



Between Fisher
and Monte Carlo:



Mapping the distribution of the maximum-likelihood estimator for GW source parameters

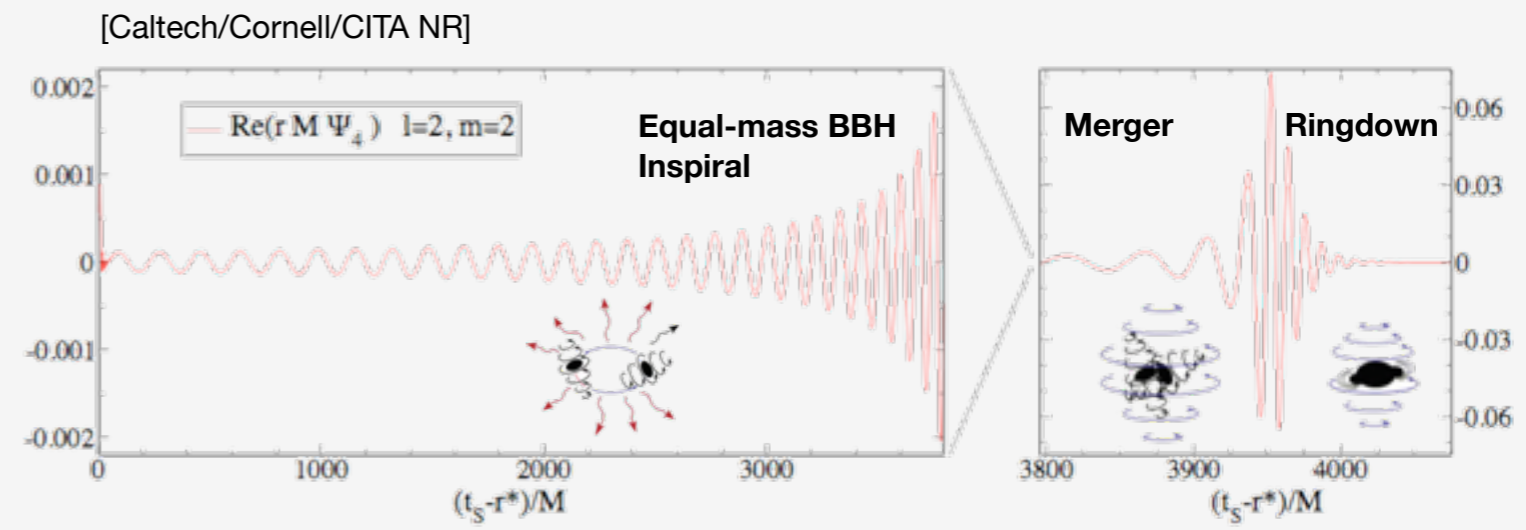
Michele Vallisneri

Jet Propulsion Laboratory

Copyright 2012 California Institute of Technology

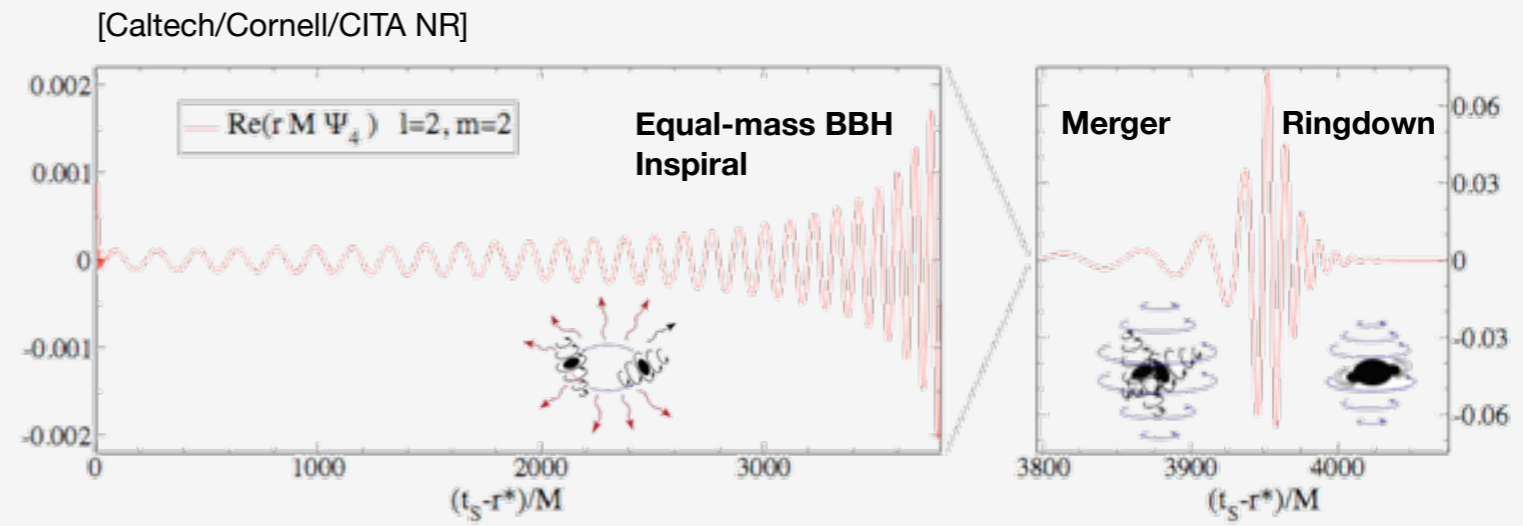
There's so much information
in a gravitational waveform!

BH and NS astrophysics, cosmology,
strong-field GR, BH structure, nuclear
physics, alternative theories...

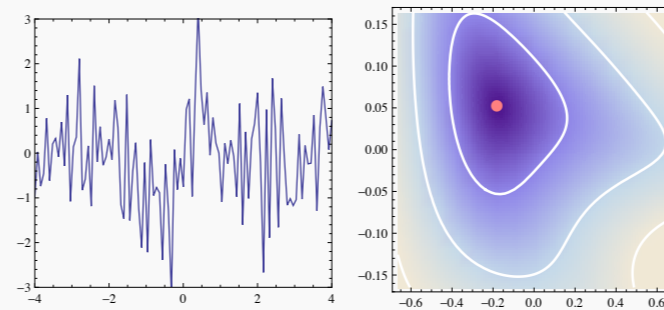


There's so much information
in a gravitational waveform!

BH and NS astrophysics, cosmology,
strong-field GR, BH structure, nuclear
physics, alternative theories...

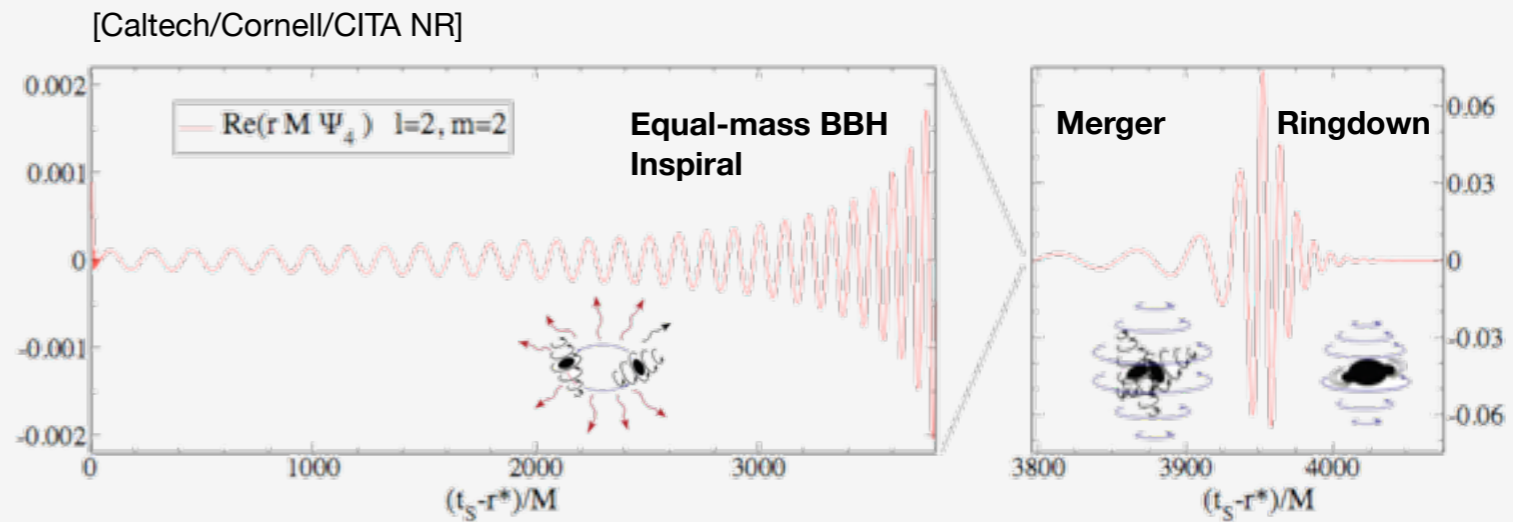


After a detection, we want
the best parameter estimates
→ Bayesian Monte Carlos



There's so much information
in a gravitational waveform!

BH and NS astrophysics, cosmology,
strong-field GR, BH structure, nuclear
physics, alternative theories...

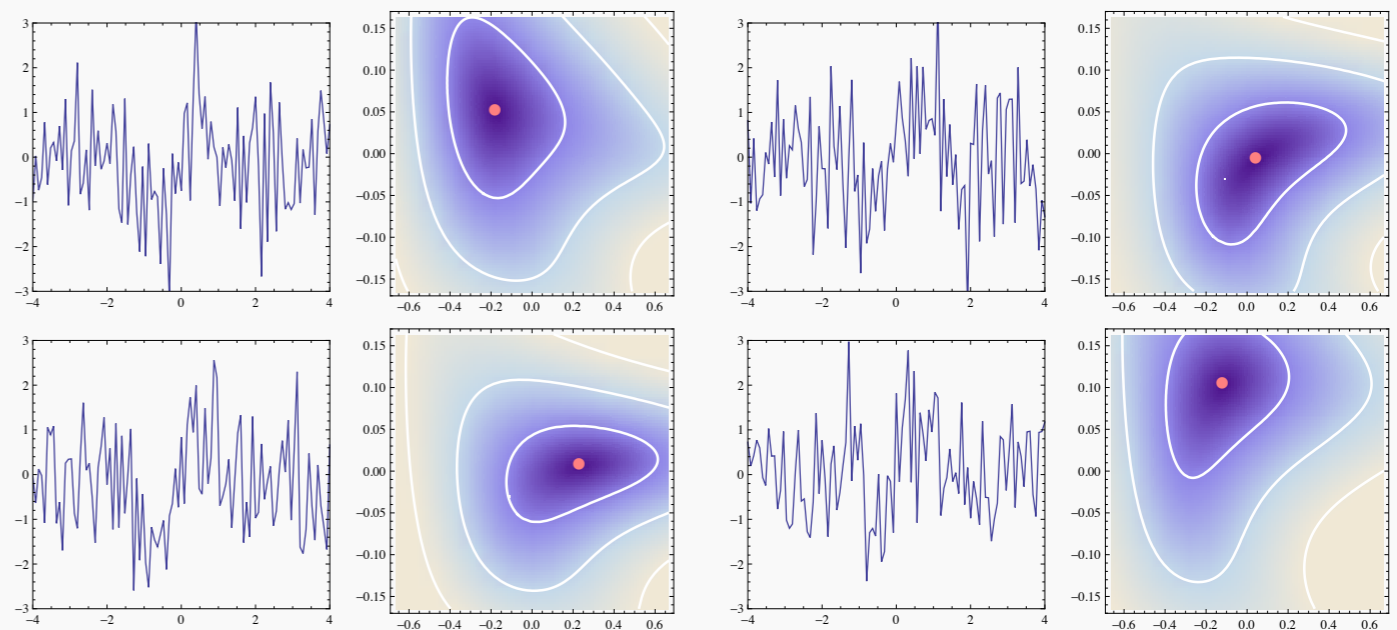


After a detection, we want
the best parameter estimates

→ Bayesian Monte Carlos

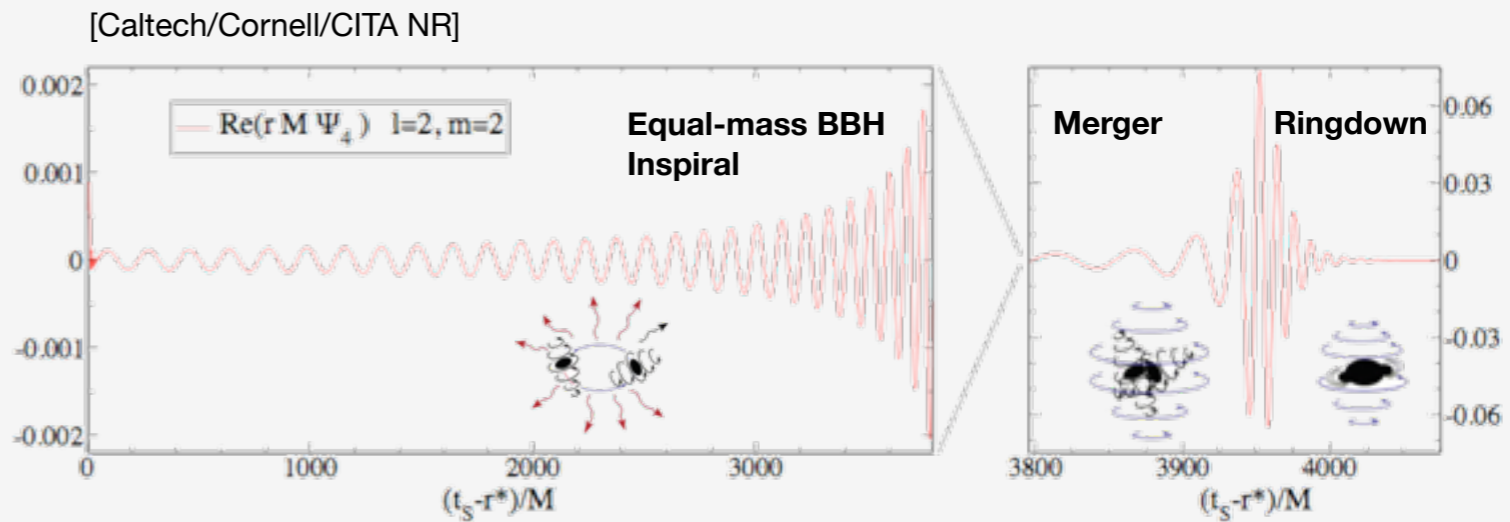
Before detections, we want
to know how well we can do

→ a Monte Carlo of Monte Carlos



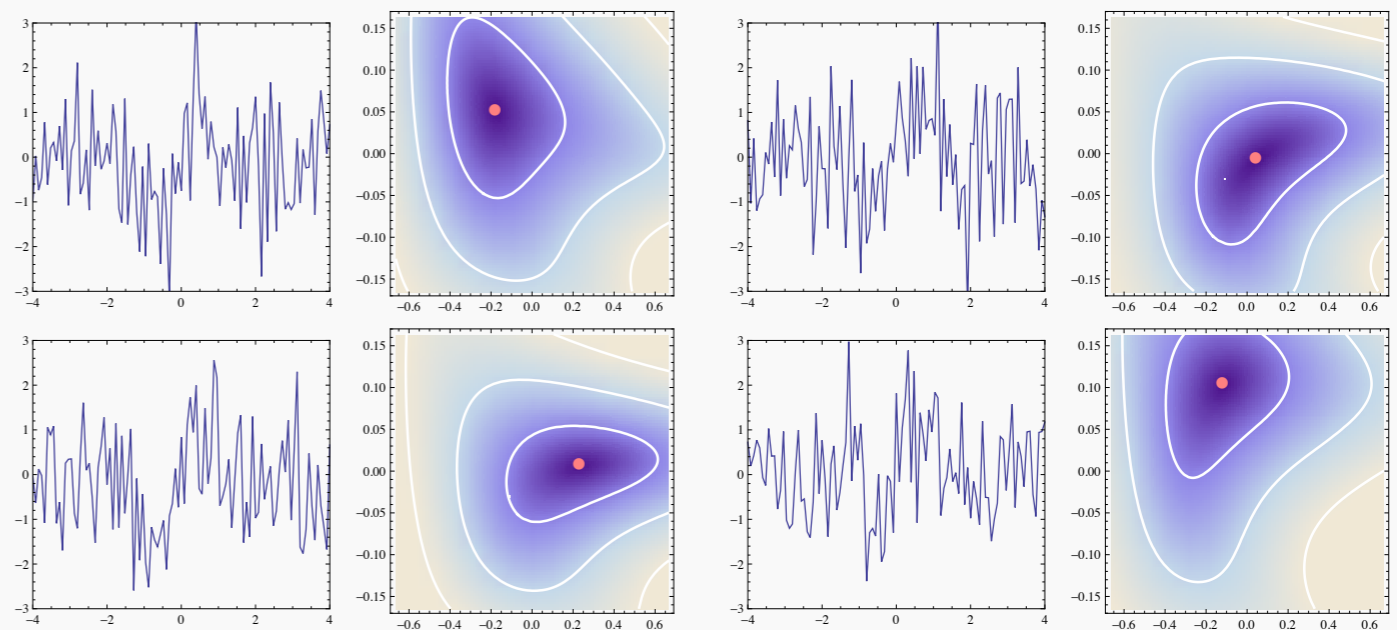
There's so much information
in a gravitational waveform!

