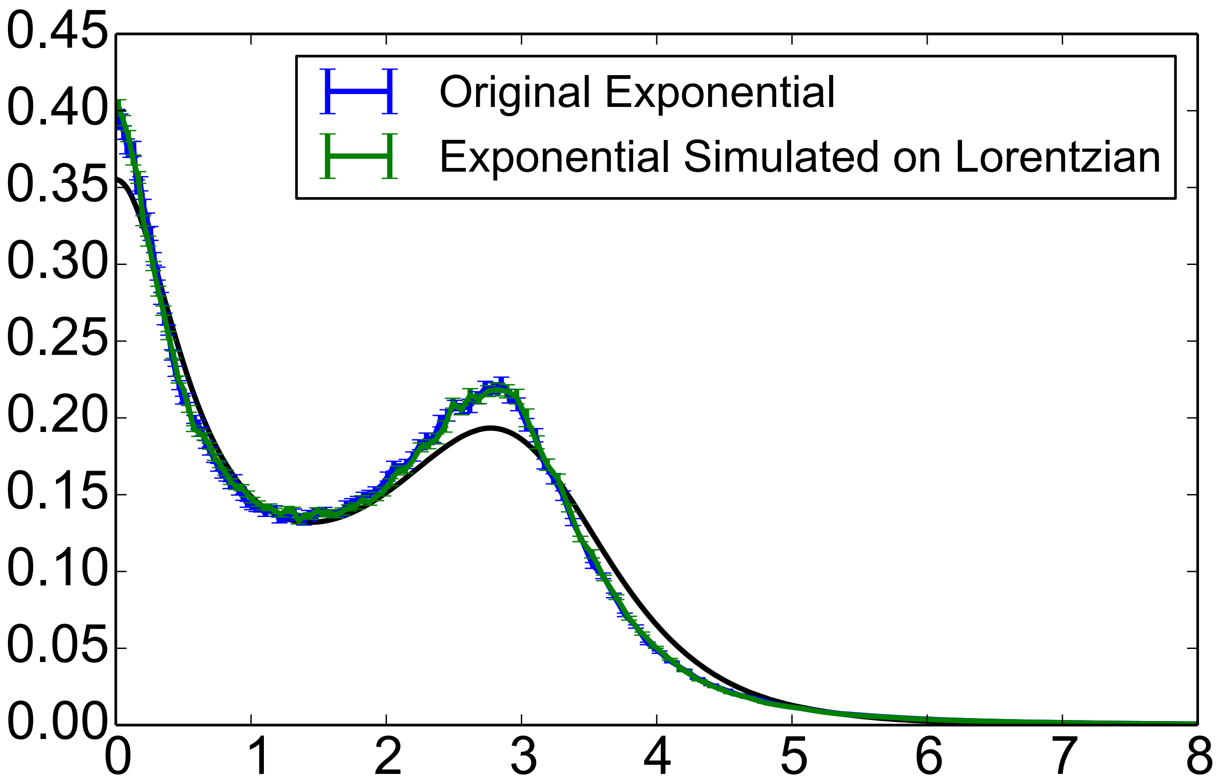
$f_2 \sim \mathrm{Gamma}(3, \lambda)$

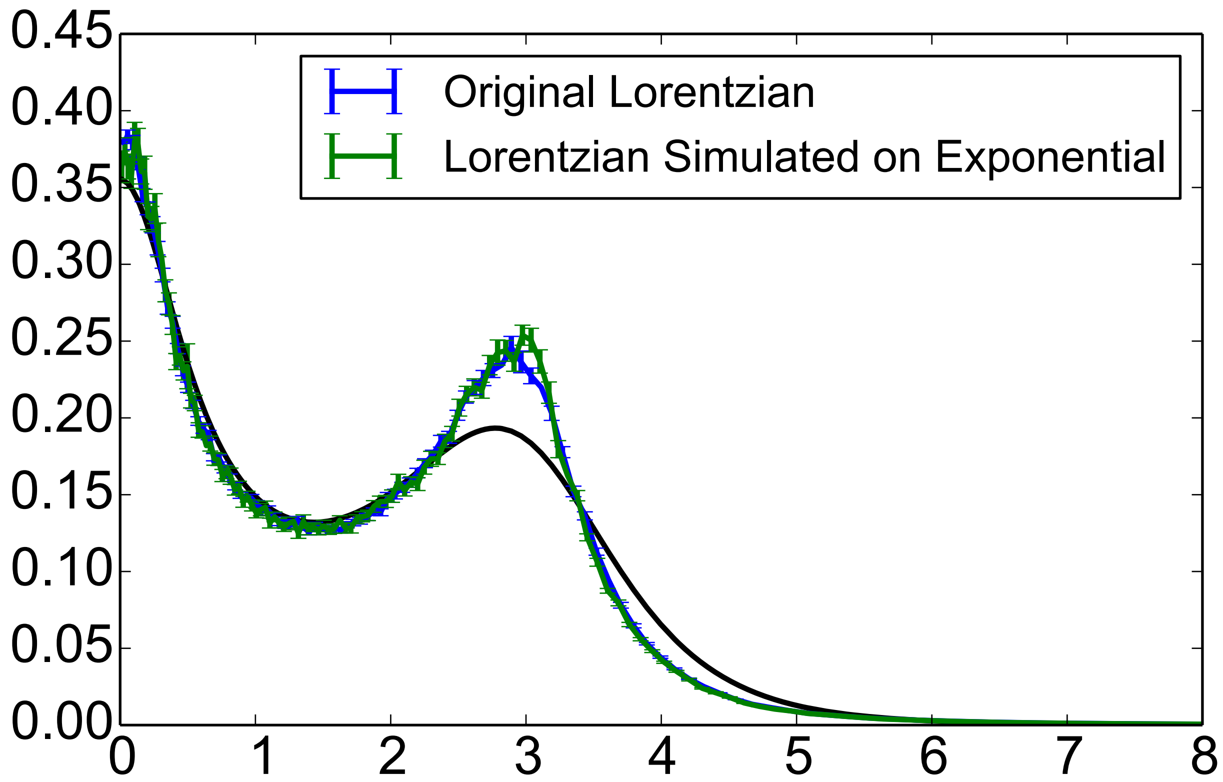
$f_3 \sim \mathrm{Gamma}(2, \lambda)$

$$\frac{p_{\text{fine}}(f_1, f_2, \ldots, f_n)}{p_{\text{other}}(f_1, f_2, \ldots, f_n)} \propto \lim_{\lambda \to \infty} \frac{\prod_{i=1}^{n} f_i^{k_i - 1} e^{-f_i/\lambda}}{\prod_{i=1}^{n} e^{-f_i/\lambda}} = \prod_{i=1}^{n} f_i^{k_i - 1}$$

simulating grids of different type

exponential and Lorentzian grid





$$\frac{p_{\mathrm{fine}}(f_1, f_2, \ldots, f_n)}{p_{\mathrm{other}}(f_1, f_2, \ldots, f_n)} \propto \lim_{\lambda \to \infty} \frac{\prod_{i=1}^{n} f_i^{k_i - 1} e^{-f_i/\lambda}}{\prod_{i=1}^{n} e^{-f_i/\lambda}} = \prod_{i=1}^{n} f_i^{k_i - 1}$$

general case

simulate grid of n_1 points with density function $\varrho_1(x)$
on grid of n_2 points with density function $\varrho_2(x)$

reweight model (f_1, f_2, ..., f_{n2}) on second grid by

$$\frac{p_2(f_1, f_2, \ldots, f_n)}{p_1(f_1, f_2, \ldots, f_n)} = \prod_{i=1}^{n} f_i^{\frac{n_1\,\rho_1(x_i}{n_2\,\rho_2(x_i)} - 1}$$

