ExPlanes: Exploring Planes in Triplet Data

Bruno Schwenk[1], Joachim Selbig[2], Yehuda Ben-Zion[3], Matthias Holschneider[1]

[1]Institut für Mathematik, Universität Potsdam, 14476 Potsdam, Germany

[2]Institut für Biochemie und Biologie, Universität Potsdam, 14476 Potsdam, Germany

[3]USC Earth Sciences, Los Angeles, CA 90089-0740

Summary

Many methods for the analysis of gene expression-, protein- or metabolite-data focus on the investigation of binary relationships, while the underlying biological processes creating this data may generate relations of higher than bivariate complexity. We give a novel method ExPlanes that helps to explore certain types of ternary relationships in a statistically robust, Bayesian framework. To arrive at an characterization of the data structure contained in triplet data we investigate 2-dimensional planes being the only linear structures that cannot be inferred from projections of the data. The key part of our methodology is the definition of a robust, Bayesian plane posterior under the assumption of an invariant prior and a Gaussian error model. A numerical representation of the plane posterior can be explored interactively. Beyond this purely Bayesian approach we can use the plane posterior to construct a family of posterior-based test statistics that allow testing the data for different plane related hypotheses. To demonstrate practicability we queried triplets of metabolic data from a plant crossing experiment for the presence of plane-, line- and point-structures by using posterior-based test statistics and were able to show their distinctiveness.

1 Introduction

Over the last twenty years, significant advances in experimental technologies have provided the prerequisites for a better understanding of cellular regulatory and metabolic processes, cf. [5, 3, 13].

These data on gene expression, metabolite concentrations, protein abundances, and also fluxes allow the improvement of pathway and network modeling and simulation with the goal of better understanding the dynamics of cellular processes, cf. [6, 8, 16].

A variety of methods to analyze these data is based on the detection of local *pair-wise* (or *binary*) relationships that also can be used to generate networks which represent defined aspects of structure within the data such that methods from graph theory can be applied to make further inferences on the underlying biology. The traditional way of defining local relations that give rise to global graph-like structure consists in considering pair-wise relationships described by linear or non-linear distance measures such as Pearson correlation or mutual information, cf. [2, 11, 17]. These approaches allow the identification of highly interconnected sets of molecules that ostensibly form functional units from dependencies that are defined only between pairs of variables.

While these methods only investigate pair-wise relations it is a fundamental fact that interactions between the building blocks of bio systems give rise to relationships of higher than bivariate complexity. One example is the regulation of enzyme activity that induces dependencies not only between product and substrate of an enzyme, but also with possible regulatory active compounds, another is the oligomerization of proteins influencing their biologic activity.

Recently the information processing inequality, partial correlation and conditional mutual information are taken into account in order to represent ternary relationships between triplets of variables and especially the concept of direct causation, cf. [12, 19, 20].

Here we generalize a method published by Kose e.a. [11] for investigating binary relationships and propose an approach, we call **ExPlanes** (**Ex**ploring **Planes** in triplet data), for investigating ternary relationships relying on the Bayesian formalism. It differs from the previous approaches in specific aspects and gives a method to investigate otherwise inaccessible types of interrelations.

While the method in [11] deals with the investigation of (multiple) 1-dimensional linear relationships in 2-space (geometrically: lines) we investigate 2-dimensional linear relationships 3-space (geometrically: planes) as the only linear generalization being non trivial in the sense that it cannot be inferred from the knowledge of 2-dimensional projections of the data alone.

The core of our method is the numeric approximation and interactive visualization of the Bayesian posterior of the set of 2-planes with respect to data in 3-space and a robust error model.

In contrast to conventional regression methods we explicitly require that the results of our method do *not* depend on the ordering of the variables a-priori. This refers to a situation in which a causal direction between the quantities under investigation is not known a-priori or even impossible to define on an conceptual level. The a-priori independence on ordering is incorporated by creating a translationally and rotationally invariant Bayesian posterior in the space of planes, which is subsequently processed and visualized by techniques building on these properties.

From the method in [11] the characteristics inherited are robustness against outliers in the data, use of estimates for correlated measurement errors, capability of exploring multiple linear relationships by use of test statistics and testing for specific structures in the data.

The generalizations compared to [11] consist in following points: treatment of ternary relationships, interactive visualization of the posterior and invariance against rotation and translation.

Our method ExPlanes has two parts: First we compute and interactively visualize the Bayesian posterior in the space of all possible 2-planes in 3-space (two dimensional linear relations) in a translationally and rotationally invariant way. Since the space on which the plane posterior is defined is itself a 3-dimensional manifold it cannot be overseen visually by plotting a single diagram, as it was possible for the case of binary relationships [11]. The method of visualization we propose uses the mathematical structure of this space, which is the Cartesian product of the surface of the unit sphere and the positive, real numbers to display and explore the posterior interactively. Second we use the posterior to detect specific ternary relationships by defining a family of test statistics as functionals of this posterior distribution.

In contrast to the methods described by [2, 12, 20] that estimate mutual information as differ-

ence of entropies indirectly calculated from histogram- or kernel-estimates of the probability density function generating the data, ExPlanes avoids the estimation of entropies. Moreover it incorporates an error model that allows the accommodation of its sensitivity to outliers. While the methods based on (conditional) mutual information do not allow the specification of the type of relation, ExPlanes allows the testing for specific ternary relationships. Because Ex-Planes is rotationally invariant the results do not depend on the ordering of the variable triplets and is especially suited in situations, in which no causal ordering can be assumed, in contrast to regression methods.

ExPlanes was applied exemplarily to metabolite profile data generated by microchip-based nanoflow-direct-infusion QTOF mass spectrometry [15]. In addition to the first results derived by a combination of principal and independent component analysis, we show here that far more detailed information on the level of triplets of single metabolites, in contrast to the processing of aggregated quantities like PCA and ICA, can be obtained.

2   Method

ExPlanes relates 3-dimensional data $\vec{x}_i \in \mathbb{R}^3$ with given error covariances $\Sigma_i \in \mathbb{R}^{3\times3}$ that have to be calculated in advance, e.g. from technical replicates, to the set of all possible 2-planes in 3-space. It does so by calculating, visualizing and processing a robust Bayesian posterior defined in the set of all possible planes. Further details of mathematical derivations and visualization can be found in the supplementary material (Appendices A - E).

2.1   2-Planes in 3-Space

A plane E is defined as the set of points $\vec{x} \in \mathbb{R}^3$ that fulfill the condition:

$$E = \left\{ \vec{x} \in R^3 | (\vec{x} \cdot \vec{n} - \beta) = 0 \right\} \tag{1}$$

Therein $\vec{n}$ is the normal-vector of the plane and $\beta \in \mathbb{R}^+$ its distance to the origin. The set of all possible 2-planes in 3-space is itself a 3-dimensional manifold. It can be characterized by two angular coordinates θ and ϕ and the distance β (Appendix A).

To enable a Bayesian treatment we have to define a probability measure in plane-space that reflects the invariance structure of the problem [9, 10] (Appendix B). We assume invariance against rotation and translation of the 3-space, in absence of any given data, and calculate the invariant measure, that results in the space of planes to:

$$d\mu(E) = \text{const} \cdot \sin[\theta] \, d\theta \, d\phi \, d\beta. \tag{2}$$

It reflects our prior knowledge of the problem, so that with respect to this measure the prior is uniform:

$$p(E) = const. \tag{3}$$

2.2　Error Model

Until now planes are just geometrical entities but not models in the statistical sense. To use them as a tool for (Bayesian) statistical analysis we augment them by an *error model*. More precisely, we have to make an assumption that if a given quantity in reality has the value $\vec{y}$ with a given probability $p(\vec{x}|\vec{y}, \Sigma)$ we observe the value $\vec{x}$ instead.

