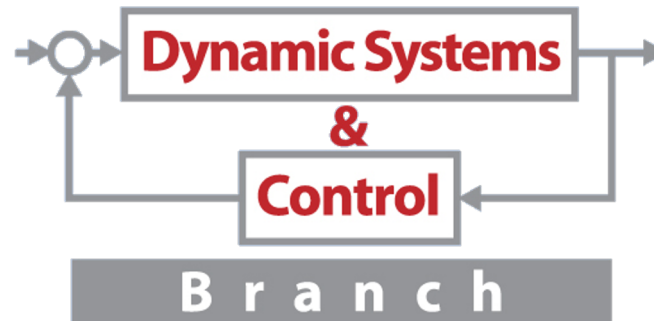


Robust Estimation of Sliced-Exponential Distributions

Luis G. Crespo, Brendon K. Colbert, Tanner Slagel, Sean P. Kenny

NASA Langley Research Center

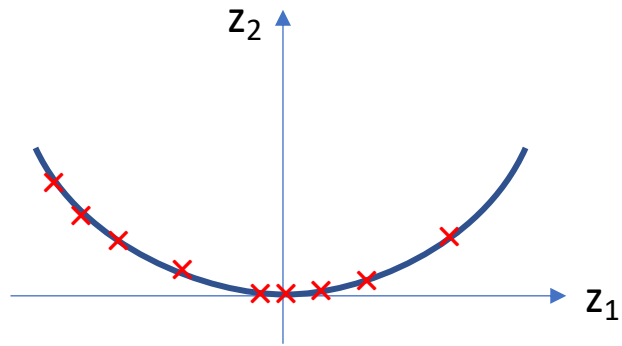
60th IEEE CDC



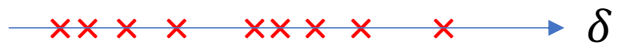
Sliced-Distributions (SD)

- SD enable the characterization of multivariate data as either a vector of continuous random variables or as a data-enclosing set
- SD can naturally model complex datasets: multimodal densities with non-monotonic dependencies
- SD are dense: they can approximate any distribution with finite support arbitrarily close

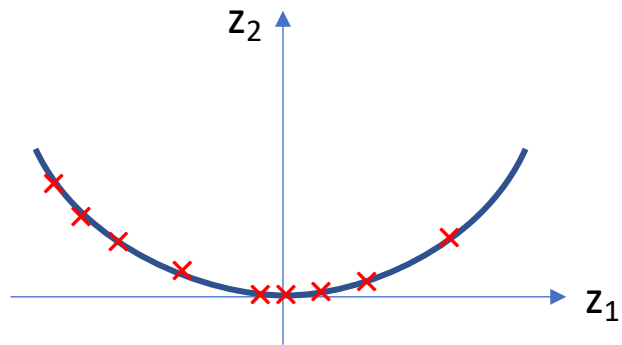
Estimation



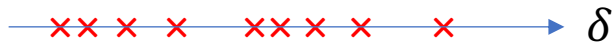
Transformation to
feature space: $z(\delta)$



