1 Curved Gaussian Models with Application to the Modeling of Foreign Exchange Rates

Juan K. Lin
Department of Mathematics
Massachusetts Institute of Technology
jklin@math.mit.edu

Peter Dayan
Gatsby Computational Neuroscience Unit
University College London
dayan@gatsby.ucl.ac.uk

Gaussian distributions lie at the heart of popular tools for capturing structure in high dimensional data. Standard techniques employ as models arbitrary *linear* transformations of spherical Gaussians. In this paper, we present a simple extension to a class of non–linear, volume preserving transformations which provides an efficient local description of curvature. The resulting generalized Gaussian models give a simple statistical tool for measuring deviations from multivariate Gaussian distributions. Remarkably, there is a computationally efficient, analytic solution for fitting the parameters of the non–linear models. The power of this approach is demonstrated in a curvature analysis of the Asian foreign exchange market.

1.1 Introduction

In many strategies for risk management and asset allocation amongst multiple investments, the expected values and covariance structure of the returns are the fundamental statistical quantities of interest (*eg* [Burmeister et al 1994]). These quantities emerge exactly as a result of fitting Gaussian probability distributions to the data – making critical the task of understanding and broadening the nature of Gaussian fits.

Principal component analysis (PCA) is one of the main tools used in fitting Gaussian distributions. The idea underlying PCA is that high dimensional data often have lower dimensional structure, and that this structure can be described in a computationally efficient manner by finding the eigenvectors associated with the largest eigenvalues of the covariance matrix of the data. PCA is a well established multivariate modeling and signal processing tool, favored for its simplicity and ease of interpretation. Recently, for example, researchers in computational finance investigated the use of PCA for isolating the driving factors in the market (*eg* [Utans et al 1997]).

From a statistical perspective, PCA can be seen as having two linked aspects –

one which models the data with a multivariate Gaussian distribution, and the other which finds the coordinate directions that optimizes the reconstruction of the data by minimizing the variance of the data components that are not modeled. Many researchers have worked to extend the applicability of PCA to more general situations. This includes work on neural network extensions of PCA (*eg* [Karhunen and Joutsensalo 1995], [Oja 1995]), local versions of PCA (*eg* [Tipping 1997], [Kambhatla and Leen 1997]), work on Principal Curves [Hastie and Stuetzle 1989] and the support vector machine based kernel PCA [Schölkopf et al 1998]. Many of these extensions have lost the simple analytic nature of the PCA solution, relying instead on non–linear optimization algorithms. Furthermore, none have retained the dimension reduction, least squares reconstruction, and maximum likelihood density estimation aspects of PCA. Our new generalization of PCA extends the multivariate linear Gaussian models to non–linear Gaussian models with curvature parameters, while successfully retaining the desirable analytical parameter estimation, least squares reconstruction and maximum likelihood density modeling properties.

Section 1.2 discusses the basic Curved Gaussian model; section 1.3 presents the curvature analysis of the Asian foreign exchange market; and section 1.4 considers natural extensions of the model.

1.2 Curved Gaussian Model

1.2.1 Non–Linear transformation

In PCA, data are linearly projected onto a lower dimensional hyperplane, and the resulting reduced dimensional data modeled with a multivariate Gaussian distribution. Since all such models can be described by a symmetric covariance matrix which can be diagonalized through an orthogonal transformation, PCA can be considered as a model of the reduced dimensional data with a dilated then rotated spherical multivariate Gaussian distribution. To extend the linearly transformed multivariate Gaussian models, we consider subsequent compositions with non–linear volume preserving transformations, as seen in Figure 1.1.

Consider a volume preserving non–linear transformation $\mathcal{T} : \Re^n \mapsto \Re^n$ with $\mathcal{T}(\mathbf{x}) = \varphi = (\varphi_1, \varphi_2, ..., \varphi_n)$ of the form:

$$\begin{aligned}
\varphi_1 &= x_1 + k_1 \\
\varphi_2 &= x_2 + \sum_i k_{2i} m_{2i}(x_1)
\end{aligned}$$

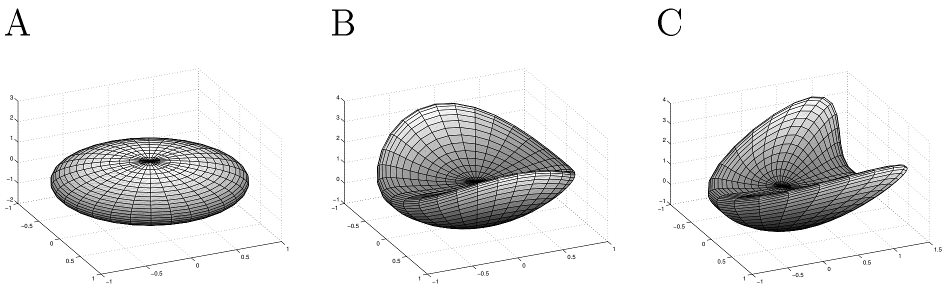

Figure 1.1
Contrasting flat and curved multivariate Gaussian models. (A) Flat Gaussian model, as defined by Principal Component Analysis and represented with an iso–density surface. (B) Curved Gaussian model, with one curved coordinate φ_3; $k_{31} = 3$, $k_{32} = 1$, all other $k_{ij} = 0$. (C) Curved Gaussian model with two curved coordinates φ_2 and φ_3; $k_{21} = 1$,$k_{31} = 3$, $k_{32} = 1$.