BH and NS astrophysics, cosmology,
strong-field GR, BH structure, nuclear
physics, alternative theories...



After a detection, we want
the best parameter estimates
→ Bayesian Monte Carlos

Before detections, we want
to know how well we can do
→ a Monte Carlo of Monte Carlos

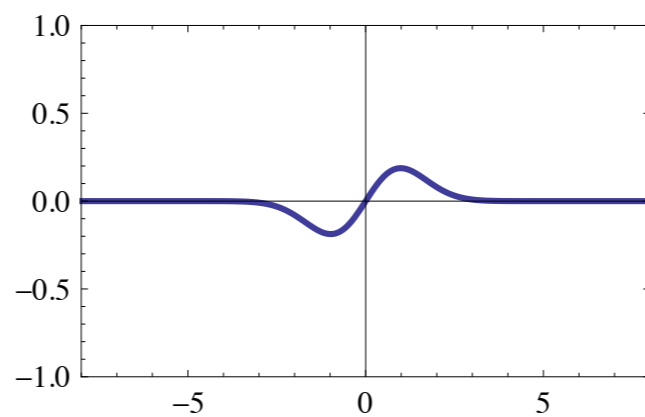
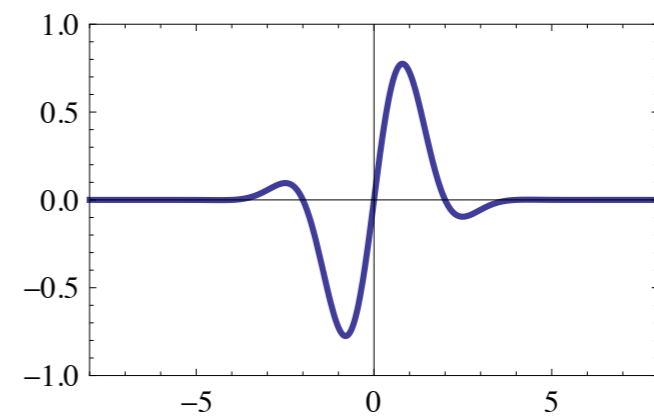
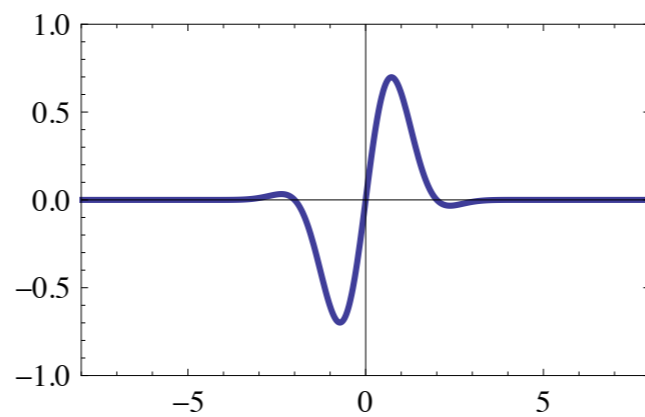
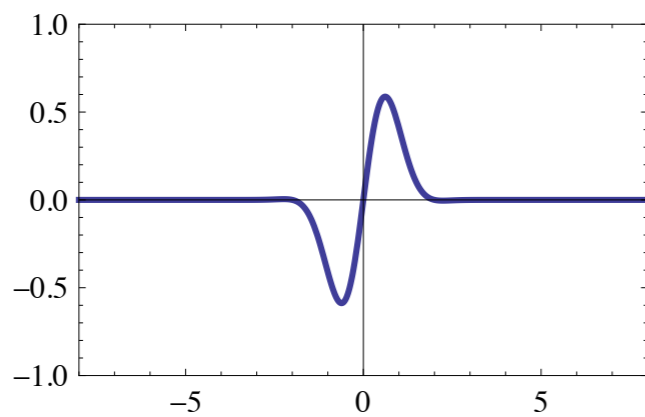
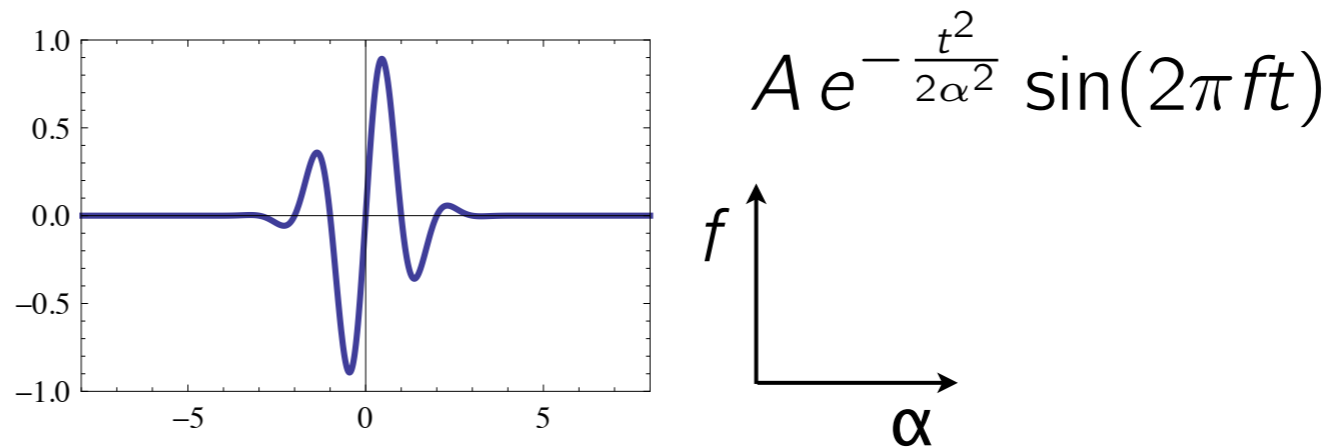


In this talk: a new efficient way
to map the **distribution of the
maximum-likelihood estimator (●)**
across noise realizations

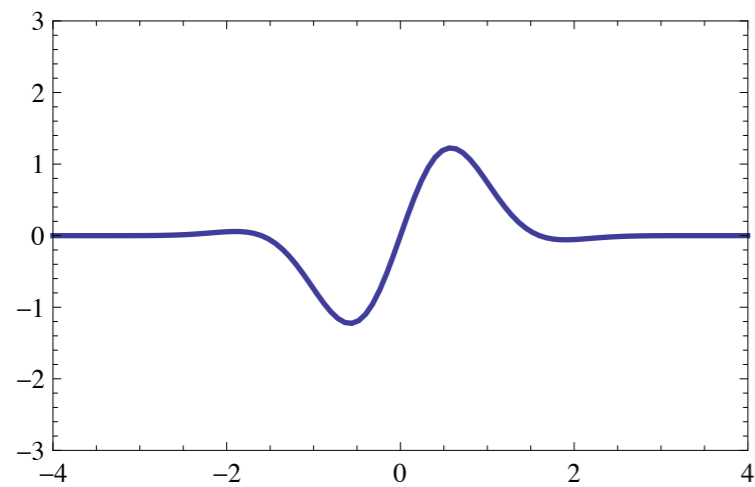
Bonus: a gentle intro to ML estimation

$$p(\theta) = \frac{e^{-(\Delta h, h_i)(F^{-1})^{ij}(\Delta h, h_j)/2}}{\sqrt{(2\pi)^d |F_{ij}|}} \times \frac{1}{\sqrt{(2\pi)^{d(d-1)/2} |D_{\mu\nu}|}} \\ \times \int |F_{ij} + (\Delta h, h_{ij}) - M_{(ij)}| e^{-M_\mu (D^{-1})^{\mu\nu} M_\nu / 2} dM_\mu$$

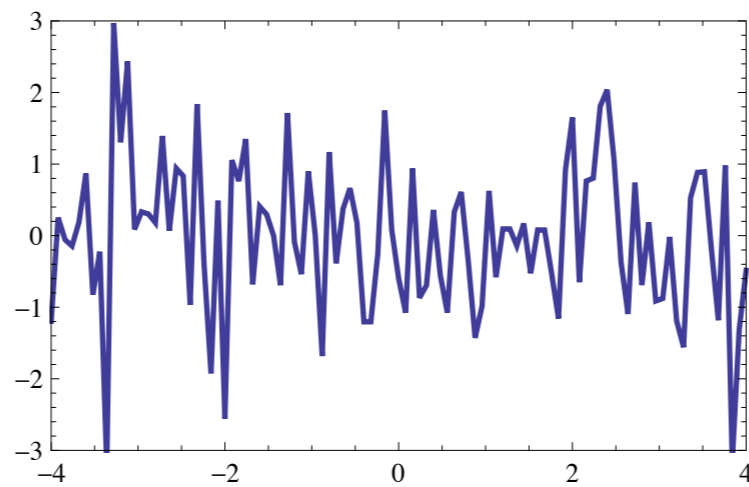
GW science = learn about sources by estimating their parameters



