

Intrinsic variables, in compressed sensing

R. Coifman A. Singer

Yale, Princeton.

- Extraction of intrinsic calibrated information from random compressed sensing measurements can be achieved through integration of local whitened variables.
- Intrinsic variables , slow variables and independent component analysis.
- Examples , Hyperspectral data universal calibrations, black box non linear ICA , nonlinear chemical reactions with slow variables
- Extensions of coordinates beyond the data

Remark

It is quite obvious that the Mahalanobis distance between data points in Euclidean space is invariant under affine transformations of the space . Specifically given a cloud of data points, it can be whitened by diagonalizing the covariance matrix to provide new coordinates , followed by a rescaling of each coordinate to have variance one .

The Mahalanobis distance is the distance in this system .

Observe that if the data lies in a subspace of dimension d of a high dimensional space , we can project it through a random projection on d random orthogonal vectors, since the projection is an affine transformation from the original space onto the random subspace the Mahalanobis distance between data points remains the same .

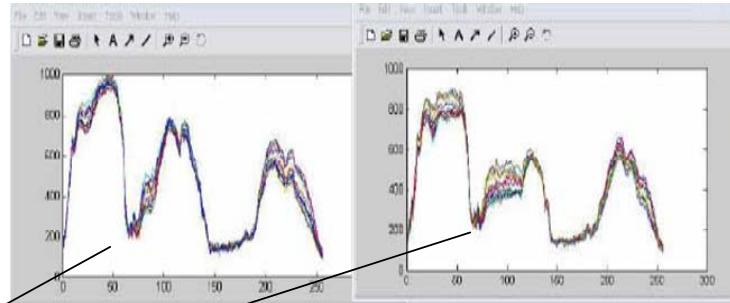
Therefore any reconstruction of the original data which relies only on distances between points ,can be done in compressed sensing domain.

Moreover as we will see under simple assumptions the mutual local Mahlanobis distances between points can be used to parametrize and represent complex data based on its local structure , These parameters are invariant under general nonlinear tranformations.

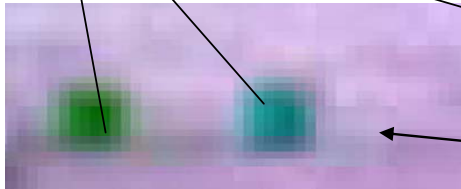
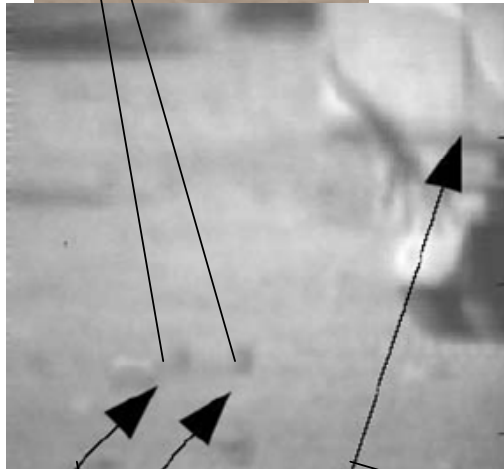
As an example consider a hyperspectral image for which we collect 20 randomized spectral bands (knowing the data depends only on say 15 independent ingredients or independent parameters) .

We claim that this can be used to obtain an image where each pixel is annotated by the concentrations of the 15 ingredients .

Moreover the result is independent of the nonlinear distortions of the instrument and choice of randomized bands.