$$
\begin{aligned}
\varphi_3 &= x_3 + \sum_i k_{3i} m_{3i}(x_1, x_2) \\
&\vdots \\
\varphi_n &= x_n + \sum_i k_{ni} m_{ni}(x_1, x_2, \ldots, x_{n-1}) \ ,
\end{aligned}
$$

where the non–linear functions $m_{ij}(x_1, ..., x_{i-1})$ are suitably well-behaved and *fixed* basis functions of $x_1, ..., x_{i-1}$. In particular, we require that the functions $m_{ij}(x_1, ..., x_{i-1})$ and their inverses have continuous partial derivatives. The Jacobian of the transformation is lower diagonal everywhere with 1's along the diagonal, so the determinant is 1, and the transformation is volume preserving. Note that an ordering of the coordinates is assumed, with different orderings leading to different transformations.

1.2.2 Transformation Model

Since the non–linear transformation defined above and its partial derivatives are continuous by our constraint on the m_{ij}'s, a simple induction proof shows that the inverse transformation $\mathcal{T}^{-1}$ exists and also has continuous partial derivatives. By the change of variables theorem, the non–linear transformation of a multivariate Gaussian probability density function will already be normalized since the transformation considered is volume-preserving.

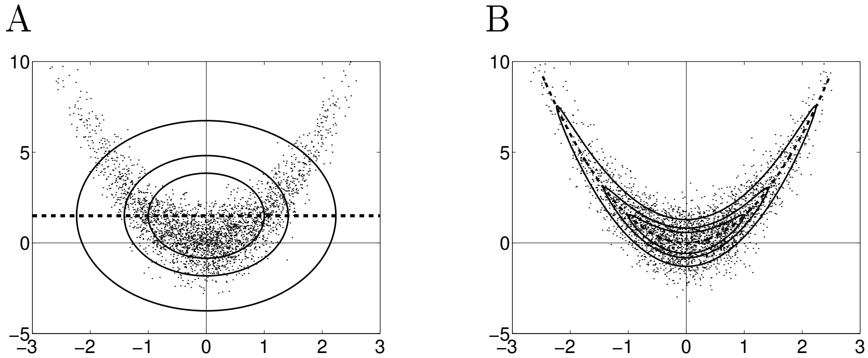

Figure 1.2
Data with a naturally curved distribution in x_1, x_2 is modeled with both a conventional two-dimensional Gaussian and a more flexible curved Gaussian model. The solid contours correspond to the σ, $\sqrt{2}\sigma$ and $\sqrt{5}\sigma$ density contours. (A) Conventional two-dimensional Gaussian model. The dashed line is the line $x_2 = 1.5$. (B) Curved two-dimensional Gaussian with $w_1 = 1$, $w_2 = 3$ and $k_{21} = 1.5$. The dashed line is the image of the line $\varphi_2 = x_2 + k_{21}x_1^2 = 0$; 3000 points were drawn from the distribution.

We are now able to define a "curved" multivariate Gaussian likelihood model of the form

$$P(\mathbf{x}; \mathbf{w}, \mathbf{k_{ij}}) \propto (w_1...w_n)^{1/2} \exp[-w_1\varphi_1^2... - w_{n-1}\varphi_{n-1}^2 - w_n\varphi_n^2], \tag{1.1}$$

where $\mathbf{x} = (x_1, x_2, ..., x_n)$ is the principal component coordinate system of the dataset. In essence, the data is modeled by the non–linear transformation $\mathcal{T}^{-1}$ of a multivariate Gaussian as defined in Eqn. 1.1 in the $\{\varphi\}$ coordinate system. A few non–linear transformations of Gaussians are shown in Figure 1.1. These curved Gaussian models provide a much larger class of parametric models for density estimation. Remarkably, it is easy to fit these models to data.