In this paper we use a multivariate Gaussian error model with covariance matrix Σ and zero systematical error:

$$p(\vec{x}|\vec{y}, \Sigma) \sim \exp\left[-\frac{1}{2}(\vec{y} - \vec{x})^T \Sigma^{-1}(\vec{y} - \vec{x})\right] \tag{4}$$

The error model gives us the probability that a single data point $\vec{x}$ results from a single, true point $\vec{y}$. We now want to compute the probability distributions of a single observation that comes from some location on a fixed plane E. For this, we need to specify the distribution of points on the plane $dp(\vec{y})$ on the plane. The distribution of a single observation coming from the plane E is then computed as:

$$p(\vec{x}|E, \Sigma) = \int_E p(\vec{x}|\vec{y})\, dp(\vec{y}) \tag{5}$$

Since we do not have any a priori information about the localization of the points $\vec{y}$ on the plane we assume an uniform and hence improper distribution. Thus $dp(\vec{y})$ is simply a Euclidean surface measure. Remembering that the plane E itself is parameterized by $E = E(\vec{n}, \beta)$ we arrive at a result analogous to eqn. (14) in [11] for the Likelihood of plane $E = E(\vec{n}, \beta)$:

$$p(\vec{x}|E, \Sigma) \sim \exp\left[-\frac{1}{2}\frac{(\vec{x} \cdot \vec{n} - \beta)^2}{\vec{n}^T \Sigma \vec{n}}\right] \tag{6}$$

For fixed E as a function of $\vec{x}$, this density is an improper Gaussian distribution with degenerated covariance matrix. It shows that the probability distribution that a given point $\vec{x}$ stems from a certain plane E depends only on its normal distance.

By the theorem of Bayes and the assumption of a uniform prior with respect to the invariant measure (2) this is proportional to the posterior of a plane E given a single data point $\vec{x}$:

$$p(E|\vec{x}, \Sigma) = p(\vec{x})^{-1}\, p(\vec{x}|E, \Sigma)\, p(E) \sim p(\vec{x}|E, \Sigma) \tag{7}$$

2.3　Robust Posterior

After obtaining the expression (6) for the probability of a given plane $E(\vec{n}, \beta)$ with respect to a single data point $\vec{x}$ we now have to find an expression for the posterior $p(E|\{\vec{x}_i\}, \Sigma, ...)$ of a plane E with respect to a set of data points $\{\vec{x}_i\}$ where $i \in \{1, 2, ..., N\}$.

A principle desideratum in data analysis is *robustness*. This means that the conclusions drawn from a data set do not change strongly if one or even few of the data points, called *outliers*, result from processes that have nothing to do with the phenomenon generating the variability under investigation. Examples are mistakes in the measurements or sample handling or genetic or environmental deviations concerning single biological organisms in the sample. An example of a method of analysis that is not robust in this sense is Pearson's coefficient of correlation for

small sample sizes, since a single outlying data point sufficiently distant from the meaningful rest of the population can twist the result away from the biologically correct number.

To arrive at an intrinsically robust formula for the posterior we follow the concept of [11] and assume implicitly that with a given plausibility any of the values in a set of data points $\{\vec{x}_i\}$ *may be* an outlier. For the posterior density $p(\vec{n}, \beta | \{\vec{x}_i\}, \Sigma, c)$ [11] arrive at following formula:

$$\ln[\, p(\vec{n}, \beta | \{\vec{x}_i\}, \Sigma, c)\,] \sim \sum_{i=1}^{N} \ln\,[\, p(\vec{n}, \beta | \vec{x}_i, \Sigma)\,+\,c\,] \tag{8}$$

$$= \sum_{i=1}^{N} \ln\,\left[\exp\left[-\frac{1}{2}\frac{(\vec{x}_i \cdot \vec{n} - \beta)^2}{\vec{n}^T \boldsymbol{\Sigma}_i \vec{n}}\right] \,+\, c\right] \tag{9}$$

Therein c, called the "decoupling term", reflects the prior that a given data point is in fact an outlier. It ensures the robustness of the formula since vanishing of a single term $p(\vec{n}, \beta | \vec{x}_i, \Sigma)$ does, for values of $c \approx 1$, not result in infinitely negative value of the logarithm, and so limits the maximum influence of a single data point on the sum to $\ln[c]$. In our subsequent analyzes we set the numerical value to unity $c = 1$, in accordance with [11].

It was shown by [11] that this formulation of the posterior (8) can be utilized to mixtures of data consisting of multiple linear relations, given approximate values of c. In a Bayesian sense this can be understood in the following way: The summation of single point posterior $p(\vec{n}, \beta | \vec{x}_i, \Sigma)$ and decoupling term c in (8) describes a situation in which the individual data point can either be attributed to the plane $E(\vec{n})$ or the completely un-informative alternative represented by an constant, improper likelihood. The possibility of investigating multiple linear relationships now corresponds to the assumption, that any of the other possible planes might be "absorbed" in this isotropic alternative.

For details on numerical discretization see Appendix C.

2.4 Visualization

Mathematically the posterior $p(\vec{n}, \beta | \{\vec{x}_i\}, \Sigma, c)$ is a probability density over the distance β and the unit vector $\vec{n}$. For visualization of its sphere part, this is the dependence on the unit vector $\vec{n}$ given a *fixed* value of the radius β, in analogy to geographical *map projections* we use an (almost) area conserving mapping of the coordinates θ and ϕ onto the screen. For simplicity of the formula we used the Kavarayskiy VII projection [18]. The logarithm of the posterior is then displayed on this map by a color coding.

To inspect the dependence of the posterior on the radius we integrate out the angular dependence and obtain the *marginal distribution* over β, the "β-marginal":

$$p(\beta | \{\vec{x}_i\}, \Sigma, c) \,=\, \int_0^{2\pi} \int_0^{\pi} p(\phi, \theta, \beta | \{\vec{x}_i\}, \Sigma, c) \sin[\theta] d\theta \, d\phi \tag{10}$$

In our software prototype we realized the simultaneous visualization of sphere part and β-marginal by arranging the color coded map of the sphere-part for a given value of β above a white bar-chart of the logarithm of the β-marginal (10). To indicate the actual value of β the respective bar in the bar chart is marked in red. We included two forms of interactivity: First we made it possible to explore the values of β by by mouse clicking on the respective position

of the β-marginal bar. Second we allowed the user to select planes by clicking on the sphere part map. The selected planes are then stored in a list and displayed as blue numbers in the diagram (Appendix D).

2.5 Posterior Based Test Functions

In addition to the visual inspection of the calculated posterior we developed a statistical test for the presence or absence of plane related structures in the data. Therefore we supplemented the Bayesian approach with conventional frequentist test statistics [14]. The idea behind this is the intuition that test functions should get advantageous properties if they are somehow related to the hypotheses to be tested for. We did not rely on a fully Bayesian treatment of the test situation since the computation of the marginal posterior likelihoods (Bayes factors) of the model classes acting as alternative hypotheses would involve their fully Bayesian treatment which is computationally demanding. We defined a posterior based *test function* as functional of the posterior:

$$T(\{\vec{x}_i\}) \,=\, T\,\{p(E|\{\vec{x}_i\}, \Sigma, c)\} \tag{11}$$

The statistical test is then constructed in the common way: We state a statistical *null hypothesis* H_0 as well as an *alternative hypothesis* H_1 by defining a probability densities for the data

$$p(\{x_i\}|H_0) \ \text{ and } \ p(\{x_i\}|H_1) \tag{12}$$

which, on a geometrical level, are related to distinct plane related structures. Then we calculate, by Monte Carlo methods, the distributions of the test function:

$$p(T|H_0) \ \text{ and } \ p(T|H_1) \tag{13}$$

After determining the empirical value of the test function $T^\star = T(\{\vec{x}_i\})$ from the experimental data $\{\vec{x}_i\}$ we compare it to the distributions and draw conclusions on the validity of the hypotheses, i.e. the presence and absence of the underlying plane related structures [14].