Estimation

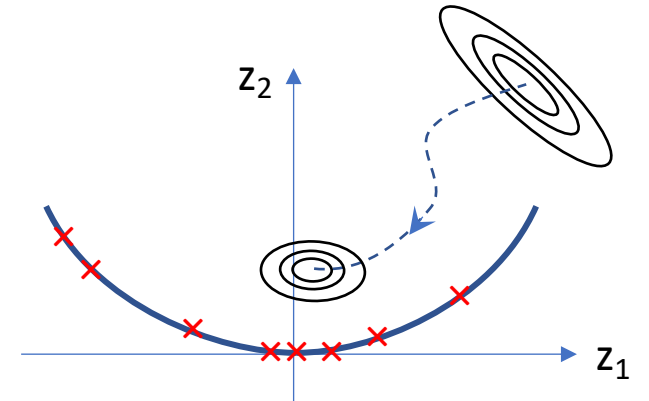


Transformation to
feature space: $z(\delta)$

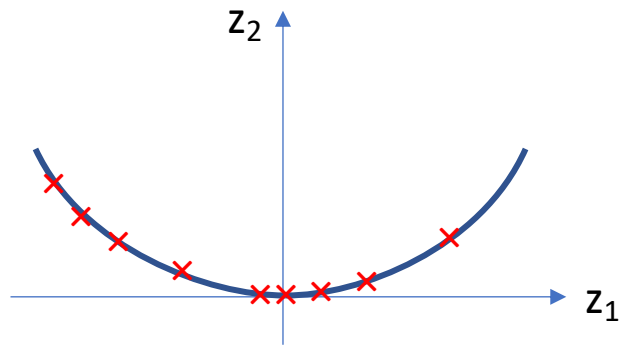


$$f_{\delta} \sim e^{-\frac{1}{2}\phi(z; \theta)}$$

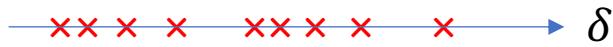
Search for an optimal distribution



Estimation

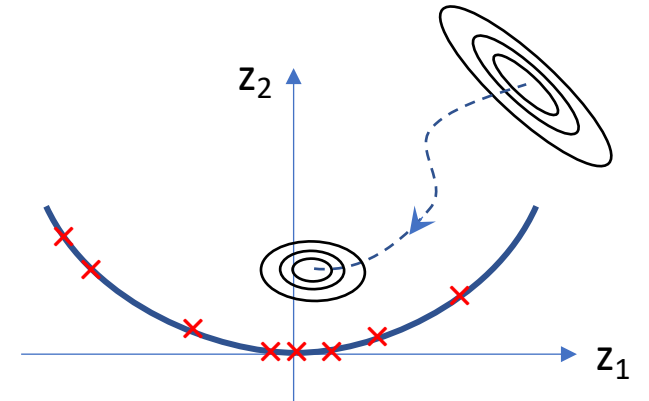


Transformation to
feature space: $z(\delta)$

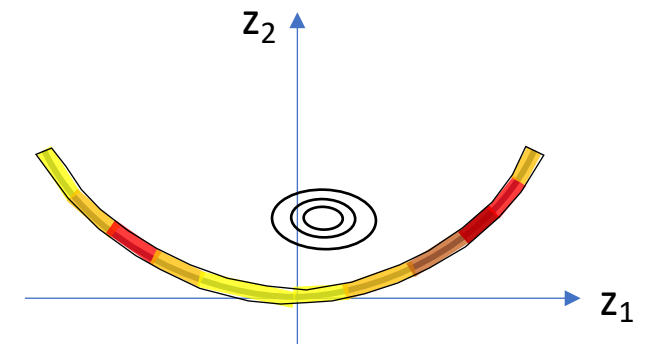


$$f_{\delta} \sim e^{-\frac{1}{2}\phi(z; \theta)}$$

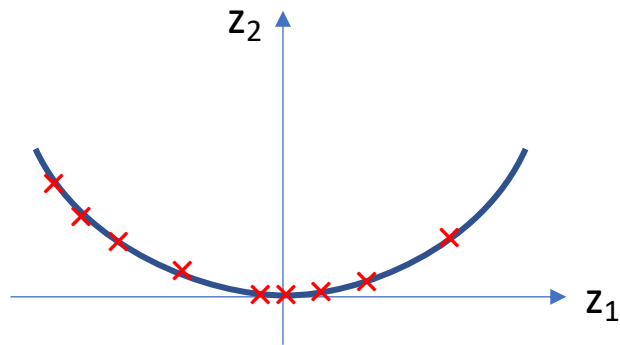
Search for an optimal distribution



Evaluate joint density on manifold

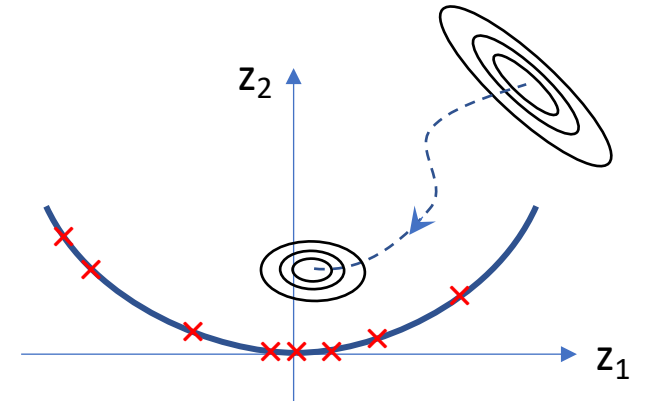


Estimation

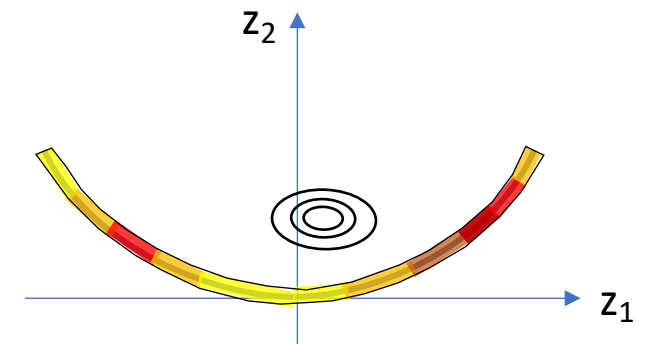


$$f_{\delta} \sim e^{-\frac{1}{2}\phi(z; \theta)}$$

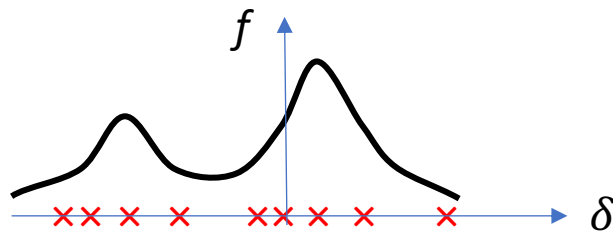
Search for an optimal distribution



Evaluate joint density on manifold



Transformation to
feature space: $z(\delta)$



Transformation to
physical space

Sliced-Distributions

$$z = t(\delta; d)$$

Transformation from physical to feature space

$$f_{\delta}(\delta; d, \theta, \Delta) = \begin{cases} \frac{1}{c_{\delta}(\theta)} \exp\left(-\frac{1}{2}\phi(z; \theta)\right) & \text{if } \delta \in \Delta, \\ 0 & \text{otherwise,} \end{cases}$$

Joint density function

$$c_{\delta}(\theta) = \int_{\Delta} \exp\left(-\frac{1}{2}\phi(z; \theta)\right) d\delta$$

Normalization constant

Sliced-Distributions

$$z = t(\delta; d)$$

Transformation from physical to feature space

- We focus on polynomial transformations to feature space
- This choice yields semi-algebraic data-enclosing sets: robust control

Sliced-Normal (SN) Distributions

$$f_{\delta}(\delta; d, \theta, \Delta) = \begin{cases} \frac{1}{c_{\delta}(\theta)} \exp\left(-\frac{1}{2}\phi(z; \theta)\right) & \text{if } \delta \in \Delta, \\ 0 & \text{otherwise,} \end{cases}$$

Joint density function

1. Sliced-Normals

- Argument: $\phi_{\mathbf{N}}(z; \mu, P) = (z - \mu)^{\top} P (z - \mu)$
- Feasibility condition: $P \in \mathcal{S}_{++}^{n_z}$
- Optimality criteria: focus on density estimation

SN: Maximum Likelihood (ML) Estimates

- ML estimate in feature space

$$\mu^\star = \frac{1}{n} \sum_{i=1}^n z^{(i)},$$

$$P^\star = \left(\frac{1}{n} \sum_{i=1}^n (z^{(i)} - \mu^\star) (z^{(i)} - \mu^\star)^\top \right)^{-1}$$

Closed-form solution

- ML estimate in physical space

$$\langle \mu^\star, P^\star \rangle = \operatorname{argmin}_{\mu \in \mathbb{R}^{n_z}, P \in S_{++}^{n_z}} \left\{ n \log c_\delta(\mu, P) + \frac{1}{2} \sum_{i=1}^n \phi_N(z^{(i)}; \mu, P) \right\}$$

Non-convex in
 $n_\theta = (n_z^2 + 3n_z)/2$
decision variables

SN: Maximum Likelihood (ML) Estimates

- ML estimate in feature space

$$\mu^\star = \frac{1}{n} \sum_{i=1}^n z^{(i)},$$

$$P^\star = \left(\frac{1}{n} \sum_{i=1}^n (z^{(i)} - \mu^\star) (z^{(i)} - \mu^\star)^\top \right)^{-1}$$