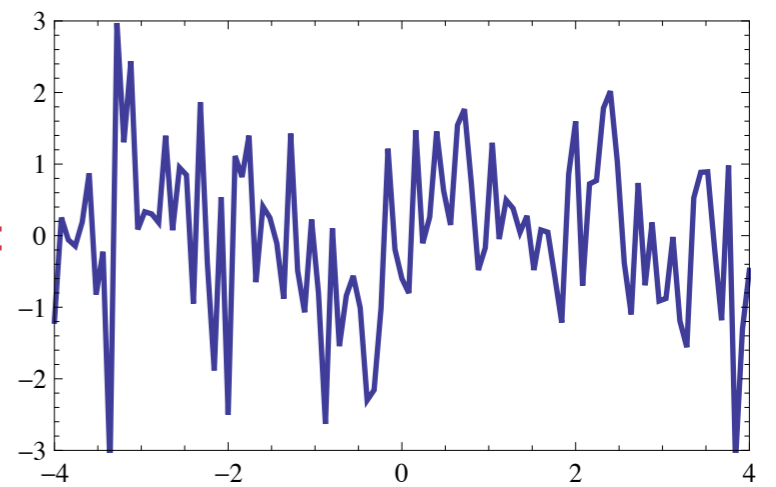
data = signal + noise



+

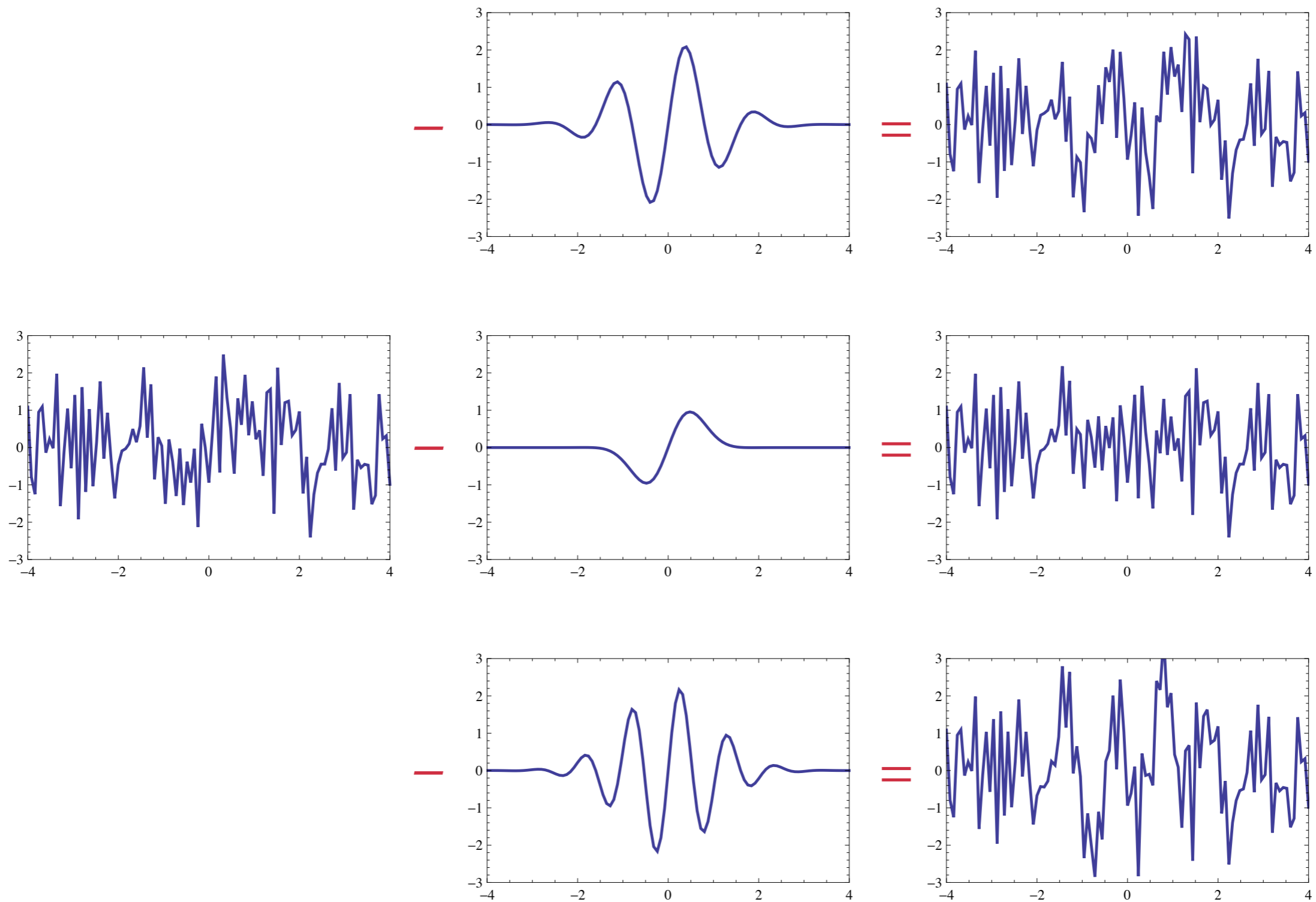


=



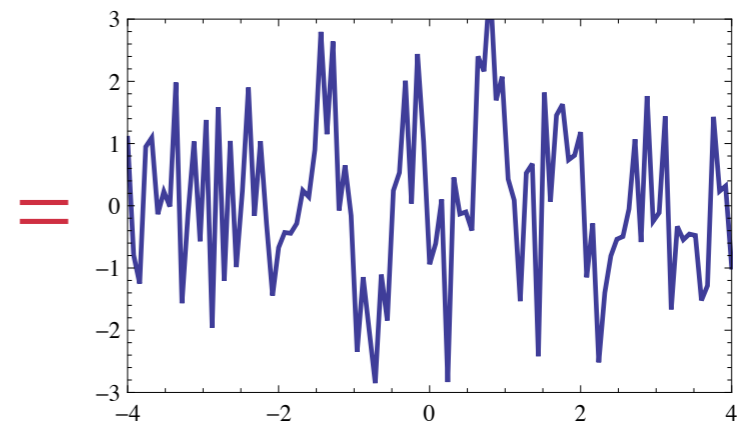
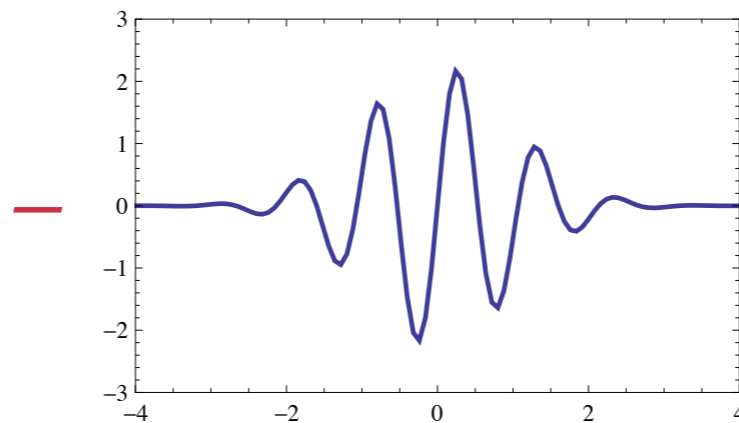
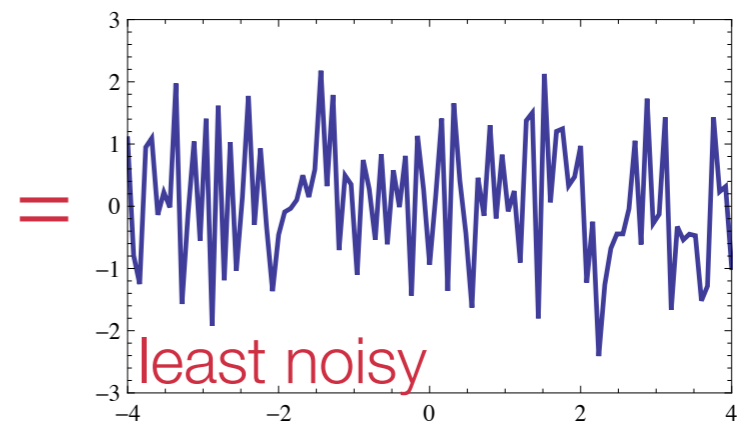
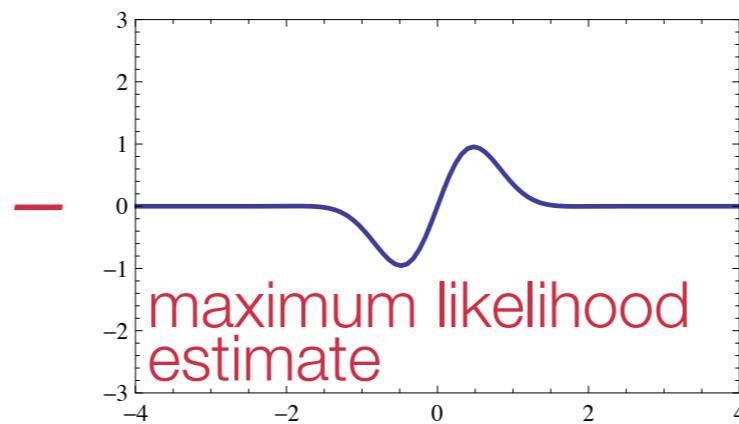
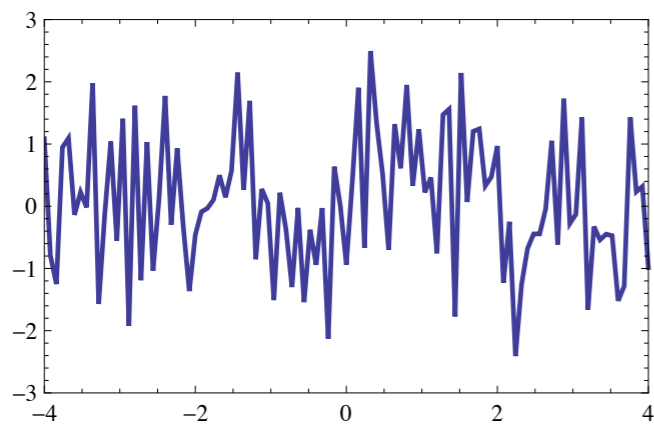
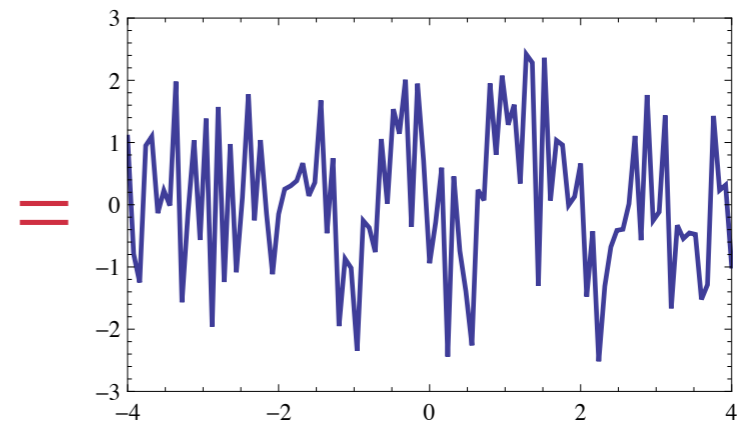
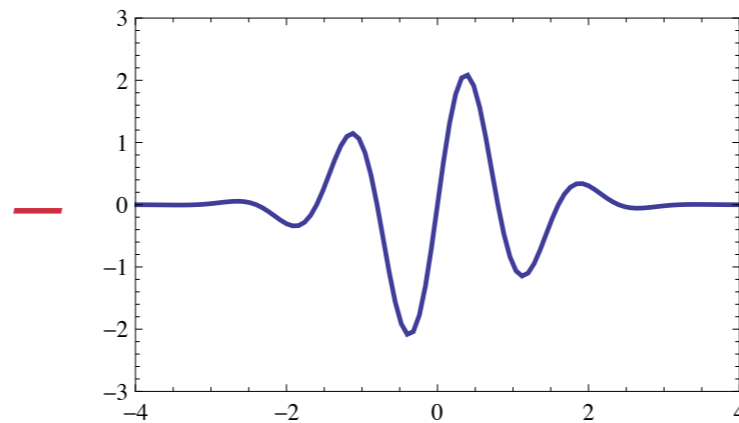
(SNR = 5)

noise = data - signal



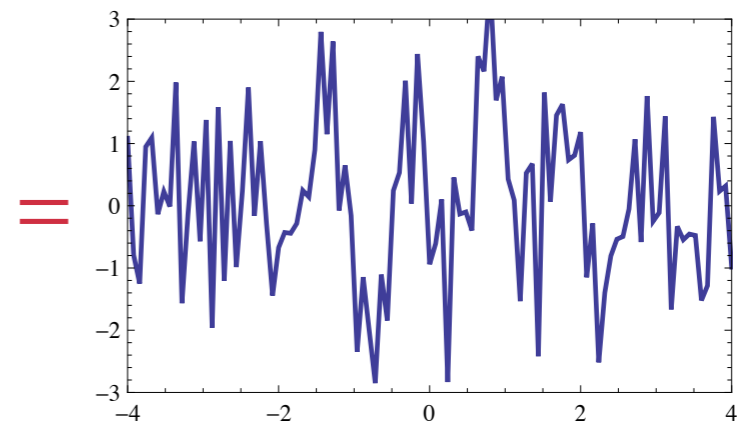
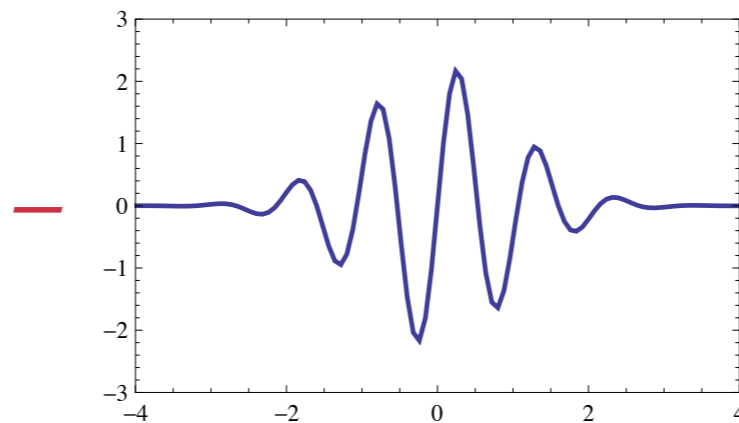
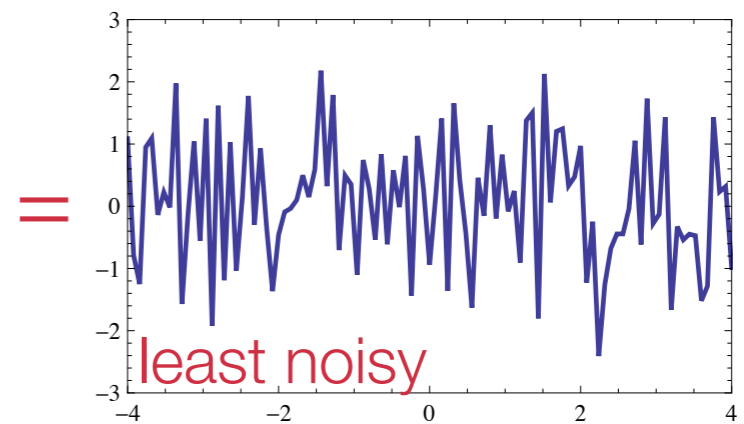
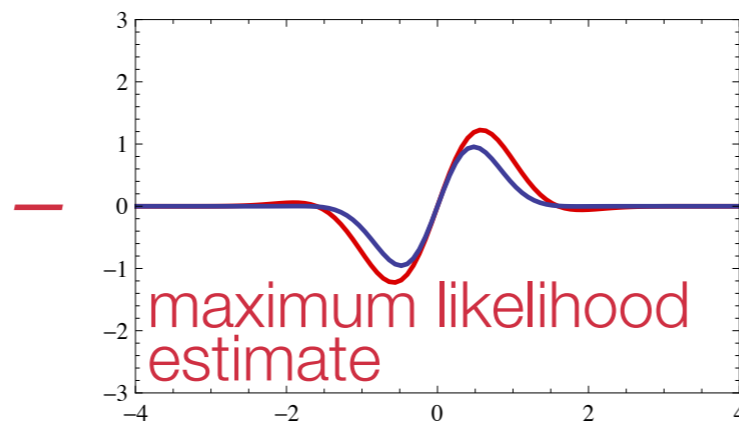
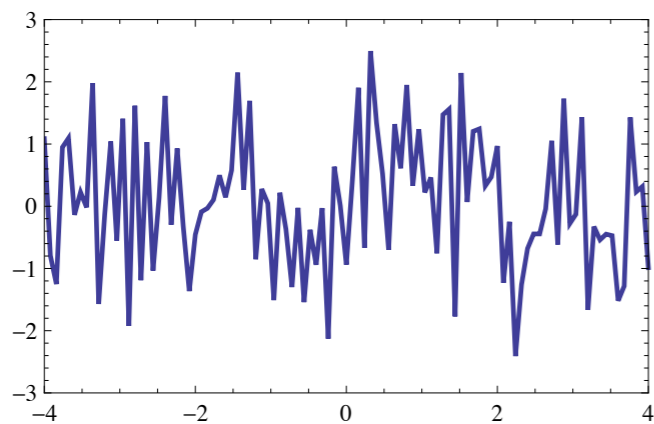
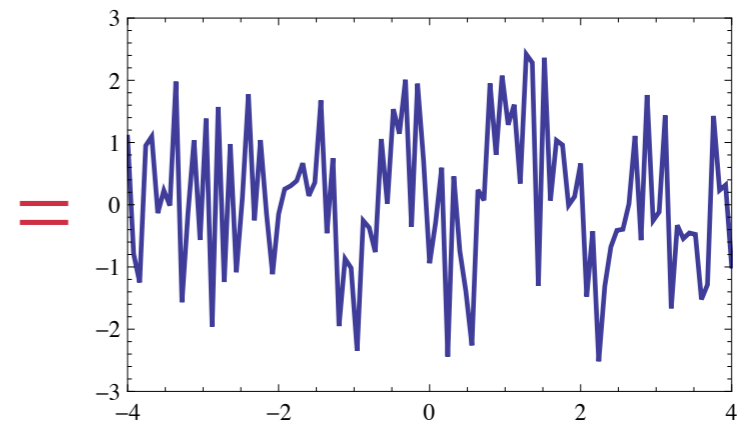
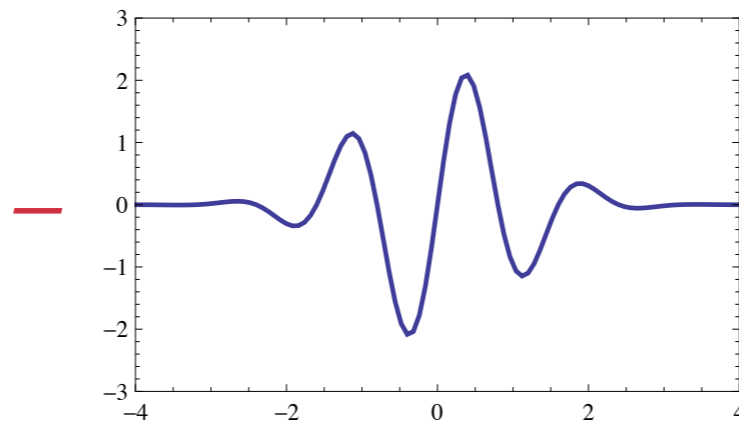
noise = data - signal

hence $p(\text{signal parameters}) = p(\text{noise residual})$

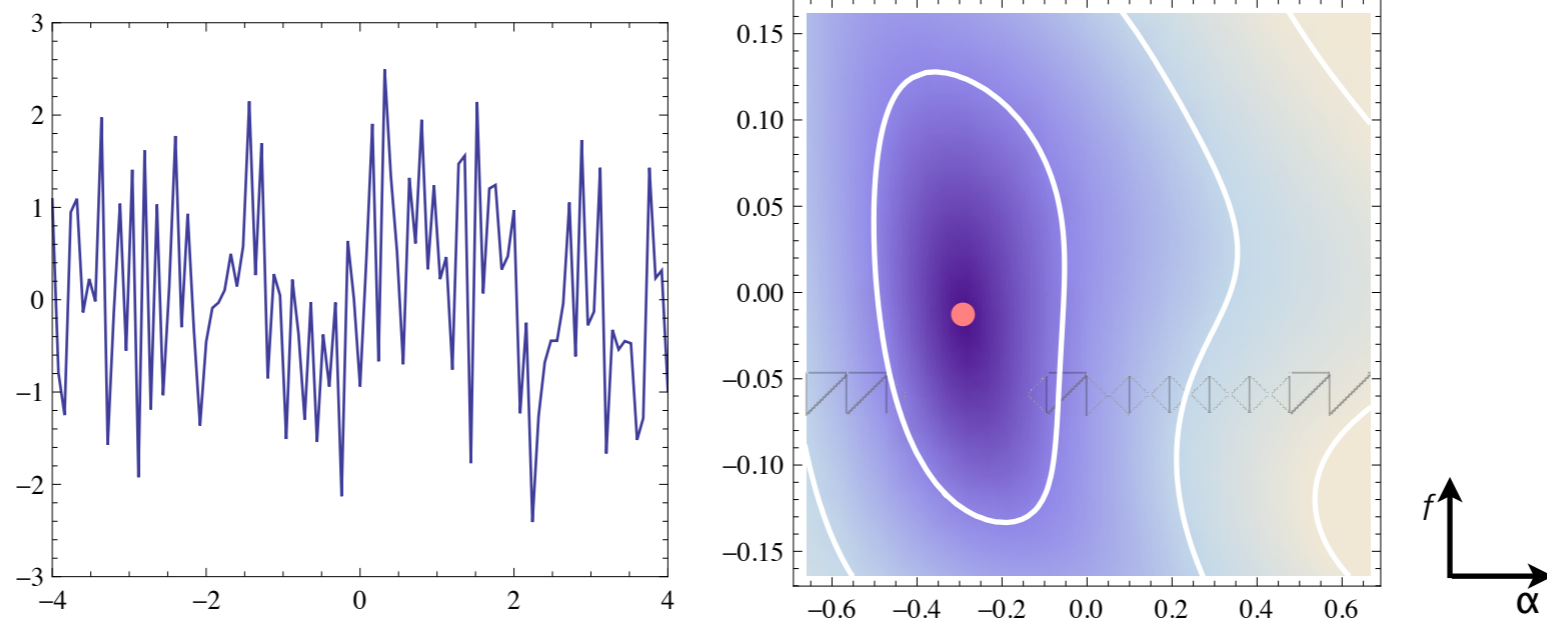


noise = data - signal

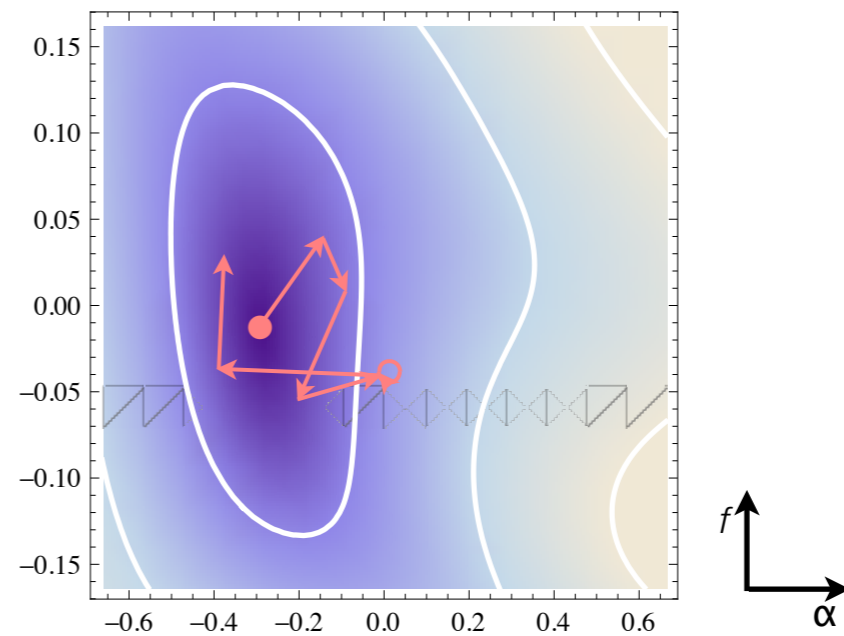
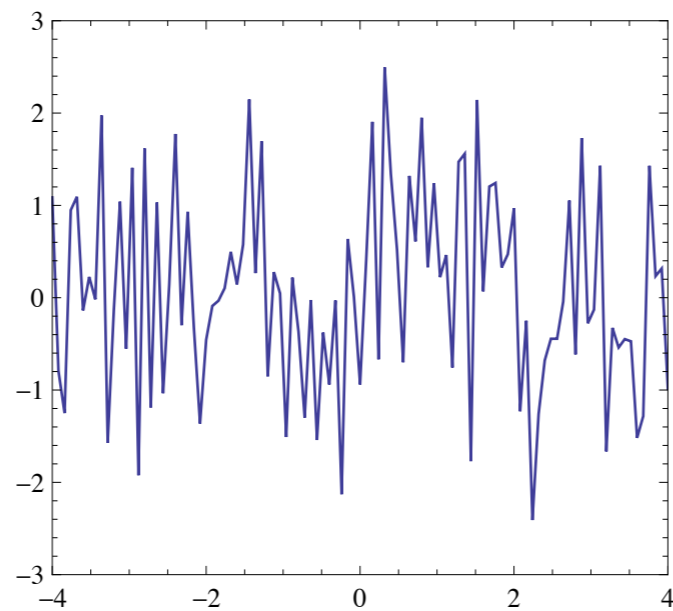
hence $p(\text{signal parameters}) = p(\text{noise residual})$



for a single noise realization, we make a likelihood map...