We present three different posterior based test functions. Two of them, the *entropy*

$$S(\{\vec{x}_i\}) \,=\, \int p(E|\{\vec{x}_i\}, \Sigma, c) \cdot \ln p(E|\{\vec{x}_i\}, \Sigma, c) \ \ d\mu(E) \tag{14}$$

and the *polynomial measure of concentration* (PMOC)

$$PMOC(\{\vec{x}_i\}) \,=\, \int p(E|\{\vec{x}_i\}, \Sigma, c)^q \ \ d\mu(E) \tag{15}$$

are measuring the concentration of the posterior density (Appendix E.1).

The third one, the *curvature based test function*

$$T_c(\{\vec{x}_i\}) \,=\, \frac{G}{M^2} \tag{16}$$

is defined as the ratio of Gaussian curvature G to the squared mean curvature M^2 at the maximum a-posteriori point. Mean- and Gaussian- curvature have to be calculated numerically from the posterior via the eigenvalues of the local curvature tensor (Appendix E.2).

3 Plant Metabolite Data

We apply the method described above to triplets of metabolic data, generated in an experiment published earlier in literature [15]. We first describe the experiment and the data, then we apply the methods described in section 2.5 to extract the presence of plane-like structures and visualize them exemplarily.

3.1 Description of Experiment and Data

In the experiment parent plants from the two *Arabidopsis thaliana* inbred lines $Col10$ and $C24$ were mated, resulting in four possible combinations of parent genotypes. For each of this crossings three F1 hybrids were grown (so called *biological replicates*), after harvesting and preparation for approximately eight samples (*technical replicates*) the intensities of 736 metabolites, identified by the mass to charge ratio (m/z), were measured. Since no identification of the chemical species was done in the mass-spectrometric data, we identify the metabolites by arbitrary indices.

Before we start the Bayesian analysis itself, we explore the data by investigating different measures for technical and biological variability:

A quantitative description of the precision of the data is the *technical variability* defined as the standard deviation of technical replicates made from one biological replicate.

We define the *estimated biological value* as the mean value of the technical replicates made from one biological replicate, it serves as an estimator for the true value of the metabolite concentration in the plant.

For a given set of plants, not necessarily being biological replicates of a single crossing, we define the *biological variability* as the standard deviation of their estimated biological values.

Based on these quantities we chose a sample combined out of the biological replicates of the $C24 \times Col10$ and the $C24 \times C24$ crossings for further investigation, because it showed the most promising ratio of (large) biological to (small) technical variability. With respect to the relatively high technical noise it was necessary to include only the most informative variables in our analysis.

We sorted the metabolites according to the ratio of biological to technical variability and selected the first six metabolites, #278, #322, #523, #380, #290 and #715, from that list. Those we combined in all possible, non redundant ways to triplets, which were then the 3-dimensional input to Bayesian method described above.

3.2 Testing Plane Related Hypotheses

We used the *empirical covariance matrix* of the technical replicates as Σ in the Gaussian error model (4) for each data point resulting from the underlying biological replicate. In the set under investigation the 6 independent entries of the symmetric positive definite 3×3 covariance matrices Σ had to bes estimated from 8 technical replicates. We controlled the accuracy of this estimate by using Bayesian estimation, see Appendix F.
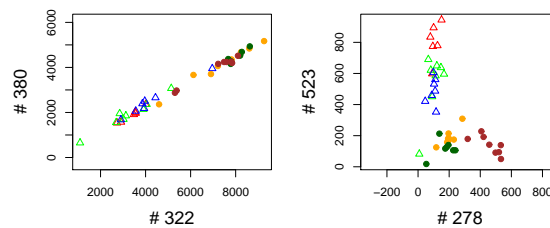
Figure 1: Pair plots of metabolite intensities, the axes are labeled with the respective metabolite indices. Triangles mark data points from the $Col10 \times Col10$ crossing, circles from the $C24 \times Col10$ crossing. The different colors indicate the individual biological replicates, in addition the more "earthly" colors are redundantly attributed to the $C24 \times Col10$ crossing. A strong correlation (up to 0.98) can be seen in the technical replicates.

The necessity for using a correlated error model is evident if one looks at the *pair plots* of the metabolite intensities as well as the numerical values of the covariances, as exemplified in Figure 1 for metabolite pairs $\#380 \times \#322$ and $\#523 \times \#278$. This a very strong correlation in the technical replicates (up to 0.98). The diagrams also indicate that the technical and biological variabilities are *related*, which can be interpreted in the sense that the same biochemical mechanisms acting in the intact plants are responsible for the variability occurring during preparation and measurement.

We defined three test hypotheses, qualitatively related to the geometrical structures point, line and plane. For each metabolite triplet under investigation the parameters of the probability distribution forming the statistical hypotheses was then set depending on numerical values characterizing the data.

The *center of the data* was defined as the median of the biological estimated values of the biological replicates under investigation. The *mean technical covariance* was defined as the average of the covariance matrices of the technical replicates over the biological replicates.

The hypotheses were defined algorithmically by:

Point Gaussian distribution with mean technical covariance around the center of the data.

Line Draw a line from the origin through the center of the data. Distribute the points on that line with a Gaussian distribution of zero mean and variance equal to the sum of the squared biological variability of the metabolites involved. Add centered Gaussian noise with technical covariance.

Plane Scatter the points on a plane trough the center of the data, employing a radial symmetrical Gaussian distribution with a variance equal to the mean of the squared biological variability of the metabolites involved. Add centered Gaussian noise with technical covariance.

From these distributions we drew 30 sample points and calculated entropy (32), polynomial measure of concentration (33) and curvature based test statistics (38). In that way we obtained small size samples of these quantities given the test hypotheses that we compared with the data measured in the experiment.
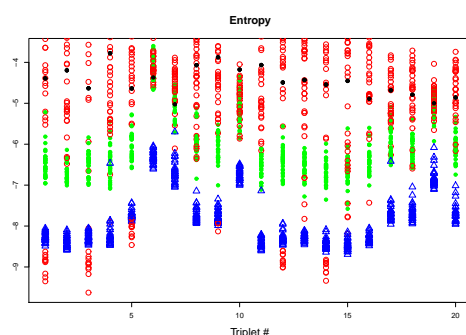
Figure 2: Entropy of the observed data (black spot) and 10 draws from the hypotheses plane (green spots), line (red circles) and point (blue triangles). The different metabolite triplets under investigation are indexed on the abscissa.
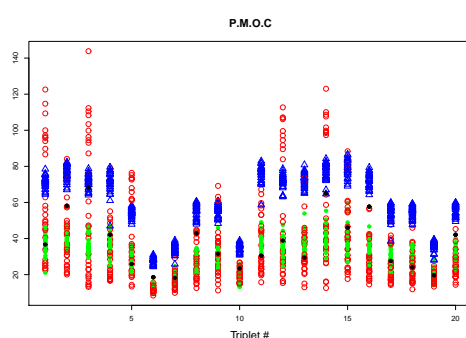

Figure 3: Polynomial measure of concentration of the observed data (black spot) and 10 draws from the hypotheses plane (green spots), line (red circles) and point (blue triangles). The different metabolite triplets under investigation are indexed on the abscissa.
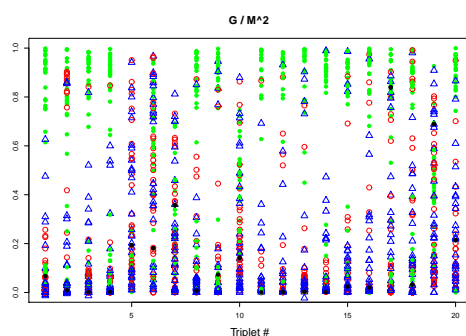

Figure 4: Curvature based Test Statistics of the observed data (black spot) and 10 draws from the hypotheses plane (green spots), line (red circles) and point (blue triangles). The different metabolite triplets under investigation are indexed on the abscissa.