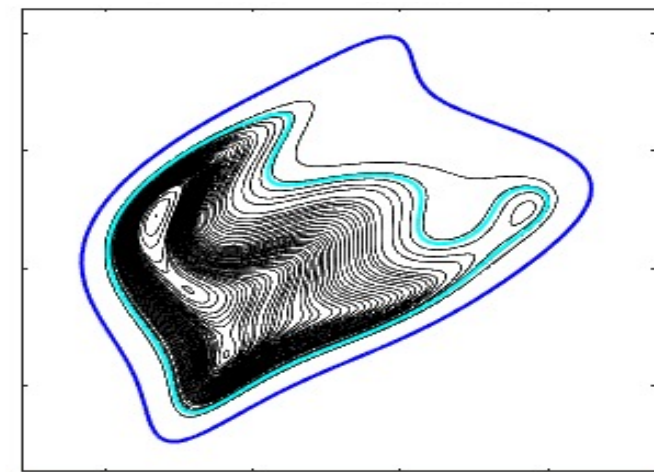
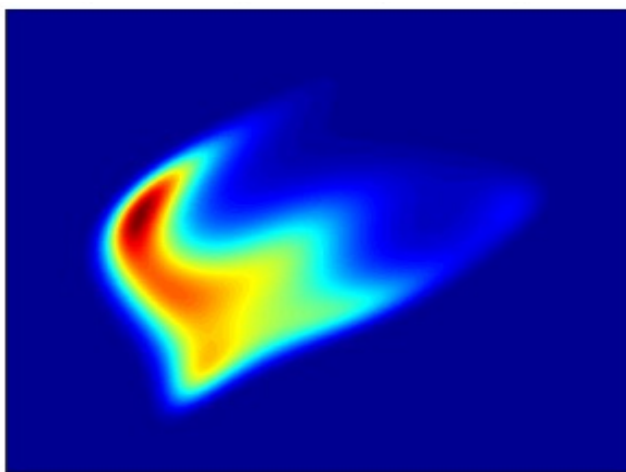
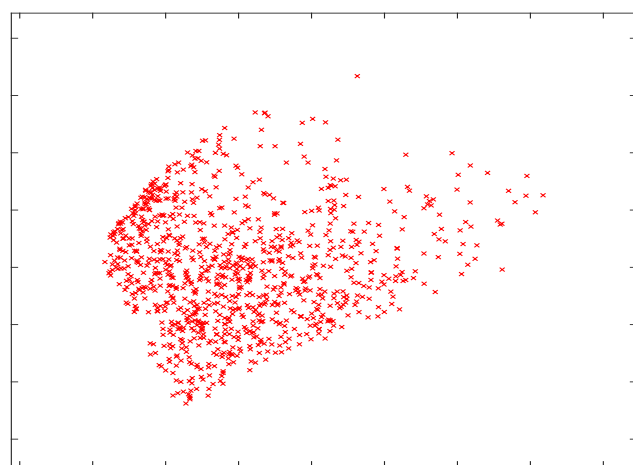
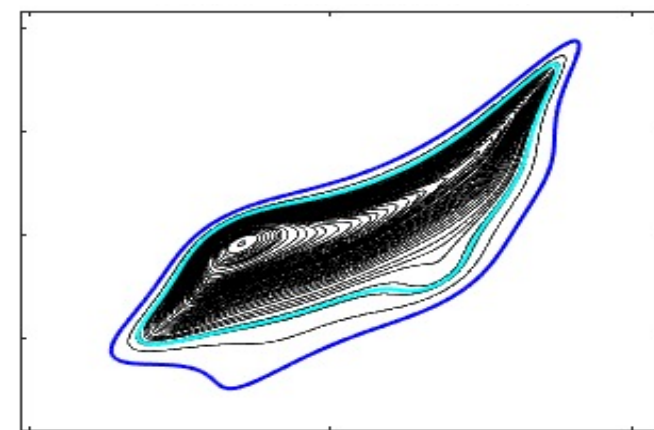
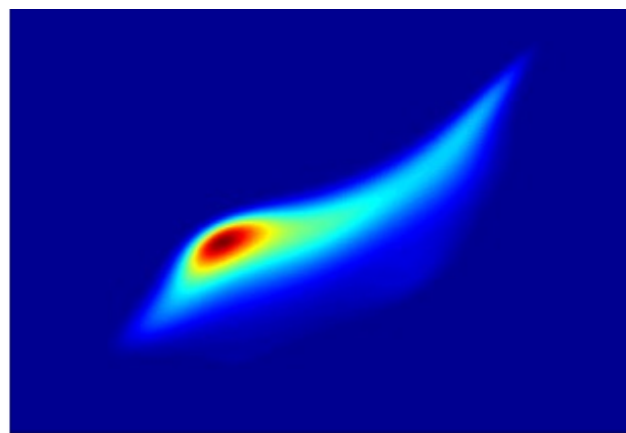
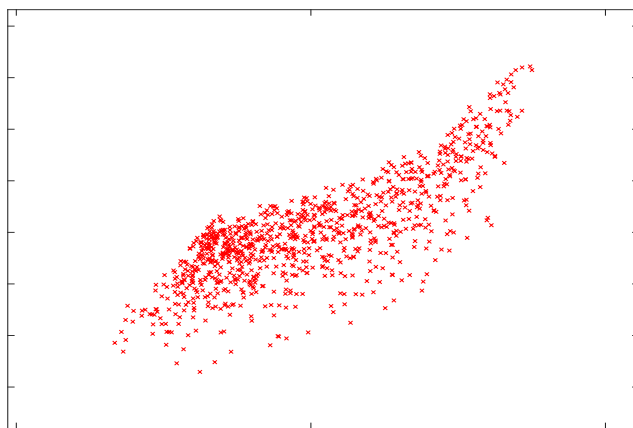
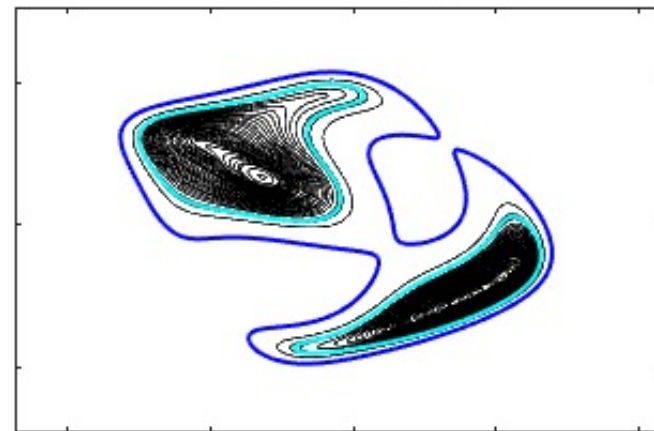
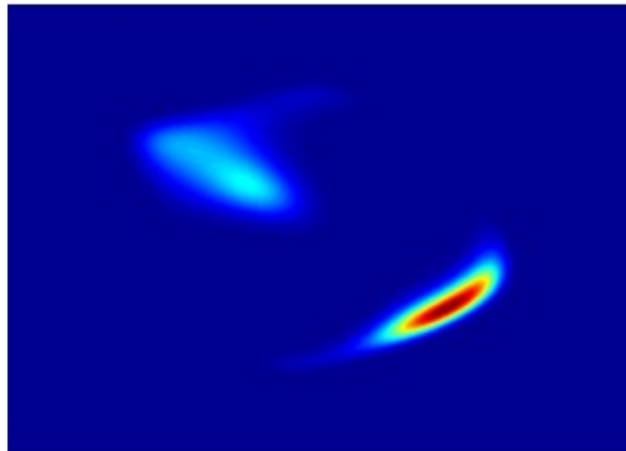
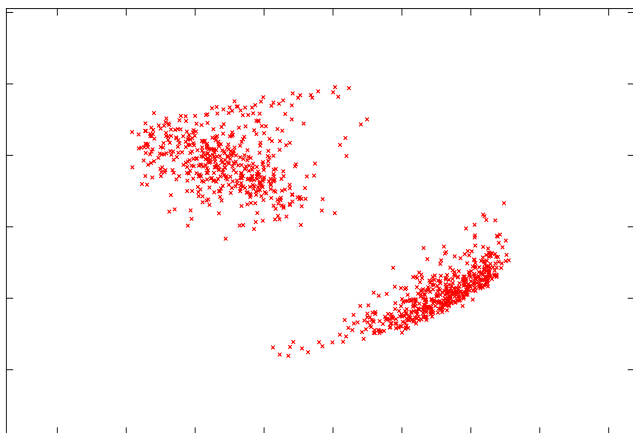
Closed-form solution

- ML estimate in physical space

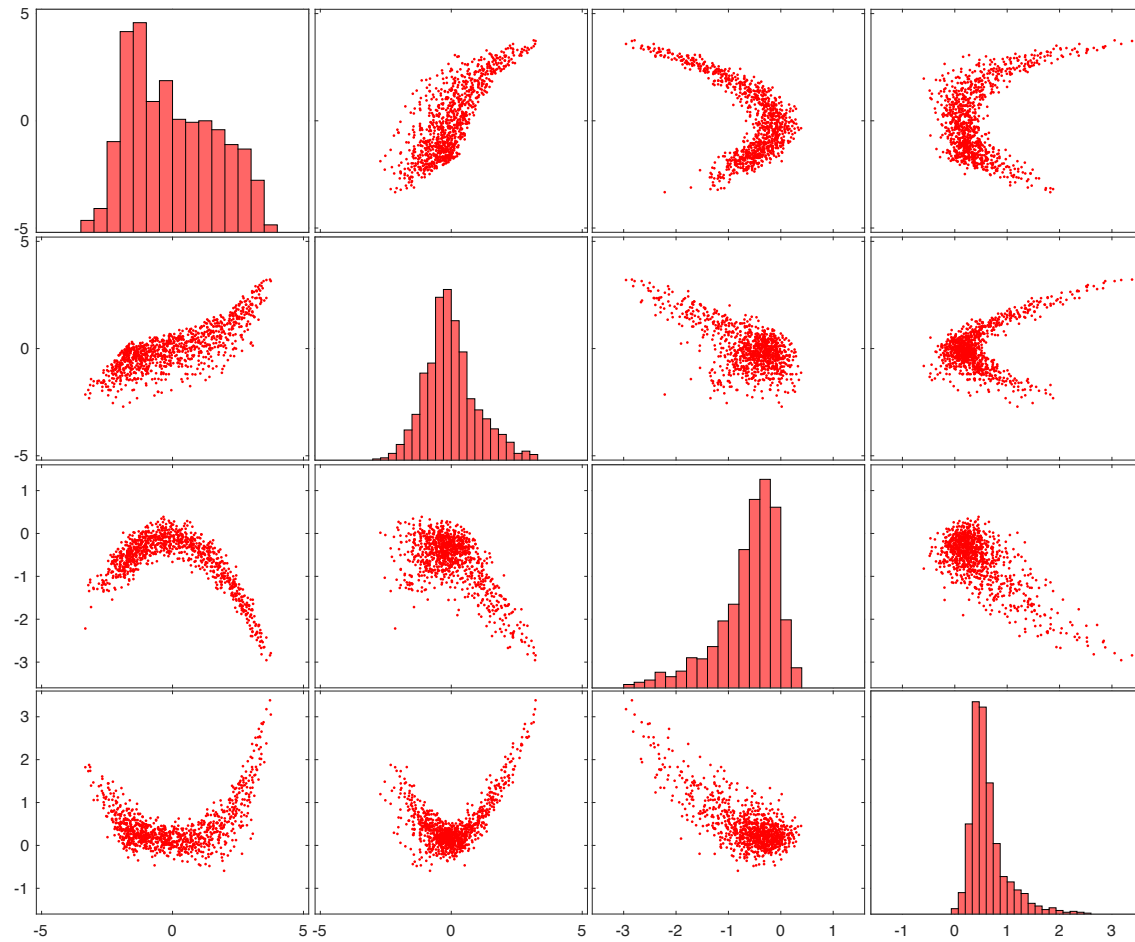
$$\langle \mu^\star, P^\star \rangle = \operatorname{argmin}_{\mu \in \mathbb{R}^{n_z}, P \in S_{++}^{n_z}} \left\{ n \log c_\delta(\mu, P) + \frac{1}{2} \sum_{i=1}^n \phi_N(z^{(i)}; \mu, P) \right\}$$

