FAST CONVERGENT FACTORIAL LEARNING OF THE LOW-DIMENSIONAL INDEPENDENT MANIFOLDS IN OPTICAL IMAGING DATA

Penio S. Penev

The Rockefeller University
1230 York Avenue
New York, NY 10021
PenevPS@IEEE.org
http://venezia.rockefeller.edu/

Manuela Gegiu and Ehud Kaplan

Mount Sinai School of Medicine
One Gustave L. Levy Place
New York, NY 10029
kaplane@rockvax.rockefeller.edu
http://camelot.mssm.edu/

ABSTRACT

In many functional-imaging scenarios, it is a challenge to separate the response to stimulation from the other, presumably independent, sources that contribute to the image formation. When the brain is optically imaged, the typical variabilities of some of these sources force the data to lie close to a low-dimensional, nonlinear manifold. When an initial probability model is derived by the Karhunen-Loève Transform (KLT) of the data, and some factors of this manifold happen to be accessibly embedded in suitably chosen KLT subspaces, vector quantization has been used to characterize this embedding as the locus of maximum likelihood of the data, and to derive an improved probability model, in which the factors—the dynamics on this locus and away from it—are estimated independently. Here we show that such a description can serve as the starting point for a convergent procedure that alternatively refines the estimates of the embedding of, and the dynamics on, the manifold. Further, we show that even a very crude initial estimate, from a heavily mixed subspace, is sufficient for convergence in a small number of steps. This opens the possibility of hierarchical semi-blind separation of the independent sources in optical imaging data, even when their contributions are nonlinear.

1. INTRODUCTION

Functional imaging is an increasingly popular tool for the study of biological systems. An important category is *optical imaging of the brain,* in which the light reflected from a piece of brain tissue is correlated with some parameter of a stimulus that is concurrently presented to the animal (see, *e.g.,* [4]). It is a big challenge to separate, from the overwhelming variability due to other sources, the part of the signal that is due to the response of the system to stimulation; this is a *source separation* problem.

Part of this research was made possible by the William O' Baker Fellowship, so generously extended to, and gratefully accepted by, PSP. He is also indebted to M. J. Feigenbaum for his hospitality, encouragement, and support—scientific and otherwise. EK is the *Jules & Doris Stein Research-to-Prevent-Blindness Professor* at MSSM. The quality of this paper has benefited substantially from the detailed comments of B. W.Knight, which are gratefully acknowledged. We are thankful to A. J. Libchaber, J. D. Victor, A. Fairhall, D. W. Dong, A. K. Prashanth, C. Sailstad, and A. T. Sornborger, for useful discussions and comments, to D. Samber and B. Z. Goldsteen, for invaluable technical support, and to E. D. Mandel, for help with some of the computer programming.

A current method of choice is the *indicator function method* [3, 4], which is a variant of *Linear Discriminant Analysis* and can be traced back to Fisher [6]. It is optimal under a *multidimensional Gaussian* assumption about the probability distribution of the mixed sources, and is based on the *Karhunen-Loève Transform (KLT)* [9, 12].

For non-Gaussian sources, the higher-order moments of the distributions may be used to aid in the separation; this has given rise to a variety of methods under the broad name of *Independent Component Analysis (ICA),* reviewed recently in [1, 7]. In ICA, one tries to decompose the total signal into a sum of "components" that span 1-dimensional *linear subspaces,* not necessarily orthogonal, and whose temporal dynamics are mutually *independent.*

It has been shown recently [16] that the probability density of the dynamics of the optical signal clusters near low-dimensional *curved manifolds,* embedded in suitably chosen higher-dimensional KLT subspaces. *Vector quantization (VQ)* has been used to characterize these distributions and to decompose the signal into a *factorial* representation in terms of: *nonlinear* parameters within these *independent manifolds,* and a residual that is amenable to subsequent analysis [13, 16].