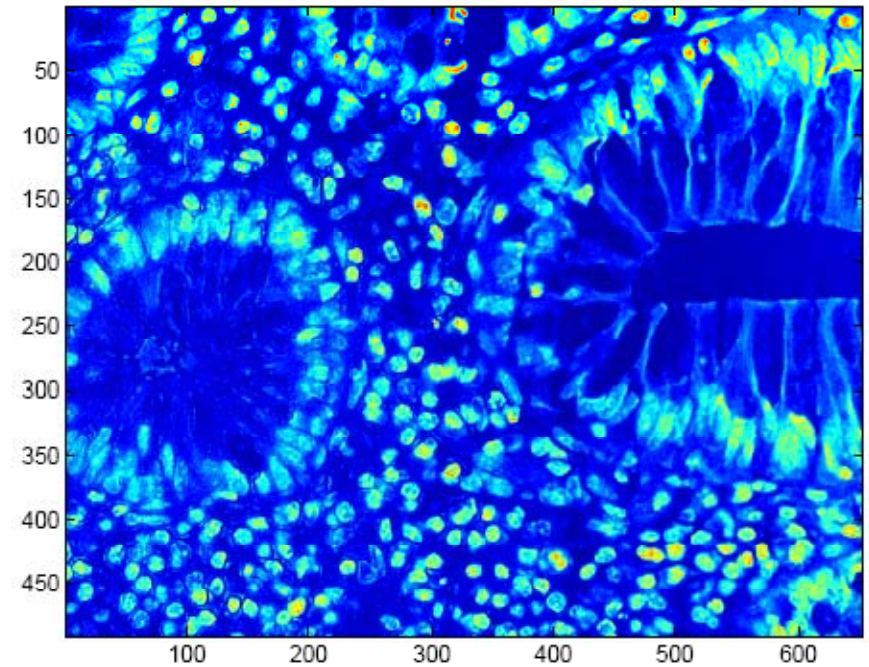
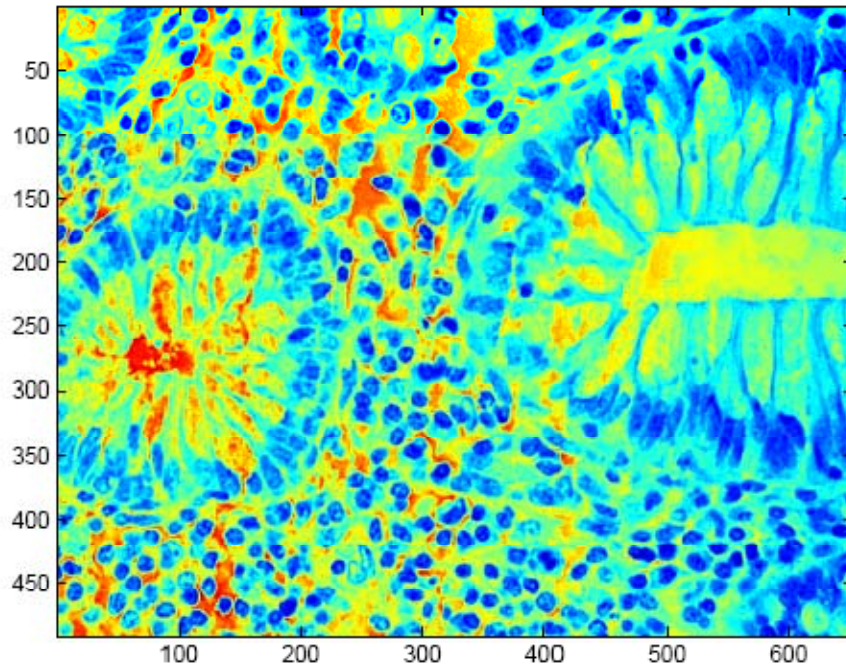
Experimental setup for concept validation



- Coded aperture digital hyperspectral camera can measure spectrum through pseudo random correlation for efficient throughput and compressed sensing
- Camera can do direct measure of spatio spectral features
- The images were taken at a China Lake two bottles containing different gases have different spectral responses shown above .
- Three spatio spatial features were measured directly and displayed in red green blue, below.

truck

Three top discrimination features displayed in rgb



The image on the left is an rgb representation of the encoded measurements, on the right they have been organized to by diffusion geometry to provide intrinsic biological parameters quantifying tissue constituents . The image on the right is independent of the selection of spectral encodings .

The basic model provides for an intrinsic organization of data driven by a stochastic process or by a black box f which is a nonlinear mapping from an unknown parameter space (independent parameters) into high dimensions, we assume that f is smooth and invertible on its range.

The goal is to extract a parametrization which is independent of f or the mode of observations.

We achieve this by locally undistorting the observations using the inverse of the square Jacobian of f , as computed through the local covariance matrix of the data

Asymptotically the local covariance of the data is computable through the Jacobian of f , this information suffices to construct the inverse map specifically:

The basic idea is that any smooth map between smooth manifolds $f : \mathcal{M}_X \mapsto \mathcal{M}_Y$ can be linearly approximated in a local neighborhood of any given point by its differential. The first-order Taylor expansion near x_0 reads

$$y = f(x) = y_0 + J_f(x_0)(x - x_0) + O(\|x - x_0\|^2), \quad (38)$$

where $J_f(x_0)$ is the Jacobian of f at x_0 and $y_0 = f(x_0)$.³ This gives a first-order approximation for the distances

$$\|y - y_0\|^2 = \|J_f(x_0)(x - x_0)\|^2 + O(\|x - x_0\|^3). \quad (39)$$

Similarly, for the inverse map $f^{-1} : \mathcal{M}_Y \mapsto \mathcal{M}_X$ we have

$$\|x - x_0\|^2 = \|J_{f^{-1}}(y_0)(y - y_0)\|^2 + O(\|x - x_0\|^3), \quad (40)$$

