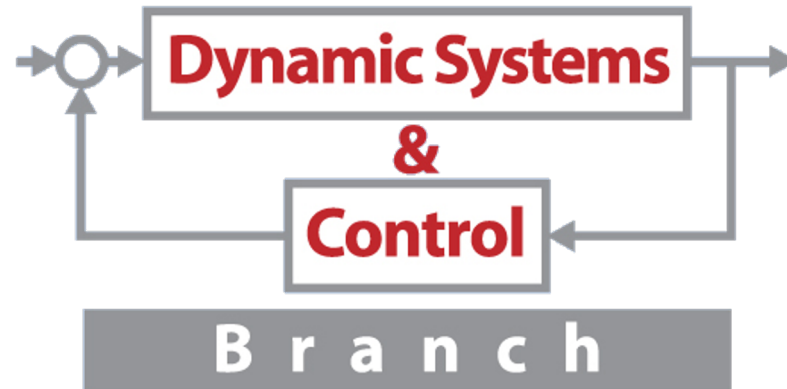


Dimensionality Reduction of Sliced-Normal Distributions

Luis G. Crespo, Brendon K. Colbert, Sean P. Kenny, Daniel P. Giesy
NASA Langley Research Center

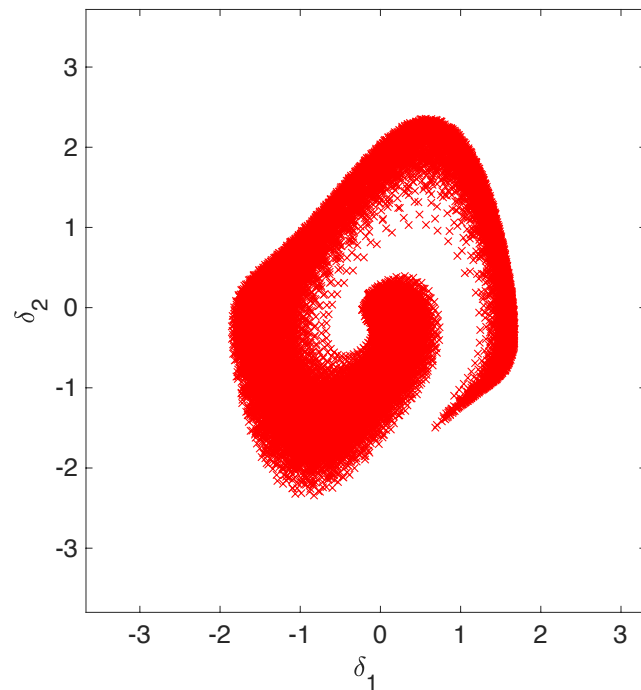


Motivation

- Sliced-Normal (SN) distributions enable the characterization of complex multivariate data as a vector of possibly dependent random variables or as a tightly enclosing semi-algebraic set

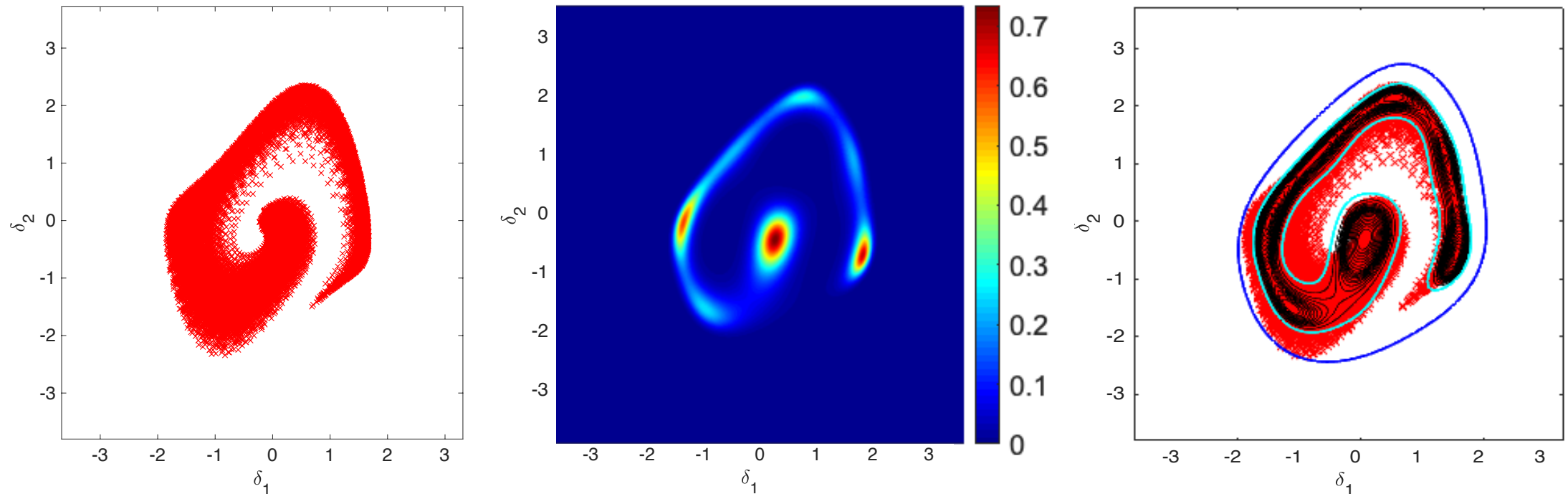
Motivation

- Sliced-Normal (SN) distributions enable the characterization of complex multivariate data as a vector of possibly dependent random variables or as a tightly enclosing semi-algebraic set

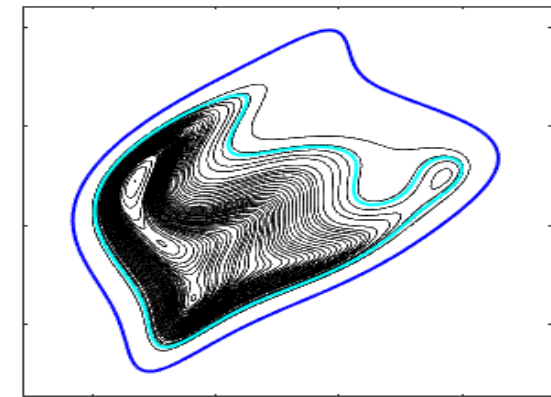
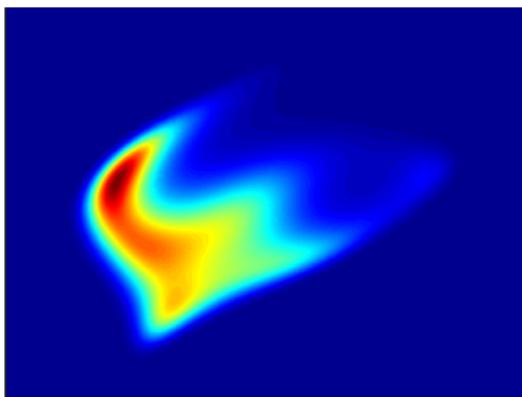
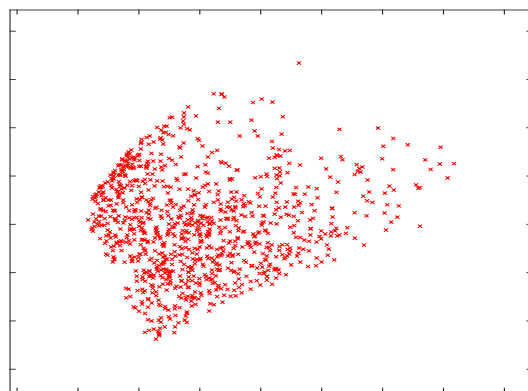
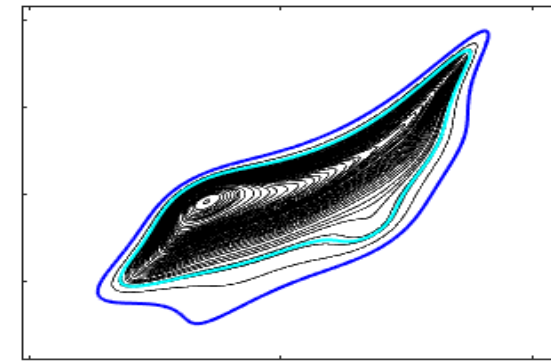
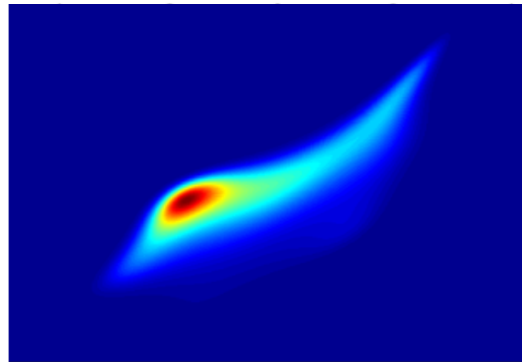
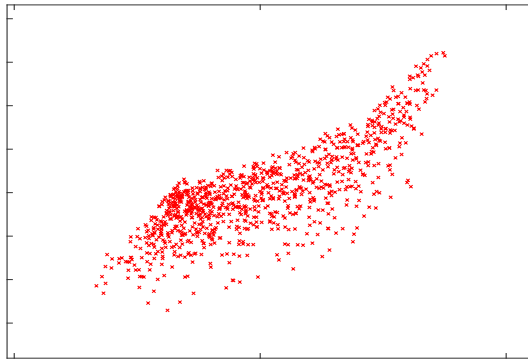
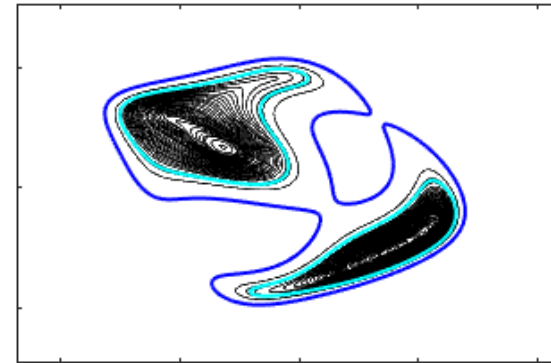
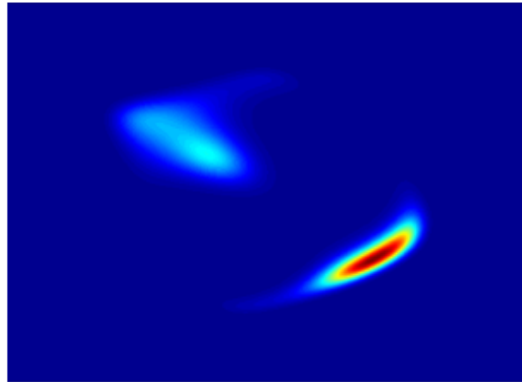
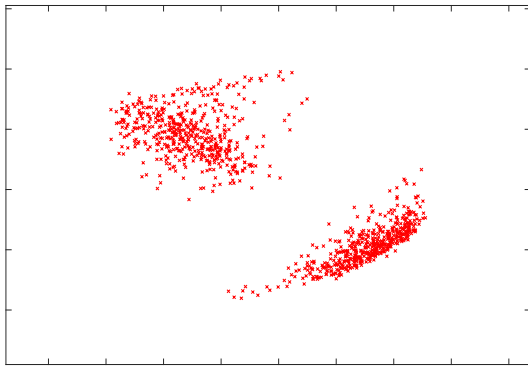