make result in absencence of data proportional to $\varrho_1(x)$

separate grid from default model

recipe for discretization

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

ω ≥ 0: $$D(\omega) = \frac{1}{\mu}\, e^{-\omega/mu}$$

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

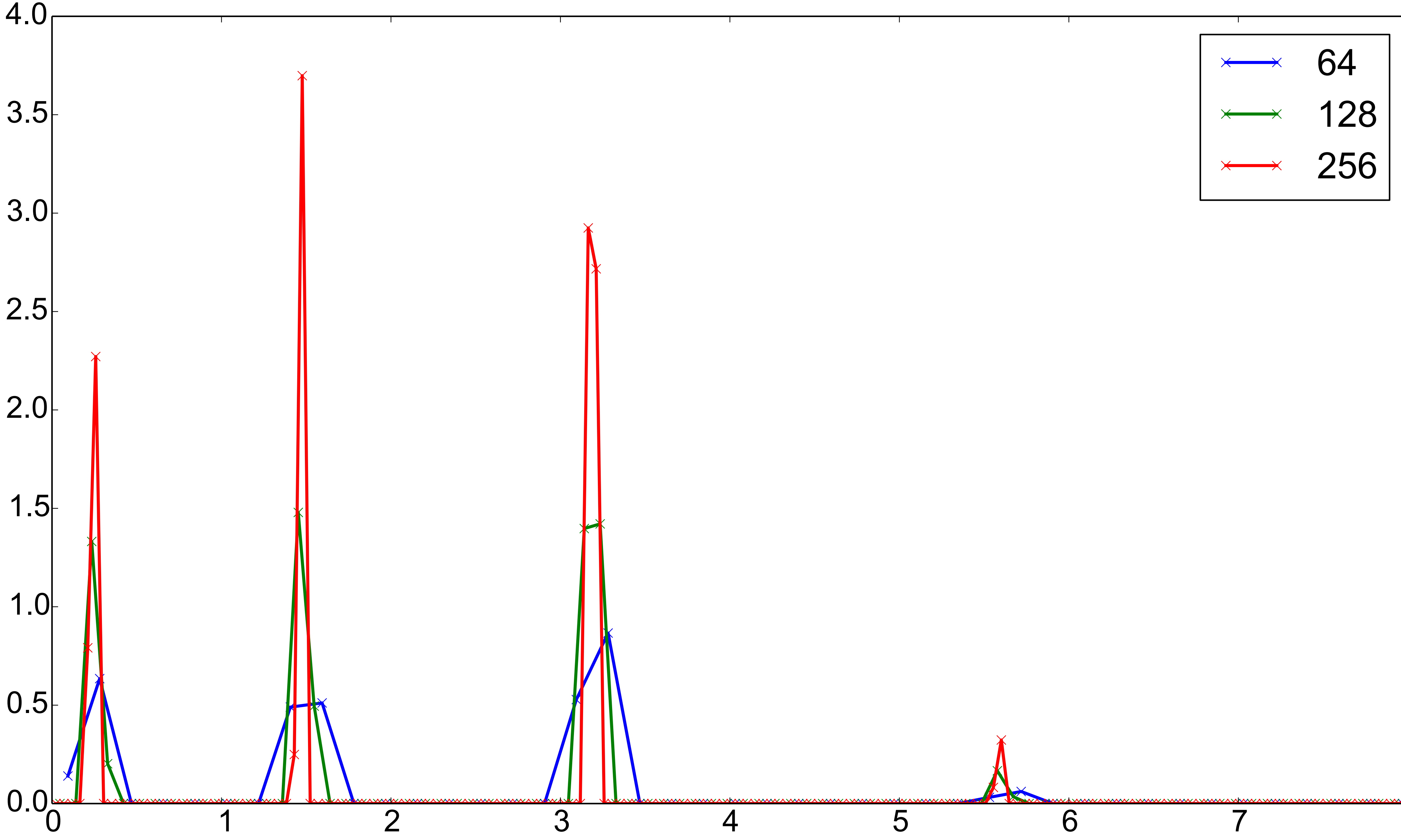
grid: make numerical error in evaluating Kernel smaller than noise in data