Non-convex in
 $n_\theta = (n_z^2 + 3n_z)/2$
decision variables

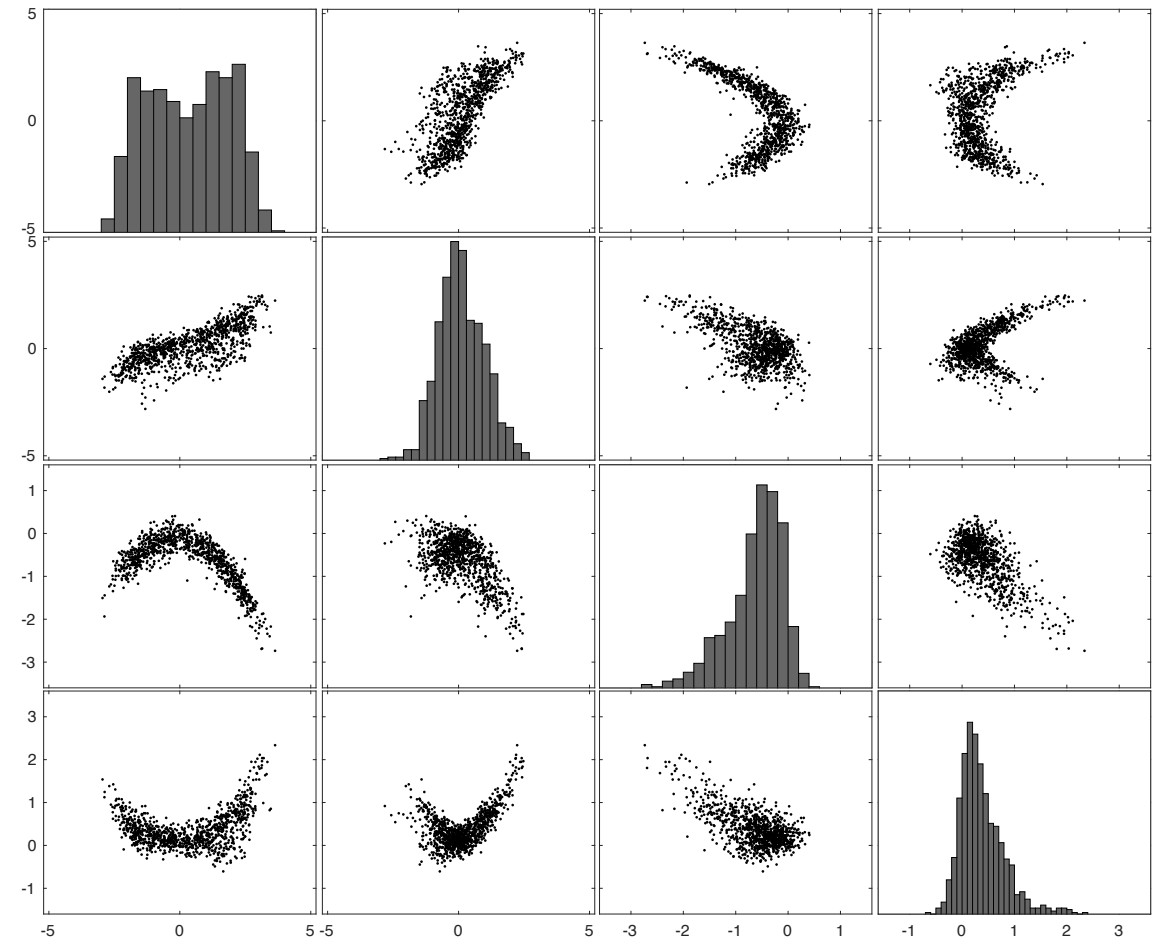
This limits its applicability



Examples



DGM

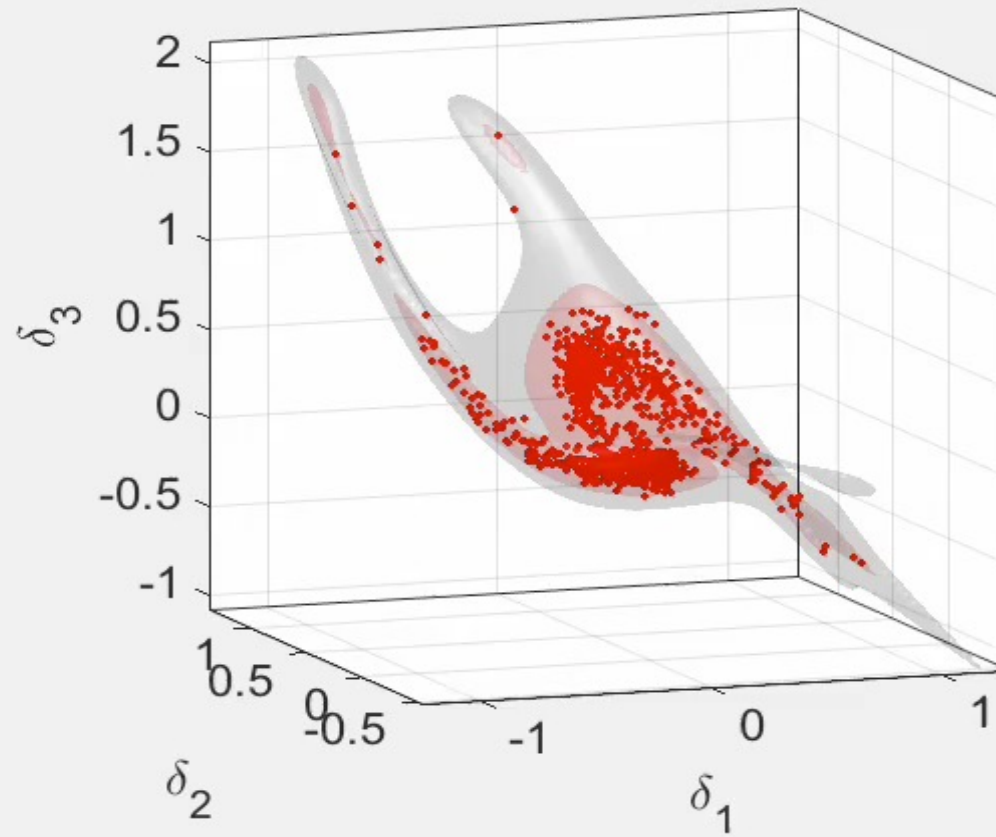


Samples of estimated SN

Examples

$$\text{Vol}_{100} = 0.683$$

$$\text{Vol}_{95} = 0.202$$



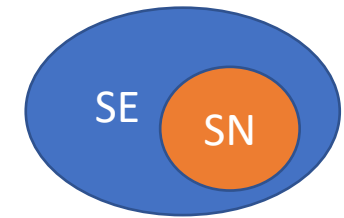
Sliced-Exponential (SE) Distributions

$$f_{\delta}(\delta; d, \theta, \Delta) = \begin{cases} \frac{1}{c_{\delta}(\theta)} \exp\left(-\frac{1}{2}\phi(z; \theta)\right) & \text{if } \delta \in \Delta, \\ 0 & \text{otherwise,} \end{cases}$$

Joint density function

2. Sliced-Exponentials

- Argument: $\phi_E(z; \lambda) = \lambda^T z$
- Feasibility condition: $c_{\delta}(\lambda)$ being finite (convex function)
- Optimality criteria:



SE: Maximum Likelihood (ML) Estimates

- ML estimate in feature space: performed poorly, discontinued
- ML estimate in physical space

$$\lambda^{\star} = \operatorname{argmin}_{\lambda \in \mathbb{R}^{n_z}} \left\{ n \log c_{\delta}(\lambda) + \frac{1}{2} \sum_{i=1}^n \phi_E(z^{(i)}; \lambda) \right\}$$

Convex program in a number of decision variables that grows linearly with n_z

Normalization Constant

$$c_{\delta}(\theta) = \int_{\Delta} \exp \left(-\frac{1}{2} \phi(z; \theta) \right) d\delta$$