Nonlinear independent components Analysis and the discrete graph laplacian L

$$L = D^{-1}W - I,$$

$$W_{ij} = \exp \left\{ -\frac{\|J^{-1}(\mathbf{y}^{(i)})(\mathbf{y}^{(j)} - \mathbf{y}^{(i)})\|^2 + \|J^{-1}(\mathbf{y}^{(j)})(\mathbf{y}^{(j)} - \mathbf{y}^{(i)})\|^2}{4\epsilon} \right\}$$

$$\frac{1}{N} \sum_{j=1}^N W_{ij} q_Y(\mathbf{y}^{(j)}) \rightarrow \tag{4.2}$$

$$\int_Y \exp \left\{ -\frac{\|J^{-1}(\mathbf{y}^{(i)})(\mathbf{y} - \mathbf{y}^{(i)})\|^2 + \|J^{-1}(\mathbf{y})(\mathbf{y} - \mathbf{y}^{(i)})\|^2}{4\epsilon} \right\} q_Y(\mathbf{y}) p_Y(\mathbf{y}) d\mathbf{y},$$

Leading to $\mathcal{L}q = \Delta q - \nabla U \cdot \nabla q,$

This Fokker Plank operator in the \mathbf{x} coordinates ,is a sum of independent operators in each variable whose first eigenvectors are monotone in the \mathbf{x} variables, these are the intrinsic variables .

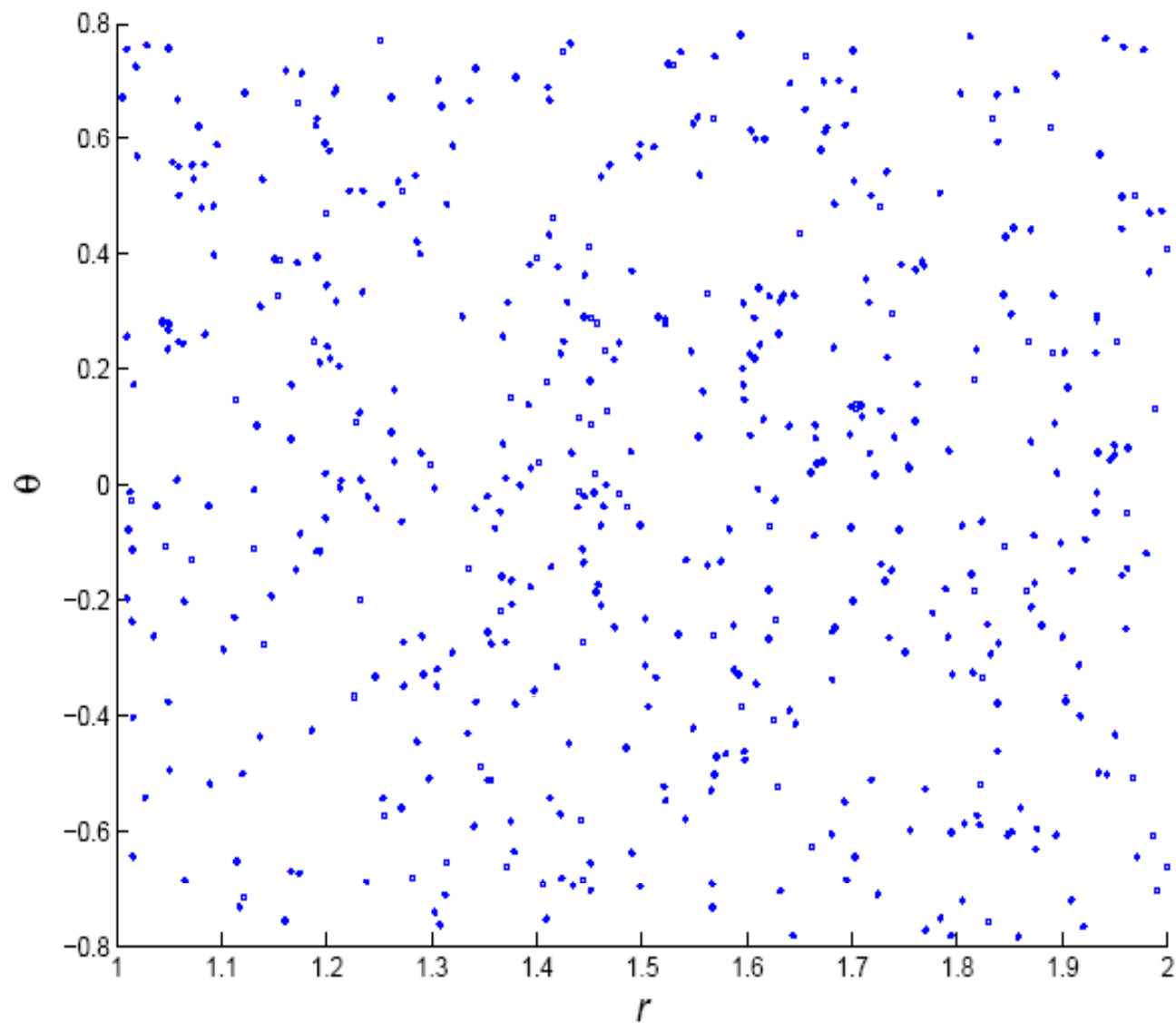


Figure 1: Data sample of $N = 500$ points drawn randomly from the stationary distributions of r (5.1) and θ (5.2).