ω ≥ 0: Lorentzian at 0 with half-width μ

otherwise: Lorentzian at μ with width σ

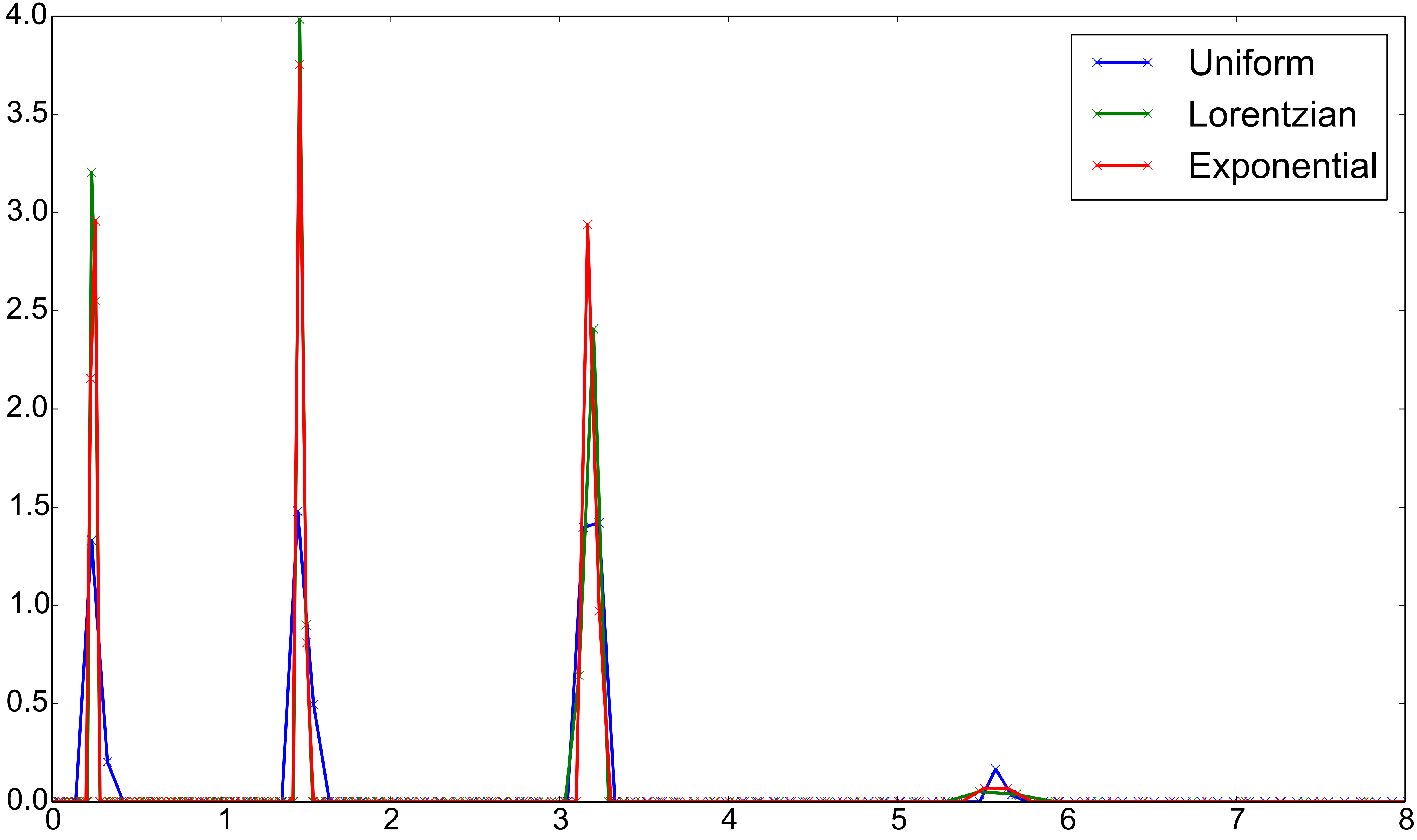
does NNLS depend on grid?