Normalization constant

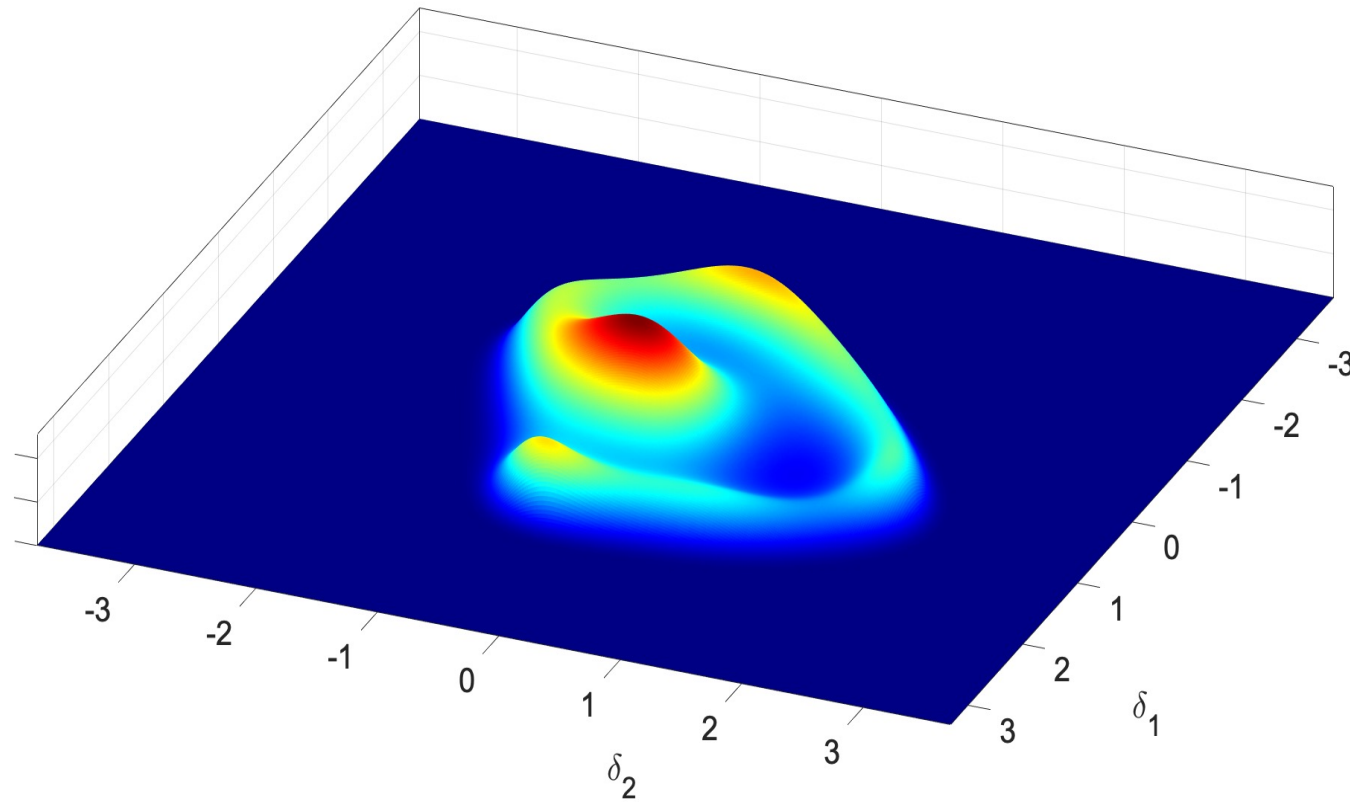
- Normalization constant must be approximated in practice

$$c_{\text{sam}}(m, \theta) = \frac{\text{Vol}(\Delta)}{m} \sum_{i=1}^m \exp \left(-\frac{1}{2} \phi \left(t(\delta_u^{(i)}; d); \theta \right) \right)$$

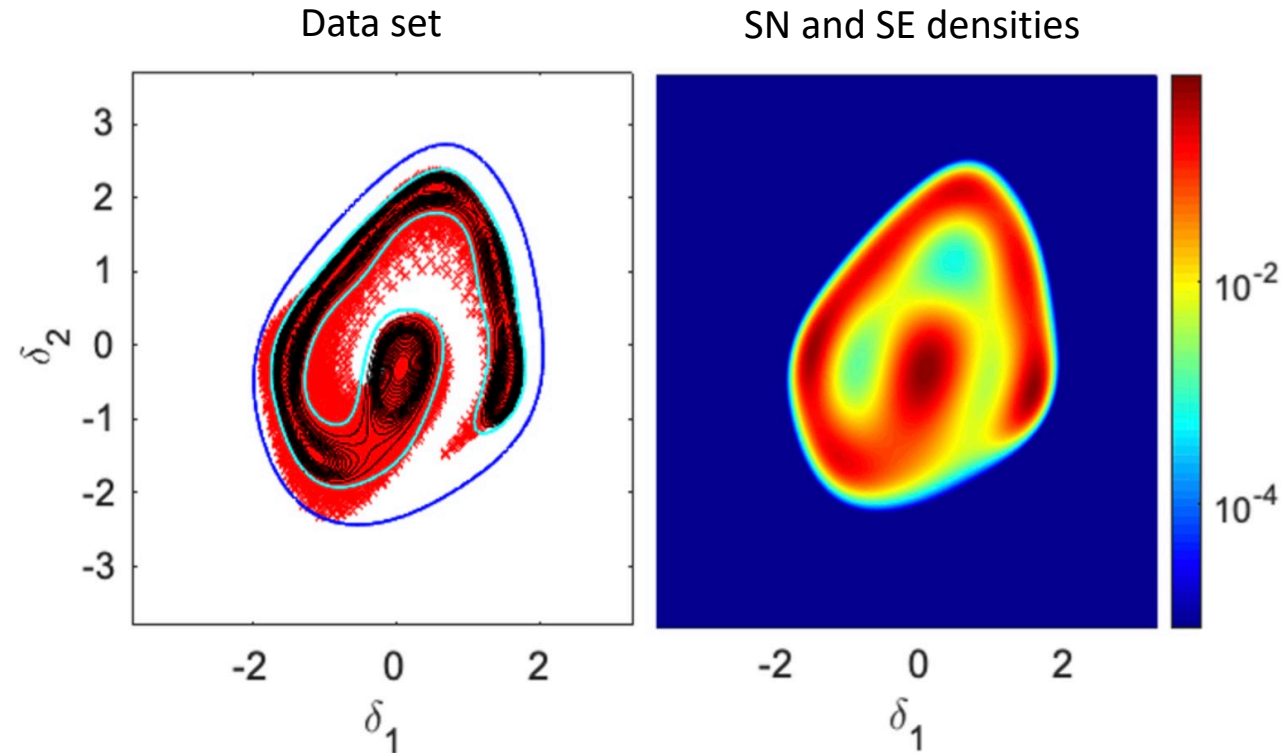
Sampling-based approximation

- This approximation preserves the convexity of the original ML program
- However, insufficiently large m might yield spuriously large likelihood values/distributions
- The developments that follow seek to mend for these deficiencies