Figure 2 to 4 show this results. On the abscissa the metabolite triplet under consideration is indexed, on the ordinate the value of the test statistics is displayed. *Green*, *blue* and *red* points mark the values generated by the test hypotheses plane, point and line, while the value originating from the experimental data are indicated in *black*. In addition we computed p-values using student's t-statistics, which were in good accordance with the diagrams of the samples of test functions Figure 5.

For the entropy Figure 2 and the polynomial measure of concentration Figure 3, we notice that they are in principle able to separate the three different hypotheses, in contrast to the curva-
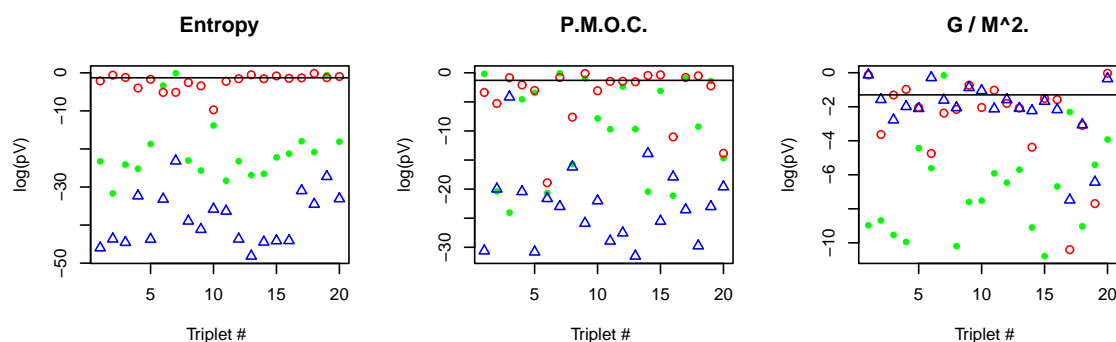
Figure 5: Logarithmic p-values as estimated from the distributions of the test functions Entropy, Polynomial Measure of Concentration and C/M^2. The 5% confidence interval is marked by a black, horizontal line. Hypotheses tested for are plane (green spots), line (red circles) and point (blue triangles).
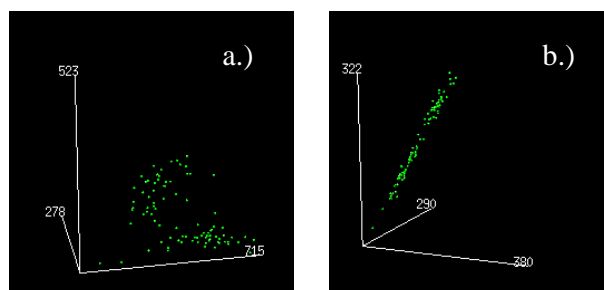

Figure 6: Part a.) Metabolite intensities of triplet #7, consisting of metabolites #278, #523 and #715, showing a pattern that visually resembles a plane in 3-space. Part b.) Metabolite intensities of triplet #14, consisting of metabolites #322, #380 and #290. Here the data are distributed along a line through the origin in 3-space.
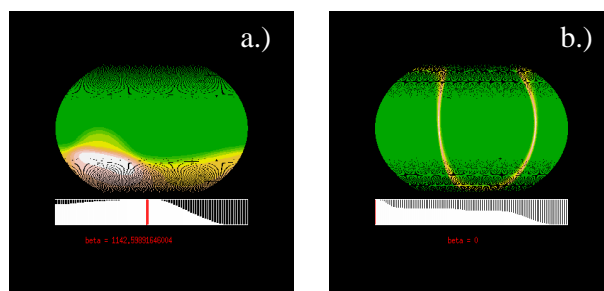

Figure 7: Posteriors for the triplets in Figure 6. Part a.) Posterior of triplet #7: The beta marginal as well as the map for the sphere part indicate an uni-modal posterior that is characteristic for the presence of a single plane. Part b.) Posterior of triplet #14: The beta marginal reaches its maximum at $\beta = 0$ while in the sphere part map a ring-like structure can be seen, both indications compatible with the "line"-hypothesis.

ture based test function Figure 4, that cannot distinguish between points and lines but more distinctively highlights true plane like structures.

For the majority of the triplets the results suggest that they can be described by the "line" hypothesis.

Further visual inspection of the data supported this findings as can be seen in Figure 6 showing typical patterns for "plane-" (Triplet #7, part a.)) and "line-" (Triplet #14, part b.)) data that visually convincing resemble the geometrical meaning of these hypotheses. However care

must be taken in comparing the visual appearance of the data to the statistical hypotheses which are based on the error model and not on the distance in 3-space directly. Therefore we used our methodology that incorporates this information to visualize the posterior . Figure 7 shows screen shots of the interactive visualization of the posterior at the respective maximal points of the β-marginal. In part a.) the posterior for triplet #7 indicates an uni-modal posterior that is characteristic for the presence of a single plane. In part b.) the beta marginal for triplet #14 reaches its maximum at $\beta = 0$ while in the sphere part a ring like structure can be seen, both indications compatible with the "line"-hypothesis; also compare Figures. 2, 3 and 4.

4   Conclusions

We were able to enrich the toolbox of the biological data analyst with the novel method Ex-Planes for the exploration of ternary relationships in profile data that combines following features: Invariance against rotations and consequently relabeling of the axes, which makes it especially suitable in cases where no a-priori ordering of the variables can be assumed. In this it differs from usual regression methods. Possibility to accommodate robustness against outliers and incorporation of an individual error estimate for each data point. The possibility of investigating multiple linear relationships. Different test functions enable the detection of a variety of plane related ternary structures, to be defined by appropriate statistical hypotheses. We demonstrated all these aspects on a metabolite data set.

Since ExPlanes combines certain functionalities, there are similarities as well as differences to established methods: With PCA it has in common, that it can be used for the detection of lower dimensional subspaces in the data. However PCA is vulnerable against outliers and does not allow for the investigation of multiple linear relationships. It does not incorporate knowledge about measurement errors, which are especially of interest in investigations of metabolic data. Projection Pursuit (PP) is a larger class of algorithms that all have in common that they search for a linear projection of the data (which by definition has reduced dimensionality) which maximizes certain objective functions [7]. PP-methods have in common with ExPlanes, that they can be used for the detection and exploration of linear relationships, and can be made robust against outliers. For the special case of 2-planes in 3-space, which is of interest here, they do not incorporate the invariant prior and the interactive method of visualization.

References

[1] I.N. Bronstein, K.A. Semendjajew, and ea. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun und Frankfurt am Main, 2001.

[2] A. Butte and I. S. Kohane. Mutual information relevance networks: Functional genomic clustering using pair-wise entropy measurements. *Pac. Symp. Biocomput.*, 5:415–426, 2000.

[3] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. Trethewey, and L. Willmitzer. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.*, 18:1157–1161, 2000.

[4] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere: With Applications to Geomathematics*. Claredon Press, Oxford, 1998.

[5] B. Grünenfelder and E.A. Winzeler. Treasures and traps in genome-wide datasets: Case examples from yeast. *Nat. Rev. Genetics*, 3:653–661, 2002.

[6] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, New York, 1996.

[7] P.H. Huber. Projection pursuit. *The Annals of Statistics*, 13(4):435–475, 1985.

[8] N. Jamshidi and B. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Molecular Systems Biology*, 4:171, 2008.

[9] E.T. Jaynes. The well-posed problem. *Foundations of Physics*, 3(4):477–491, 1973.

[10] E.T. Jaynes and G.L. Bretthorst. *Probability Theory. The Logic of Science: Theory and Elementary Applications Vol. 1*. Cambridge University Press, Cambridge, 2003.

[11] F. Kose, J. Budczies, M. Holschneider, and O. Fiehn. Robust detection and verification of linear relationships to generate metabolic networks using estimates of technical errors. *BMC Bioinformatics*, 8:162, 2007.