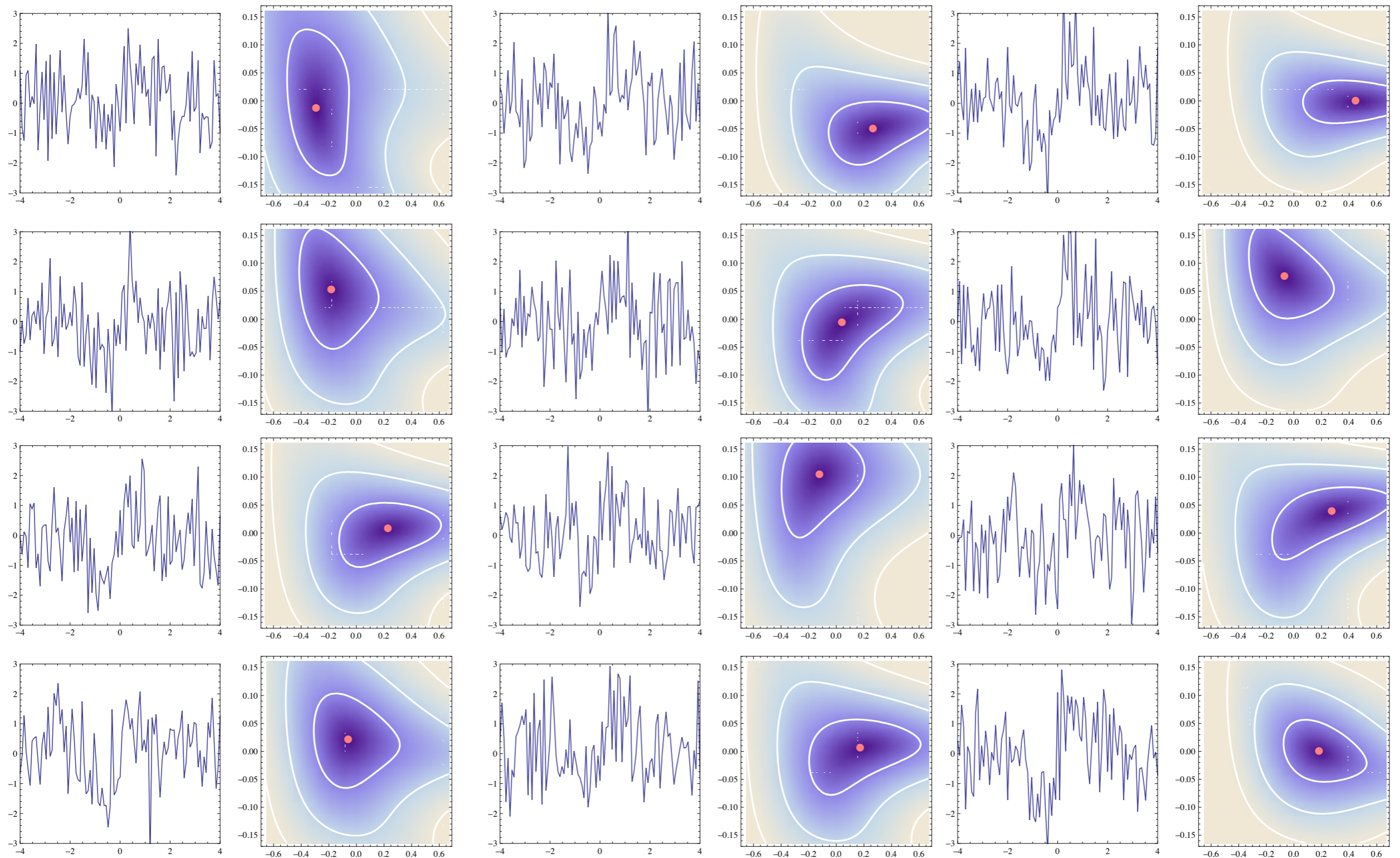
for a single noise realization, we make a likelihood map...



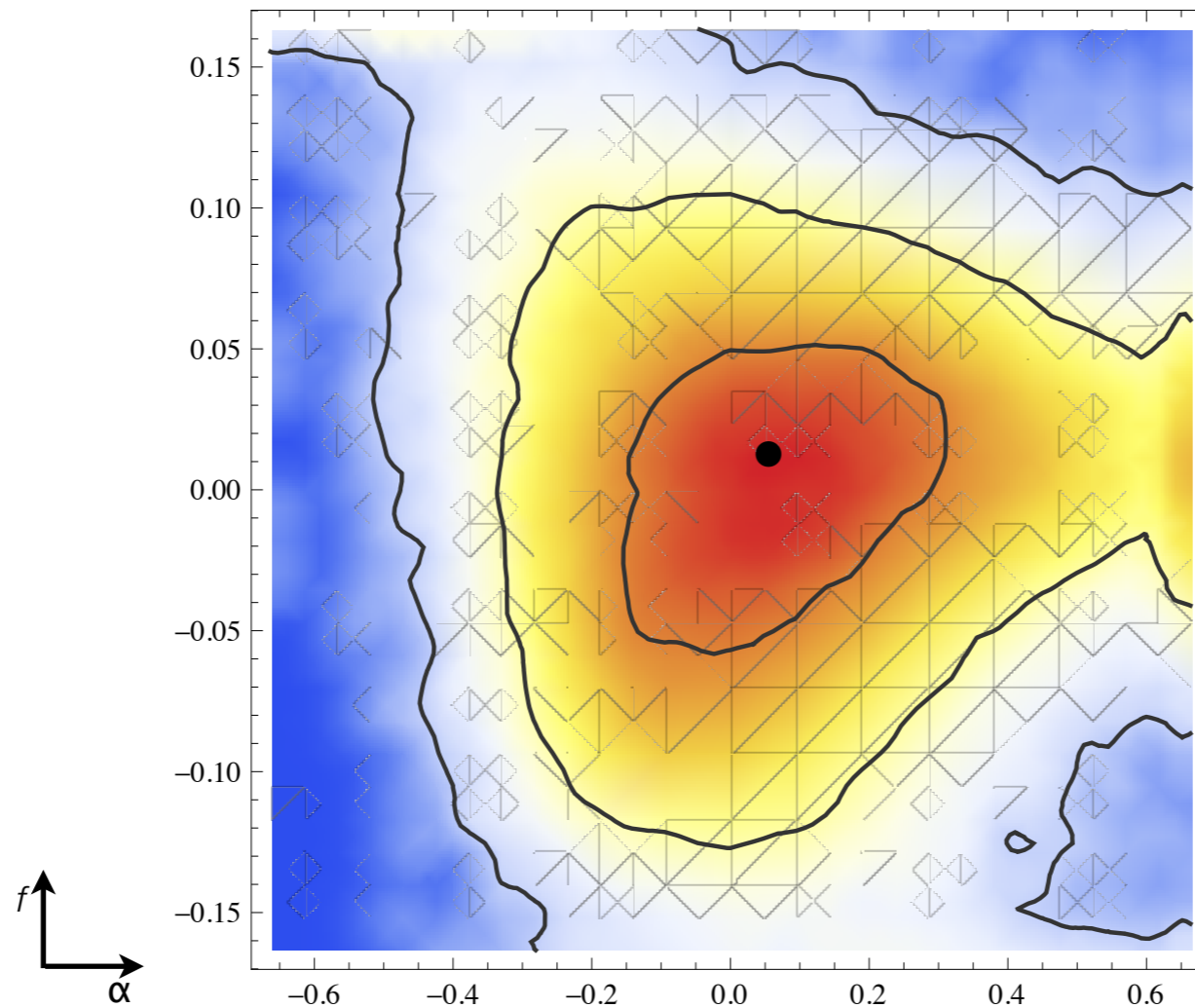
Markov-chain Monte Carlos
explore probability density
with a controlled random walk

the Holy Grail: a Monte Carlo of Monte Carlos

[10^6 maps x 10^6 parameter sets x 10^6 -point FFT = exaproblem]

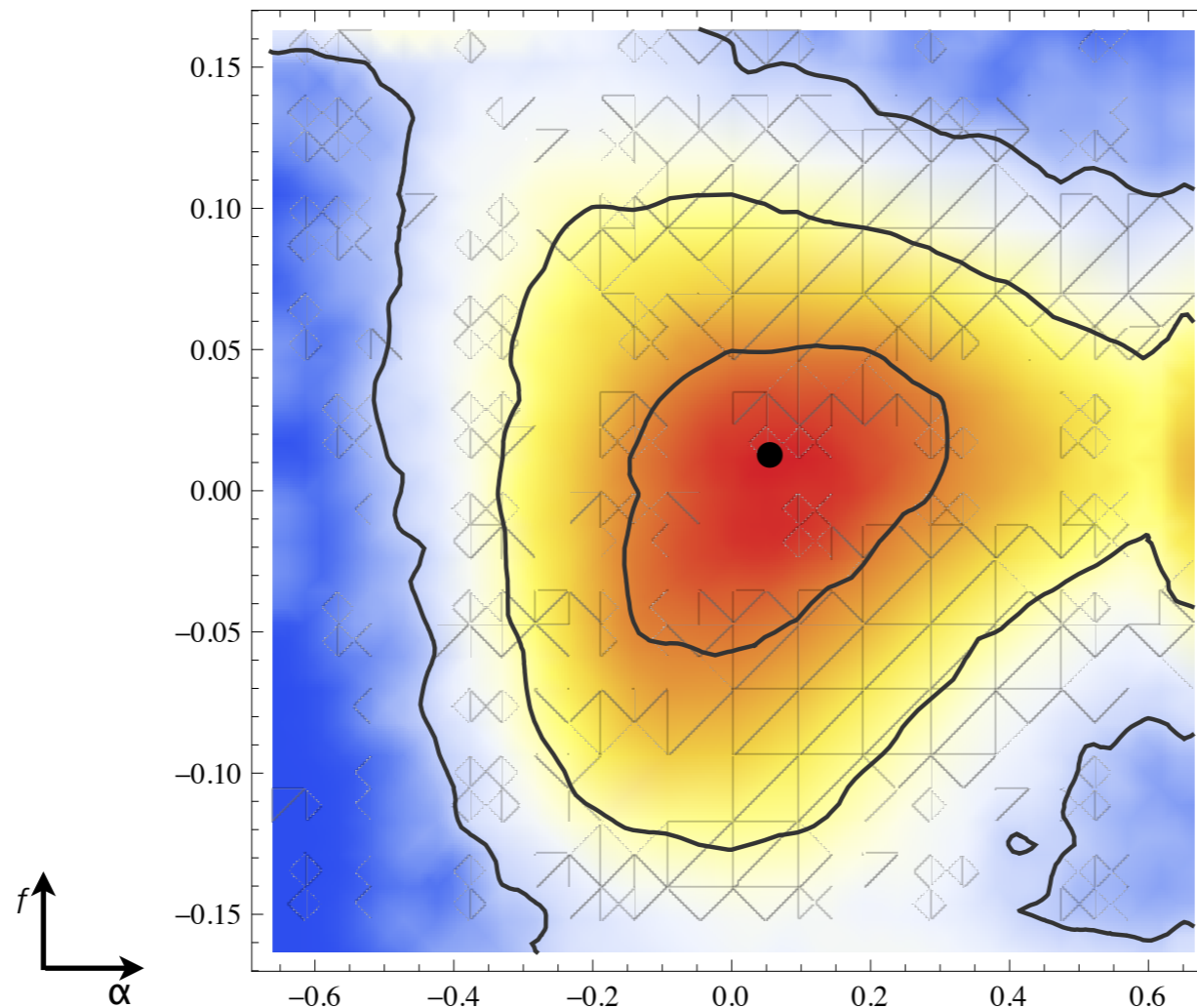


not as holy, still a challenge: map the distribution of the maximum-likelihood estimator (●)

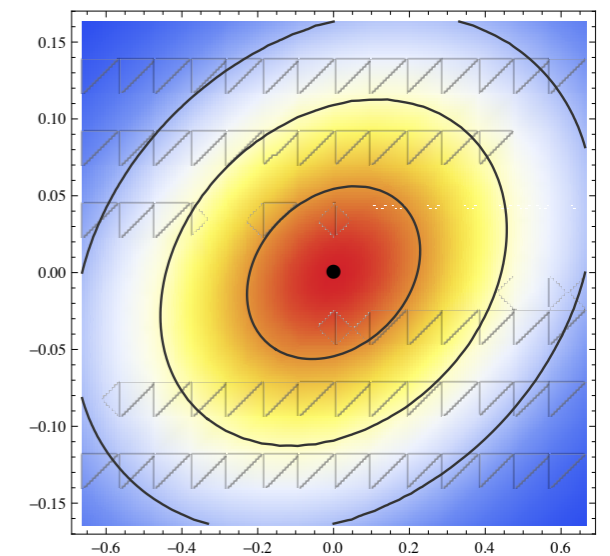


numerical searches over 100,000
explicit realizations of the noise

not as holy, still a challenge: map the distribution of the maximum-likelihood estimator (●)



numerical searches over 100,000 explicit realizations of the noise



Fisher-matrix “prediction”

$$p(\Delta\theta) = \frac{e^{-\Delta\theta^i F_{ij} \Delta\theta^j / 2}}{\sqrt{(2\pi)^d |F_{ij}^{-1}|}}$$

$$F_{ij} = (\partial_i h, \partial_j h)$$

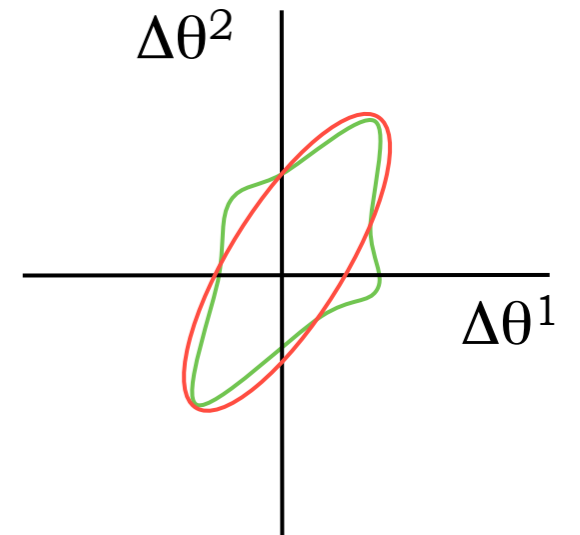
aside: how you check if you can use the Fisher matrix!

[MV, PRD 77, 042001 (2008)]

aside: how you check if you can use the Fisher matrix!

[MV, PRD 77, 042001 (2008)]

- explore the 1σ probability surface predicted by the Fisher matrix
- evaluate the ratio of the exact likelihood to the linearized likelihood

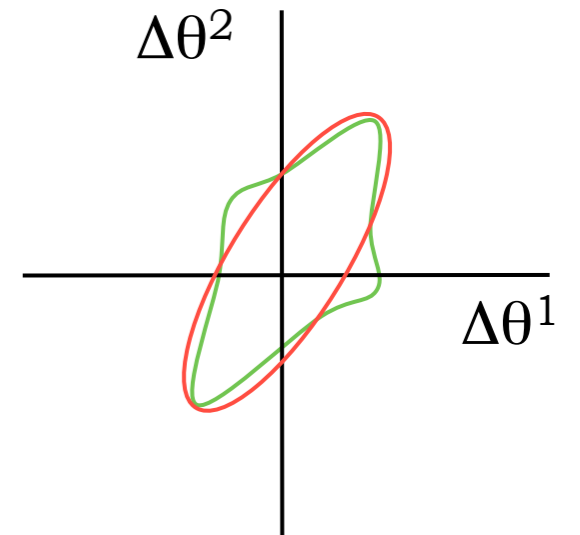


$$|\log r(\theta, A)| = (\theta_j h_j - \Delta h(\theta), \theta_k h_k - \Delta h(\theta)) / 2$$

aside: how you check if you can use the Fisher matrix!