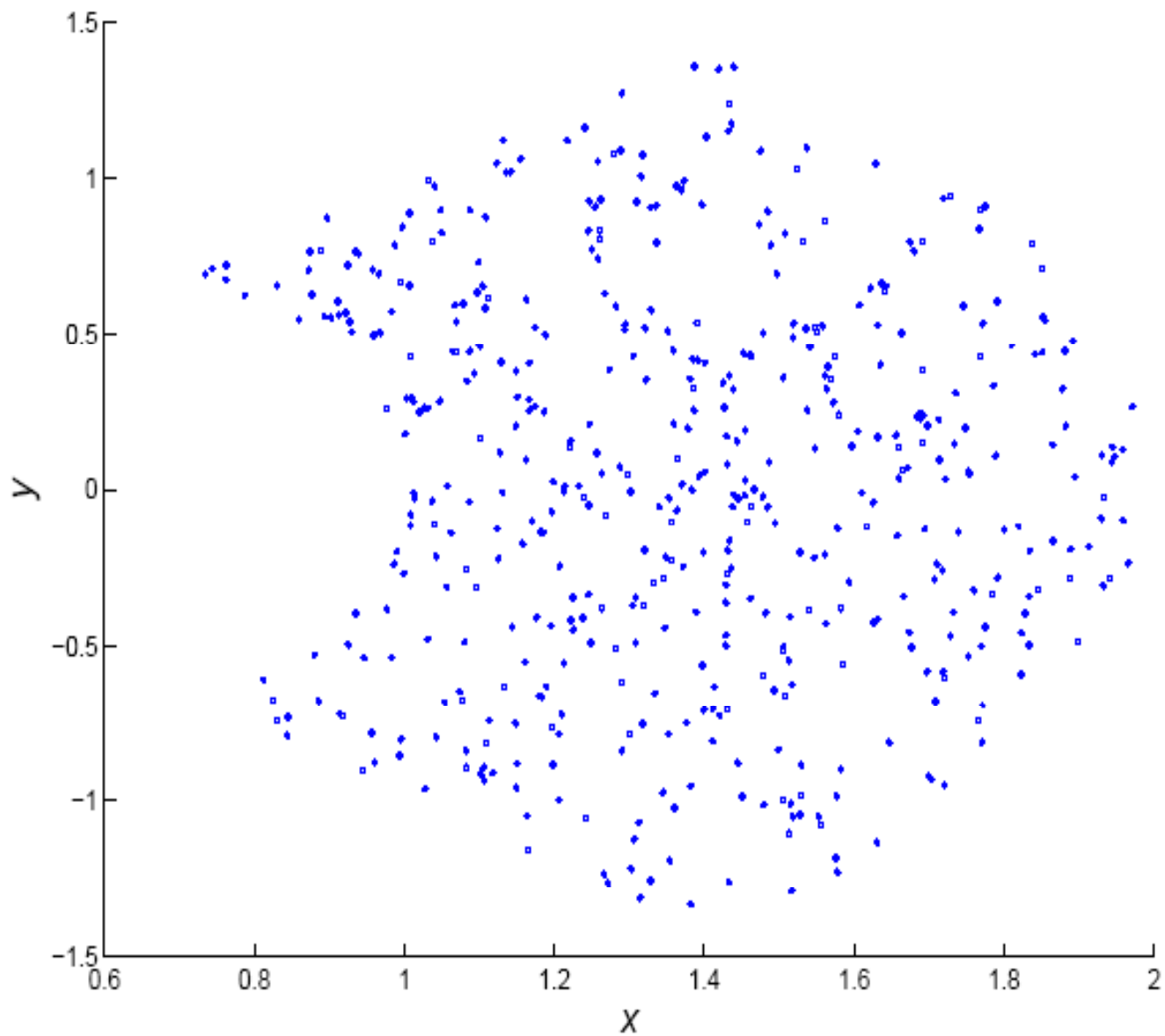


Figure 2: Cartesian coordinates of data points $x^{(i)} = r^{(i)} \cos \theta^{(i)}$, $y^{(i)} = r^{(i)} \sin \theta^{(i)}$, $i = 1, \dots, N$.