Here we introduce an iterative method to refine the estimates of the embedding parameters of the manifold on one hand, and the dynamics along it, on the other. We treat the VQ description as a starting point in this procedure, and show that it converges rapidly. Further, in a more generally applicable scenario, we start from a KLT subspace in which the nonlinear manifold is heavily mixed with other sources of variability, and the initial VQ model can be treated as no more than an educated guess. We show that, even under such adverse conditions, the estimation of the manifold converges in a small number of steps to the correct solution.

Finally, we apply the procedure hierarchically to the residual, in a step towards the building of a complete, non-linear, generative model of the dynamics of the optical signal, in which all sources of variability—vegetative, equipment generated, response to stimulation, and Gaussian noise—are modeled as such manifolds, and the parameters of those models are estimated from the data set through an iterative *analysis by synthesis* approach.

The independent-manifolds framework presented here is related to ICA, and also to two other approaches that have been proposed recently. With the hypothesis that the dynamics of a certain source occupies a particular *temporal frequency band,* short-term *multi-taper spectral estimation* [20] can be used in the context of *space-frequency PCA* [14] to find an estimate of the linear subspace in which the manifold is embedded (*e.g.,* [11]). Also, relatively good separation has been achieved by the enforcing of exact spatio-temporal decorrelation across a whole range of time lags [2]. In the framework presented here, the results of these approaches can serve as good starting points for the convergent learning of the curved manifolds and their associated dynamics.

2. THE GAUSSIAN MODEL

The optical signal due to the cortical activity[1] will be represented by the image intensity values $\phi(\mathbf{x})$, where $\{\mathbf{x}\}$ is a pixel grid that contains V pixels. An *ensemble* of T snapshots will be denoted by $\{\phi^t(\mathbf{x})\}_{t\in T}$. Briefly (see, *e.g.,* [15] for details), its KLT representation is given by

$$\phi^t(\mathbf{x}) = \sum_{r=1}^{M} a_r^t \sigma_r \psi_r(\mathbf{x}) \tag{1}$$

where $M \leq \min(T, V)$ is the rank of the ensemble, $\{{\sigma_r}^2\}$ (arranged in non-increasing order) is the *eigenspectrum* of the two correlation matrices

$$\begin{aligned} R(\mathbf{x},\mathbf{y}) &\stackrel{\triangle}{=} \tfrac{1}{T}\textstyle\sum_t \phi^t(\mathbf{x})\phi^t(\mathbf{y}) &&= \sum_{r=1}^{M} \psi_r(\mathbf{x}){\sigma_r}^2\psi_r(\mathbf{y}) \\ C^{tt'} &\stackrel{\triangle}{=} \tfrac{1}{V}\textstyle\sum_{\mathbf{x}} \phi^t(\mathbf{x})\phi^{t'}(\mathbf{x}) &&= \sum_{r=1}^{M} a_r^t{\sigma_r}^2 a_r^{t'} \end{aligned} \tag{2}$$

and $\{\psi_r(\mathbf{x})\}$ and $\{a_r^t\}$, the respective orthonormal eigenvectors.[2]

With the standard *multidimensional Gaussian* model for the probability density $\mathcal{P}[\phi]$ [15], the *information content* of ϕ_N^t—the reconstruction with the first $N \leq M$ terms in (1), which is optimal in the amount of captured signal power [12, 8]—is

$$-\log \mathcal{P}[\phi_N^t] \propto \sum_{r=1}^{N} |a_r^t|^2. \tag{3}$$

Notably, this model is *spherical*—the KLT coefficients are of *unit variance* (2), and all contribute equally to the information.

How much of this information is useful? For the purpose of this study, following [14, 16], we take a safe upper bound of the signal dimensionality, $N = 300$, which is well beyond the crossover to the noise regime in the eigenspectrum of the ensemble, shown in Fig. 1. This choice is supported by the *sample kurtoses* of the KLT coefficients, shown in Fig. 2, and by the spatial structure of the eigenmodes $\psi_r(\mathbf{x})$ (2) and the temporal structure of the KLT coefficients (not shown).