[MV, PRD 77, 042001 (2008)]

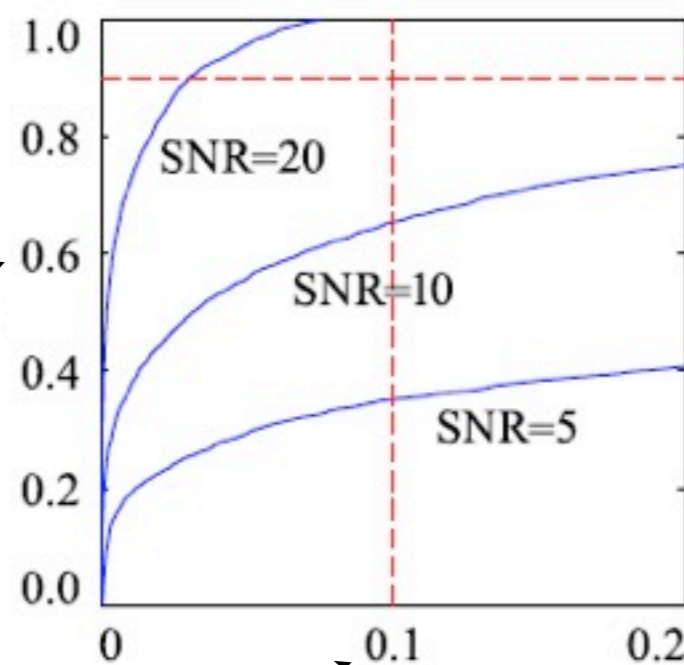
- explore the 1σ probability surface predicted by the Fisher matrix
- evaluate the ratio of the exact likelihood to the linearized likelihood



$$|\log r(\theta, A)| = (\theta_j h_j - \Delta h(\theta), \theta_k h_k - \Delta h(\theta)) / 2$$

BNS inspiral

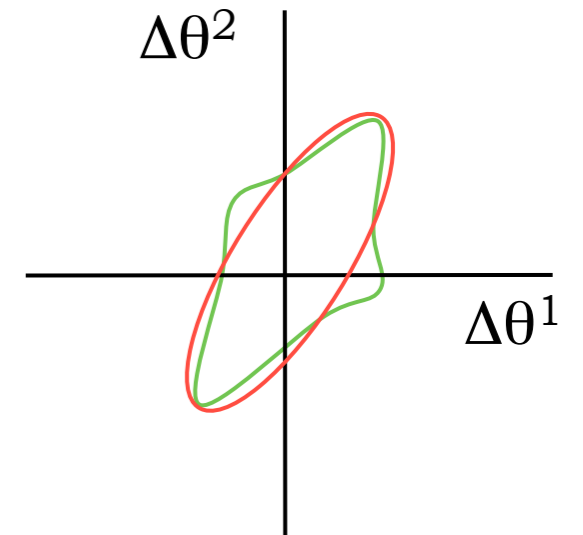
fraction of 1σ surface
with likelihood exact to



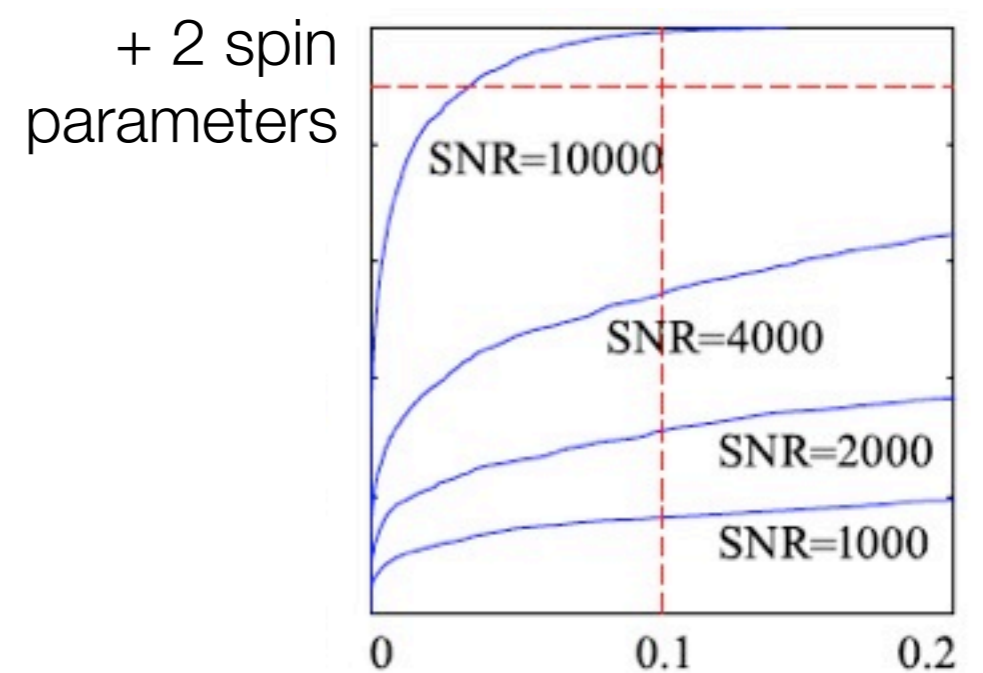
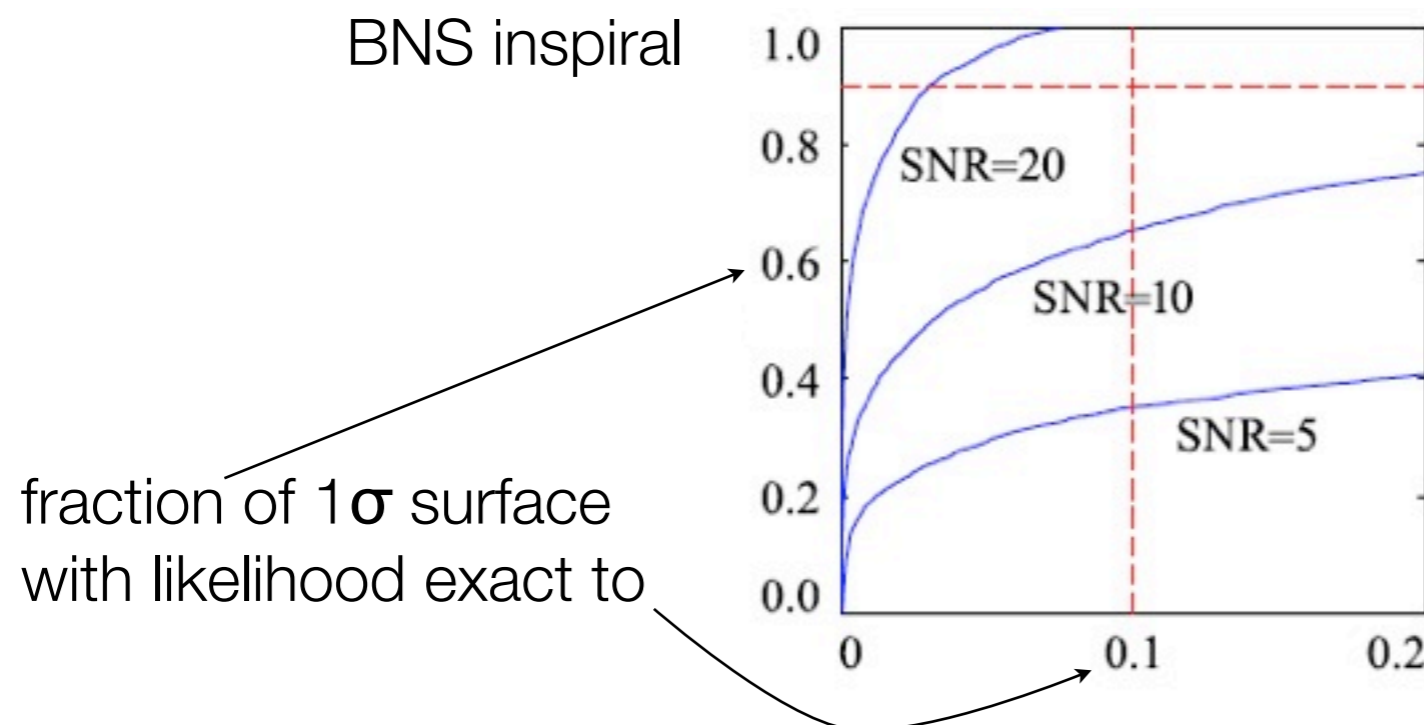
aside: how you check if you can use the Fisher matrix!

[MV, PRD 77, 042001 (2008)]

- explore the 1σ probability surface predicted by the Fisher matrix
- evaluate the ratio of the exact likelihood to the linearized likelihood



$$|\log r(\theta, A)| = (\theta_j h_j - \Delta h(\theta), \theta_k h_k - \Delta h(\theta)) / 2$$



let's find the distribution of the ML estimator
from first principles:

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -\left(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta)\right)/2$

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -\left(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta)\right)/2$
- max-likelihood equation: $(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta)) = 0$

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -\left(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta)\right)/2$
- max-likelihood equation: $\underbrace{\left(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta)\right)}_{\text{ML}_i(\theta; n, \theta_{\text{true}})} = 0$

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta))/2$
- max-likelihood equation: $\underbrace{(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta))}_{ML_i(\theta; n, \theta_{\text{true}})} = 0$
- formal solution of ML equation: $\theta_{\text{ML}}(n, \theta_{\text{true}})$

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta))/2$
- max-likelihood equation: $\underbrace{(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta))}_{\text{ML}_i(\theta; n, \theta_{\text{true}})} = 0$
- formal solution of ML equation: $\theta_{\text{ML}}(n, \theta_{\text{true}})$
- formal distribution of ML estimator: $p(\theta) = \int \delta(\theta_{\text{ML}}(n, \theta_{\text{true}}) - \theta) p(n) dn$

let's find the distribution of the ML estimator
from first principles:

- log likelihood: $\log p(\theta) = -(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta))/2$
- max-likelihood equation: $\underbrace{(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta))}_{\text{ML}_i(\theta; n, \theta_{\text{true}})} = 0$
- formal solution of ML equation: $\theta_{\text{ML}}(n, \theta_{\text{true}})$
- formal distribution of ML estimator: $p(\theta) = \int \delta(\theta_{\text{ML}}(n, \theta_{\text{true}}) - \theta) p(n) dn$
- but it's much easier to solve the ML equation for n than for θ : $\frac{\delta(\theta_{\text{ML}}(n, \theta_{\text{true}}) - \theta)}{|\partial \text{ML}_i / \partial \theta_j|} = \delta(\text{ML}_i(\theta; n, \theta_{\text{true}}))$

let's find the distribution of the ML estimator from first principles:

- log likelihood: $\log p(\theta) = -(n + h(\theta_{\text{true}}) - h(\theta), n + h(\theta_{\text{true}}) - h(\theta))/2$
- max-likelihood equation: $\underbrace{(\partial_i h(\theta), n + h(\theta_{\text{true}}) - h(\theta))}_{\text{ML}_i(\theta; n, \theta_{\text{true}})} = 0$
- formal solution of ML equation: $\theta_{\text{ML}}(n, \theta_{\text{true}})$
- formal distribution of ML estimator: $p(\theta) = \int \delta(\theta_{\text{ML}}(n, \theta_{\text{true}}) - \theta) p(n) dn$
- but it's much easier to solve the ML equation for n than for θ : $\frac{\delta(\theta_{\text{ML}}(n, \theta_{\text{true}}) - \theta)}{|\partial \text{ML}_i / \partial \theta_j|} = \delta(\text{ML}_i(\theta; n, \theta_{\text{true}}))$
- thus: $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$

which is really an integral over $\sim d^2$, not N , dimensions!

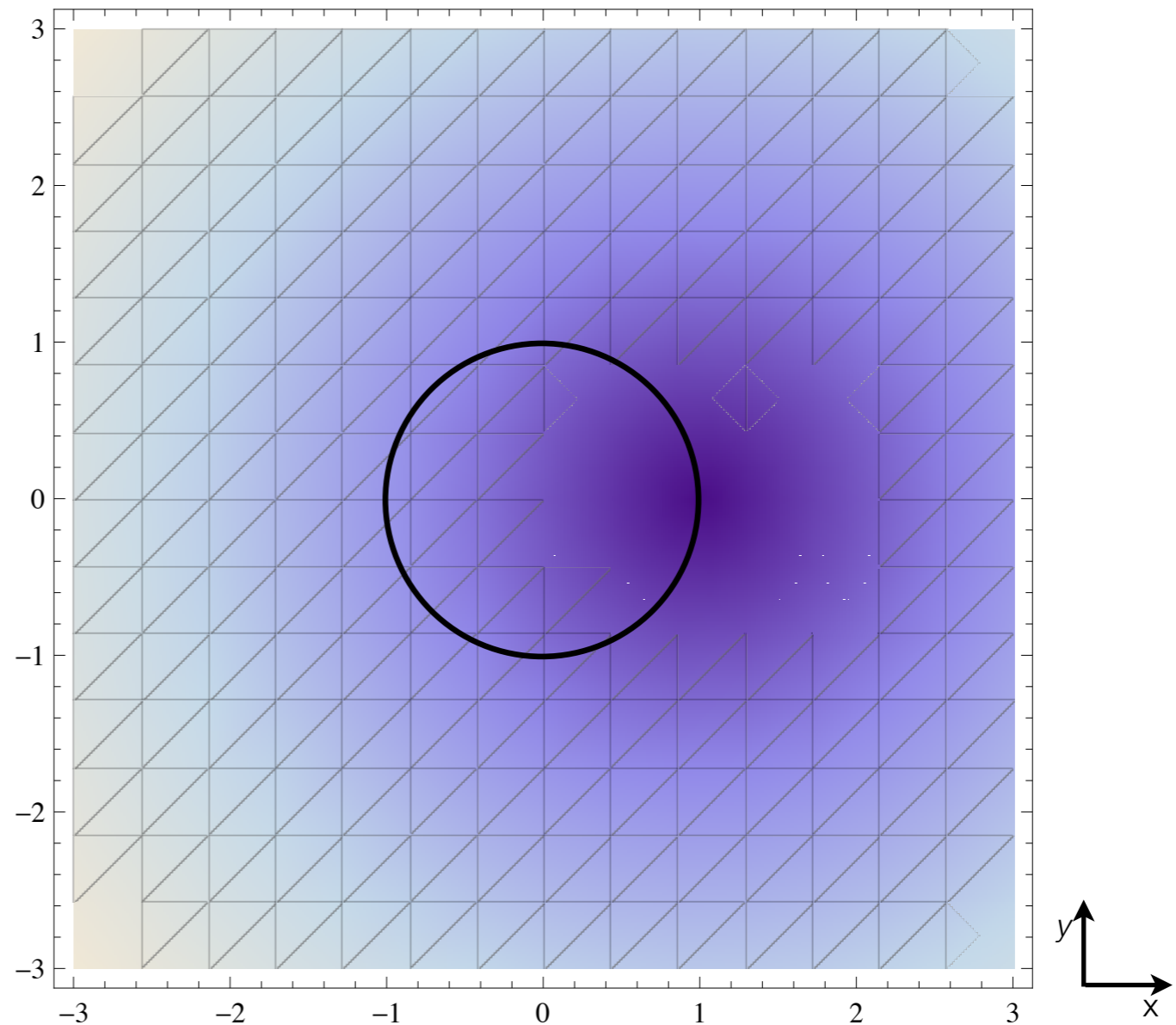
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



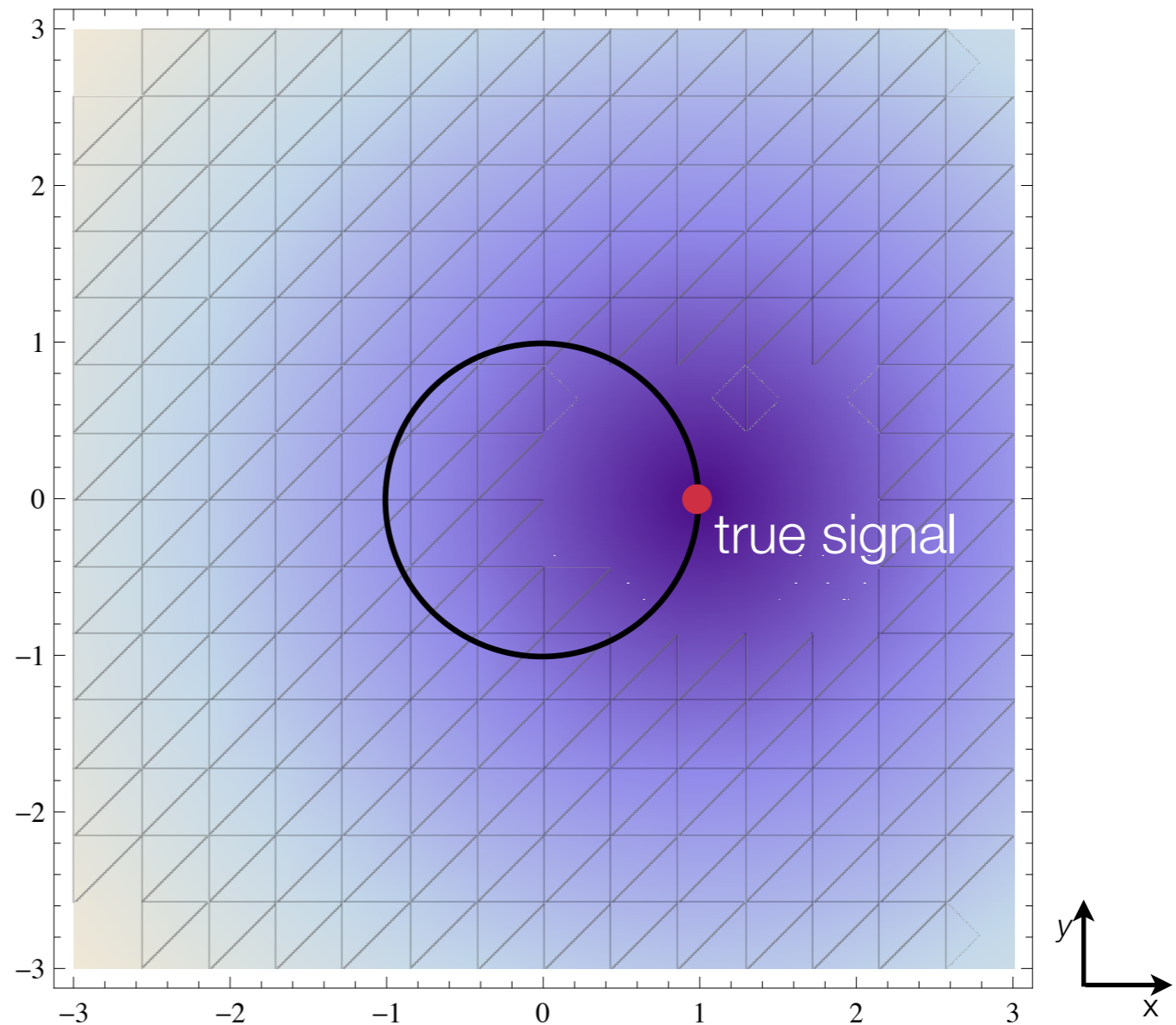
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