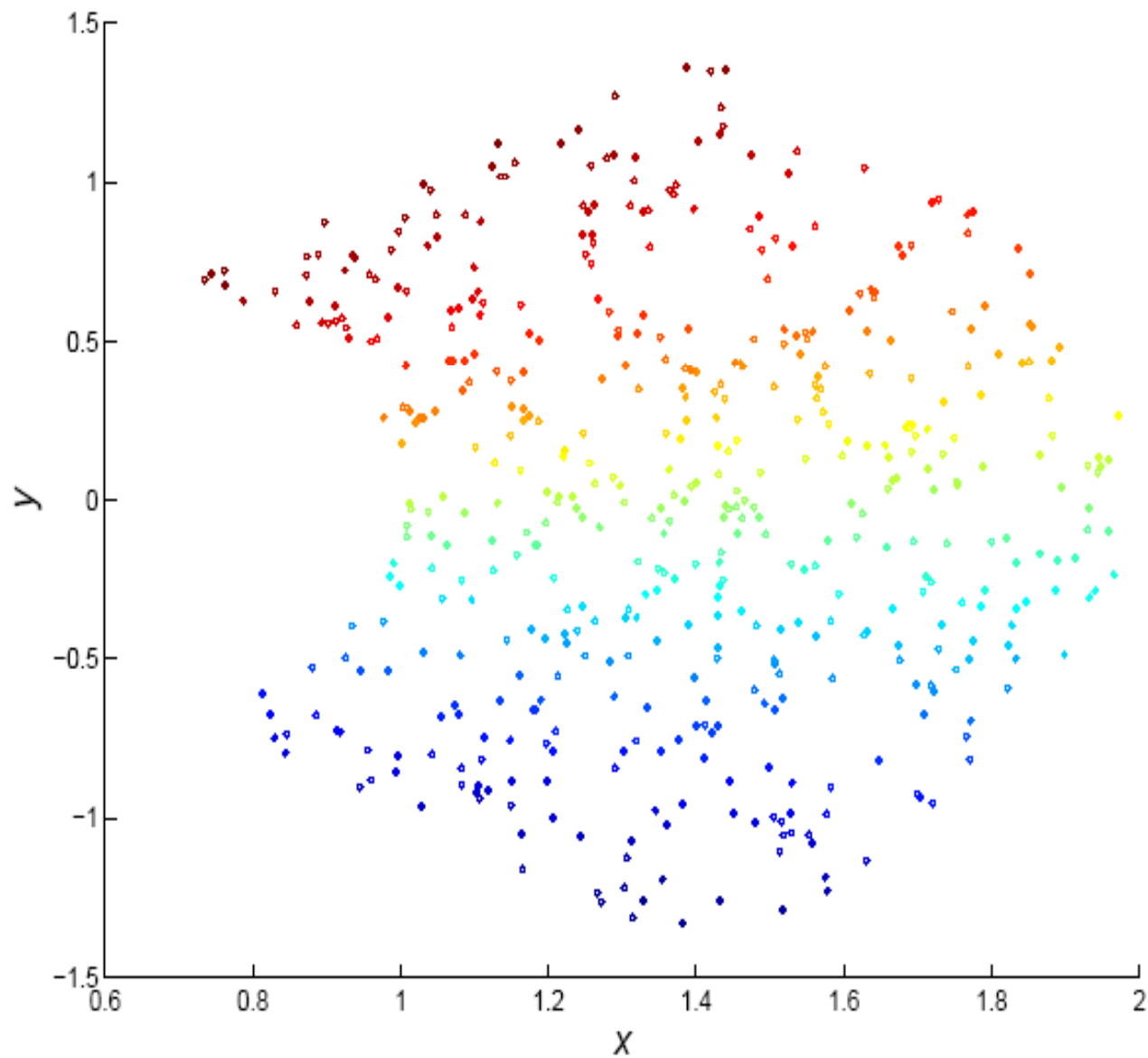


Figure 3: Color map plot of ϕ^1 shows its level lines to be rays emanating from the origin, thus ϕ^1 reveals the θ coordinate.