[1]For the illustrations in this study, the activity of a cat cortex was imaged optically as previously described [4]. Briefly, the cortex was illuminated uniformly with monochromatic light and the reflected intensity was recorded with a 12-bit charge-coupled device every 200 ms, for $T = 8500$ snapshots $\approx$ 28 min, which were subsequently cropped to $V = 378 \times 282 = 106,596$ pixels. Additionally, all of the discussed phenomena were invariably observed in all five data sets that we have analyzed so far, which were collected over a period of more than one year, in both cat and monkey cortices, and some lasting as long $T \approx 12$ hours.

[2]When $M = T < V$, the diagonalization of $\mathbf{C}$ is easier. For most of the ensembles in this paper, $M << T << V$, and substantial additional savings have been achieved, by the diagonalization of an $M \times M$ matrix.

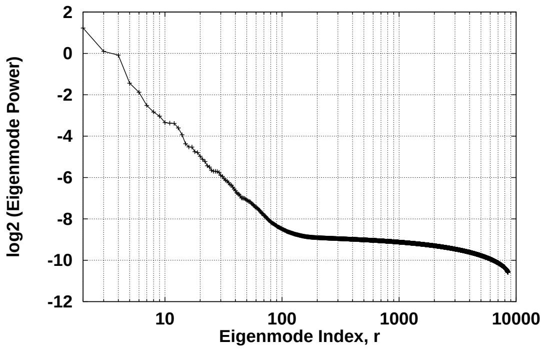

Figure 1: The eigenspectrum of the full ensemble. Two regimes are immediately obvious: an initial power-law regime, which we attribute to the signal; and a final plateau-like regime. For an infinitely long experiment, this plateau would be flat, at the level of the white-noise power. For a finite experiment, the eigenspectrum of the noise is not flat; its shape is determined by the ratio V/T [18, 17], and this knowledge can be used to recover the true spectrum of the signal through Bayesian estimation [5].

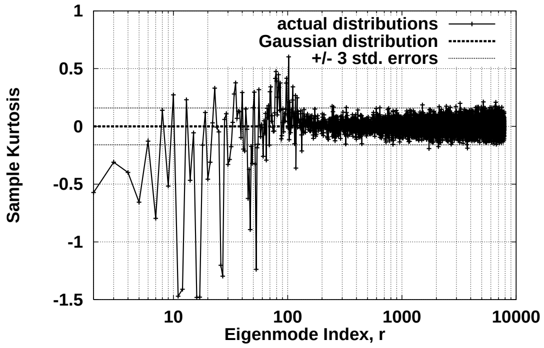

Figure 2: Sample kurtoses of the distributions of KLT coefficients, $\langle a_r^4\rangle - 3$. Notably, $\langle a_r^2\rangle \equiv 1$. The expectation for a Gaussian distribution, 0, and a corresponding confidence interval of ± 3 *standard errors,* $\sqrt{24/T}$ (*e.g.,* p. 137 of [19]), with $T = 8500$ independent samples, are also shown. Evidently, the $r \in [100, 200]$ regime still has some systematic deviations from Gaussianness, and traces of the signal can be expected there, but not much further.

3. THE FACTORIAL MODEL OF NOISY TRAJECTORIES DERIVED BY VECTOR QUANTIZATION

In Section 2, KLT was used to model successfully the white noise in the ensemble, in the regime $r > 300$. In the $r \in [1, 200]$ regime, the Gaussian assumption does not hold well—evidently from Fig. 2, most of the distributions are either sub- or super-Gaussian. Of interest are not only the marginal distributions of the individual KLT coefficients, but also their joint distributions [13, 16]. Evidently from Fig. 3, some exhibit drastic departure from the 2-dimensional Gaussian cloud of points, with highest density at the origin (3). In extreme cases, such as (2:7), (11:12), and (11:27) the actual density vanishes there.