1.2.3 Analytic Parameter Estimation

Consider fitting a model of the form Eqn. 1.1 to data $\mathcal{D} = \{\mathbf{x}^l\}$. Maximizing the log likelihood with respect to the coefficients of m_{ij} gives optimizing equations

$$\frac{\partial}{\partial k_{ij}} \langle \varphi_i^2 \rangle = 0. \tag{1.2}$$

To make the notation more compact, let $\mathbf{k_i} = (k_{i1}, k_{i2}, ..., k_{i\hat{i}})^T$ denote the coefficients in the transformation for the φ_i component. Using the expansion for φ_i given above, this implies $\langle \varphi_i m_{ij} \rangle = 0$ for $j = 1, ..., \hat{i}$. Define $\mathbf{l_i} \in \Re^{\hat{\mathbf{i}}}$, and

$\mathbf{\Gamma(i)} \in \Re^{\hat{i}} \times \Re^{\hat{i}}$ as follows

$$\mathbf{l_i} = -(\langle x_i m_{i1}\rangle, \langle x_i m_{i2}\rangle, ..., \langle x_i m_{i\hat{i}}\rangle)^T, \tag{1.3}$$

$$\mathbf{\Gamma(i)}_{jk} = \langle m_{ij} m_{ik}\rangle. \tag{1.4}$$

The maximum likelihood condition can be expressed as $\mathbf{\Gamma(i)k_i} = \mathbf{l_i}$, with solution $\mathbf{k_i} = \mathbf{\Gamma(i)^{-1}l_i}$. Or more compactly

$$k_{ij} = \left[\mathbf{\Gamma(i)}^{-1}\mathbf{l_i}\right]_j. \tag{1.5}$$

This solution, if it exists, is the unique global maximum of the log likelihood.

The maximum likelihood solution with respect to the inverse variance parameters are

$$w_i = \frac{1}{2\langle\varphi_i^2\rangle}. \tag{1.6}$$

Surprisingly, just like in Principal Component Analysis, we have analytic solutions for all of the model parameters. Furthermore, the determination of these parameters only involve simple linear algebra – this is a significant advantage over the numerous extensions of PCA which rely on non–linear optimization algorithms. Figure 1.2 compares a two dimensional curved Gaussian model to a flat Gaussian model of data drawn from a naturally curved distribution.

In selecting the non–linear functions $m_{ij}(x_1, ..., x_{i-1})$, it is simplest to consider the case where they are products of non-negative powers of the variables $\{x_1, ..., x_{i-1}\}$. More specifically, let m_{ij}'s be multinomials

$$m_{ij} = \prod_{k=1}^{i-1} x_k^{c_k}, \tag{1.7}$$

with $c_k \geq 0$ for all k. The restriction to positive powers in the polynomial ensures that there are no singularities in the transformation. In the numerics presented in this paper, we considered quadratic non–linear transformations where $m_{ij} = x_j^2$ for $0 < j < i$ and $m_{i0} = 1$. The parameters k_{ij} in this case have an intuitive interpretation as curvatures.

1.2.4 Least Squares Reconstruction

The optimizing Eqns. 1.5–1.6 reveal a direct connection with least squares. From Eqn. 1.2, the maximum likelihood solution coincides with an extremum for $\langle\varphi_i^2\rangle$, viewed as a function of the multinomial coefficients $\mathbf{k_i}$. Only minima exist since $\langle\varphi_i^2\rangle$ is unbounded above, so the maximum likelihood solution is the least squares solution. Parameter estimation for the model thus corresponds to a set of n uncoupled least squares fits.

By specifying that $m_{i0} = 1$, we impose the condition $\langle \varphi_i \rangle = 0$. Because the coefficients $\mathbf{k_i}$ minimize $\langle \varphi_i^2 \rangle$, this implies $\langle \varphi_i^2 \rangle \leq \langle x_i^2 \rangle$, and hence $Var(\varphi_i) \leq Var(x_i)$ since $\langle \varphi_i \rangle = \langle x_i \rangle = 0$. Applying the curved Gaussian model to dimension reduction therefore guarantees an improved least squares reconstruction of the data over conventional PCA approaches.

1.2.5 Sequential Flattening

Finally, because $\mathbf{k_i}$ and $\mathbf{k_{i'}}$ are uncoupled for $i \neq i'$, the multinomial coefficients $\mathbf{k_i}$ can be computed in parallel, completely independently of each other. Alternatively, consider single coordinate non–linear transformations $\mathcal{T}^j : \Re^n \mapsto \Re^n$, $\mathcal{T}^j(\mathbf{x}) = \rho = (\rho_1, \rho_2, ..., \rho_n)$ defined as follows:

$$\begin{aligned}
\rho_j &= x_j + \sum_i k_{ji} m_{ji}(x_1, x_2, \ldots, x_{j-1}), \\
\rho_i &= x_i, \;\; i \neq j.
\end{aligned}$$

The composition $\mathcal{T}^n \circ \mathcal{T}^{n-1} \circ ... \circ \mathcal{T}^1$ of single coordinate transformations couples the $\mathbf{k_j}$'s and sequentially transforms the data starting with the x_1 component. After a simple rewriting of the equations, the resulting composite transformation is given by $\varphi = \mathcal{T}^n \circ \mathcal{T}^{n-1} \circ ... \circ \mathcal{T}^1(\mathbf{x})$ with

$$\begin{aligned}
x_1 &= \varphi_1 - k_1 \\
x_2 &= \varphi_2 - \sum_i k_{2i} m_{2i}(\varphi_1) \\
&\vdots \\
x_n &= \varphi_n - \sum_i k_{ni} m_{ni}(\varphi_1, \varphi_2, \ldots, \varphi_{n-1}) \;.
\end{aligned}$$

It is now clear that we are in fact directly parameterizing the transformation of the multivariate Gaussian instead of its inverse, and that we can model data with this direct parameterization of the transformation of a multivariate Gaussian distribution by solving for $\mathbf{k_1}$,$\mathbf{k_2}, ..., \mathbf{k_n}$ in sequence. Intuitively, this analysis "sequentially flattens" the data one coordinate at a time, in the given ordering of the coordinates.