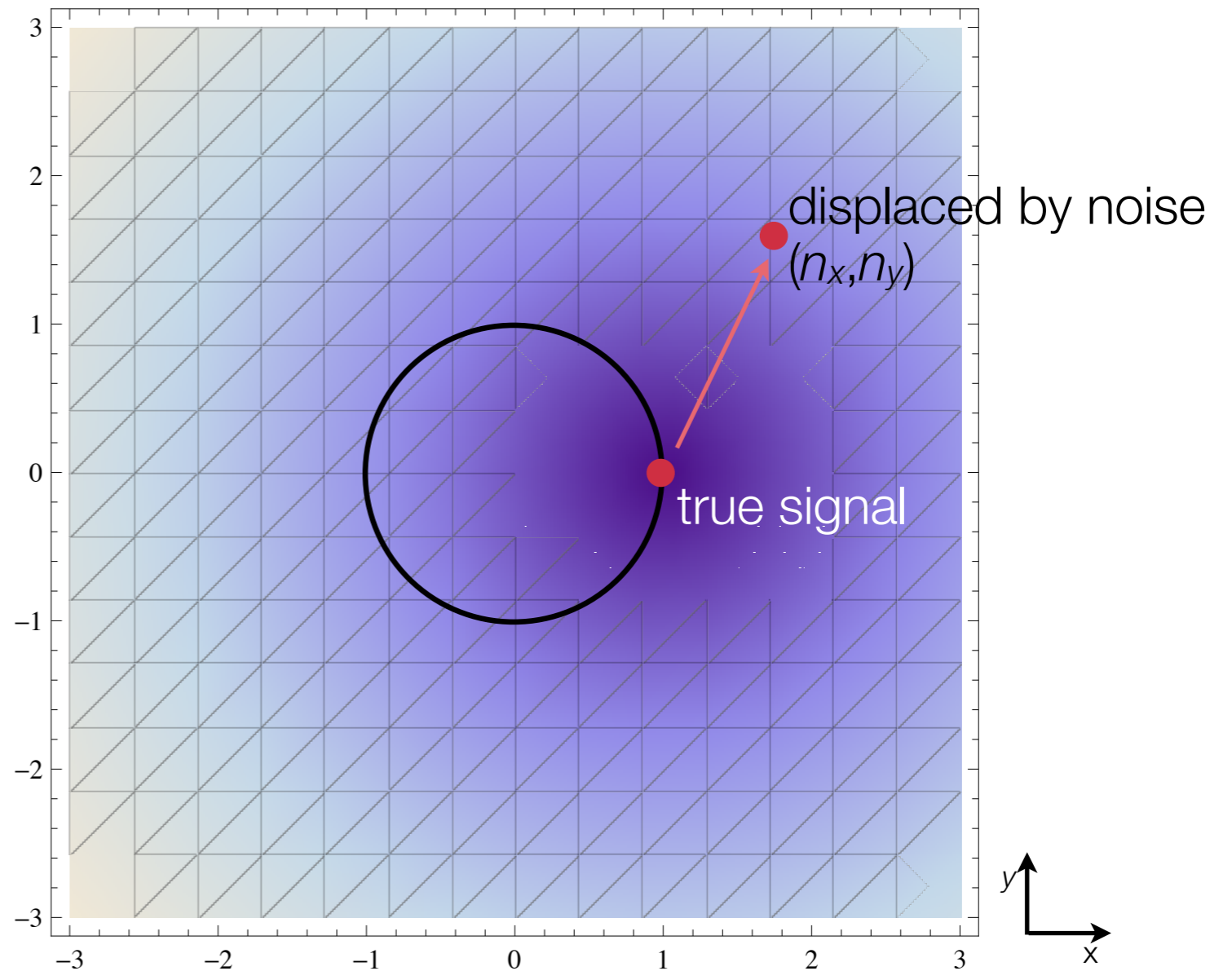
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



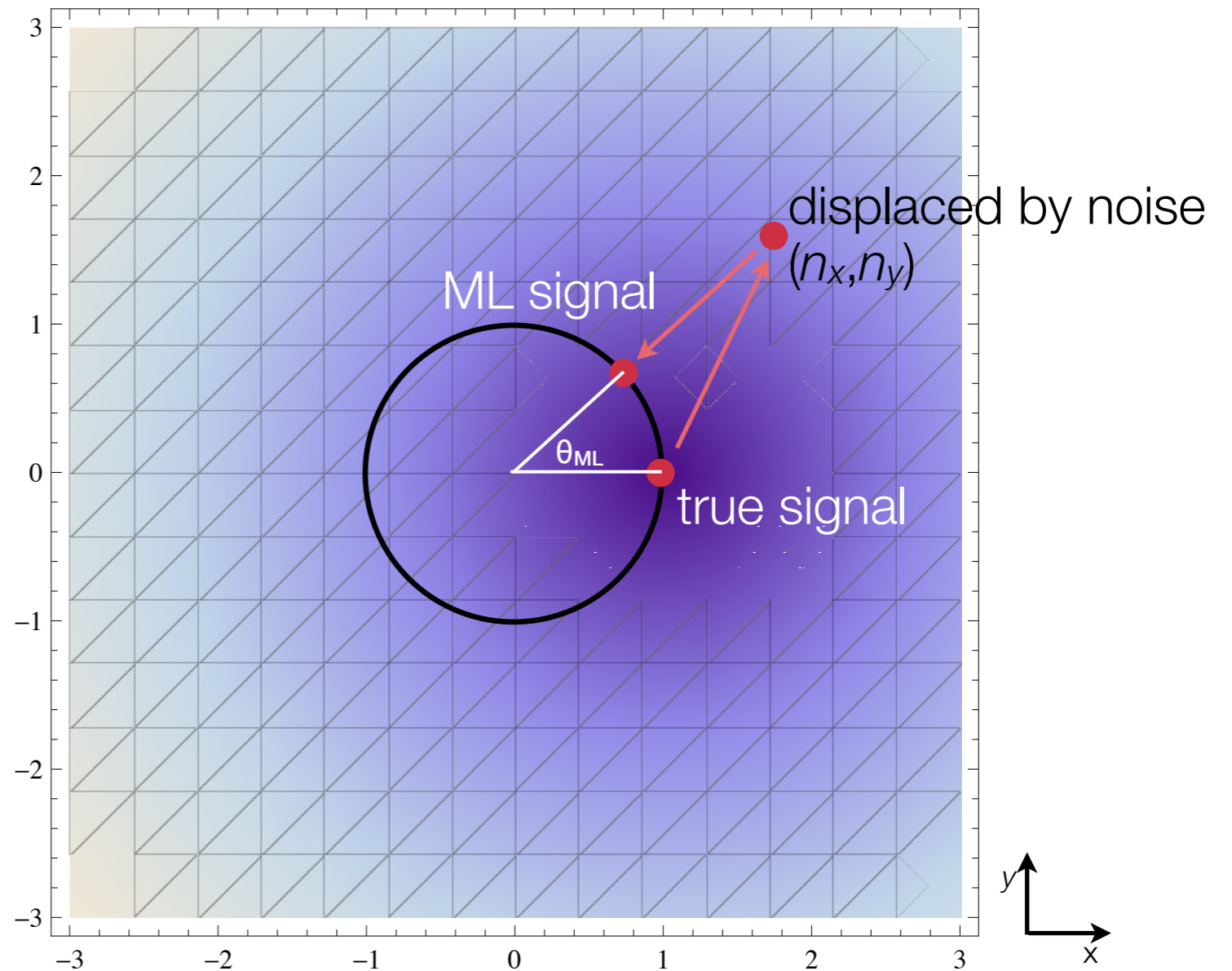
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



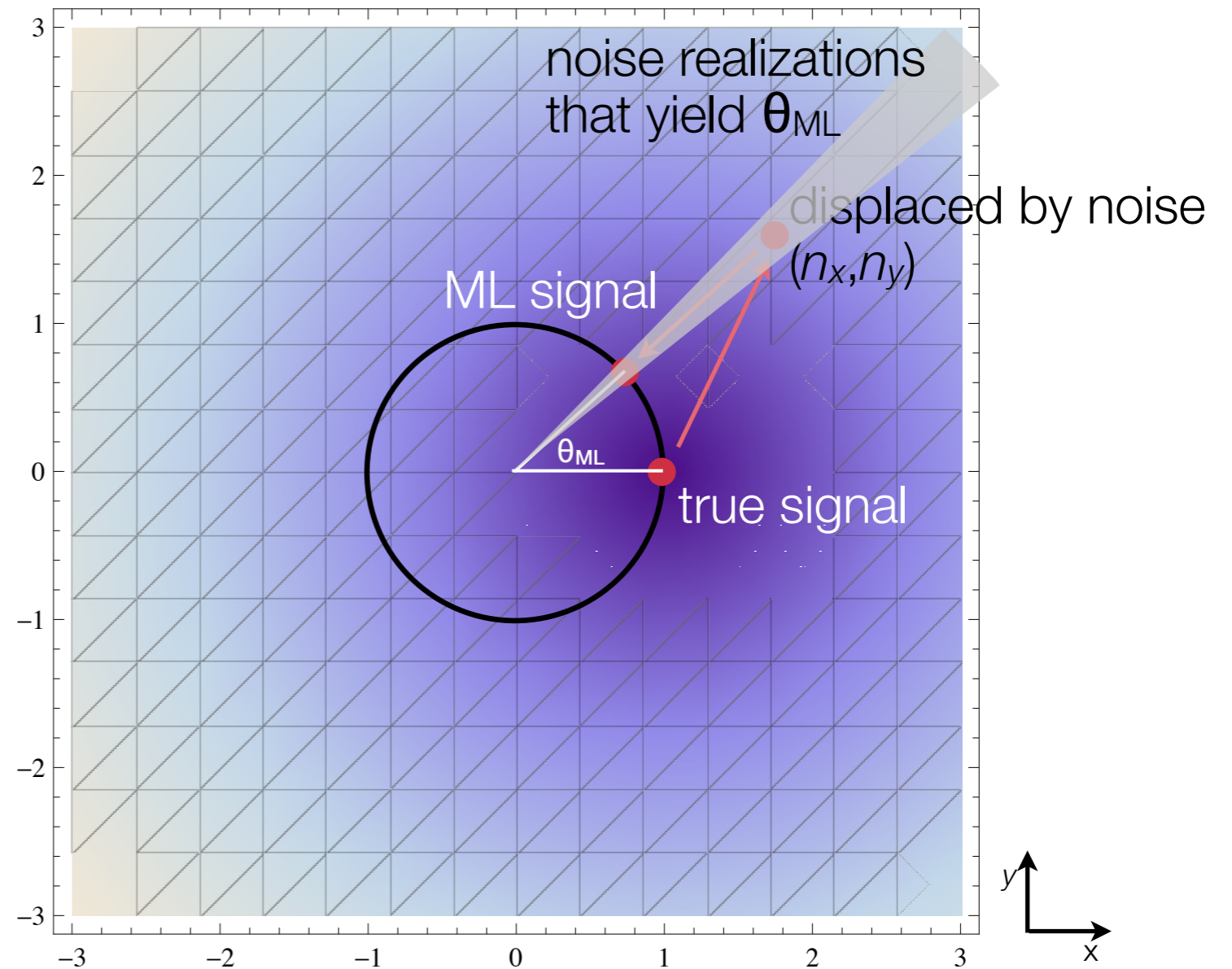
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



$$p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) \left| \frac{\partial \text{ML}_i}{\partial \theta_j} \right| e^{-(n,n)/2} dn$$

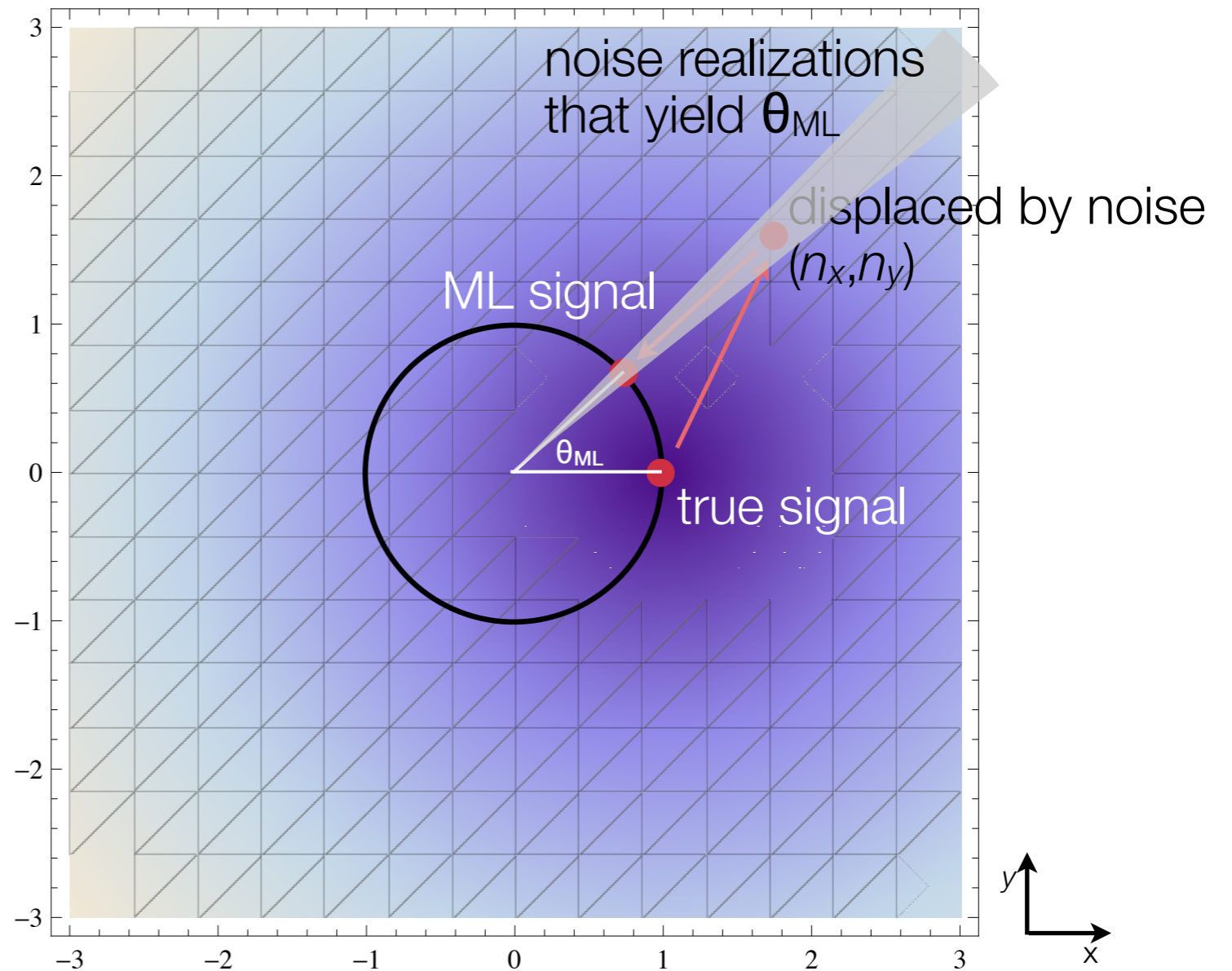
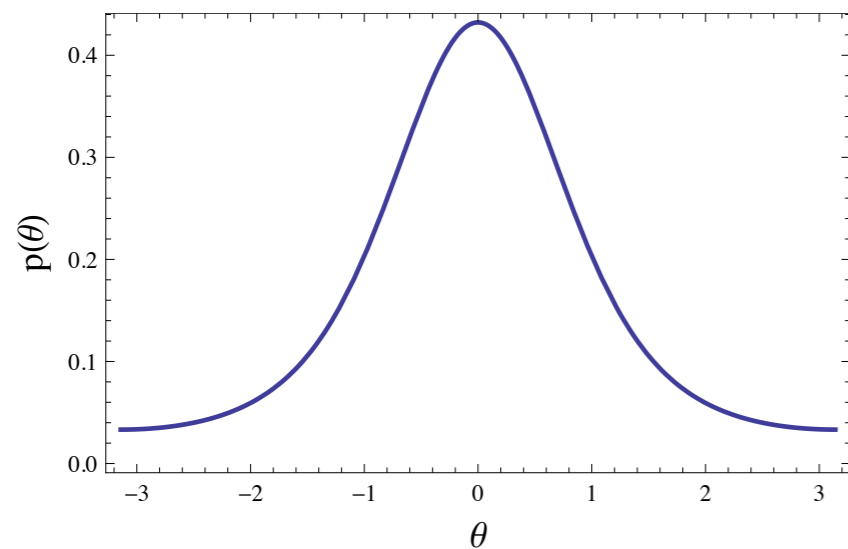
for instance, with one parameter and two noise dimensions:

$$h(\theta) = (\cos \theta, \sin \theta)$$

$$\theta_{\text{true}} = 0$$

$$s = (n_x + 1, n_y)$$

$$p(n) = \mathcal{N} \exp -(n_x^2 + n_y^2)/2$$



$$p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) \left| \frac{\partial \text{ML}_i}{\partial \theta_j} \right| e^{-(n,n)/2} dn$$

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
is really low-dimensional!