Some KLT modes with non-Gaussian marginal distributions,

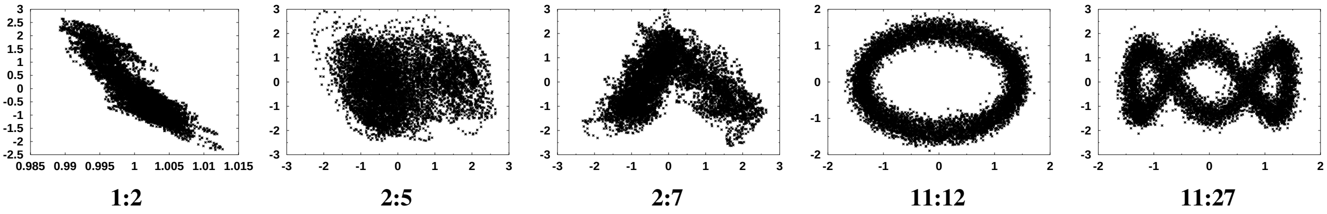

Figure 3: Phase diagrams of the dynamics, $\{a_r^t\}_{t \in T}$ (2), projected on the 2-dimensional KLT subspaces with the indicated pairs of r (cf. Fig. 2, Fig. 4, Fig. 9). For the purposes of this presentation, we will call "L2" the source that is associated with the 1:2 KLT subspace, because it is roughly along a line, and is embedded in two dimensions, and "C6," the cyclical 11:12:16:17:26:27 source (cf. Fig. 2).

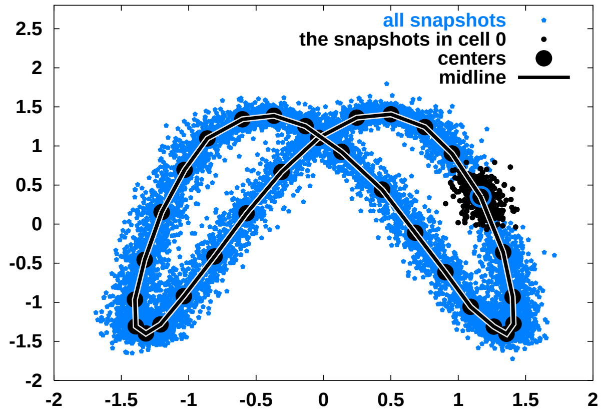

Figure 4: The 11:16 subspace projection of the 1D trajectory of the C6 source (cf. Fig. 3), estimated by the LBG vector quantization algorithm [10], according to the probability metric (3) restricted on the 4D 11:12:16:17 KLT subspace [16]. The number of anchors $Q = 31$ was chosen at the end of the 1D regime of its rate-distortion curve [16]. The centers of the Voronoi cells are shown with circles, and the piece-wise linear model of the trajectory, with a line. The snapshots in one of the cells are shown with black points, and all the rest, with gray.

such as $r \in \{2, 5, 7\}$, may or may not participate together in coordinated motions in higher dimensions. Strikingly, some distributions, such as (11:12:27), seem to be charactered well by a one-dimensional trajectory in a high-dimensional KLT subspace "fattened" by noise and/or contributions from extraneous source of variability.

Explicit in the formulation of the KLT model (3) is the assumption that the origin has *maximum likelihood.* A generalization has been suggested [16] to account for the actual locus of maximum likelihood—modeled as the "midline," $C(\theta, \mathbf{x})$, of the probability density, where θ is a 1-dimensional parameter (cf. Fig. 4). In order to estimate this locus from the data, *vector quantization (VQ)* has been applied previously [13, 16] to tessellate suitably-chosen KLT subspaces into Q *Voronoi cells,* whose centers have the property of *minimum distortion;* in the entropy-based metric (3), this is equivalent to *maximum likelihood.* Further, $C(\theta)$ has been parameterized as a piecewise-linear curve anchored at these centers. Such a model of the C6 source (cf. Fig. 3) is shown in Fig. 4.