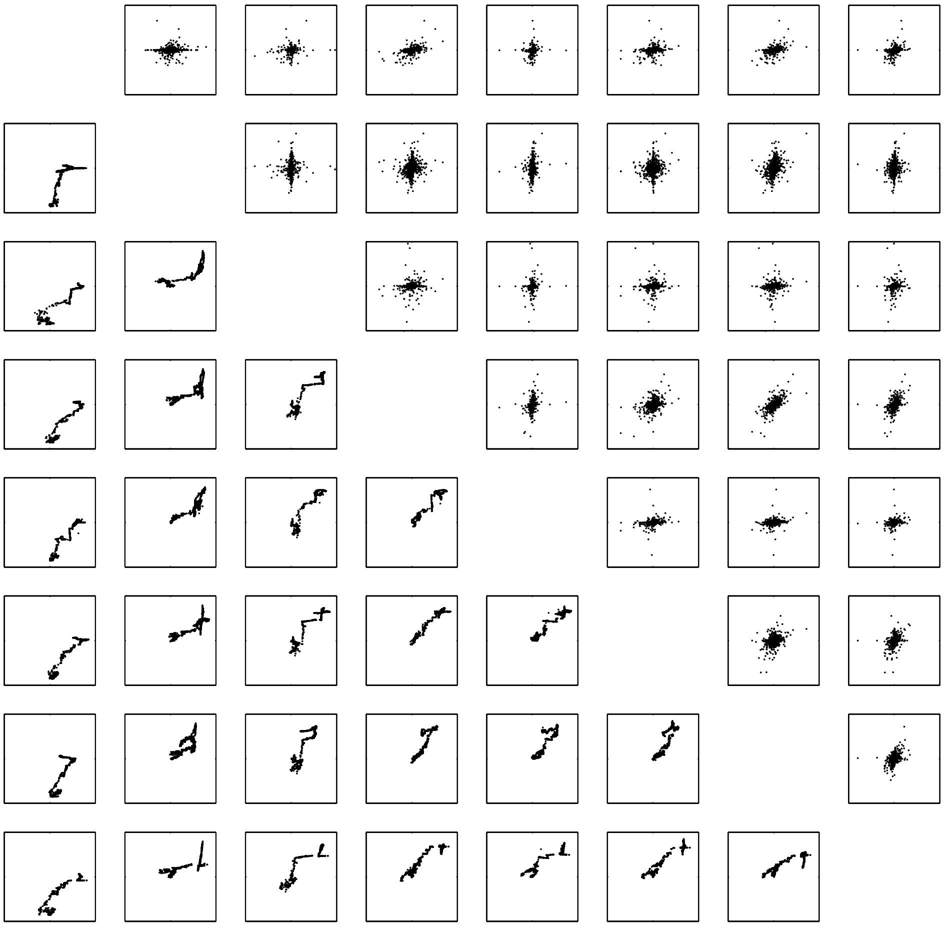

Figure 1.3
Pairwise plots of the value in U.S. Dollar of various Asian currencies. Lower diagonal and upper diagonal plots correspond to price and return data respectively. Beginning with the first column/row from the top left corner, the currencies are the Indonesian Rupiah, Japanese Yen, Korean Won, Malaysian Ringgit, Taiwan Dollar, Philippine Peso, Singapore Dollar, and the Thai Baht. The corresponding column identification is plotted as a function of the row identification.