[12] A. Margolin, K. Basso I. Neemenmann, C. Wiggins, G. Stolovitzky, R. Della Favera, and A. Califano. Arcane: An algorithm for the reconstruction of gene regulatory networks in a mammalina cellular context. *BMC Bioinformatics*, 7, Suppl.1:S7, 2006.

[13] A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405:837–846, 2000.

[14] H. Rinne. *Taschenbuch der Statistik 2. überarbeitete und erweiterte Auflage*. Verlag Harri Deutsch, Thun und Frankfurt am Main, 1997.

[15] M. Scholz, S. Gatzek, A. Sterling, O. Fiehn, and J. Selbig. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics*, 20(15):2447–2454, 2004.

[16] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, 2002.

[17] R. Steuer, J. Kurths, C.O. Daub, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18, Suppl 2:231–240, 2002.

[18] Wikipedia. Kavrayskiy vii projection. $http//en.wikipedia.org/wiki/Kavrayskiy\_VII$.

[19] Y. Yang, A.P. Tashman, J.Y. Lee, S. Yoon, W. Mao, K. Ahn, W. Kim, N.R. Mendell, D. Gordon, and S.J. Finch. Mixture modelling of microarray gene expression data. *BMC Proc*, 1 Suppl. 1:S 50, 2007.

[20] W. Zhao, E. Serpedin, and E.R. Gougherty. Inferring connectivity of genetic regulatory networks using information-theoretic criteria. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5:262–274, 2008.
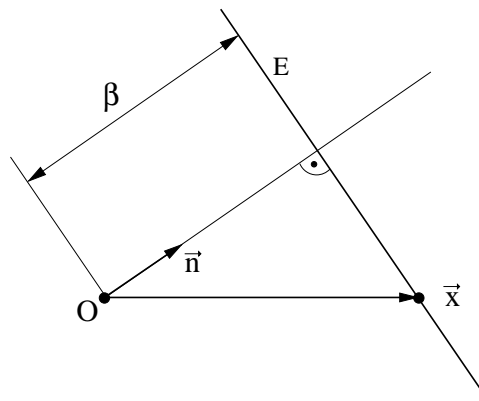
Figure 8: Definition of a plane E as the set of vectors $\vec{x}$ that have the same projected distance β in direction of a given unit vector $\vec{n}$.

A   The Set of Planes

Consider the 3-dimensional Euclidean space $\mathbb{R}^3$. A 2-plane is a two dimensional affine subspace. The set of all 2-planes is a 3 dimensional manifold. Unfortunately, there is no global parameterization of this space. However, there is a parameterization that is good for all practical purposes. Consider now a unit-vector $\vec{n}$ and a real number $\beta \geq 0$. The set of points $\vec{x}$ that have in the direction $\vec{n}$ the distance β is a 2 dimensional plane. In formulas we may write (see Figure 8):

$$E = \left\{ \vec{x} \in R^3 | (\vec{x} \cdot \vec{n} - \beta) = 0 \right\} \tag{17}$$

Vice versa all planes that do not contain the origin can be written uniquely in that way. The planes that go through the origin can still be written in that way, however, the representation is not unique anymore, since $\vec{n}$ and $-\vec{n}$ give rise to the same plane. Nevertheless we identify

$$E \leftrightarrow (\beta, \vec{n}).$$

The unit vector itself may be written in polar coordinates using two angles $\theta \in [0; \pi]$ and $\phi \in [0, 2\pi)$. We then may write [1]

$$\vec{n} = \begin{pmatrix} \sin[\theta] \cdot \cos[\phi] \\ \sin[\theta] \cdot \sin[\phi] \\ \cos[\theta] \end{pmatrix} \tag{18}$$

where the usual care has to be taken for $\theta = 0$ and $\theta = \pi$ since there the parameterization degenerates. In that way we obtain a parameterization

$$E \leftrightarrow (\beta, \theta, \phi) \in [0, \infty) \times [0, \pi] \times [0, 2\pi),$$

where care has to taken at the points $\theta \in \{0, \pi\}$ and $\beta = 0$.

B   Invariant Measure

Since it is our aim to perform Bayesian analysis on the set of possible planes, which itself is a 3-dimensional manifold, that can not be continuously to the $\mathbb{R}^3$ itself, we have to determine

a probability measure on it. We need this measure for integrating over the prior and posterior (if we, for example want to calculate measures of concentration like entropy). Since in the absence of data it contains all the information at hand the prior density relative to this measure is uniform. The complete absence of information from data should be invariant under the natural coordinate changes: No information is gained by translating the whole space or rotating it. Therefore the measure that we are looking for should be invariant under the group generated by rotations and translations (i.e. the Euclidean group). We denote by

$$\vec{x}\prime \ = \ \mathbf{R}\vec{x} \ \text{ with } \ \mathbf{R}^{-1} \ = \ \mathbf{R}^T \tag{19}$$

the rotations and by

$$\vec{x}\prime \ = \ \vec{x} \ + \ \vec{a} \tag{20}$$

the translations of points in three dimensional space $\mathbb{R}^3$.

In our case we can show that there is an up to a multiplicative constant unique measure that satisfies these invariance properties. The proof is following the method given in [9, 10]. In the parameterization given by (18) this invariant measure in the space of planes $d\mu(E)$ can be expressed as:

$$d\mu(E) \ = \ \text{const} \cdot \sin[\theta] \ d\theta \, d\phi \, d\beta. \tag{21}$$

Note that the points, where the parameterization becomes singular only have measure 0. So our parameterization is fine for any distribution of planes which does not have a point mass at theses singular values.

C　Discretization

To represent numerically and visualize the posterior we have to find a discretization of the set of planes under consideration.

For any data set we can, before the computation of the posterior, see that the possible distance to the origin is limited to some interval $\beta \in [\beta_{min}; \beta_{max}]$. In this interval we discretize β by a set of equidistant values $\{\beta_{min}, ..., \beta_b, ..., \beta_{max}\}$.

The discretization of the sphere part of the posterior, for every of the discretized values of β_b, is taken over from [4].

The polar angle θ is represented by a set of γ equidistant points $\{\theta_0, ..., \theta_i, ...\theta_\gamma\}$ with $\theta_0 = 0$ (north pole) and $\theta_\gamma = \pi$ (south pole).

The azimuthal angle ϕ is also discretized by an equidistant lattice $\{\phi_{i,0}, ..., \phi_{i,j}, ..., \phi_{i,\gamma_i}\}$. Here the number of points of discretization γ_i depends on θ_i in a way ensuring that the points of

discretization $(\theta_i, \phi_{i,j})$ are uniformly spread over the surface of the sphere [4]:

$$\Theta_0 = 0 \quad \phi_{0,1} = 0 \quad \text{north pole} \tag{22}$$

$$\Delta\Theta = \frac{\pi}{\gamma} \quad \text{with} \quad \gamma \in N \tag{23}$$

$$\Theta_i = i \cdot \Delta\Theta \quad \text{with} \quad i \in \{1, 2, ..., \gamma - 1\} \tag{24}$$

$$\gamma_i = \left\lceil \frac{2\pi}{\arccos\left[(\cos[\Delta\Theta] - \cos^2[\Theta_i])/\sin^2[\Theta_i]\right]} \right\rceil \tag{25}$$

$$\phi_{i,j} = \left(j - \frac{1}{2}\right)\left(\frac{2\pi}{\gamma_i}\right) \quad \text{with} \quad j \in \{1, 2, ..., \gamma_i\} \tag{26}$$

$$\Theta_\gamma = \pi \quad \phi_{\gamma,1} = 0 \quad \text{south pole} \tag{27}$$

The angular resolution depends on the integer parameter γ which gives the number of lattice points in Θ direction, so before the calculation of the posterior, after specifying the data and the parameters of the error model, we have to find an numerically adequate value for it. To obtain a practical estimate we look again at the posterior of a single data point (6):

$$p(\vec{n}, \beta | \vec{x}, \Sigma) \sim \exp\left[-\frac{1}{2}\frac{(\vec{x} \cdot \vec{n} - \beta)^2}{\vec{n}^T \Sigma \vec{n}}\right]$$

We consider the generic case of a data point $\vec{x} = x \cdot \vec{n}_z$ situated at distance $|\vec{x}|$ in "north" direction $\vec{n}_z = \vec{n}(\theta = 0, \phi = 0)$. The error model is assumed to be isotropic with $\Sigma = \sigma^2 \cdot \mathbf{I}$.