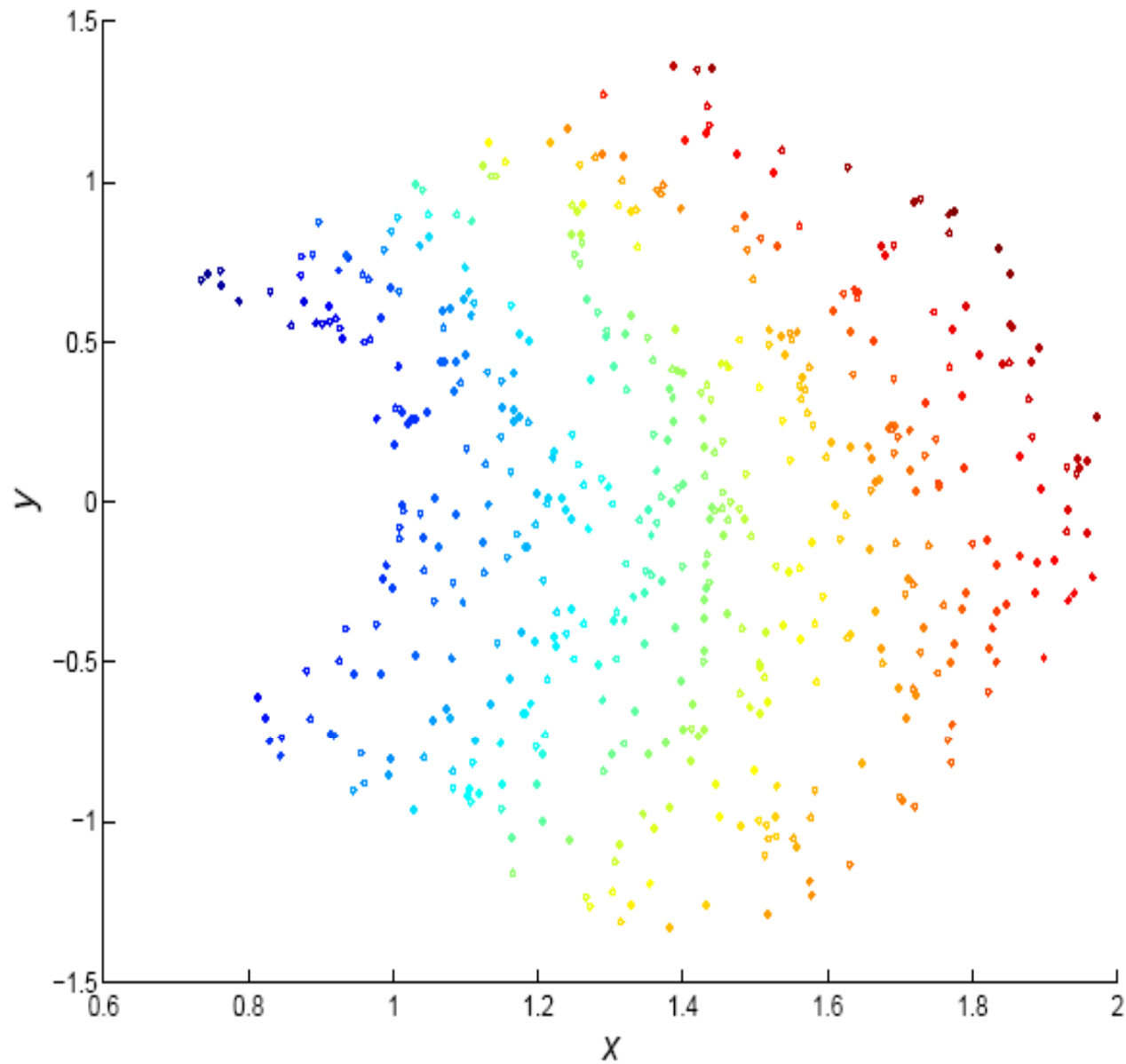


Figure 4: Color map plot of ϕ^2 shows its level lines to be concentric circles, thus ϕ^2 reveals the r coordinate.

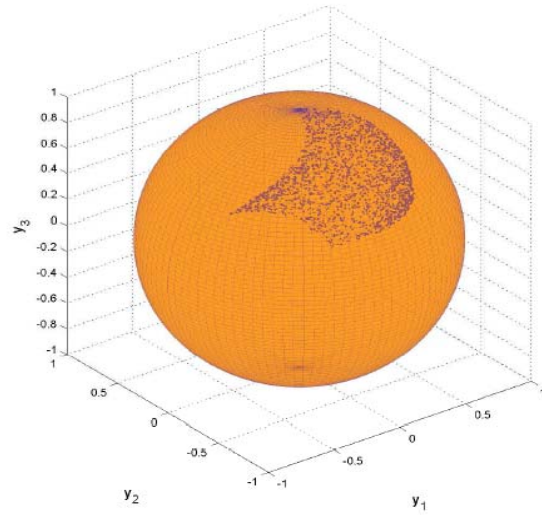
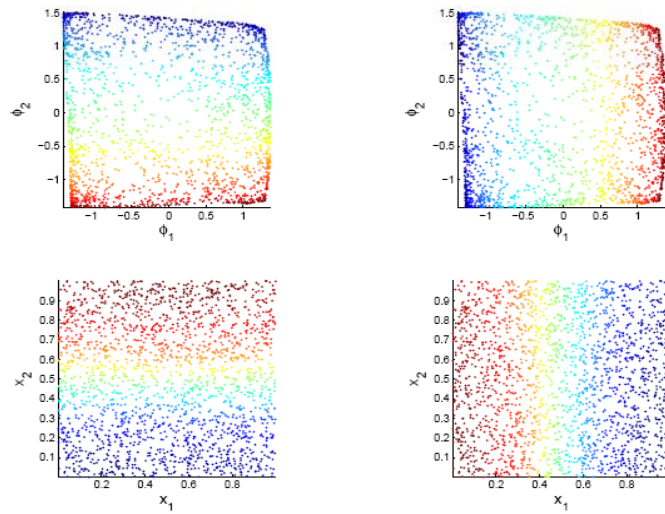


Fig. 6. Points (y_1, y_2, y_3) on the unit sphere are obtained by mapping uniformly sampled points in the unit square by (52).