Motivation

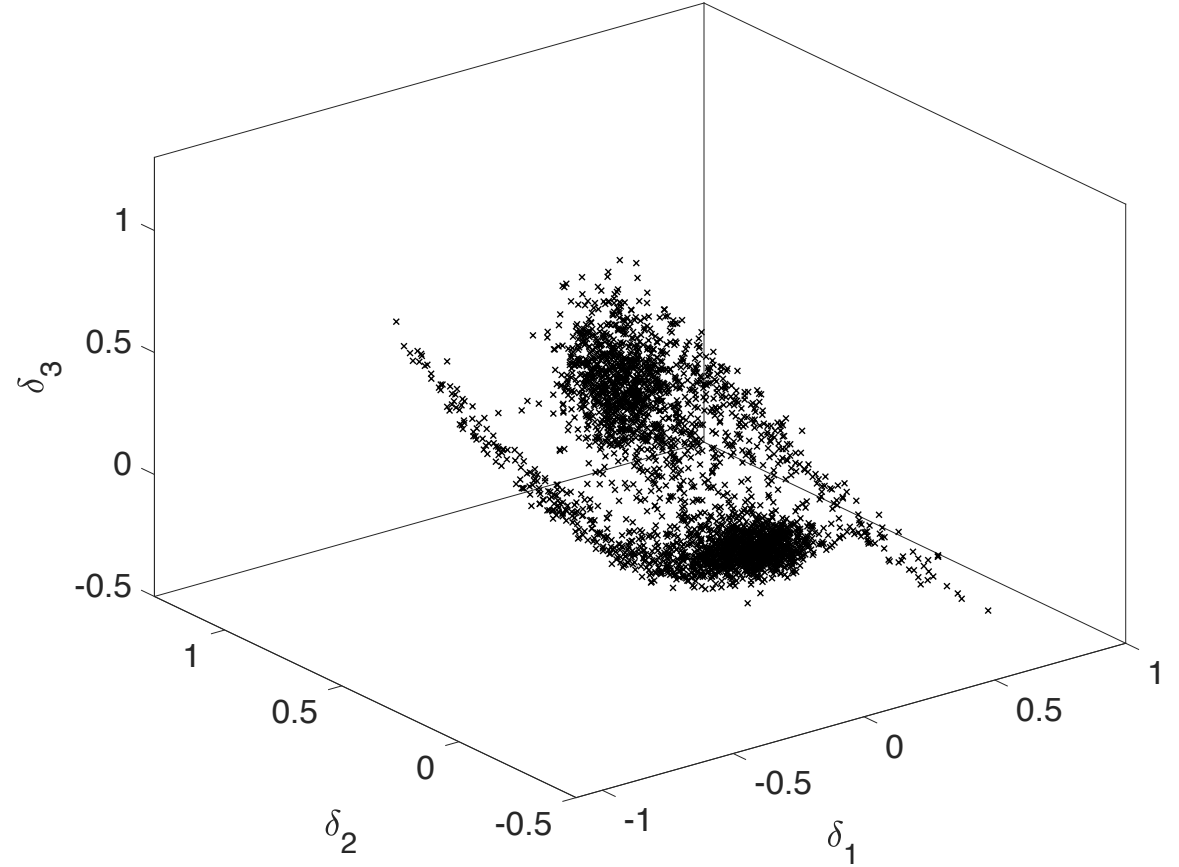
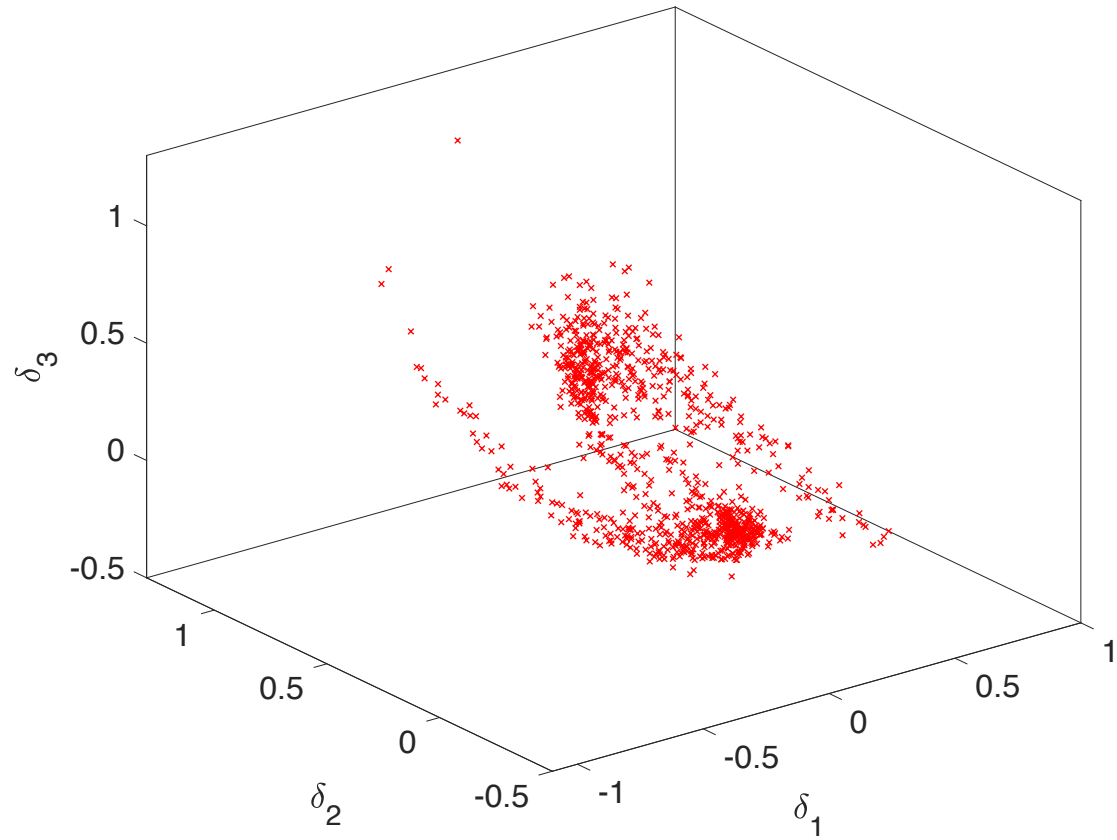
- Sliced-Normal (SN) distributions enable the characterization of complex multivariate data as a vector of possibly dependent random variables or as a tightly enclosing semi-algebraic set