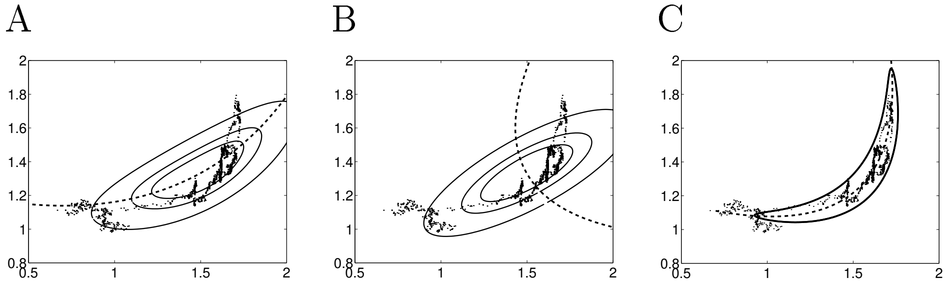

Figure 1.4
Analysis of Japanese Yen vs. Korean Won price data. Plotted are curved Gaussian models starting from the principal component coordinate system with (A) the smallest variance principal component taken as the curvature direction, (B) the largest variance component as the curvature direction. In units of inverse U.S. Dollar, the curvature constants are $k = .35$ in A, and 1.03 in B, corresponding to dimensionless curvatures of .30 and .03 respectively. The dimensionless curvature gives a much better handle on the significance of the curvatures because of the normalization relative to the σ_i length scales. The solid lines are the σ, $\sqrt{2}\sigma$ and $\sqrt{5}\sigma$ contours of the model, while the dashed line shows the transformation of the $\varphi_2 = 0$ line. The analytic fit in (A) captures some of the curvature, though qualitatively there appears to be more curvature. This is partly due to the asymmetric concentration of the data. (C) Additional parameterization of the coordinate system, as described in section 1.4, results in a better curvature fit of the data.

1.3 Curvature Analysis of Asian Foreign Exchange Rates

1.3.1 Data Preprocessing

We investigated the Asian foreign exchange market consisting of the Indonesian Rupiah, Japanese Yen, Korean Won, Malaysian Ringgit, Taiwan Dollar, Philippine Peso, Singapore Dollar, and the Thai Baht. Daily price values of the various currencies as measured in U.S. Dollar were taken from October 23, 1993 to August 27, 1998, corresponding to a dataset of 1769 eight-dimensional datapoints. They were first normalized relative to their values on August 27, 1998, resulting in a multi–currency portfolio with components weighted equally according to the exchange rates of that date. The relative return data was calculated by taking differences of the logarithms of the normalized price data. Two–dimensional projections of both the price and return data are shown in Figure 1.3.

1.3.2 Japanese Yen vs. Korean Won

From Figure 1.3, curvature is clearly present is some pairwise price data. A simple curved Gaussian model of the Japanese Yen vs. Korean Won data is presented in Figure 1.4, with significant curvature fits of $k = 0.35$ and 1.03 in units of inverse U.S. Dollar for the two possible choices of the curved coordinate (*ie* the two possible orderings of the principal components). From the plot, we see that the magnitude of k is not directly indicative of its importance in the model fit. In Figure 1.4A the curvature in the fitted model describes a significant non–linear warping in the data of the curved coordinate relative to the standard deviation of the curved coordinate. In comparison, for the model in Figure 1.4B, the large value of k is due to the small variance in the non–curved coordinate. Clearly the $\{k_{ij}\}$ parameters need to be considered relative to the standard deviation length scales, as determined by the $\{\mathbf{w}\}$ parameters. In order to better represent the magnitudes of the curvature coefficients, we define the dimensionless curvature as $k^*_{ij} = k_{ij}\sigma_j^2/\sigma_i$, where the standard deviation is related to the $\{\mathbf{w}\}$ parameters by $w_i = \frac{1}{2\sigma_i^2}$. The numerator $k_{ij}\sigma_j^2$ is a length scale along the i-th coordinate due to the curvature parameter k_{ij}. The dimensionless curvature measures this length scale relative to the standard deviation of the i-th coordinate σ_i. Justifying our intuitive definition, the dimensionless curvature in the model depicted in Figure 1.4A is 30% while it is only 3% for the model in Figure 1.4B.

To demonstrate an application of the curved Gaussian model to dimension reduction, after principal component analysis and curved Gaussian analysis of the data, we discarded the minimal variance principal component, and reduced the dimensionality of the data to seven. The PCA and curved Gaussian reconstructions of the Japanese Yen price data are then compared. The cumulative least squares reconstruction errors for the two reconstructions are shown in Figure 1.5. As expected, the least squares reconstruction error over the entire dataset of the curved Gaussian reconstruction is lower than the PCA reconstruction error.

1.3.3 Single component curved Gaussian model

In addition to the curvature analysis of a two dimensional projection of the price data, we also performed a multidimensional curvature analysis on the relative return data. Here we first performed PCA on the full eight dimensional data, then successively chose each of the eight PCA directions as the direction in which to fit a single component curved Gaussian model. Each curved model has seven parameters, one for each of the other PCA directions. Out of the total of 56 curvature parameters, dimensionless curvature values with magnitudes as high as four and five percent were found.

For the eight separate curved Gaussian models, we found the standard deviation of the curved coordinate φ_8 relative to the flat coordinate x_8 to be .916, .948, .977, .979, .989, .993, .996 and .999 respectively. As discussed in section 1.2, this result-