Inherently non-linear chemical reactions

We consider the following set of chemical reactions



The first two reactions (24) are production and degradation of Z (catalyzed by X and Y , respectively). The production and degradation of Z is assumed to be happening on a fast time scale. Reactions (25) are production and degradation of Y . They are assumed to occur on an intermediate time scale (i.e. slower than the fast time scale, but faster than the slow time scale). The reaction (26) is production of X which is assumed to be slow. We choose the values of the rate constants as

$$k_1 = 1000, \quad k_2 = 1, \quad k_3 = 40, \quad k_4 = 1, \quad k_5 = 1. \quad [27]$$

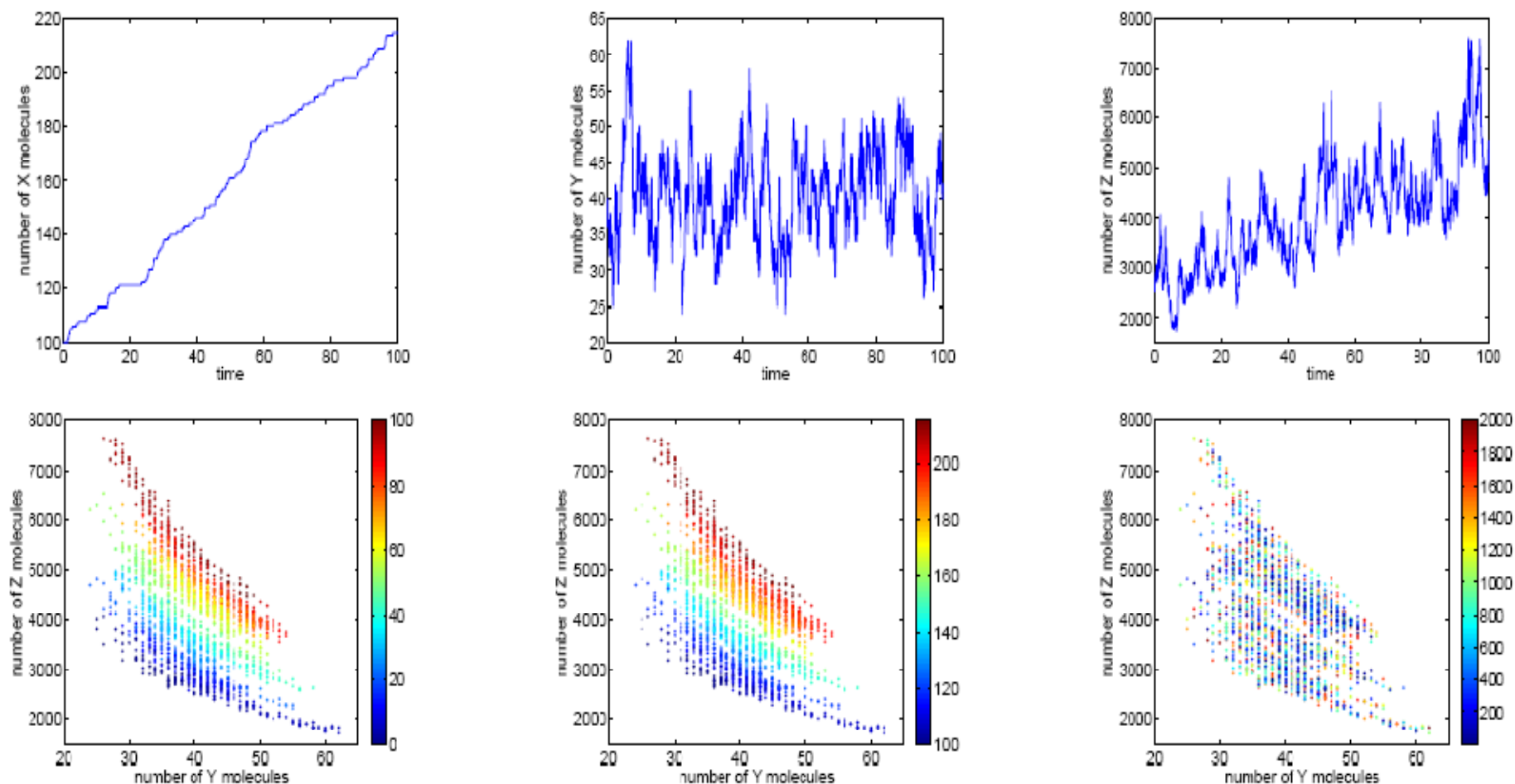


Fig. 5: *Inherently non-linear chemical reactions: the time evolution of X (top left panel), Y (top middle panel) and Z (top right panel) given by the stochastic simulation of the chemical system (24) – (26). The same trajectory (2000 data points, saved at equal time intervals $\Delta t = 0.05$ apart) plotted in the Y-Z plane is shown in the bottom panels. We color the points according to time (bottom left panel) and according to the number of X molecules (bottom middle panel). To emphasize the strength of our approach, we randomize the order of the data points – we color the resulting data set according to the order in the new list in the right panel (bottom right panel).*