Versatility 2d



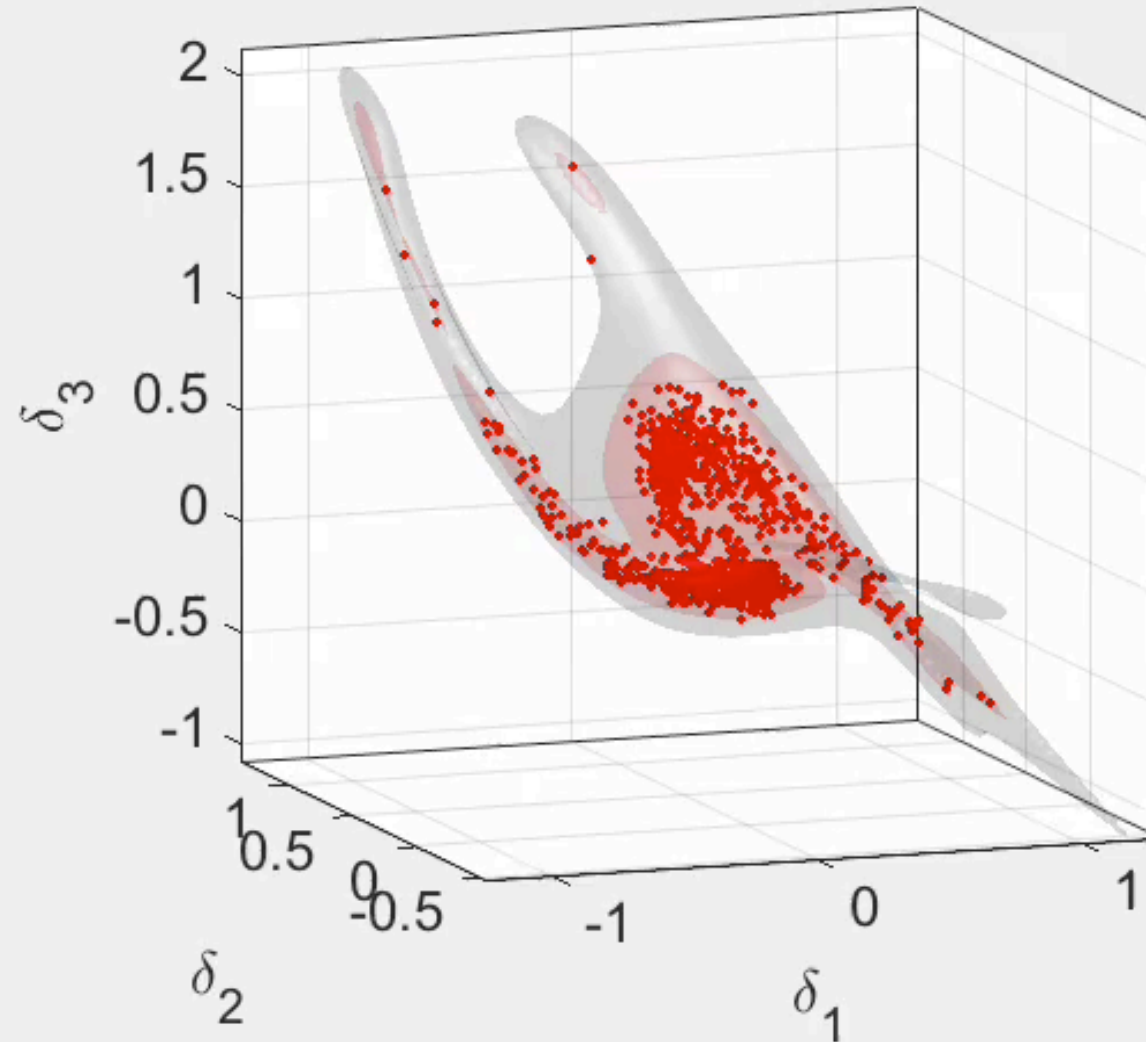
Versatility 3d



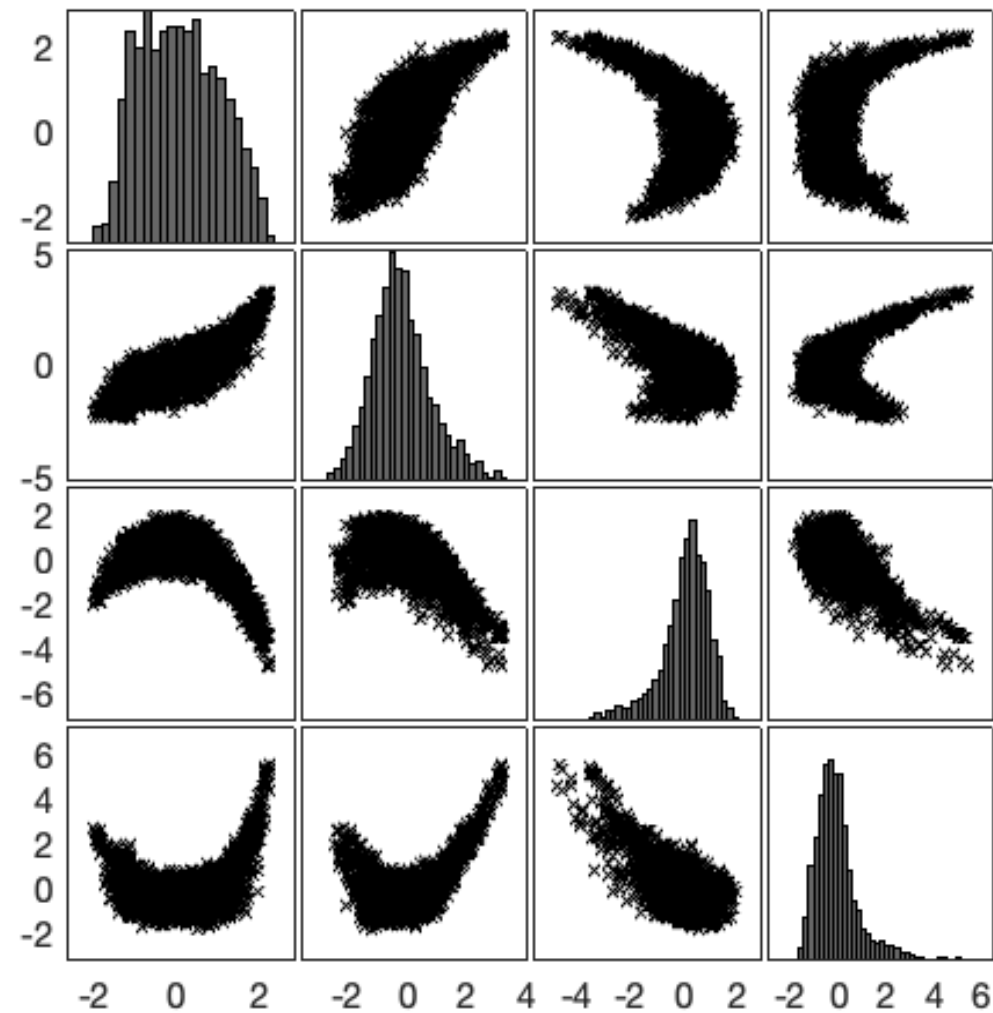
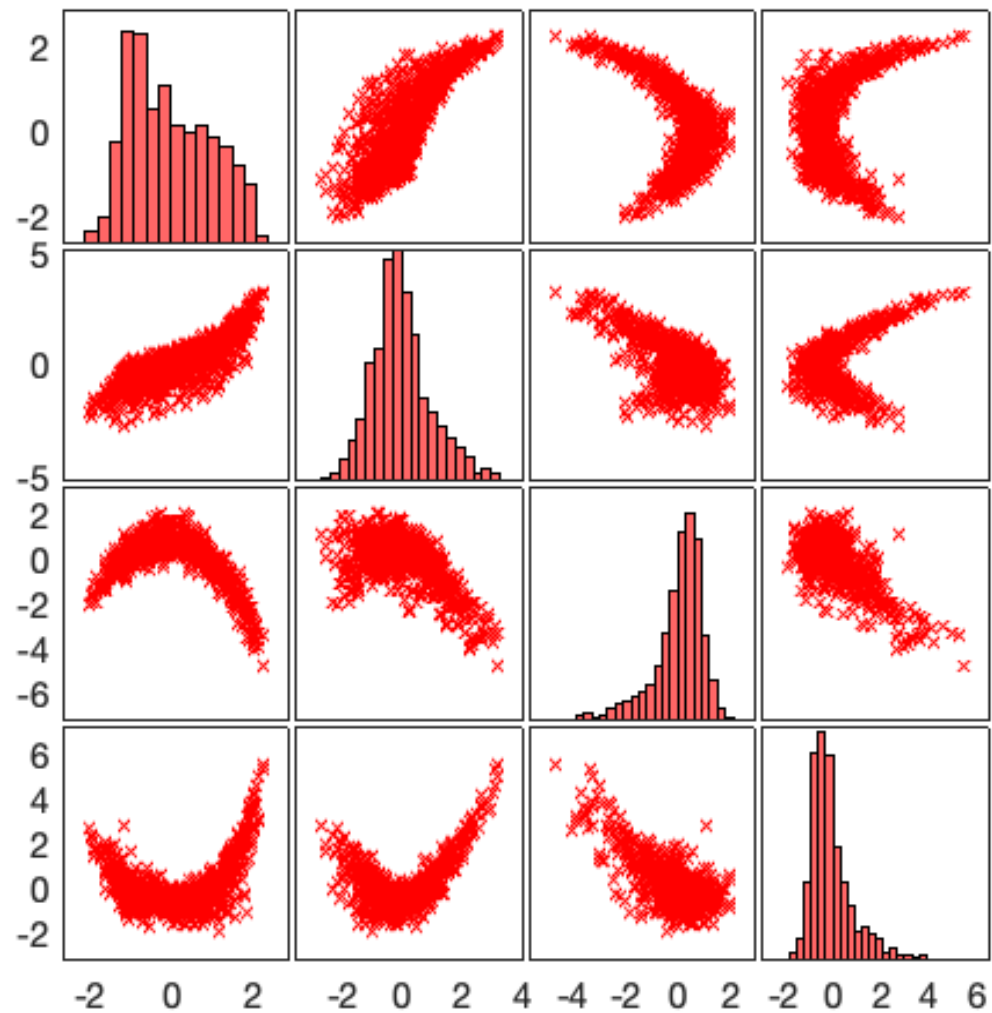
Versatility 3d