Example: Multimodal Dataset



Example: Negligible Error

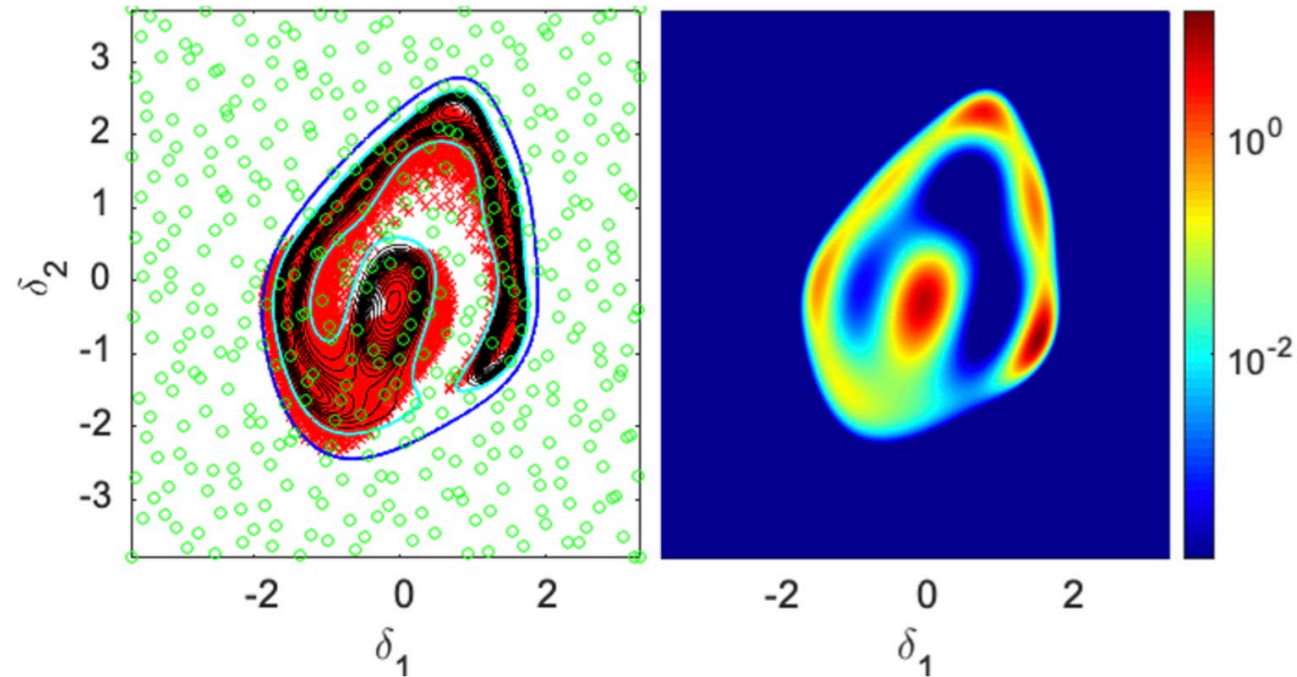


LOG-LIKELIHOOD OF THE DATA, $\mathcal{L}(\theta, \mathcal{D}_\delta)$, FOR $m = 1e5$

	$m = 8e4$ & $\beta = \infty$	$m = 400$ & $\beta = \infty$	$m = 360$ & $\beta = \infty$	$m = 360$ & $\beta = 1$
$\theta = \{\mu^*, P^*\}$	-0.987e5	-1.549e5	-7.038e5	-1.049e5
$\theta = \{\lambda^*\}$	-0.987e5	-1.549e5	-9.837e5	-1.049e5
$\theta = \{\lambda_{\text{sos}}^*\}$	-1.033e5	-1.037e5	-1.054e5	-1.054e5

Example: Sizable Error

- Optimization algorithm converges
- Performance of the distribution is suboptimal
- Spurious spikes appear in under-sampled regions

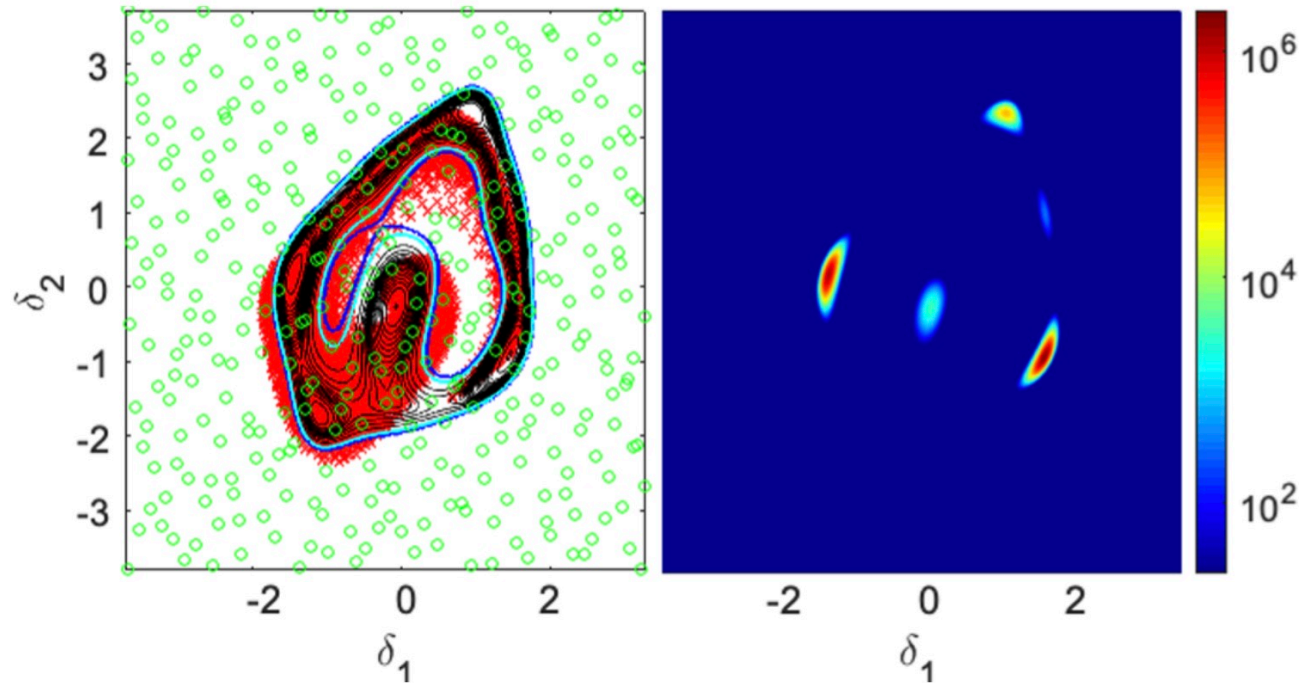


LOG-LIKELIHOOD OF THE DATA, $\mathcal{L}(\theta, \mathcal{D}_\delta)$, FOR $m = 1e5$

	$m = 8e4$ & $\beta = \infty$	$m = 400$ & $\beta = \infty$	$m = 360$ & $\beta = \infty$	$m = 360$ & $\beta = 1$
$\theta = \{\mu^\star, P^\star\}$	-0.987e5	-1.549e5	-7.038e5	-1.049e5
$\theta = \{\lambda^\star\}$	-0.987e5	-1.549e5	-9.837e5	-1.049e5
$\theta = \{\lambda_{\text{sos}}^\star\}$	-1.033e5	-1.037e5	-1.054e5	-1.054e5

Example: Sizable Error

- Optimization algorithm converges
- Performance of the distribution is suboptimal
- Spurious spikes appear in under-sampled regions



LOG-LIKELIHOOD OF THE DATA, $\mathcal{L}(\theta, \mathcal{D}_\delta)$, FOR $m = 1e5$

	$m = 8e4$ & $\beta = \infty$	$m = 400$ & $\beta = \infty$	$m = 360$ & $\beta = \infty$	$m = 360$ & $\beta = 1$
$\theta = \{\mu^\star, P^\star\}$	-0.987e5	-1.549e5	-7.038e5	-1.049e5
$\theta = \{\lambda^\star\}$	-0.987e5	-1.549e5	-9.837e5	-1.049e5
$\theta = \{\lambda_{\text{sos}}^\star\}$	-1.033e5	-1.037e5	-1.054e5	-1.054e5

Mitigating the Effects of Sampling Error

- Increase the number of samples
- Constrain the decision space to a preferential subspace/subset