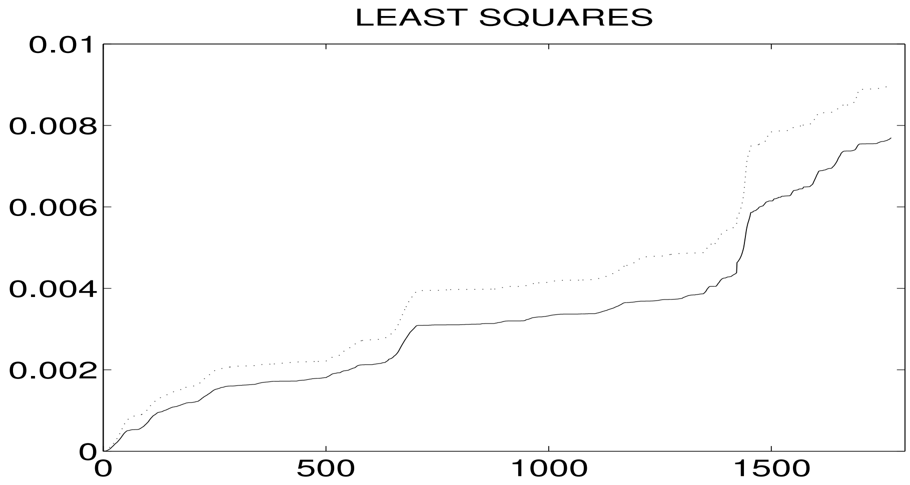

Figure 1.5
Cumulative least squares reconstruction error of the Japanese Yen price data. The minimal variance coordinate was chosen as the curved coordinate x_8 and thrown away to reduce the dimensionality of the data to 7. The dotted line corresponds to the cumulative square error of the conventional PCA reconstruction, and the solid line the lower reconstruction error achieved with the curved Gaussian reconstruction.

ing variance reduction from the volume preserving non–linear transformations is indicative of the better fits achievable with the introduction of curvature parameters.

1.3.4 Sequential flattening

Finally, we proceeded with the full multi–dimensional curved Gaussian model of the price data with a direct parameterization of the non–linear transformation of the Gaussian. Since ordering of the principal components matters, there is a total of 8! models — one for each order in which the components are flattened. A compilation of the dimensionless curvatures of all 8! models of the foreign exchange data are shown in Figure 1.6AB for the price and return data. Because the dataset consists of less than two thousand datapoints, it is important to perform the same analysis on data sets of the same size sampled from the Gaussian distribution. The resulting curvature distribution of the normally distributed data is shown in Figure 1.6C. The curvatures in the price data are an order of magnitude larger than what is expected of a Gaussian dataset of the same size. In contrast, curvatures in the return data are comparatively smaller, indicative of the sparse nature of the return data.

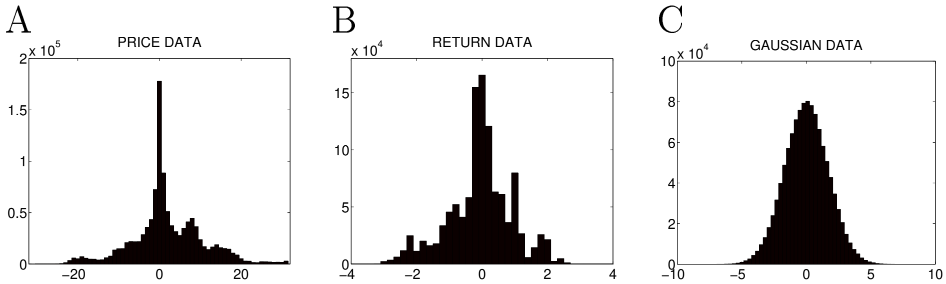

Figure 1.6
Histogram of the dimensionless curvature parameters (quoted in percentage values) in the curved Gaussian models of the price and return data. (A) Dimensionless curvatures for models corresponding to all 8! coordinate permutations of the price data. (B) Similar plot for the return data. (C) Curvature histogram for 5040 datasets of 1768 data points sampled from eight dimensional Gaussian distributions with unit variance. Eight random coordinate permutations are considered for each dataset.

1.4 Extensions

1.4.1 Linear Coordinate Transformations

The lesson from Figure 1.4B and the lack of a principled approach to ordering the flat coordinates is that the coordinate system itself should be parameterized and optimized.

Drawing inspiration from Independent Component Analysis (*eg* [Bell and Sejnowski 1995, Amari et al 1996, Cardoso and Laheld 1996, Lin et al 1997, Lin 1998]), one way to do this is to consider a curved, source datapoint $\mathbf{x}$ being used to generate an observed datapoint $\mathbf{y}$ through a linear transformation:

$$\mathbf{y} = \mathbf{W}^{-1}\mathbf{x} - \mathbf{b} \tag{1.8}$$

where $\mathbf{b}$ is a translation, and $\mathbf{W}^{-1}$ is an invertible 'mixing' transformation. Given a dataset $\mathcal{D} = \{\mathbf{y}^l\}$, the task is to find the unmixing linear transformation parameters $\mathbf{b}$ and $\mathbf{W}$, together with the parameters $\mathbf{w}$ and $\mathbf{k_j}$ of the curved source model so as to fit the data as tightly as possible. The contribution to the log likelihood from the datapoint $\mathbf{y}^l$ is

$$\log|\mathbf{W}| + \log p(\mathbf{W}(\mathbf{y}^l + \mathbf{b}), \mathbf{w}, \mathbf{k_{ij}}).$$