Since here the scalar product $\vec{n} \cdot \vec{x} = |\vec{x}| \cdot \cos\theta$ does not depend on ϕ the posterior is a function of θ and β alone. We use a saddle point approximation around the maximum value $\theta^\star = \arccos[\beta/|\vec{x}|]$, $\beta^\star = |\vec{x}|$ and find that it is approximately proportional to:

$$p(\theta, \phi, \beta^\star) \sim \exp\left[-\frac{1}{2}\frac{(\beta^\star)^2 \sin^2\theta^\star (\theta - \theta^\star)^2}{\sigma^2}\right] \tag{28}$$

The standard deviation in angular direction σ_θ is:

$$\sigma_\theta = \frac{\sigma}{|\vec{x}|} \cdot \frac{1}{\sqrt{1 - \beta/|\vec{x}|}} \tag{29}$$

It has its minimum at $\sigma_\theta = \sigma/|\vec{x}|$, so the angular standard deviation depends on the *relative error* $\sigma_{rel} \equiv \sigma/|\vec{x}|$ of the data.

From identical arguments we see that the standard deviation in β direction $\sigma_\beta = \sigma$, and hence depends on the *absolute error* $\sigma_{abs} \equiv \sigma$.

We use this insight and to the number of points γ in θ direction to

$$\gamma = \text{ceiling}\left\{\frac{res_\theta \cdot \pi}{\sigma_{rel}}\right\} + 1 \tag{30}$$

and the spacing of the points in β direction to:

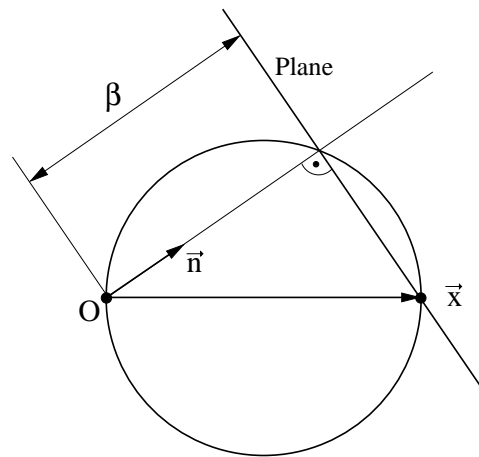$$\Delta\beta = \frac{\sigma_{abs}}{res_\beta} \tag{31}$$

Figure 9: The set of planes that *explain* **a single data point** $\vec{x}$**, that means which have zero distance to it. The vector** $\beta \cdot \vec{n}$**, called the** *descriptor* **of a plane** E**, forms a rectangular triangle with** $\vec{x}$ **as hypotenuse. This means the descriptors of all planes explaining a data point form a sphere through** $\vec{x}$ **and the origin of the coordinate system (Thales' theorem), of which here a cut in two dimensions is shown.**

Herein res_θ and res_β are the precision of the discretization. In the actual choice of their values a tradeoff between runtime of the calculation and precision has to be kept in mind. Numerical investigations showed that all ready a value of $res_\theta = res_\beta = 3.6$ gave very accurate results.

In dealing with different, anisotropic estimates for the standard deviation of the error we calculated the relative and absolute error in the data for each single point, using the eigenvalues of the covariance matrix Σ_i, and took the maximum of the resolution.

D   Visualization Example

D.1   Plane Posterior for a Single Point

To get an intuition how the plane posterior for a single point (7) looks like, we draw a figure that shows which planes are able to "explain" a given data point perfectly. This means we ask for the subset of planes that have zero distance to the data point. The distance of a point $\vec{x}$ to a plane $E(\vec{n}, \beta)$ computes to $|\vec{x} \cdot \vec{n} - \beta|$. If we set it to zero we can see that the vector $\beta \cdot \vec{n}$, called the *descriptor* of the plane, has to form an rectangular triangle with $\vec{x}$ as hypotenuse. From Thales' theorem we see that the set of descriptors belonging to planes that explain a given data point is a *sphere* through the data point and the origin of the coordinate system. Figure 9 shows this in two dimensions, rotating the drawing around $\vec{x}$ gives the sphere.

From this considerations we see that the probability density of a single data point attains a constant, maximal value on the Thales' sphere and decreases in accordance to the covariance parameters Σ of the error model.

D.2   Plane Posterior of a Mixture

In Figure 10 as an example we visualize the posterior of the Thales sphere resulting from a single data point, as discussed in section 2.2. In Figure 10.a the sphere part of zero radius
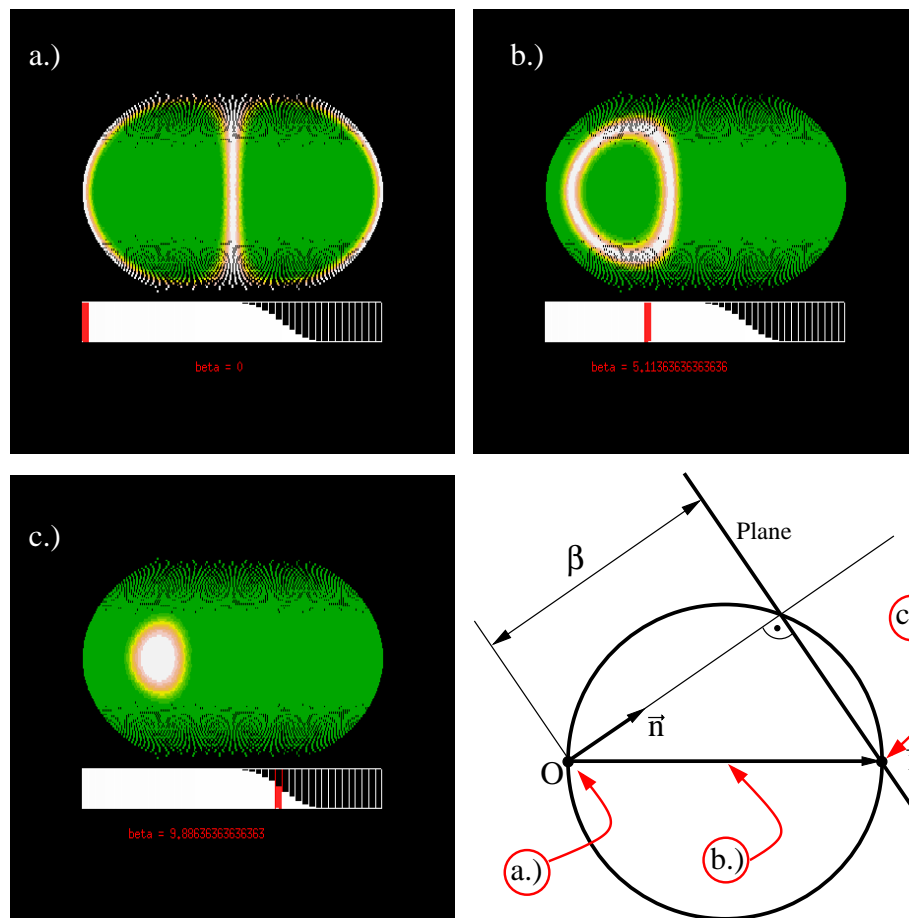
Figure 10: Visualization of the posterior resulting from a single data point $\vec{x} = (0, 10, 0)$. In three screen shots the sphere part of the posterior for values of β corresponding to the origin of the coordinate system (Part a.), the position of the data point itself (Part c.) and an intermediate value are (Part b.) shown. The red arrows in the overview diagram indicate which screen shot belongs to which position in the Thales sphere. The error model was isotropic with $\sigma = 1$.