Strategies

1. Choose a data-based square polynomial basis resulting from the ML SN in feature space and use $\lambda > 0$

$$\phi_E(z; \lambda) = \lambda^\top z_{\text{sos}}$$

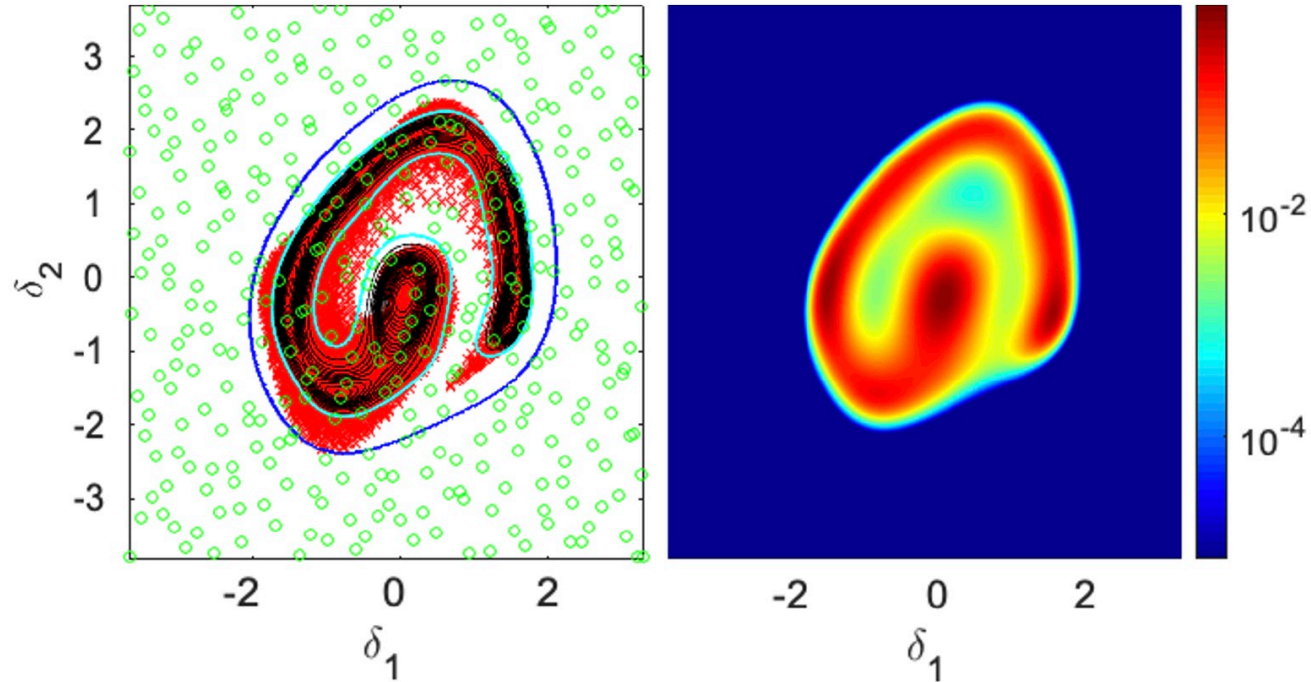
$$z_{\text{sos}} = \text{diag}(v)v$$

$$v = M(z - \mu^\star)$$

$$P^\star = M^\top M$$

- Only algebraic manipulations are needed to compute a SN
- Convex optimization with a lower-dimensional decision space
- This basis is both versatile and robust to sampling error

Example



LOG-LIKELIHOOD OF THE DATA, $\mathcal{L}(\theta, \mathcal{D}_\delta)$, FOR $m = 1e5$

	$m = 8e4$ & $\beta = \infty$	$m = 400$ & $\beta = \infty$	$m = 360$ & $\beta = \infty$	$m = 360$ & $\beta = 1$
$\theta = \{\mu^\star, P^\star\}$	-0.987e5	-1.549e5	-7.038e5	-1.049e5
$\theta = \{\lambda^\star\}$	-0.987e5	-1.549e5	-9.837e5	-1.049e5
$\theta = \{\lambda_{\text{sos}}^\star\}$	-1.033e5	-1.037e5	-1.054e5	-1.054e5

Strategies

2. Search a 1d closed convex cone connecting a robust SN with a high-performing SN using a large m

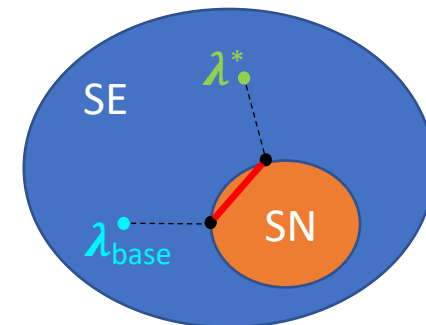
$$\text{Proj}(\lambda) = \{ \lambda : \phi_{\text{N}}(t(\delta; d); \mu, P) = \phi_{\text{E}}(t(\delta; 2d); \lambda) \}$$

$$\min_{\mu, P \in S_{++}^{n_z}} \{ r^{\top} r : \phi_{\text{E}}(t(\delta; 2d); \lambda) = \phi_{\text{E}}(t(\delta; 2d); r) + \phi_{\text{N}}(t(\delta; d); \mu, P) \}$$

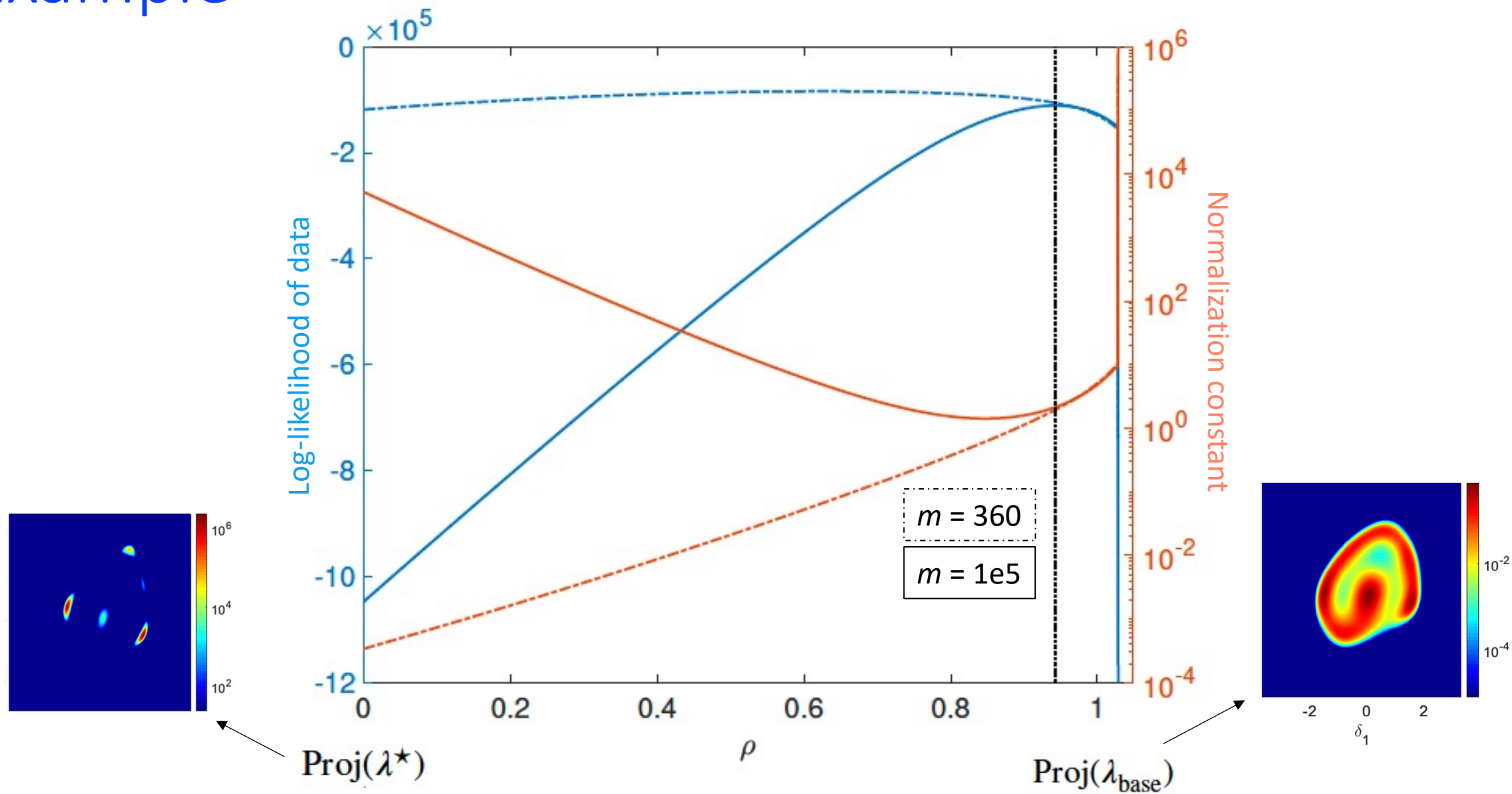
$$\Theta = \{ \lambda : \lambda \in \Lambda(\rho, \lambda^*, \lambda_{\text{base}}), \forall \rho \in [0, 1] \}$$

$$\Lambda(\rho, \lambda^*, \lambda_{\text{base}}) = (1 - \rho) \text{Proj}(\lambda^*) + \rho \text{Proj}(\lambda_{\text{base}})$$

$$\rho^* = \underset{\rho \in [0, 1]}{\text{argmin}} \left\{ n \log c_{\text{sam}}(\hat{m}, \lambda) + \frac{1}{2} \sum_{i=1}^n \phi_{\text{E}}(z^{(i)}; \lambda) : \lambda = \Lambda(\rho, \lambda^*, \lambda_{\text{base}}) \right\}$$