uniform grid, increasing density
peaks get narrower but retain their weight

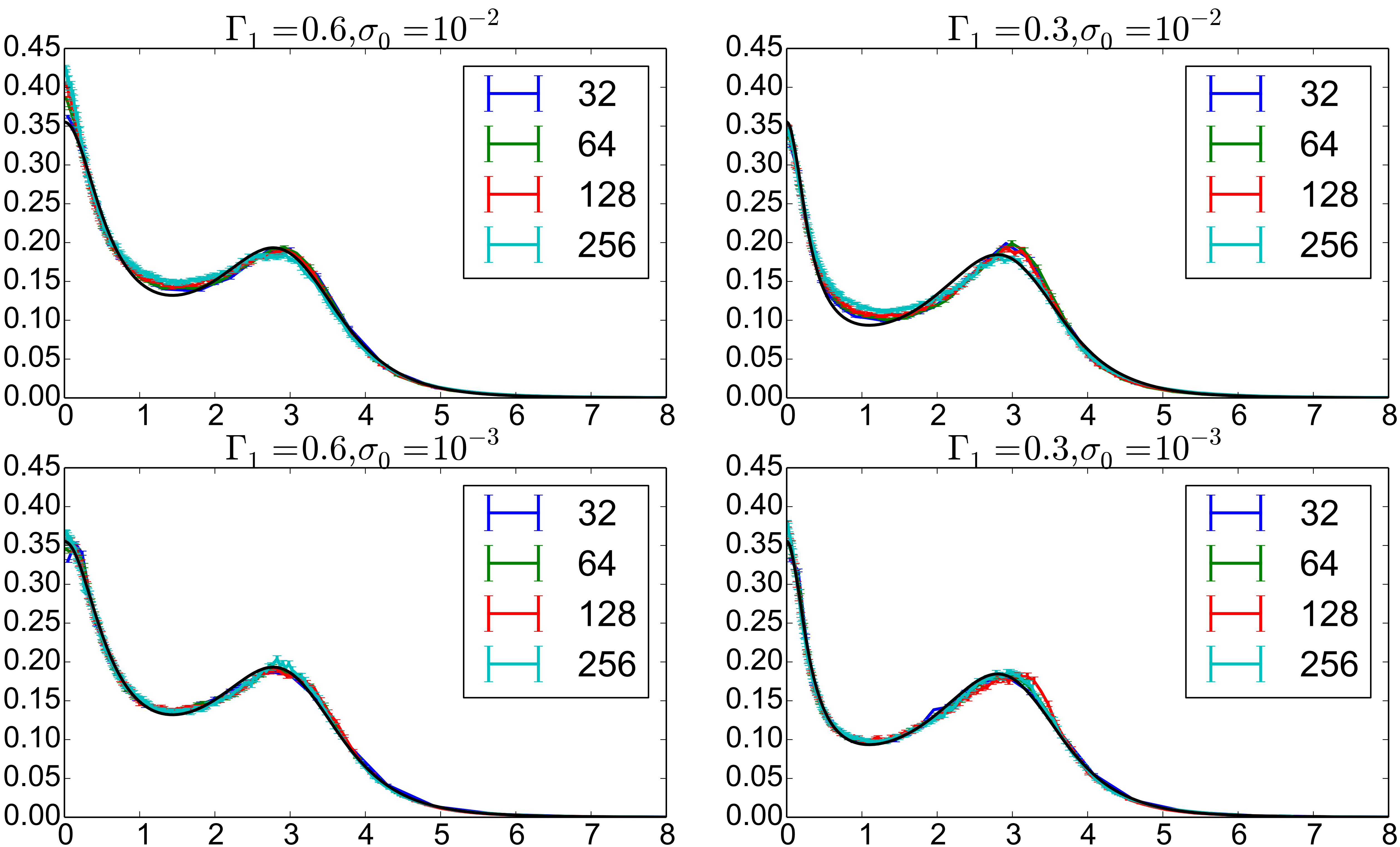


does NNLS depend on grid?

different grids, same peaks (up to resolution)