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

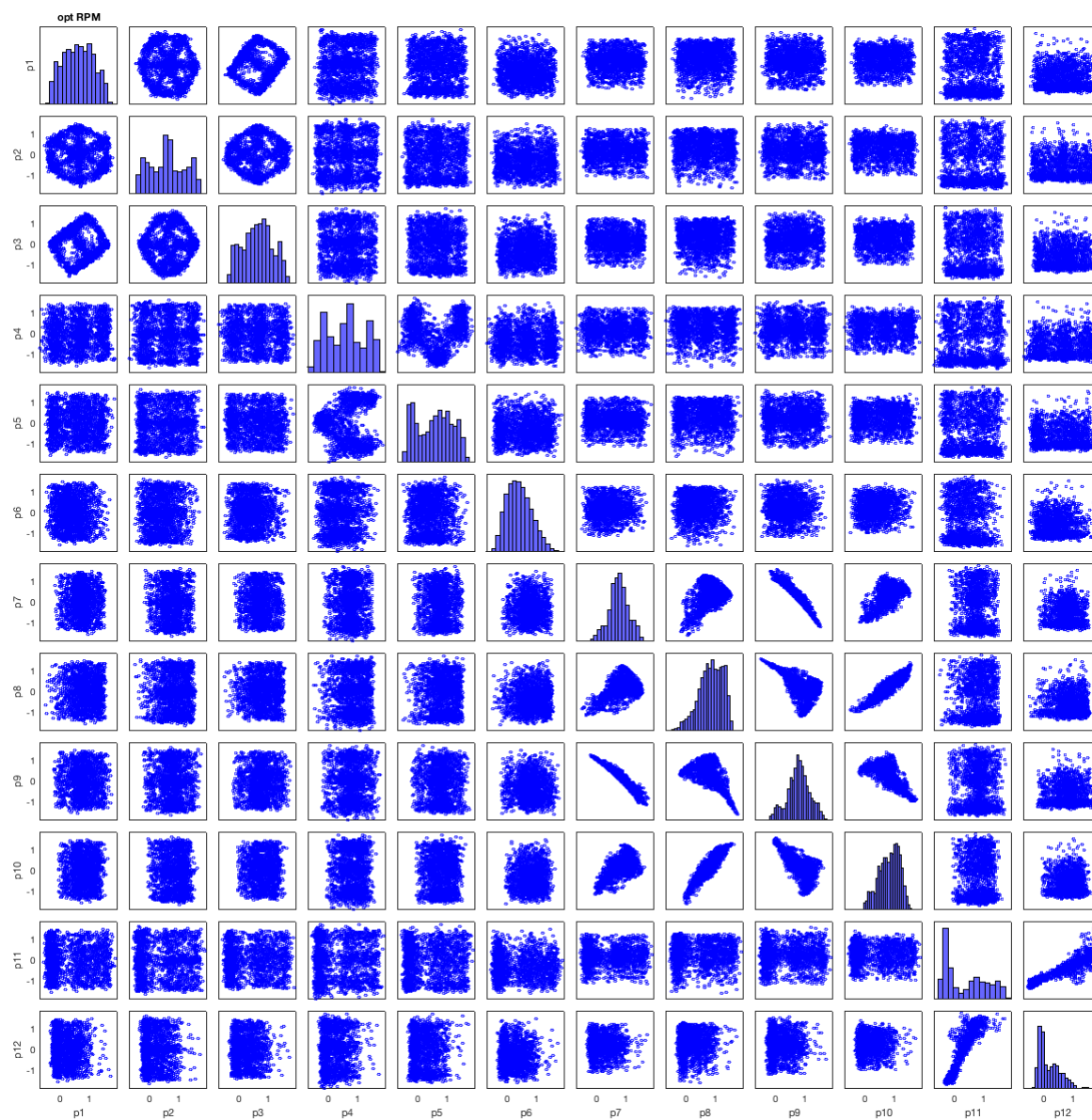
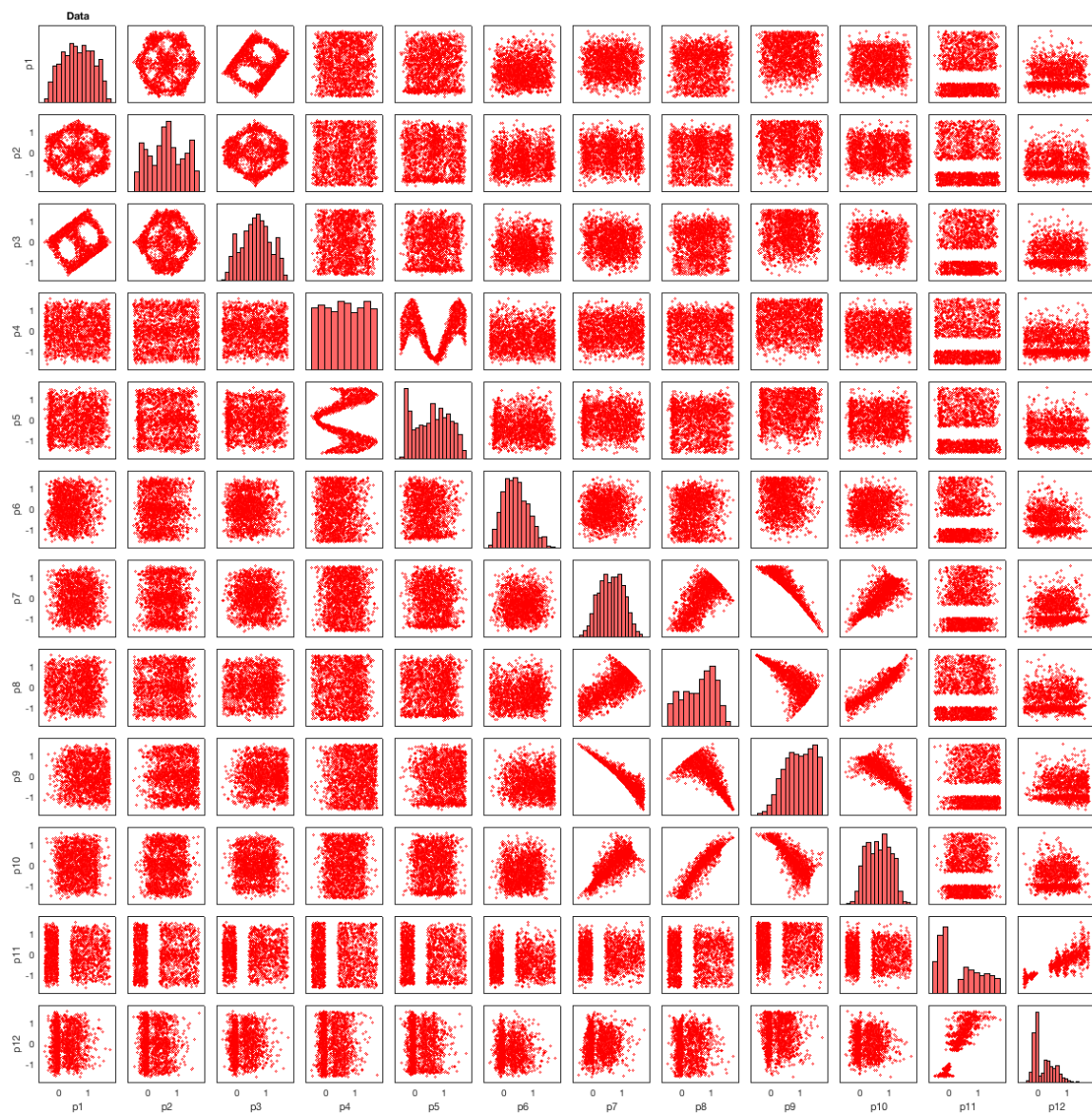
$$\text{Vol}_{95}=0.202$$



Versatility 4d

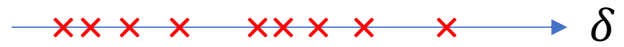


Versatility 12d

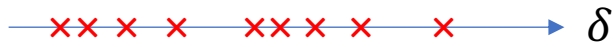
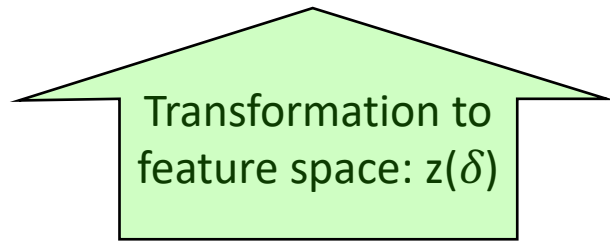
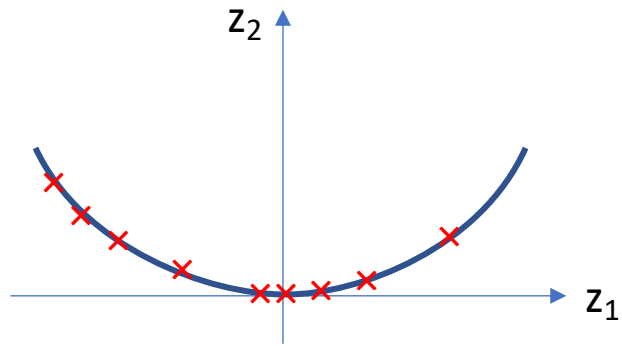


Sliced-Normal Distributions

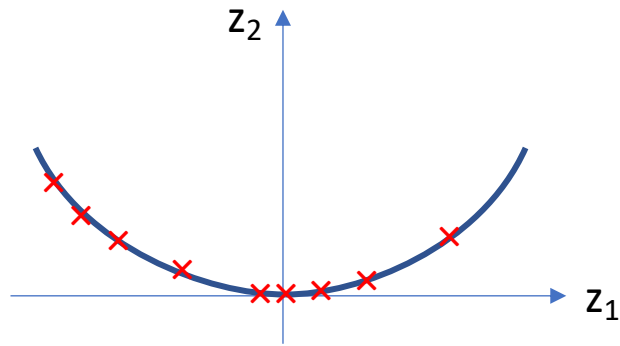
Given IID observations drawn from a
Data Generating Mechanism (**DGM**)