It does not seem possible, in general, to derive closed–form optima for the likelihood as a function of all the parameters. Instead, we consider an iterative stochastic

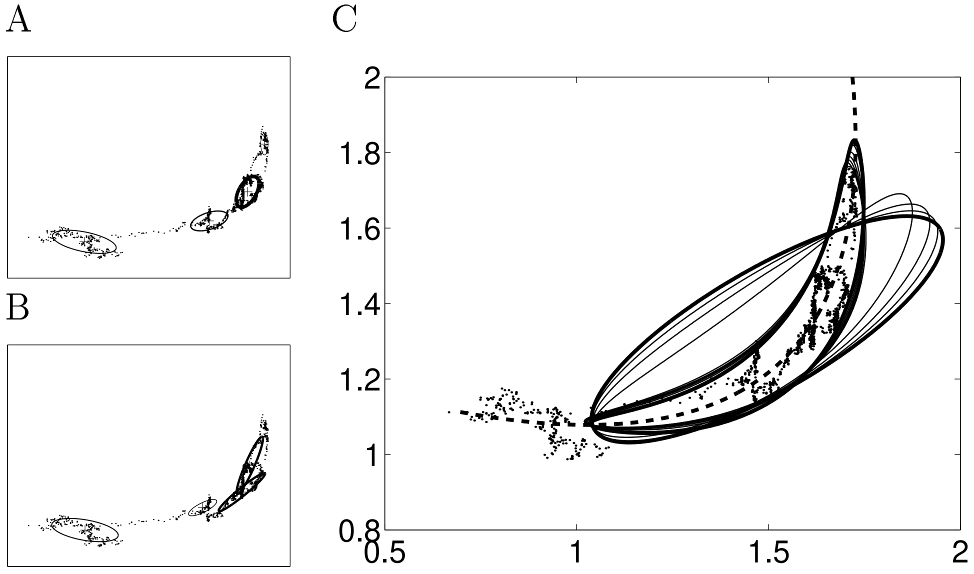

Figure 1.7
Mixture of Gaussians models for modeling the Japanese Yen *versus* the Korean Won data. (A;B) Two different mixtures of Gaussian fits to these data using four full covariance two-dimensional Gaussians. The centers of the Gaussians are marked by a cross, $\sqrt{2}\sigma$ values by the ellipses, and the mixing proportions by the widths of the outlines of the ellipses. (C) Curved model fit with coordinate system parameterization. The dashed line is the quadratic 'skeleton' of the final curved model and the thick solid line the final $\sqrt{2}\sigma$ contour for the model.

gradient ascent of the likelihood. In each step, first the global maximum of the likelihood in $\mathbf{w}$ and $\mathbf{k_{ij}}$ for the current values of $\mathbf{W}$ and $\mathbf{b}$ is calculated, then the latter variables are changed by stochastic gradient ascent. The partial derivatives of the log likelihood with respect to the mixing parameters are straightforward to calculate.

Note that there is substantial redundancy in the model. For example, the linear transformation $\mathbf{W}$ can be restricted to an arbitrary orthogonal transformation. However, the learning rule is considerably simpler with the model parameterization given above. Just as for ICA, we can use the natural gradient ascent algorithm [Amari et al 1996, Cardoso and Laheld 1996] to update the mixing transformation along the direction of steepest ascent by right multiplying the update rule for $\mathbf{W}$ by the matrix $\mathbf{W}^T\mathbf{W}$.

An example of the dynamics of the iterative curved Gaussian model fit to the Japanese Yen vs. Korean Won data is shown in Figure 1.7C. Initialized according to a linear Gaussian model, the algorithm quickly settles into an optimal curved

Gaussian fit of the data. However, as can be expected with adaptive optimization approaches, the algorithm sometimes falls into local maxima.

1.4.2 Mixture Models

Another natural extension is to consider mixtures of curved Gaussians. One of the most popular extensions of standard Gaussian models is to *mixture* models such as the mixture of Gaussians [Nowlan 1991], and in more recent work, mixtures of principal component or factor analyzers [Bregler and Omohundro 1995, Kambhatla and Leen 1997, Hinton et al 1997, Roweis and Ghahramani 1999, Tipping 1997]. Such mixture models are attractive because the expectation maximization algorithm (EM; [Dempster et al 1977]) allows them to be fit to data in a computationally efficient manner.

The curved Gaussian model can substitute exactly for flat Gaussian model in mixtures. As seen in Figure 1.7, few curved Gaussians can often capture the information contained in many linear Gaussians. The E phase of EM, in which the responsibility of each element of the mixture for each data point is assigned is straightforward because the non–linear transformation is volume preserving. The M phase of EM, in which the parameters of each element are changed to reflect their responsibilities is straightforward because of the analytical solution for the curved models presented in section 1.2.3.