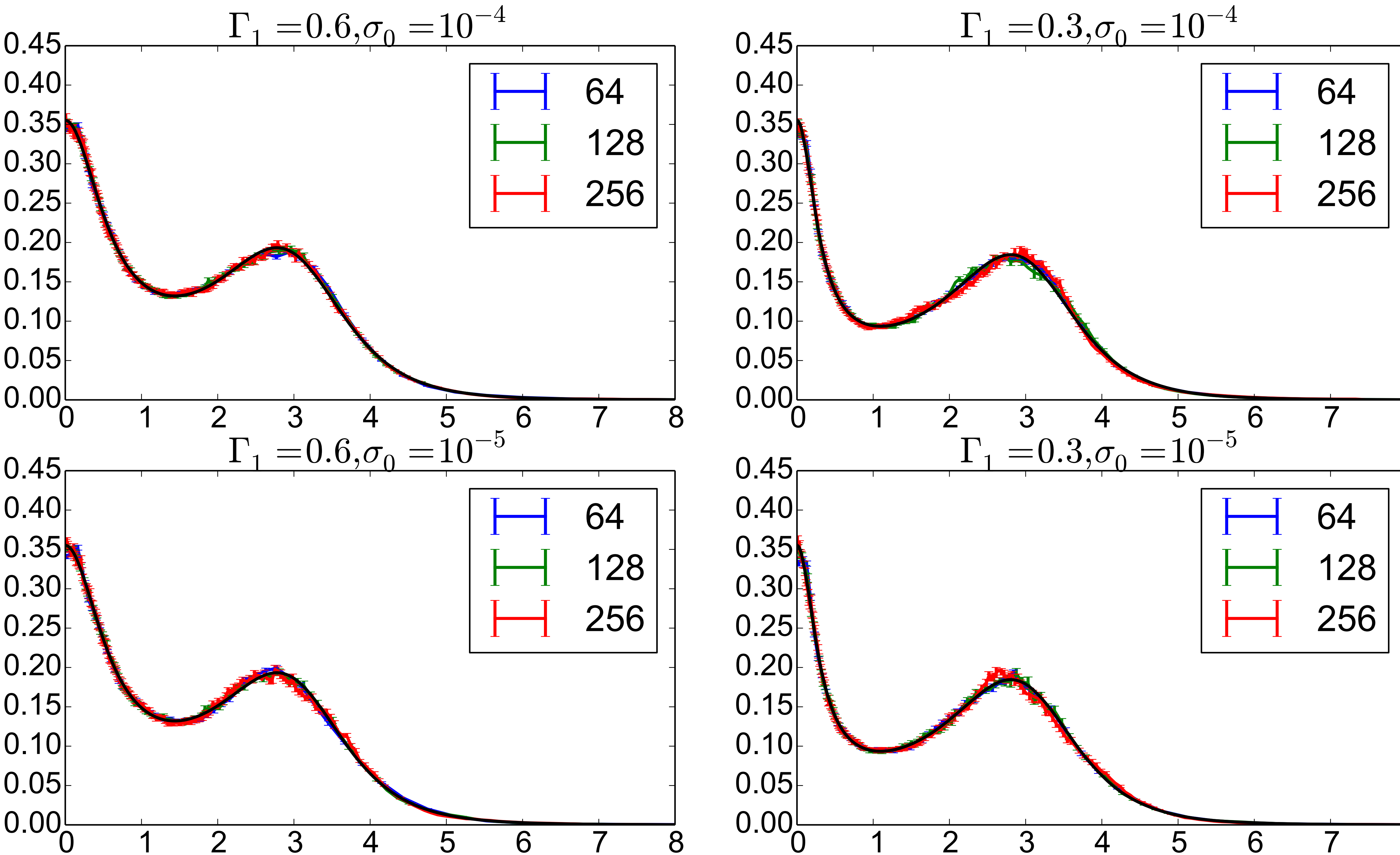


results

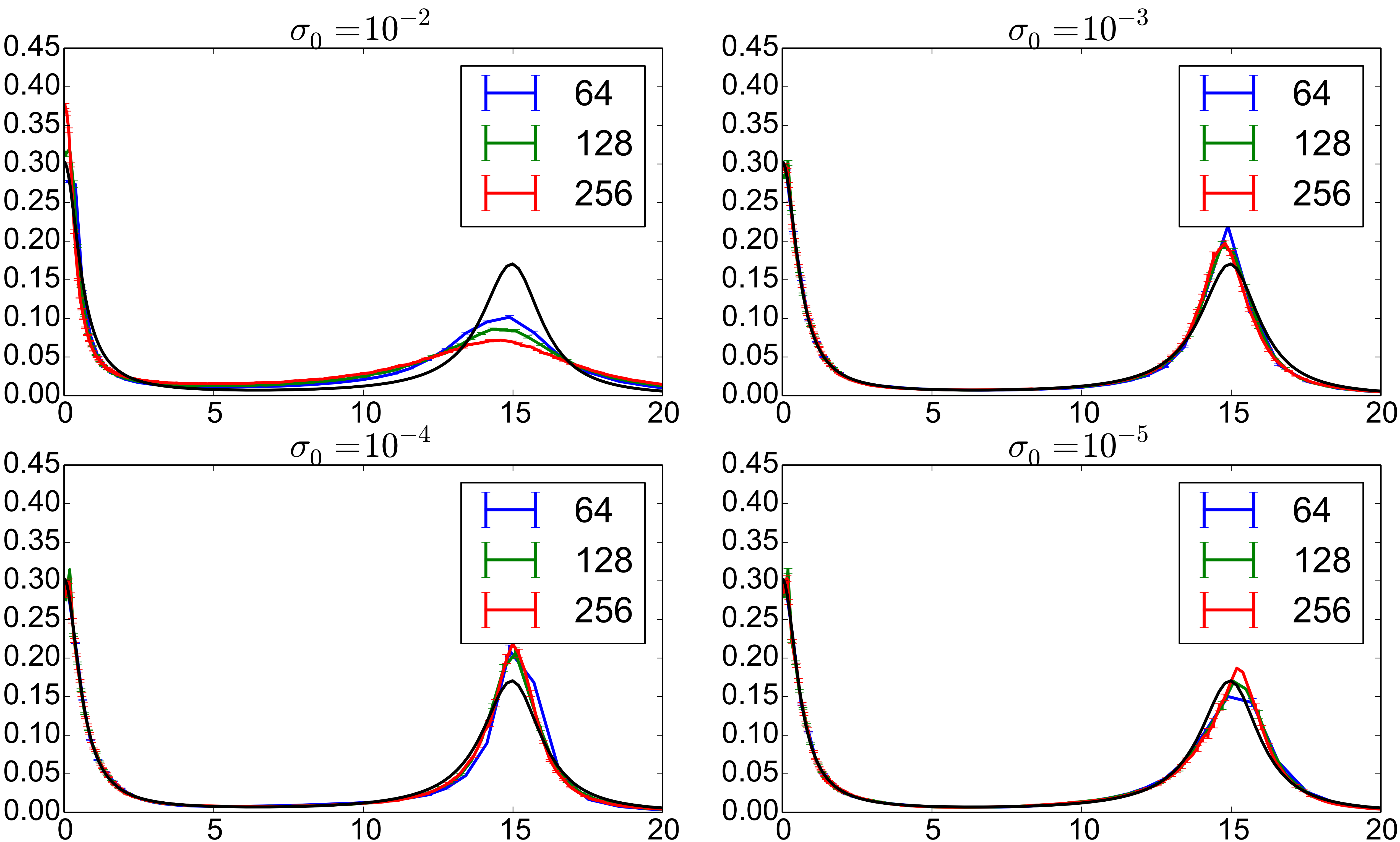


$\Gamma_1 = 0.6, \sigma_0 = 10^{-2}$
$\Gamma_1 = 0.3, \sigma_0 = 10^{-2}$
$\Gamma_1 = 0.6, \sigma_0 = 10^{-3}$
$\Gamma_1 = 0.3, \sigma_0 = 10^{-3}$
32
64
128
256
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45
0 1 2 3 4 5 6 7 8