Sliced-Normal Distributions

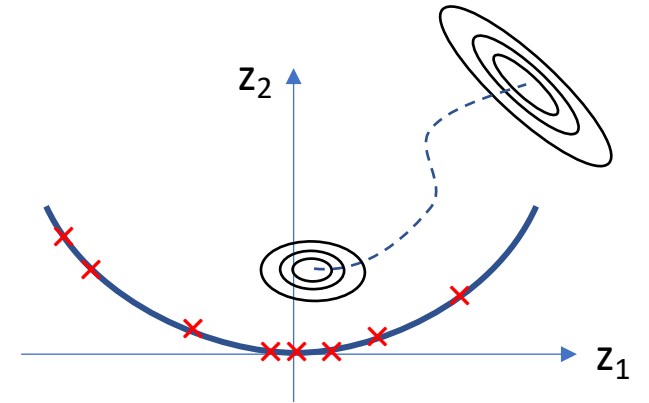


Sliced-Normal Distributions

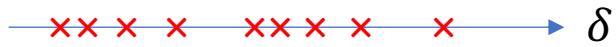


$$f_z \sim e^{-(z-\mu)^\top P(z-\mu)}$$

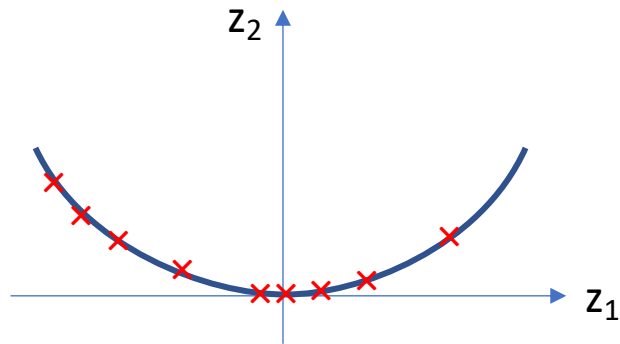
Search for an optimal Normal



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

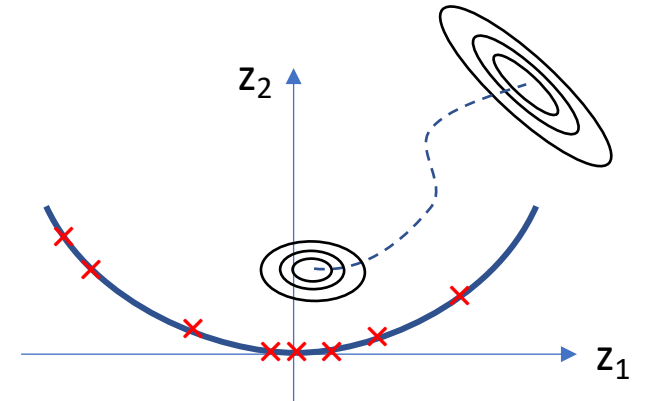


Sliced-Normal Distributions

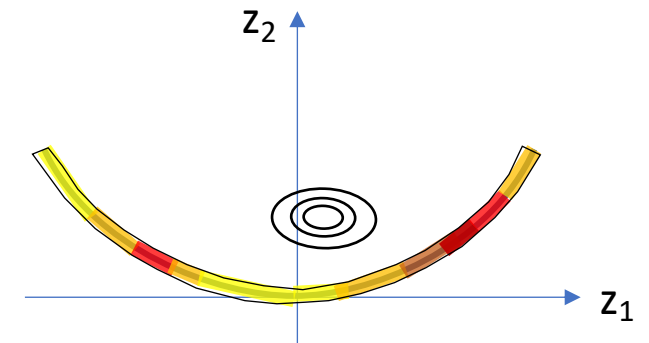


$$f_z \sim e^{-(z-\mu)^\top P(z-\mu)}$$

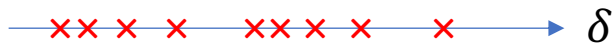
Search for an optimal Normal



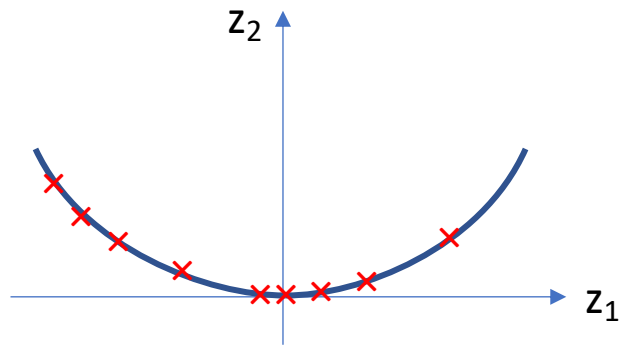
Evaluate Normal on manifold



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

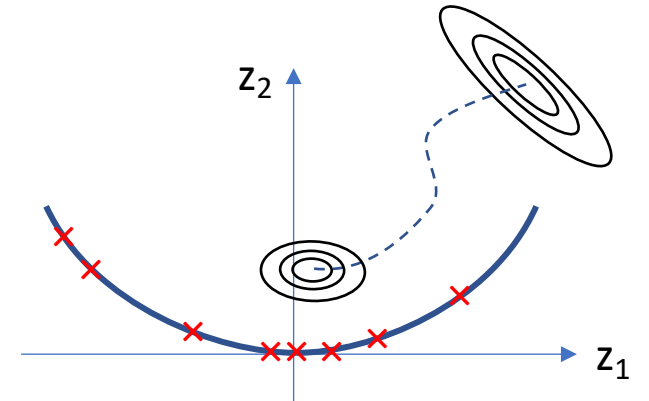


Sliced-Normal Distributions

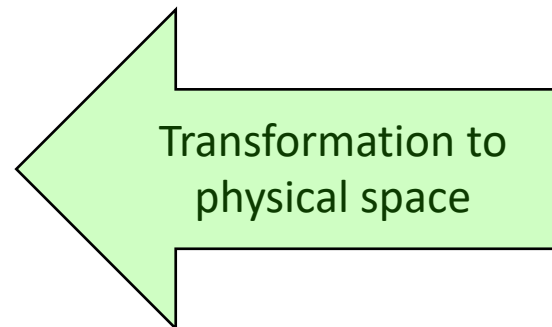
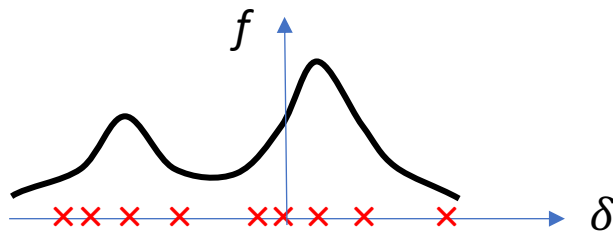
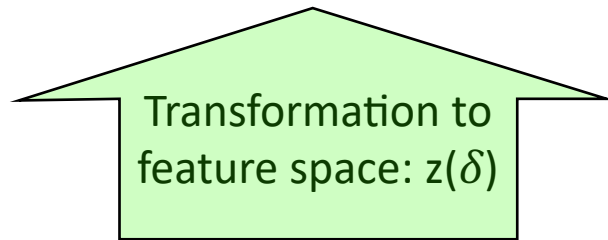
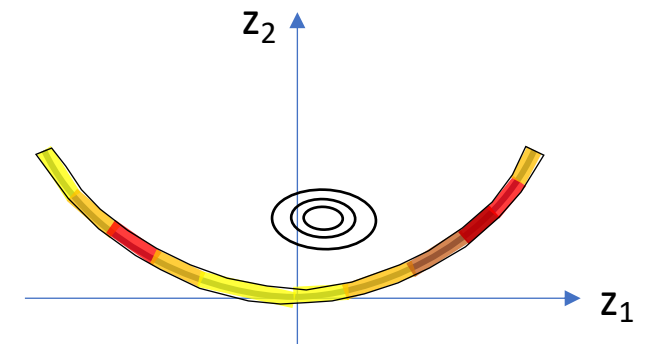