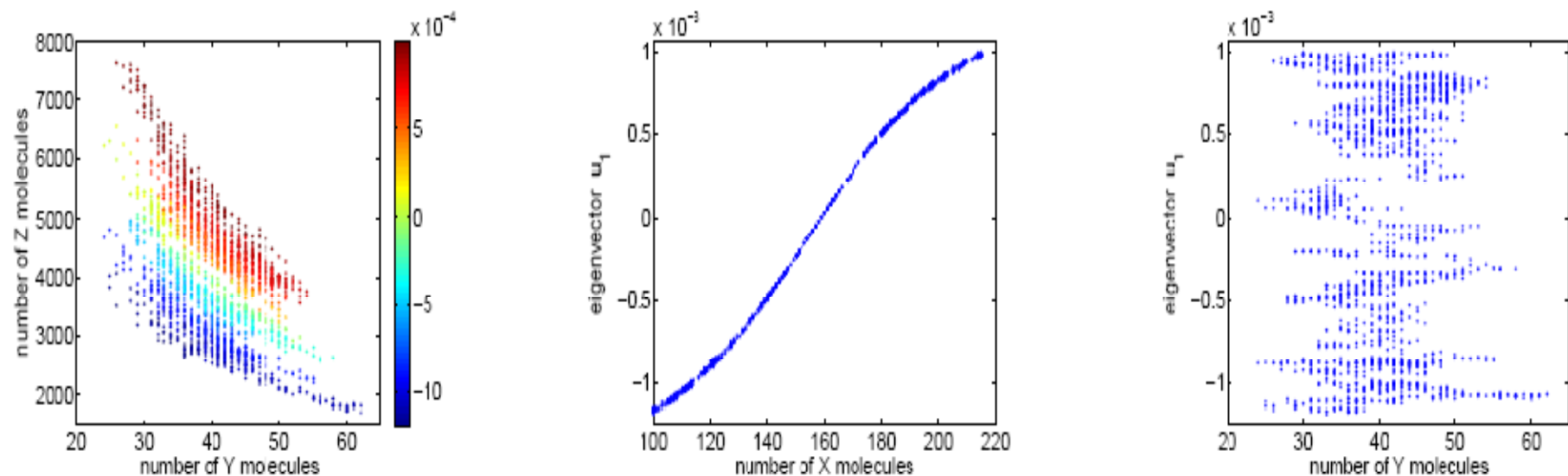


Fig. 6: *Inherently non-linear chemical reactions: the data set in the Y-Z plane with each point colored according to \mathbf{u}_1 (left panel). Vector \mathbf{u}_1 as a function of X (middle panel). Vector \mathbf{u}_1 as a function of Y (right panel).*

For these intrinsic parametrizations to be useful for modeling and prediction it is essential to extend their definition beyond the known empirical data .

This can be achieved by extending the inverse of the local data correlation matrix, viewed as an empirical matrix valued function on the data , our basic assumption is that this function is smooth and therefore can be extended by the multiscale algorithm already discussed previously .

After such an extension has been achieved , we can use the Gaussian defined by the extended Jacobian to extend the corresponding coordinates , this should be compared to a direct extension of the coordinates as described below .

Multiscale extensions of intrinsic parameters using Gaussian kernels at different scales, described below

We consider an empirical function f defined on a data set \mathcal{X} . We diagonalize various scaled Gaussian kernels restricted to the data as follows;

$$e^{-4^m |x_i - x_j|^2} = \sum_l \lambda_l^m \Phi_l^m(x_i) \Phi_l^m(x_j) \quad \text{each } \Phi_l^m \text{ has an extension " to distance } \approx 2^{-m} \text{ " given by}$$

$$\Phi_l^m(x) = \frac{1}{\lambda_l^m} \sum_i e^{-4^m |x - x_i|^2} \Phi_l^m(x_i) \quad \text{if } \lambda_l^m / \lambda_0^m \geq \delta = 0.1$$

We define the projection $\sum_{l; } \langle g, \Phi_l^m \rangle \Phi_l^m(x_i) = P_m(g)$ where l are as above.

Given f defined on the data x_i let $P_1(f)$ be the coarse scale approximation.

We then expand the residual $f - P_1(f)$ using $e^{-4|x_i - x_j|^2}$, which is half as wide, to build $P_2(f - P_1(f))$ which extends "to distance $\approx 1/2$ ". If $P_1(f) + P_2(f - P_1(f))$ is a good approximation we stop, otherwise we repeat at the next scale.

YALE