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
is really low-dimensional!

- the inner product in the noise sampling probability can be written with respect to any basis...

$$\begin{aligned} p(n) &= \mathcal{N} \exp -(n, n)/2 \\ &= \mathcal{N} \exp \left\{ - \sum_k N^k N_k / 2 \right\} \end{aligned}$$

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- the inner product in the noise sampling probability can be written with respect to any basis...
- ...such as one that **orthonormalizes** the $\partial_i h$ and $\partial_{ij} h$ (needed for $\partial \text{ML}_i / \partial \theta_j$):

$$p(n) = \mathcal{N} \exp -(n, n)/2$$

$$= \mathcal{N} \exp \left\{ - \sum_k N^k N_k / 2 \right\}$$

$$\hat{n}_1 \propto h_1,$$

$$\hat{n}_2 \propto (1 - \hat{n}_1 \otimes \hat{n}_1) h_2,$$

...

$$\hat{n}_{d+1} \propto (1 - \hat{n}_1 \otimes \hat{n}_1 - \cdots) h_{11},$$

...

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- the inner product in the noise sampling probability can be written with respect to any basis...
- ...such as one that **orthonormalizes** the $\partial_i h$ and $\partial_{ij} h$ (needed for $\partial \text{ML}_i / \partial \theta_j$):
- the integrand is a function only of the first $d + d(d+1)/2$ coefficients N^k (i.e., the projections of the noise over the first and second signal derivatives); all the others integrate out

$$p(n) = \mathcal{N} \exp -(n, n)/2$$

$$= \mathcal{N} \exp \left\{ - \sum_k N^k N_k / 2 \right\}$$

$$\hat{n}_1 \propto h_1,$$

$$\hat{n}_2 \propto (1 - \hat{n}_1 \otimes \hat{n}_1) h_2,$$

...

$$\hat{n}_{d+1} \propto (1 - \hat{n}_1 \otimes \hat{n}_1 - \cdots) h_{11},$$

...

$$n = N^k \hat{n}_k,$$

$$h_i = C_i^k \hat{n}_k,$$

$$h_{i,j} = C_{ij}^k \hat{n}_k,$$

$$\Delta h = h(\theta) - h(\theta_{\text{true}})$$

$$\text{ML}_i = C_i^k N_k - C_i^k (\Delta h, \hat{n}_k),$$

$$\text{ML}_{i,j} = C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk}$$

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- the integrand is a function only of the first $d + d(d+1)/2$ coefficients N^k (i.e., the projections of the noise over the first and second signal derivatives); all the others integrate out

$$\begin{aligned}
 n &= N^k \hat{n}_k, \\
 h_i &= C_i^k \hat{n}_k, \\
 h_{i,j} &= C_{ij}^k \hat{n}_k, \\
 \text{ML}_i &= C_i^k N_k - C_i^k (\Delta h, \hat{n}_k), \\
 \text{ML}_{i,j} &= C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk}
 \end{aligned}$$

$\Delta h = h(\theta) - h(\theta_{\text{true}})$

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- furthermore, the deltas satisfy integration over the first d N_k s: “ h_{ij} ” degrees of freedom only!

$$p(\theta) = \frac{e^{-\sum_{k=1}^d (\Delta h, \hat{n}_k)^2 / 2}}{(2\pi)^{d(d+3)/4} \prod_{k=1}^d C_k^k} \int |C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk}| e^{-\sum_m (N_m)^2 / 2} dN_m$$


- the integrand is a function only of the first $d + d(d+1)/2$ coefficients N^k (i.e., the projections of the noise over the first and second signal derivatives); all the others integrate out

$$\begin{aligned} n &= N^k \hat{n}_k, \\ h_i &= C_i^k \hat{n}_k, \\ h_{i,j} &= C_{ij}^k \hat{n}_k, \\ \text{ML}_i &= C_i^k N_k - C_i^k (\Delta h, \hat{n}_k), \\ \text{ML}_{i,j} &= C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk} \end{aligned}$$

$\Delta h = h(\theta) - h(\theta_{\text{true}})$

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- furthermore, the deltas satisfy integration over the first d N_k s: “ h_{ij} ” degrees of freedom only!

$$p(\theta) = \frac{e^{-\sum_{k=1}^d (\Delta h, \hat{n}_k)^2 / 2}}{(2\pi)^{d(d+3)/4} \prod_{k=1}^d C_k^k} \int |C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk}| e^{-\sum_m (N_m)^2 / 2} dN_m$$


$$p(\theta) = \frac{e^{-(\Delta h, h_i)(F^{-1})^{ij}(\Delta h, h_j)/2}}{\sqrt{(2\pi)^d |F_{ij}|}} \times$$


$$\frac{1}{\sqrt{(2\pi)^{d(d-1)/2} |D_{\mu\nu}|}} \int |F_{ij} + (\Delta h, h_{ij}) - M_{(ij)}| e^{-M_\mu (D^{-1})^{\mu\nu} M_\nu / 2} dM_\mu$$

- ...where the $M_\mu \equiv M_{ij}$ are normal random variables with covariance matrix given by $D_{\mu\nu} \equiv (\tilde{h}_{ij}, \tilde{h}_{kl})$, where \sim denotes projection orthogonal to the h_k

let's show that $p(\theta) = \mathcal{N} \int \delta(\text{ML}_1(n)) \cdots \delta(\text{ML}_d(n)) |\partial \text{ML}_i / \partial \theta_j| e^{-(n,n)/2} dn$
 is really low-dimensional!

- furthermore, the deltas satisfy integration over the first $d N_k$ s:

“ h_{ij} ” degrees of freedom only!

$$p(\theta) = \frac{e^{-\sum_{k=1}^d (\Delta h, \hat{n}_k)^2 / 2}}{(2\pi)^{d(d+3)/4} \prod_{k=1}^d C_k^k} \int |C_{ij}^m N_m - C_{ij}^m (\Delta h, \hat{n}_m) - C_i^k C_{jk}| e^{-\sum_m (N_m)^2 / 2} dN_m$$


$$p(\theta) = \frac{e^{-(\Delta h, h_i)(F^{-1})^{ij}(\Delta h, h_j)/2}}{\sqrt{(2\pi)^d |F_{ij}|}} \times \frac{1}{\sqrt{(2\pi)^{d(d-1)/2} |D_{\mu\nu}|}} \int |F_{ij} + (\Delta h, h_{ij}) - M_{(ij)}| e^{-M_\mu (D^{-1})^{\mu\nu} M_\nu / 2} dM_\mu$$

- ...where the $M_\mu \equiv M_{ij}$ are **normal random variables** with covariance matrix given by $D_{\mu\nu} \equiv (\tilde{h}_{ij}, \tilde{h}_{kl})$, where \sim denotes projection orthogonal to the h_k
- this $d(d+1)/2$ -dimensional integral is **trivial numerically**. All that's needed are F_{ij} and $D_{\mu\nu}$, which require $\sim d^4/8$ inner products

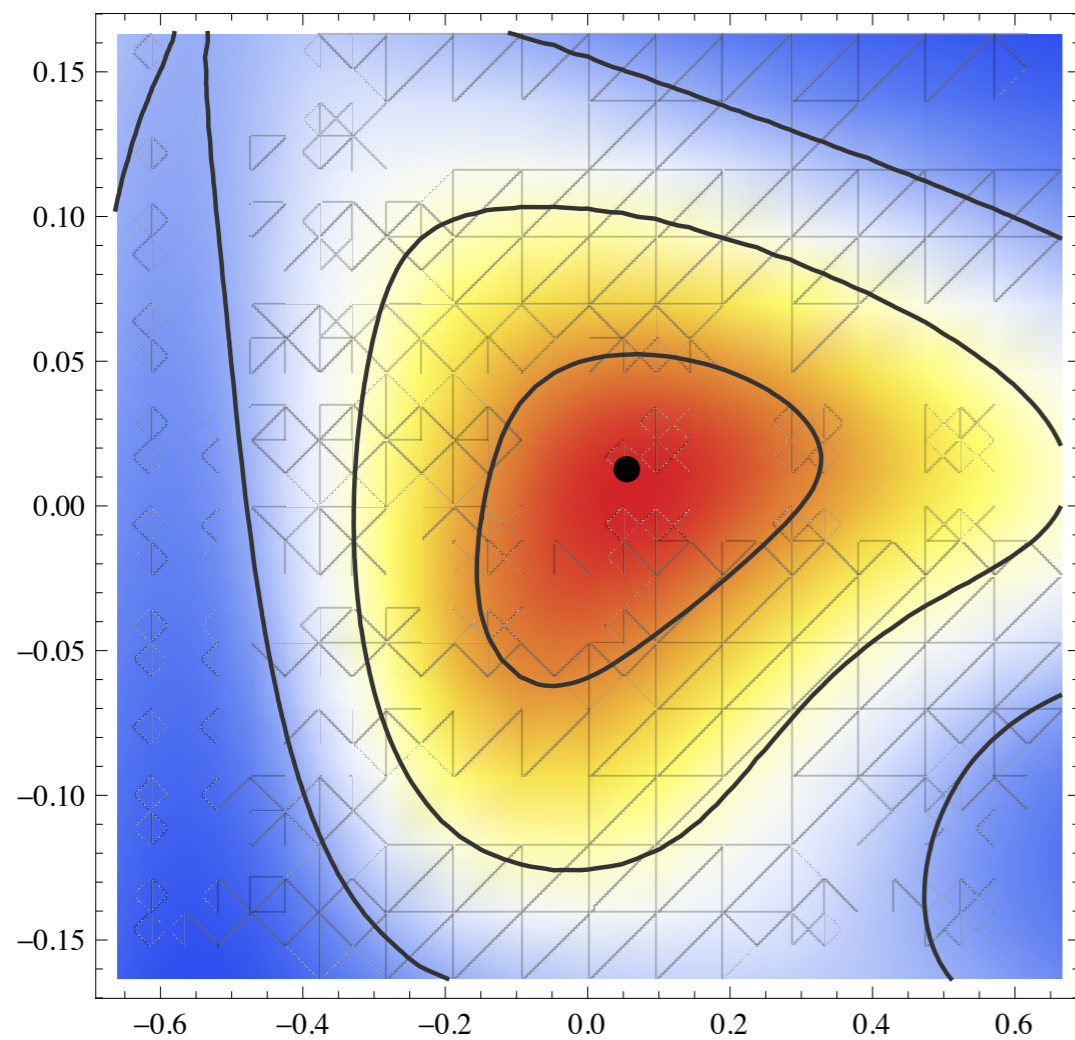
check: the leading-order Fisher-matrix result follows from neglecting the second derivatives of the signal

$$p(\theta) = \frac{e^{-(\Delta h, h_i)(F^{-1})^{ij}(\Delta h, h_j)/2}}{\sqrt{(2\pi)^d |F_{ij}|}} \times \frac{1}{\sqrt{(2\pi)^{d(d-1)/2} |D_{\mu\nu}|}} \int |F_{ij} + \cancel{(\Delta h, h_{ij})} - \cancel{M_{(ij)}}| e^{-M_\mu (D^{-1})^{\mu\nu} M_\nu / 2} dM_\mu$$

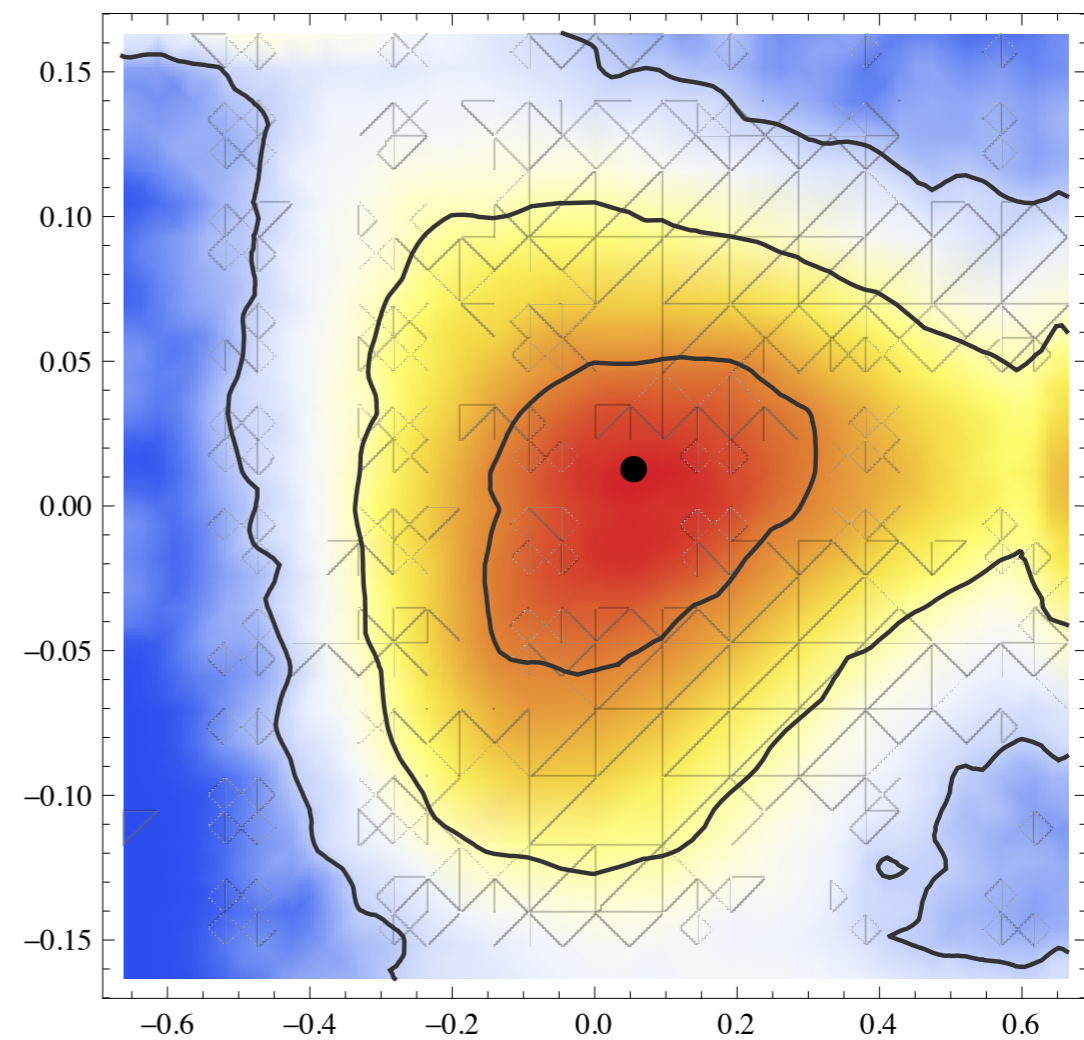


$$p(\Delta\theta) = \frac{e^{-\Delta\theta^i F_{ij} \Delta\theta^j / 2}}{\sqrt{(2\pi)^d |F_{ij}^{-1}|}}$$

check: evaluating the mapping integral recovers the distribution from the numerical search



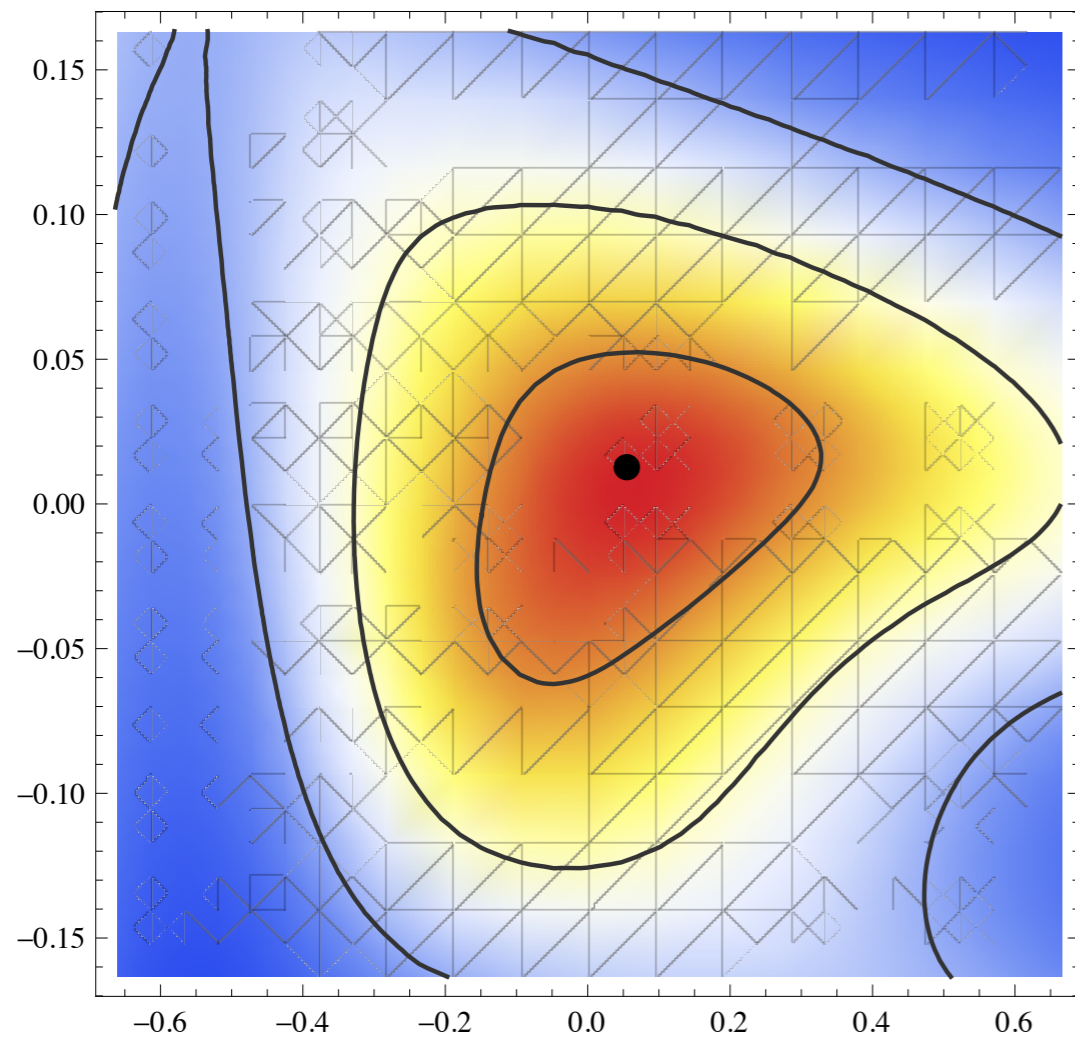
integral formula,
100x100 grid



numerical search,
100,000 noise realizations

go read about it and use it!