$$f_z \sim e^{-(z-\mu)^\top P(z-\mu)}$$

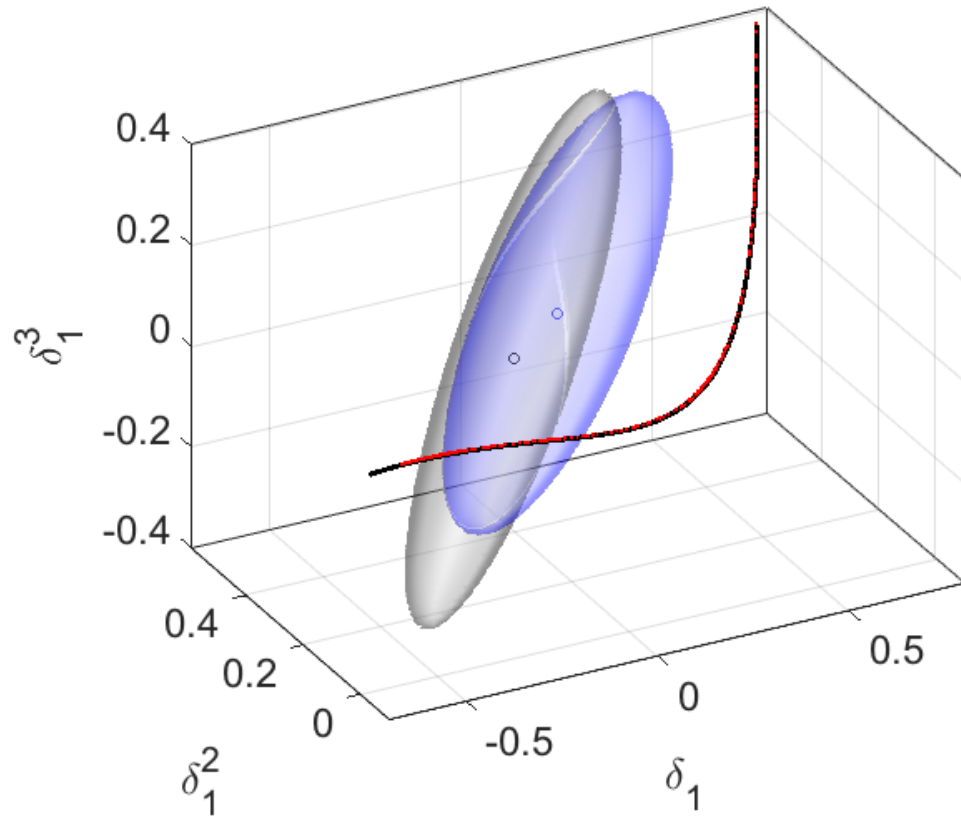
Search for an optimal Normal



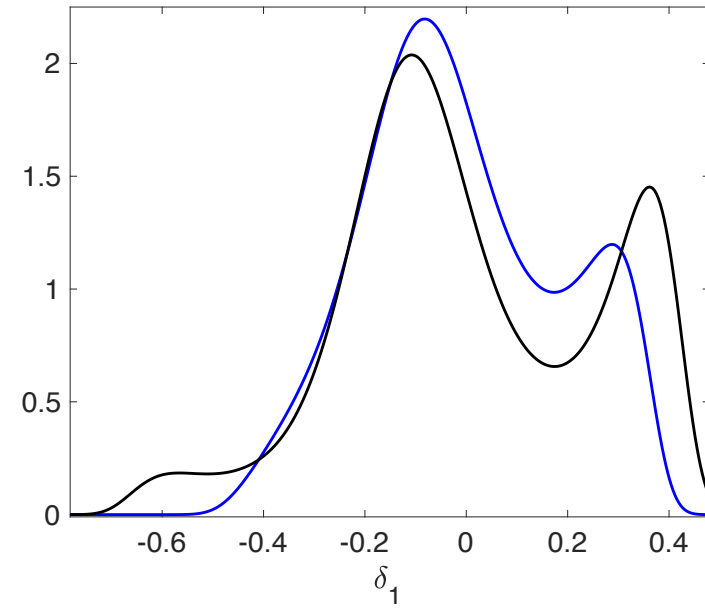
Evaluate Normal on manifold



Sliced-Normal Distributions



Ellipsoids in feature space



Corresponding densities

Sliced-Normal Distributions

$$z = Z(\delta, d)$$

$$\phi(z, \mu, P) = (z - \mu)^\top P(z - \mu),$$

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

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

Estimation of SNs from Data

- Analytical

- Estimation requires inverting a matrix
- Substandard performance
- Readily applicable to large dimensional problems

- Suboptimal

- Estimation requires inverting a matrix and solving a convex program in a single decision variable
- Readily applicable to large dimensional problems

- Optimal

- Estimation requires solving a “well-behaved” non-convex program
- As all SOS methods, the number of decision variables grows exponentially with the dimension of feature space, thereby restricting its application

Computational Complexity

Table 1. Dimension of feature space n_z .

	$d = 2$	$d = 3$	$d = 4$	$d = 5$
$n_\delta = 2$	5	9	14	20
$n_\delta = 5$	20	55	125	251
$n_\delta = 10$	65	285	1000	3002
$n_\delta = 15$	135	815	3875	15503

Table 2. Number of decision variables

	$d = 2$	$d = 3$	$d = 4$	$d = 5$
$n_\delta = 2$	20	54	119	230
$n_\delta = 5$	230	1595	8000	31877
$n_\delta = 10$	2210	41040	501500	4510505
$n_\delta = 15$	9315	333335	7513625	120194759

$$n_z = \binom{n_\delta + d}{n_\delta} - 1$$

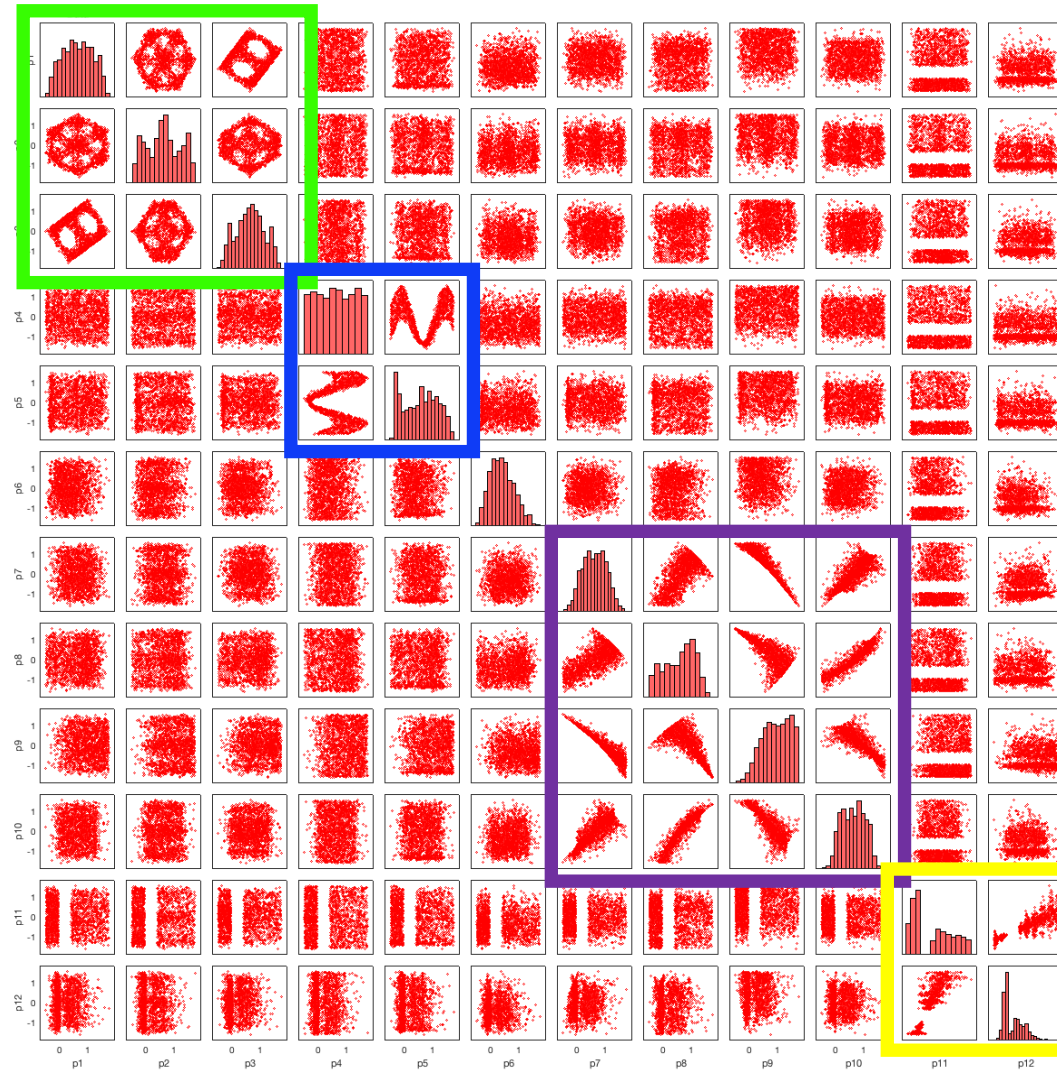
$$n_\theta = (n_z^2 + 3n_z)/2$$

- We seek to lower the computational complexity of SN estimation

Strategies Proposed

1. Dependency analysis: Identify subsets of highly interdependent parameters and estimate a SN for each of subset independently from the rest