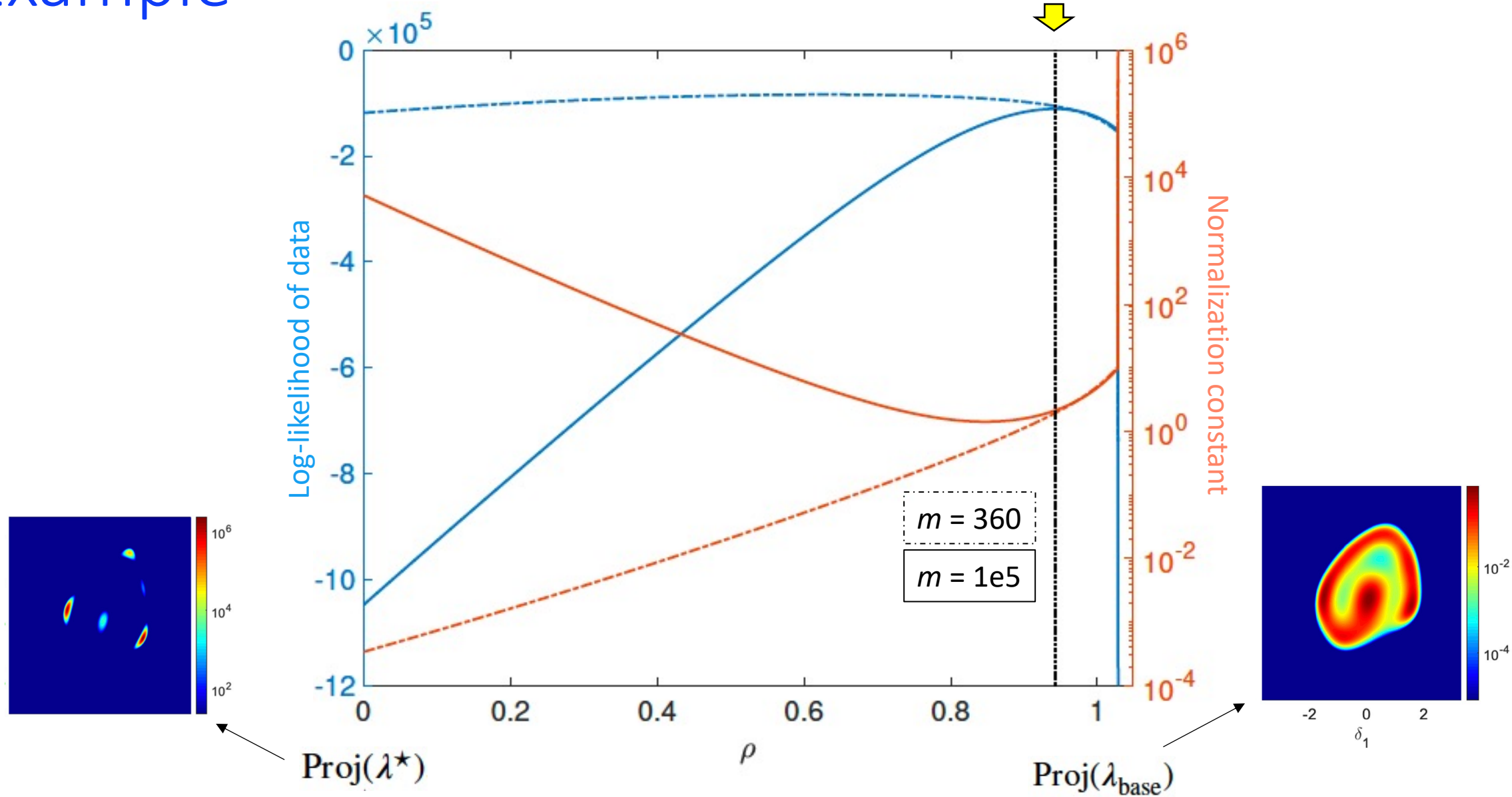


Example



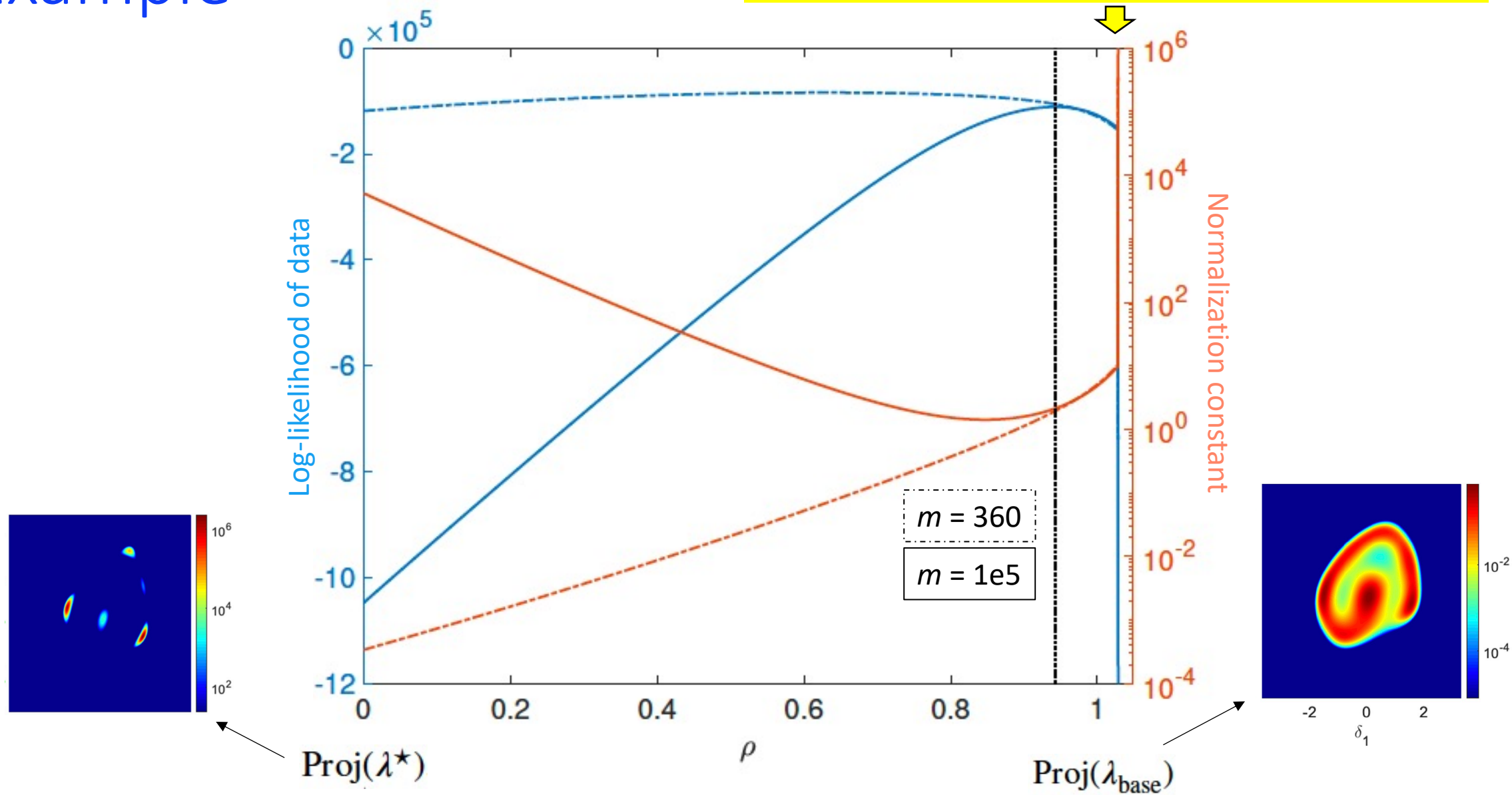
Example

Resulting distribution outperforms both SNs



Example

There is a diverging SE closer to the baseline than the optimal



Robust Estimation of Sliced-Exponential Distributions

Luis G. Crespo, Brendon K. Colbert, Tanner Slagel, Sean P. Kenny

NASA Langley Research Center

60th IEEE CDC