reduce noise on data

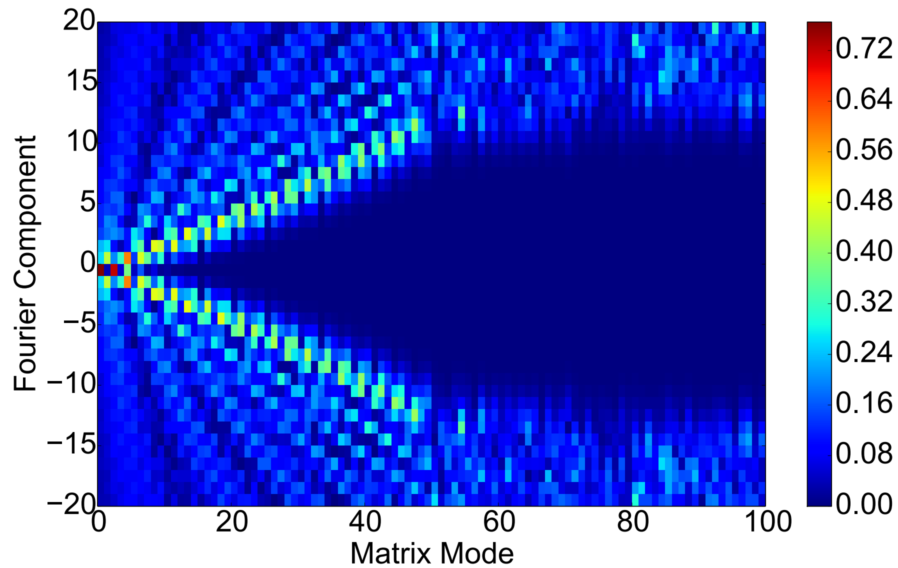


second peak further out

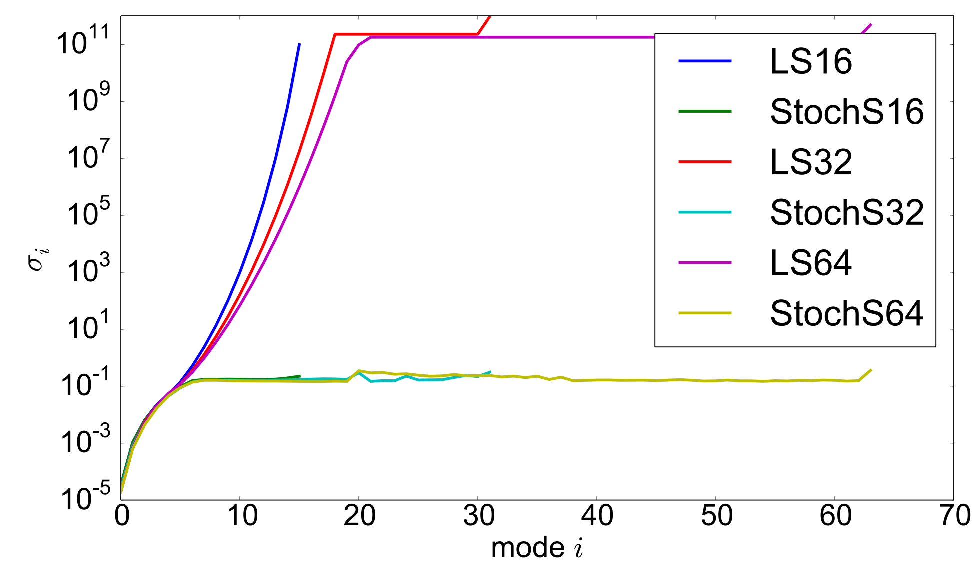


summary

singular modes



regularization by constraint



avrage spectrum method

$$|\bar{f}\rangle \propto \int_{\mathcal{F}} \mathcal{D}f \; e^{-\frac{1}{2}\langle \tilde{K}f-\tilde{g}|\tilde{K}f-\tilde{g}\rangle} \; |f\rangle$$

blocked mode sampling

separate grid & default model