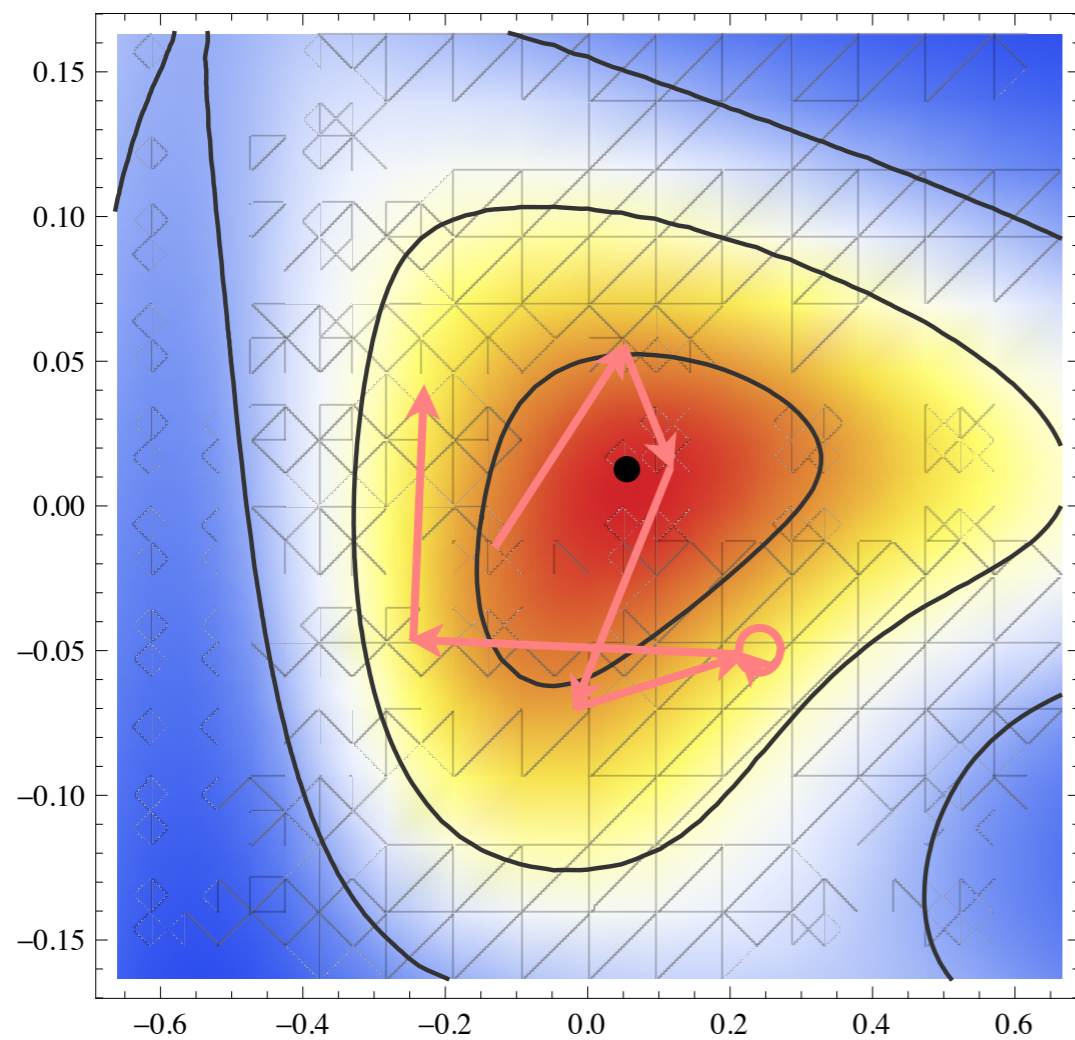
[MV, PRL 107, 191104 (2011)]



go read about it and use it!

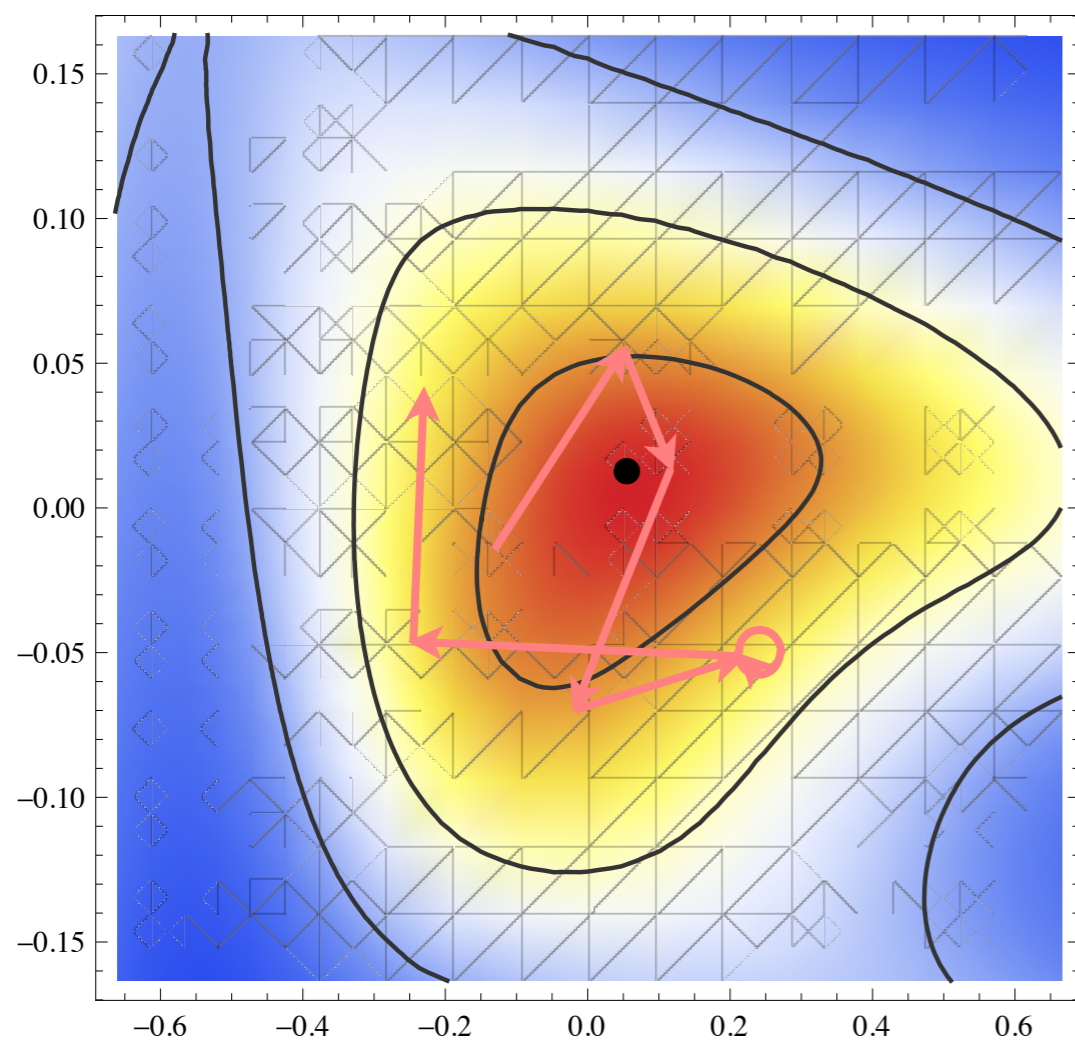
[MV, PRL 107, 191104 (2011)]

- in d dimensions, a **Markov Chain** can be used to sample $p(\theta)$



go read about it and use it!

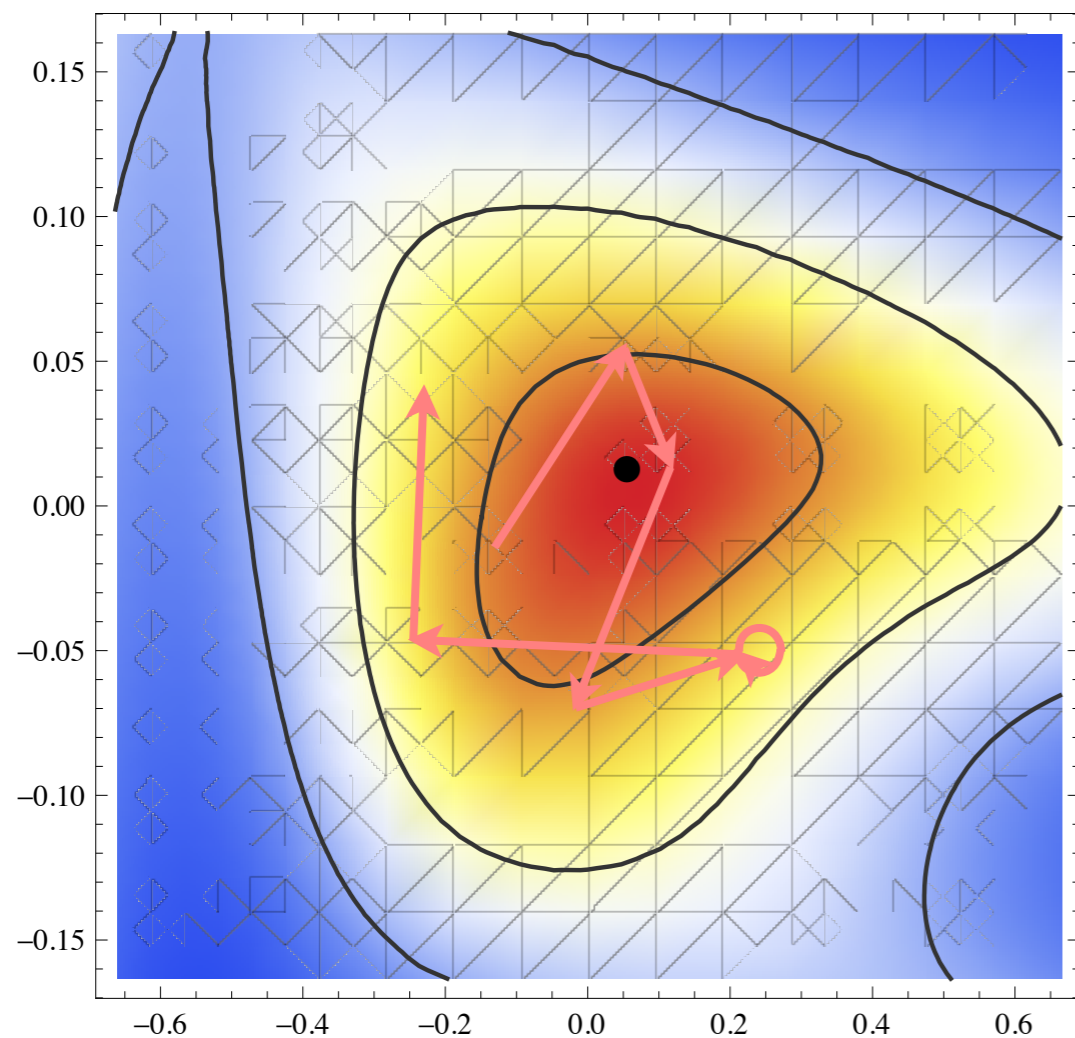
[MV, PRL 107, 191104 (2011)]



- in d dimensions, a **Markov Chain** can be used to sample $p(\theta)$
- [10^6 samples $\times d^4/8 \times 10^6$ -point FFT = **1000x cheaper** for $d = 10$, still very parallelizable]

go read about it and use it!

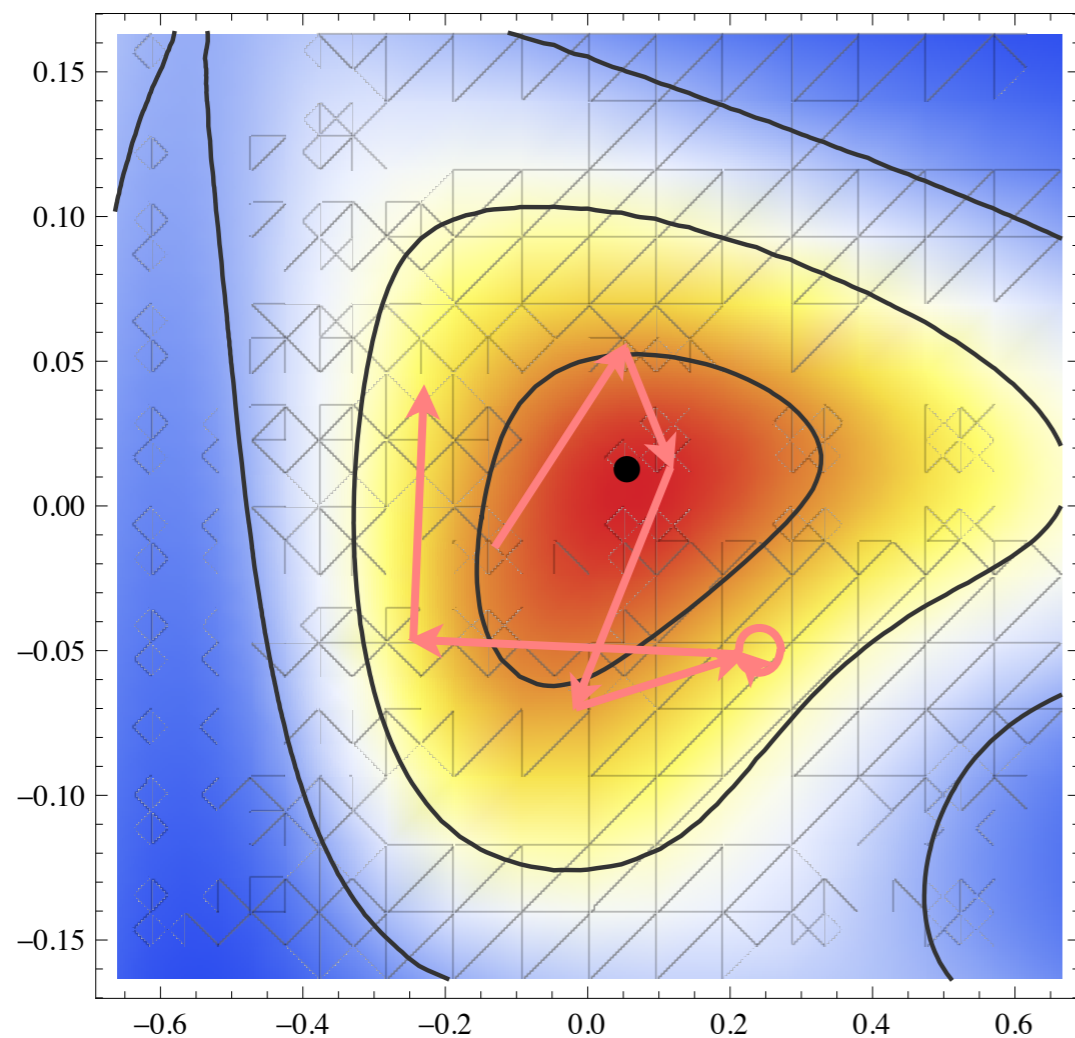
[MV, PRL 107, 191104 (2011)]



- in d dimensions, a **Markov Chain** can be used to sample $p(\theta)$
- [10^6 samples $\times d^4/8 \times 10^6$ -point FFT = **1000x cheaper** for $d = 10$, still very parallelizable]
- this technique generates **exact estimates** of the **frequentist error** for the ML estimator for any SNR

go read about it and use it!

[MV, PRL 107, 191104 (2011)]



- in d dimensions, a **Markov Chain** can be used to sample $p(\theta)$
- [10^6 samples $\times d^4/8 \times 10^6$ -point FFT = **1000x cheaper** for $d = 10$, still very parallelizable]
- this technique generates **exact estimates** of the **frequentist error** for the ML estimator for any SNR
- the resulting distribution can be used to **seed Bayesian-inference Monte Carlo** (and will include isolated secondary maxima)



Between Fisher
and Monte Carlo:



Mapping the distribution of the maximum-likelihood estimator for GW source parameters

Michele Vallisneri

Jet Propulsion Laboratory

Copyright 2012 California Institute of Technology