$\beta = 0$, which is a great circle along the "Greenwich meridian" of the map, can be seen. Figure 10.b shows the case for an intermediate value of β lying in between the data points $\vec{x}$ and the origin of the coordinate system. The intersection between the cone defined by the Thales sphere and the space of unit normals $\vec{n}$, which is the sphere part, now forms a circle. If β equals the distance of point $\vec{x}$ to the origin, as in Figure 10.c, this circle degenerates to a point that because of the error model is visible as a round blob.

One of the advantages of the robust posterior (2.3) is the possibility of investigating multiple linear relationships resulting in multi-modal posteriors. To demonstrate this in Figure 11 we look at a situation in which 100 data points are scattered on one plane (green points) and 50 data points are scattered on an other. In Figure 12.a and 12.b we see the posterior at the respective values of β. The bimodality is clearly visible from the $beta$-marginal and the sphere part.
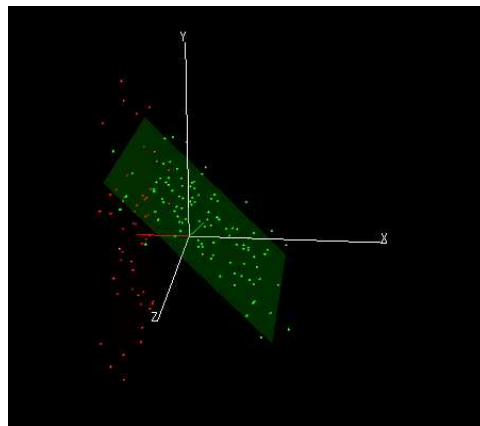
Figure 11: Synthetic data for a mixture of two planes with $\phi = \frac{\pi}{4}$, $\theta = \frac{\pi}{2}$, $\beta = 2$ (green, 100 data points) and $\phi = \pi$, $\theta = \frac{\pi}{2}$, $\beta = 5$ (red, 50 data points). The data points are normally scattered on the planes with a standard deviation of $\sigma = 5$, in addition isotropic Gaussian noise with standard deviation of $\sigma = 1$ is added.
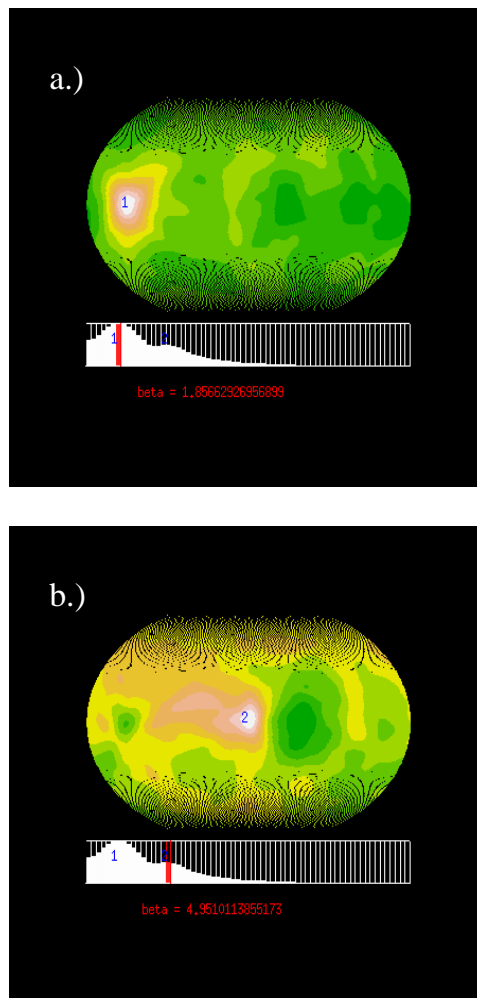
Figure 12: Posterior for the synthetic data in Figure 11. An isotropic error model with standard deviation of $\sigma = 1$ is assumed, the out lier factor is $c = 1$. Part a.) shows the sphere part for $\beta = 1.86$ and Part b.) for $\beta = 4.95$, which are approximately the values of the corresponding planes, indicated by blue integer numbers in the map. The bi-modality is clearly visible from the sphere part and the β-marginal.
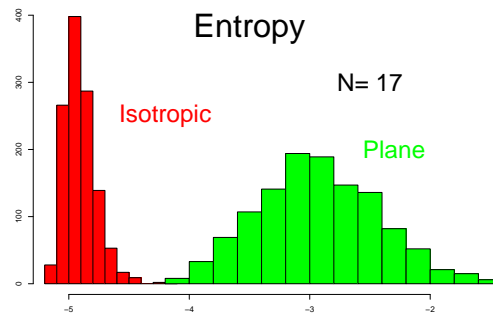
Figure 13: Distribution of the entropy (32) under the null hypothesis H_0 "data stray around plane" and alternative hypothesis H_1 "data are isotropically scattered in space", see definition (E.1), for $N = 17$ data points. Histogram sampled over 1200 repetitions.

E Posterior Based Test Functions

E.1 Concentration Based Test Functions

We present two different concentration based test functions. The first one is based on measures of concentration of the posterior like the *entropy*

$$S(\{\vec{x}_i\}) = \int p(E|\{\vec{x}_i\}, \Sigma, c) \cdot \ln p(E|\{\vec{x}_i\}, \Sigma, c) \ d\mu(E) \qquad (32)$$

or the *polynomial measure of concentration*:

$$PMOC(\{\vec{x}_i\}) = \int p(E|\{\vec{x}_i\}, \Sigma, c)^q \ d\mu(E) \qquad (33)$$

Herein q is the exponent of the polynomial, that in our study was constantly set to $q = 0.5$.

We investigated the capability of these functions to distinguish between certain plane related structures by defining the two hypotheses:

H_0: [Null Hypothesis] Data points are first casted isotropically on a square plane of size 20×20 and afterwards are subjected do a 3-dimensional Gaussian error with standard deviation $\sigma = 1$

H_1: [Alternative Hypothesis] Data points are scattered isotropically in a cube of size $20 \times 20 \times 20$

Then we repeatedly drew samples of N data points and, assuming an isotropic error with standard deviation $\sigma = 1$, we calculated the test functions, in this way obtaining a sample from the test distribution (13). As can be seen from Figure 13 and Figure 14 the entropy as well as the polynomial measure of concentration were indeed able to separate the two hypotheses under the given conditions.
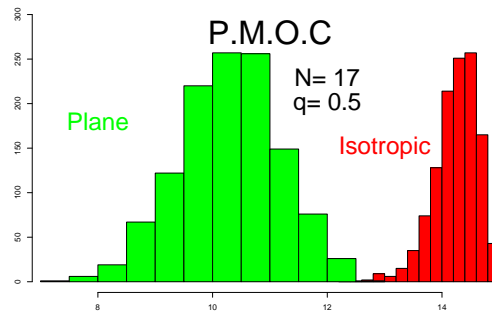
Figure 14: Distribution of the polynomial measure of concentration (33) under the null hypothesis H_0 "data stray around plane" and alternative hypothesis H_1 "data are isotropically scattered in space", see definition (E.1), for $N = 17$ data points and exponent $q = 0.5$. Histogram sampled over 1200 repetitions.

E.2 Curvature Based Test Function

The second approach for constructing test functions is based on measures of the local curvature of the sphere part of the posterior, calculated in the immediate neighborhood of its global maximum.