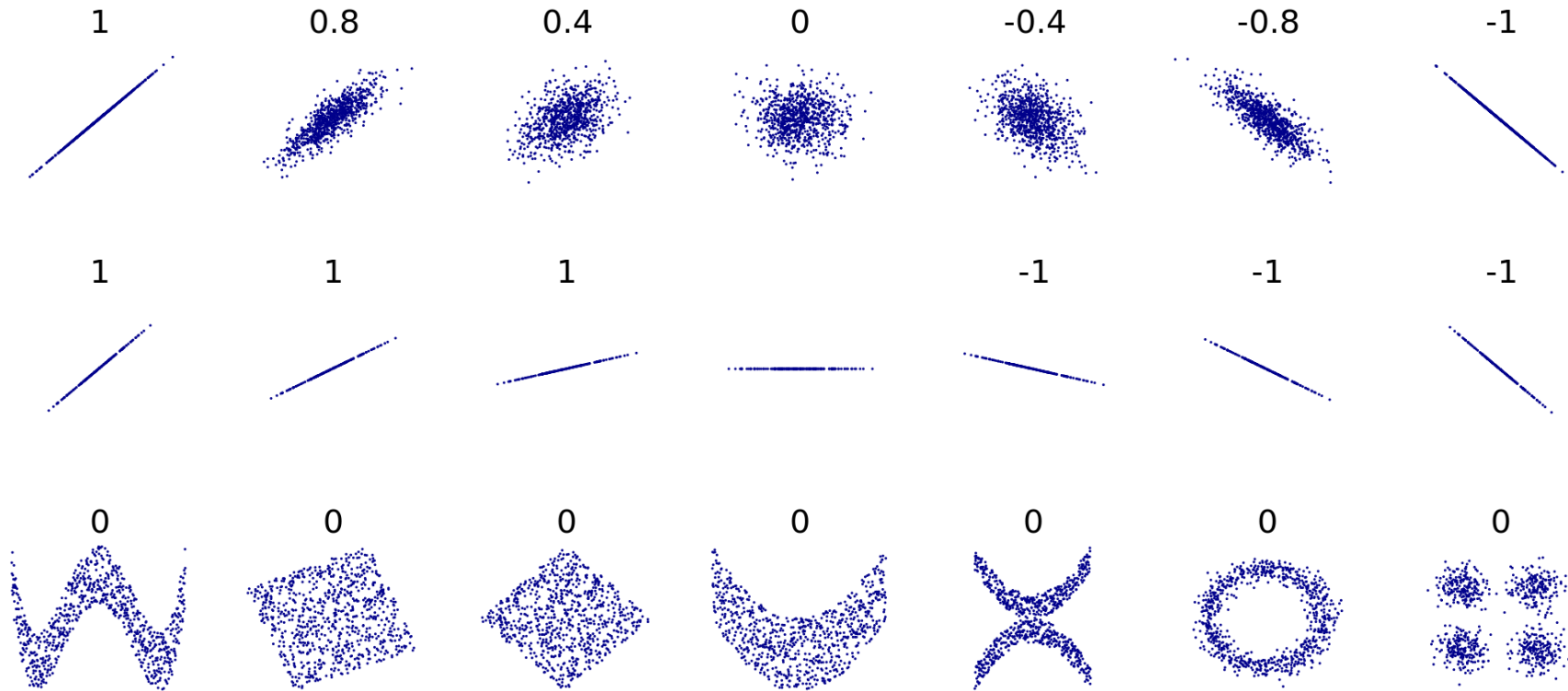
Dependency Analysis



- Enables using different levels of complexity when modeling each subgroup

Dependency Analysis

- Inadequate approaches
 - Correlation coefficient based



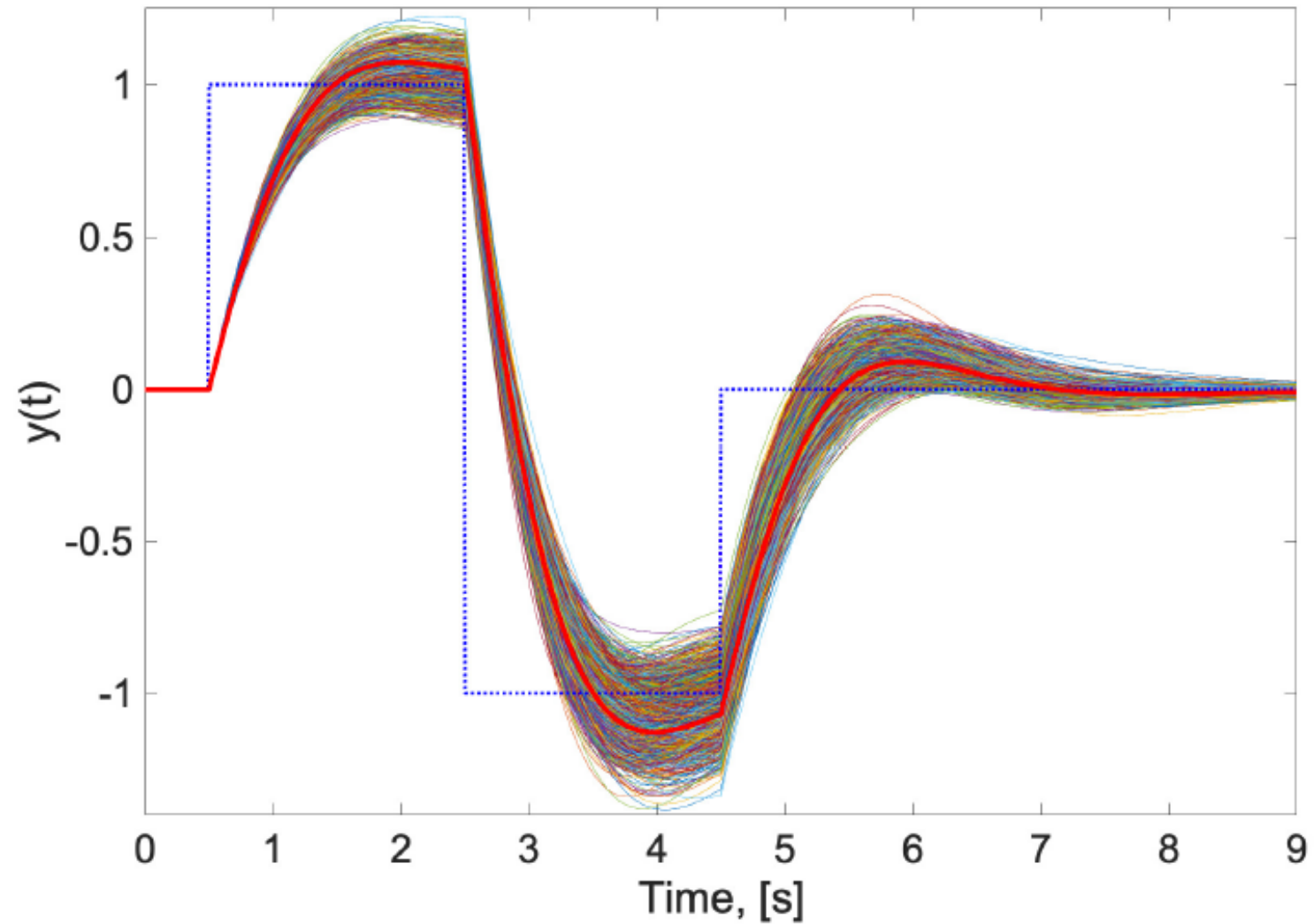
Dependency Analysis

- Inadequate approaches
 - Correlation coefficient based
 - Postprocessing a fitted distribution

Dependency Analysis Framework

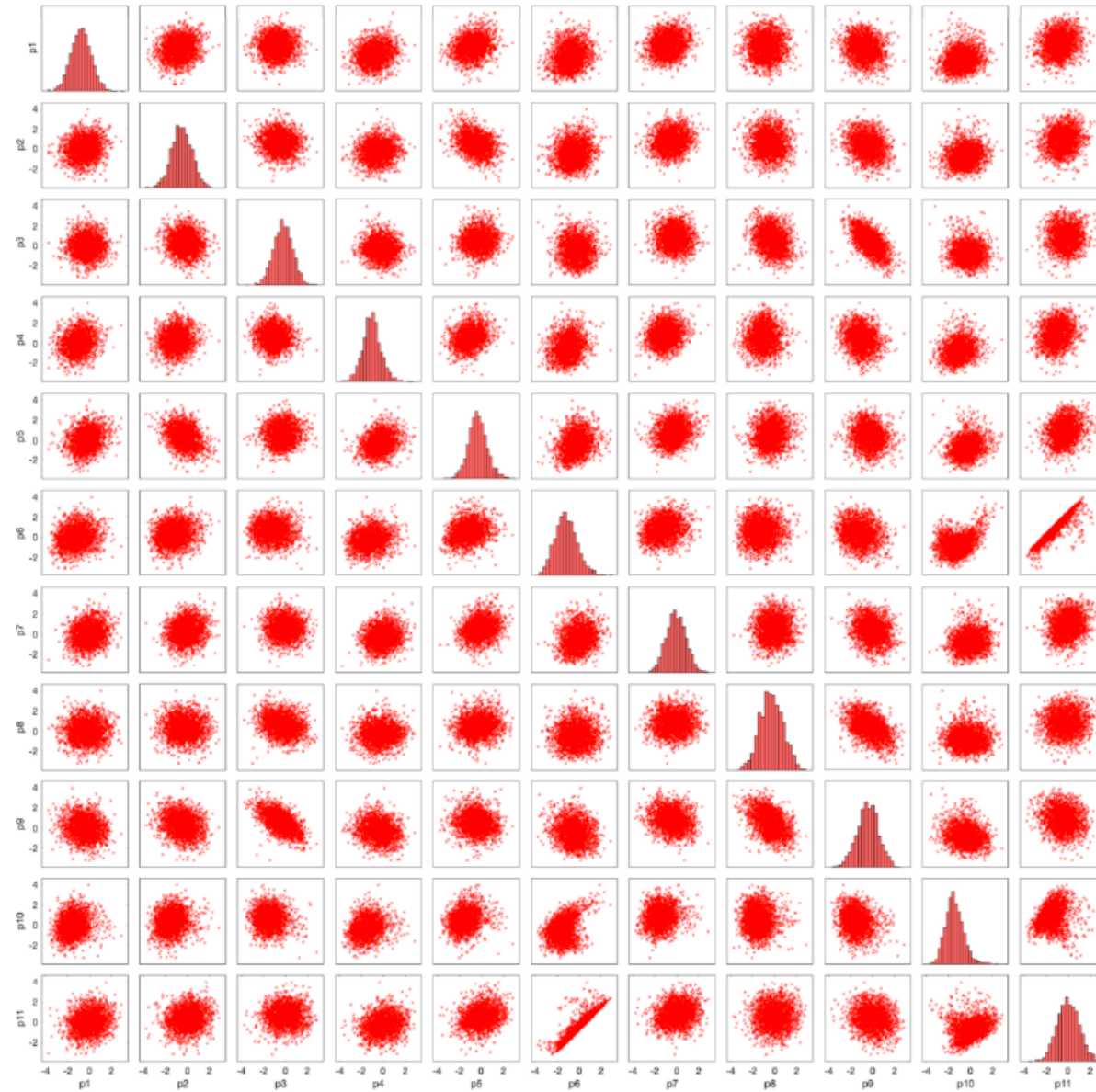
- Rationale: evaluate the offset between the empirical copula and a copula for independent parameters
- The degree of multivariate dependency between the elements of two groups is given by the largest degree of bivariate dependency among all pairs
- It is data-based and distribution-free
- It can be applied outside the context of the SNs
- It only requires algebraic manipulations