1.5 Discussion

We have presented an analytic generalization of the linear Gaussian models that captures weak non–linear correlations in the data. There are various natural extensions of the work. Particularly important is the adoption of ideas from independent component analysis to infer the appropriate coordinate system in which to fit the curved model, and the notion of using *mixtures* of the curved Gaussian distributions, in the same way that one uses mixtures of standard Gaussian distributions. This presents no conceptual or computational hurdle, and can be done simply using the expectation–maximization algorithm. Another extension is that of using the curvature information for things other than fitting a Gaussian model. For instance, the information could be used to enhance local kernel methods in a curved form of tangent distance [Simard et al 1993]. Also, more general non–linear models could be used in place of the simple quadratic form used in our analysis of the foreign exchange data. To manage the trade–offs of bias for variance in assuming more flexible parameterizations, more sophisticated cross–validation or Bayesian methods could be considered for choosing the orders of the polynomials or the forms of the non–linearities.

There is great interest in multivariate statistical analysis tools beyond linear Gaussian models. The curved Gaussian models presented in this paper provide a simple way of looking at non–linearities in the data. Since the curvature parameter fit is still analytic, all current applications of PCA will benefit from the added flexibility. Although the subsequent non–linear transformation might be very close to the identity for some datasets, the existence of a computationally inexpensive analytic solution strongly motivates the consideration of these curvature parameters. In conclusion, the curved Gaussian models provide an extremely simple tool for probing and characterizing deviations of the data from a multivariate Gaussian distribution. The alternative of considering higher order multivariate moments, which are higher order tensors, quickly runs into the curse of dimensionality. Finally, with respect to computational finance, this curvature modeling provides the fundamental basis for non–linear asset allocation strategies and new non–linear financial products. We believe curved Gaussian models will be a very useful multivariate statistical modeling and signal processing tool.

References

Amari, S., A. Cichocki, and H. Yang. 1996. A new learning algorithm for blind signal separation. In Touretzky, D. S., M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural and Information Processing Systems, 8*, pages 757–763, Cambridge, MA. MIT Press.

Bell, A. J. and T. J. Sejnowski. 1995. An information–maximizatin approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159.

Bregler, C. and S. M. Omohundro. 1995. Nonlinear image interpolation using manifold learning. In Tesauro, G., D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural and Information Processing Systems, 7*, pages 971–980, Cambridge, MA. MIT Press.

Burmeister, E., R. Roll, and S. A. Ross. 1994. A practitioner's guide to arbitrage pricing theory. *Finanzmarkt und Portfolio Management*, 8 Jahrgand Nr. 3:312–331.

Cardoso, J.-f. and B. Laheld. 1996. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 45:2:434–444.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, series B 39:1–38.

Hastie, T. and W. Stuetzle. 1989. Principal curves. *Journal of the American Statistical Association*, 84 no. 406:502–516.

Hinton, G. E., P. Dayan, and M. Revow. 1997. Modeling the manifolds of images of handwritten digits. *IEEE Transactions on Neural Networks*, 8:65–74.

Kambhatla, N. and T. K. Leen. 1997. Dimension reduction by local principal component analysis. *Neural Computation*, 9:1493–1516.

Karhunen, J. and J. Joutsensalo. 1995. Generalizations of principal component analysis, optimization problems and neural networks. *Neural Networks*, vol 8 no 4:549–562.

Lin, J. K. 1998. Factorizing multivariate function classes. In Jordan, M. I., M. J. Kearns, and S. A. Solla, editors, *Advances in Neural and Information Processing Systems, 10*, pages 563–569, Cambridge, MA. MIT Press.

Lin, J. K., D. G. Grier, and J. D. Cowan. 1997. Faithful representation of separable distributions. *Neural Computation*, 9:1303–1318.

Nowlan, S. J. 1991. Soft competitive adaptation: neural network learning algorithms based on fitting statistical mixtures. Technical Report PhD Thesis, Department of Computer Science, Carnegie–Mellon University.

Oja, E. 1995. The nonlinear pca learning rule and signal separation - mathematical analysis. Technical Report Report A26, Helsinki University of Technology, Laboratory of Computer and Information Science, Helsinki.

Roweis, S. and Z. Ghahramani. 1999. A unifying review of linear gaussian models. *Neural Computation*, 11:305–345.

Schölkopf, B., A. Smola, and K. Müller. 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319.

Simard, P., Y. LeCun, and J. Denker. 1993. Efficient pattern recognition using a new transformation distance. In Cowan, J. D., S. J. Hanson, and C. L. Giles, editors, *Advances in Neural and Information Processing Systems, 5*, pages 50–58, San Mateo, CA. Morgan Kaufmann.

Tipping, ME & Bishop, C. 1997. Mixtures of principal component analyzers. *Fifth International Conference on Artificial Neural Networks.*

Utans, J., W. T. Holt, and A. N. Refenes. 1997. Principal component analysis for modeling multi–currenty portfolios. In Weigend, A. S., Y. S. Abu-Mostafa, and A.-P. N. Refenes, editors, *Proceedings of the fourth international conference on neural networks in the capital markets*, pages 359–369, Singapore. World Scientific.