To define the invariant local measures of curvature, the *mean curvature M* and the *Gaussian curvature G*, we use the sphere part of the posterior that remains if one fixes β to the value $\beta^\star$ where the posterior attains its global maximum:

$$p(\vec{n}, \beta^\star | \{\vec{x}_i\}, \Sigma, c) \;\; \text{with} \;\; (\vec{n}^\star, \beta^\star) \;=\; \arg\max_{\vec{n},\beta}[p(\vec{n}, \beta | \{\vec{x}_i\}, \Sigma, c)] \tag{34}$$

After parameterization (18) this is a function of the two angles θ and ϕ. We concentrate our interest into the immediate neighborhood around the global maximum of the posterior, that occurs at $\theta^\star$ and $\phi^\star$. From the values of (34) in this neighborhood we can now compute the local values of the main radii of curvature R_1 and R_2 at the point $(\theta^\star, \phi^\star, \beta^\star)$ [1].

From the main radii of curvature we obtain the *invariants* of the curvature tensor, the mean curvature

$$M \;=\; \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{35}$$

and the Gaussian curvature:

$$G \;=\; \frac{1}{R_1 \cdot R_2} \tag{36}$$

The definition of the test function itself is based on the observation that, in case the data are in fact straying around a plane, the sphere part (34) forms a circular "blob" around the global maximum. This circular symmetry implies that the main radii of curvature are equal $R_1 = R_2$, which leads to following simple relationship between mean and Gaussian curvature:

$$G \;=\; M^2 \tag{37}$$

In the left part of Figure 15 we illustrate this by plotting the Gaussian against the mean curvature for data points sampled from different, synthetically generated hypotheses, namely data stray around a plane, a line or a point. The exact mathematical definition of the hypotheses is
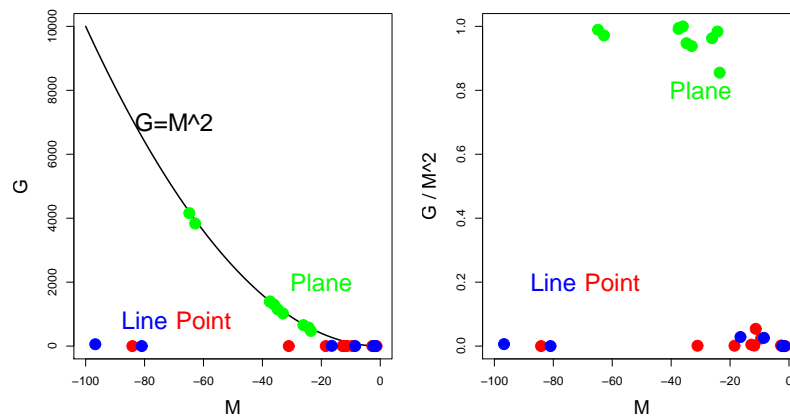
Figure 15: Left: Values of mean curvature M and Gaussian curvature G for data points sampled from different statistical hypotheses, which are: Data stray around a plane (green), a line (blue) and a point (red). Precise mathematical definitions of these hypotheses and parameters are given in section 3.2. The plane generated values clearly lie near the $G = M^2$ curve given by (37). Right: Values of the curvature based test statistics $\frac{G}{M^2}$ defined in (38) plotted against mean curvature M for the same data as in the left part.

given in section 3.2. We see, that the values for plane generated data points approximately lie on the curve predicted by (37), while for the two remaining hypotheses the Gaussian curvature vanishes.

This leads us to the definition of a *curvature based test function* by:

$$T_c(\{\vec{x}_i\}) \;=\; \frac{G}{M^2} \tag{38}$$

By definition this lies in the $[0, 1]$ interval. In the right part of Figure 15 we plot values for the given data points and notice its discriminating power under the given circumstances.

F   Quality of Covariance Estimation

The error covariances Σ have to be estimated from a number of $N = 8$ technical replicates $\{\vec{x}_n\}$, with $\vec{x}_n \in \mathbb{R}^d$. Assuming a Gaussian model for the in-replicate variablity

$$p(\{\vec{x}_n\}|\mu, \Sigma) \;\; \Pi_{n=1}^N \; N(\vec{x}_n|\mu, \Sigma) \tag{39}$$

and the prior $p(\Sigma) \propto |\Sigma|^{-1}$ we can apply the theorem of Bayes. Integrating over the mean value μ, which is unimportant in this context, we find that the posterior of Σ is an inverse Wishart function:

$$p(\Sigma|\{\vec{x}_n\}) \;=\; W^{-1}(\Sigma|\nu, \Psi) \tag{40}$$

with $\nu \;=\; N - d$, here $d = 3$, $\Psi \;=\; \Sigma_{n=1}^N X_n - \bar{X}$ with $X_n \;=\; \vec{x}_n^T o \vec{x}_n$ and $\bar{X} \;=\; N^{-1} \Sigma_{n=1}^N X_n$. Using a sample of 10.000 points from this we arrive at marginal posterior quantiles for each entry of the covariance matrices. Results given in Table 1 are the median and the ratio of the

75% quantile to the 25% quantile of the marginal posteriors for Σ for each triplet. They indicate that the quality of the estimation, in which the choice of the standard estimator for Σ is founded, is sufficient for practical purposes.

# Triplet.	$\Sigma_{1,1}$		$\Sigma_{2,2}$		$\Sigma_{3,3}$	
	75% / 25%	median	75% / 25%	median	75% / 25%	median
1	1.36	22600	1.36	5630000	1.36	79200
2	1.36	22500	1.36	5650000	1.36	1700000
3	1.36	22400	1.36	5660000	1.36	70600
4	1.36	22500	1.36	5610000	1.35	185000
5	1.37	22300	1.36	79100	1.36	1690000
6	1.36	22500	1.36	79600	1.36	70500
7	1.37	22500	1.37	79400	1.36	185000
8	1.35	22500	1.36	1690000	1.36	70500
9	1.35	22500	1.36	1690000	1.36	185000
10	1.36	22300	1.36	70200	1.36	185000
11	1.36	5640000	1.37	79300	1.35	1690000
12	1.36	5650000	1.36	79100	1.36	70200
13	1.36	5660000	1.37	79600	1.36	184000
14	1.36	5640000	1.36	1690000	1.36	70400
15	1.36	5660000	1.36	1700000	1.37	185000
16	1.36	5640000	1.36	70500	1.36	185000
17	1.36	79400	1.36	1690000	1.37	70200
18	1.35	79700	1.36	1690000	1.36	185000
19	1.36	79400	1.35	70200	1.36	184000
20	1.36	1700000	1.36	70600	1.37	186000

# Triplet.	$\Sigma_{1,2}$		$\Sigma_{2,3}$		$\Sigma_{3,1}$	
	75% / 25%	median	75% / 25%	median	75% / 25%	median
1	1.59	194000	0.683	-467000	0.62	-21700
2	1.58	194000	1.36	3080000	1.59	104000
3	1.58	193000	2.17	187000	1.4	32700
4	1.56	194000	1.38	946000	1.52	38400
5	0.618	-21600	0.677	-251000	1.58	103000
6	0.615	-21700	0.155	-10700	1.41	32600
7	0.616	-21600	0.668	-78700	1.53	38400
8	1.59	104000	2.27	97500	1.4	32600
9	1.57	104000	1.38	523000	1.52	38400
10	1.41	32500	1.9	41000	1.55	38200
11	0.687	-469000	0.682	-252000	1.35	3070000
12	0.686	-469000	0.16	-10800	2.15	186000
13	0.686	-468000	0.676	-78500	1.38	949000
14	1.36	3070000	2.22	97500	2.13	187000
15	1.37	3080000	1.38	524000	1.38	952000
16	2.17	186000	1.89	41400	1.38	947000
17	0.677	-253000	2.29	96900	0.18	-10900
18	0.687	-254000	1.37	524000	0.676	-78800
19	0.156	-10900	1.87	41500	0.675	-78600
20	2.28	97800	1.92	41600	1.38	526000

Table 1: Estimation of covariances from technical replicates. The median and the the ratio of the 75% quantile to the 25% quantile of the marginal covariance posteriors for $\Sigma_{1,1}$ ... $\Sigma_{3,1}$.