Dependency Analysis Example

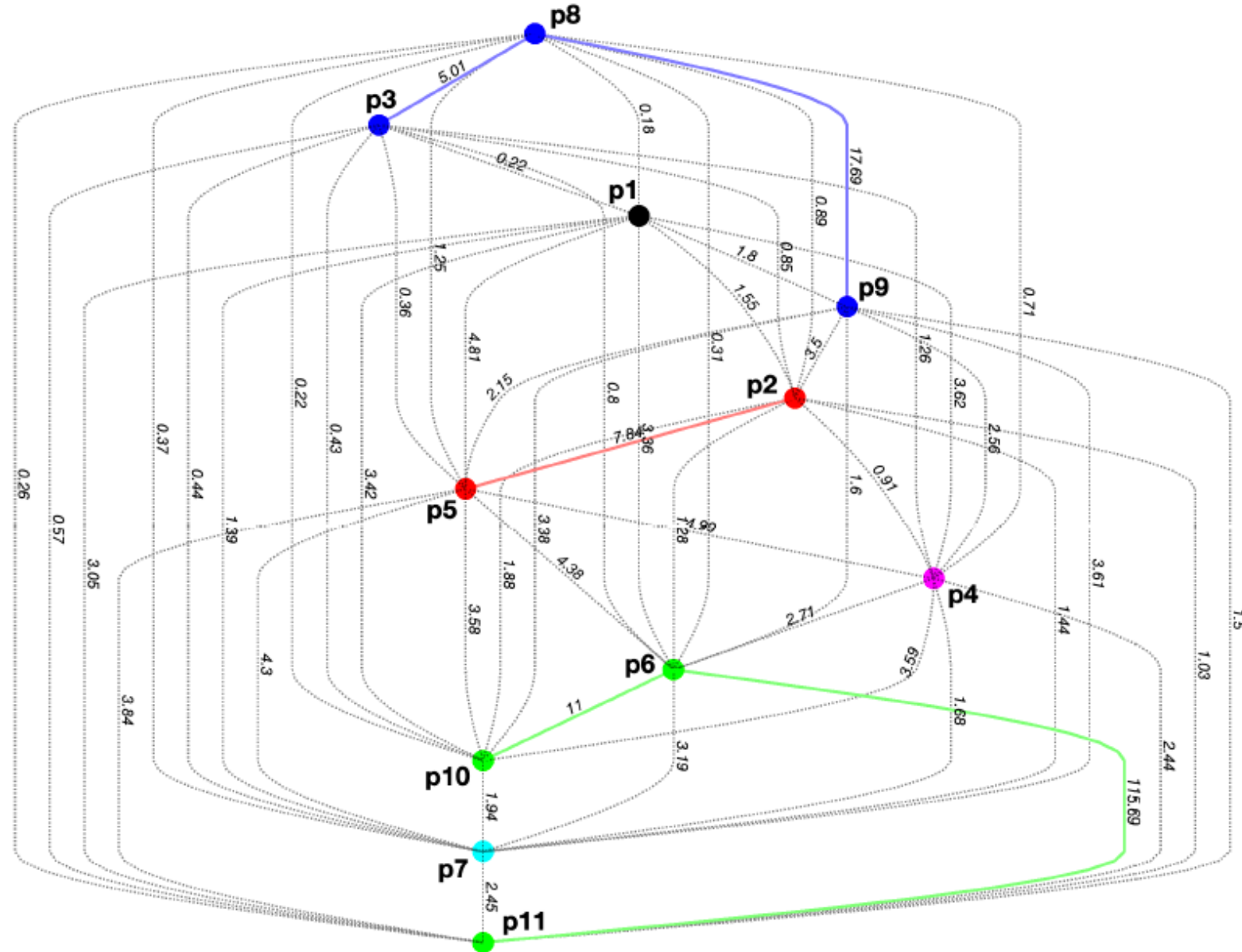


Dependency Analysis Example

$$H(s) = \frac{p_1 s^4 + p_2 s^3 + p_3 s^2 + p_4 s + p_5}{p_6 s^5 + p_7 s^4 + p_8 s^3 + p_9 s^2 + p_{10} s + p_{11}}$$



Dependency Analysis Example



Dependency Analysis Example

- Reduction of 99.96% in the number of decision variables for degree=5: one SN with 9541895 decision variables vs. 6 SNs with 1595, 20, 230, 20, 1595 and 20 decision variables
- Sampling the combination of independent SNs yields a sample set that is indistinguishable from the data

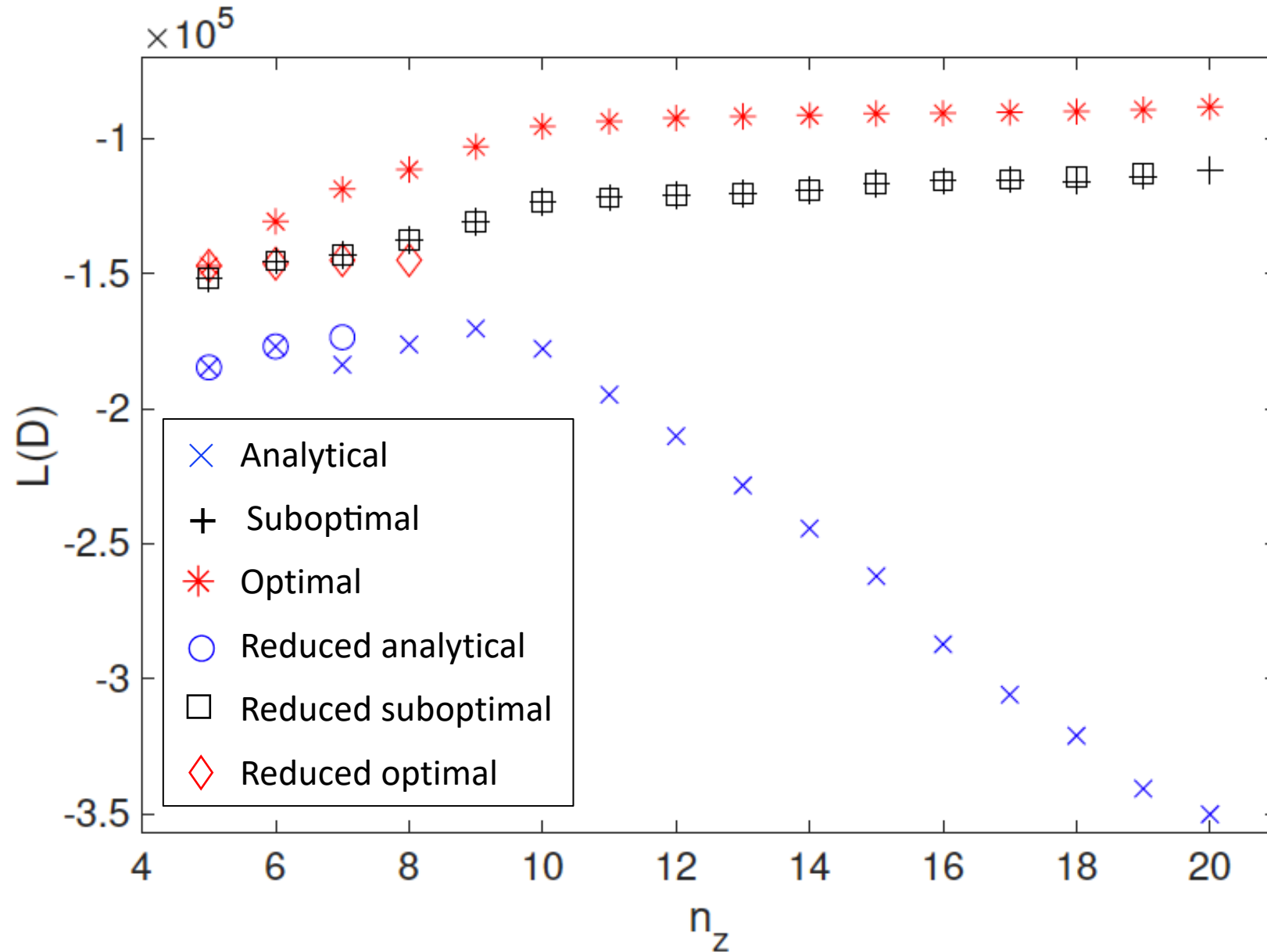
Strategies Proposed

1. Dependency analysis: Identify subsets of highly interdependent parameters and estimate a SN for each of subset independently from the rest
2. Mapping reduction: Sequentially augment the dimension of feature space by only retaining monomials that significantly increase the likelihood of the data

Mapping Reduction

- Approach uses the Schur complement to generate a higher-dimensional P from a lower-dimensional P
- Monomials that don't increase the likelihood of the data significantly are eliminated from the mapping
- Augmenting by one or by many
- Resulting SN is suboptimal

Mapping Reduction Example



Concluding Remarks

- SNs are versatile, require minimal modeling effort, and enable the analyst to evaluate and enforce desirable properties on the resulting model
- The polynomial structure of the data-enclosing sets is amenable to rigorous UQ analysis
- Applications to uncertainty quantification, model calibration, system identification and robust control
- The strategies proposed facilitate the estimation of SNs for higher-dimensional datasets

Dimensionality Reduction of Sliced-Normal Distributions

Luis G. Crespo, Brendon K. Colbert, Sean P. Kenny, Daniel P. Giesy
NASA Langley Research Center