With $C(\theta)$, for any snapshot ϕ^t, a parameter $\tilde{\theta}(t)$, of the dynamics along the manifold, can be found, such that the manifold element $C(\tilde{\theta}(t))$ is closest to ϕ^t according to the metric used for vector quantization; this is a *maximum likelihood estimate.* When the *residual*— the departure from the manifold—is defined as

$$\tilde{\phi}^t(\mathbf{x}) \triangleq \phi^t(\mathbf{x}) - C(\tilde{\theta}(t), \mathbf{x}) \tag{4}$$

the initial probability model (3) can be refined to

$$\mathcal{P}[\phi(\mathbf{x})] \triangleq \mathcal{P}[\tilde{\theta}] \times \mathcal{P}[\tilde{\phi}(\mathbf{x})] \tag{5}$$

where the two factors are estimated independently.

4. PROBABILISTIC MODELING OF THE DYNAMICS ON THE MANIFOLD

The procedure in Section 3 does not assume any temporal properties of the manifold dynamics, but rather provides an estimate, $\tilde{\theta}(t)$.

The dynamics of the C6 source, estimated from the trajectory in Fig. 4, is shown in Fig. 5. Evidently from Fig. 5*A*, constant-rate motion along the manifold is a reasonable initial model. The departure from this model, shown in Fig. 5*B*, exhibits slow, smooth variation at some relatively large characteristics time scale. An examination at a shorter time scale, shown in Fig. 5C, reveals a different structure. Notably, none of these variabilities has oscillatory nature, and spectral analysis [14] is not the natural model for them. In principle a general statistical model, possibly that incorporates domain-specific knowledge, could be built at these time scales; ideally, the estimated parameters could be related to physical causes.

Yet a different structure is revealed at the time scale of one revolution along the C6 manifold—fine oscillations, shown in Fig. 5*D*. While a slow change in the rate of advance is understandable, there is no *a priori* expectation that some points in the cycle should be somehow preferred, and that the probability density $\mathcal{P}[\{\theta\}]$, of the *fractional* part of θ, shown in Fig. 6, should be nonuniform. Indeed, the placement of the anchors produced by the VQ algorithm (Section 3) is according to just an estimate of the correct metric, and any invertible continuous distortion of θ is *a priori* just as good. Here we choose the distortion that compensates for the departure of the originally estimated probability density on the manifold (Fig. 6) from the uniform. The resulting estimate of the dynamics, shown in Fig. 5*E*, is pretty good—if any "fast" oscillation of $\tilde{\theta}(t)$ can be assumed caused by the failure to unmix the other sources, then the peak-to-peak error are $\approx 3\%$.

5. REFINEMENT OF THE EMBEDDING PARAMETERS

In Section 3, following [16], $C(\theta)$ was modeled as a piecewise-linear curve, in which the anchors were somehow special—they

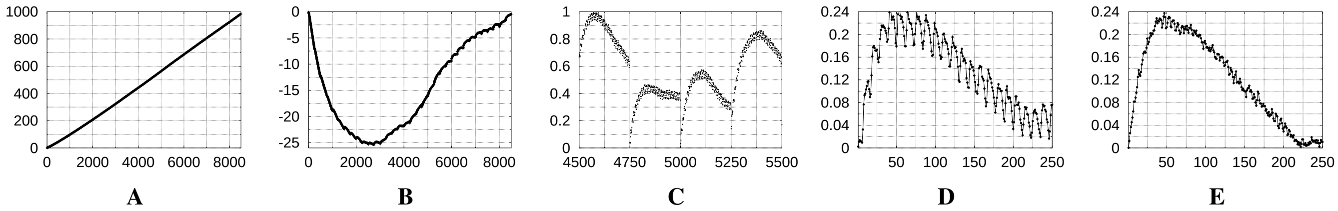

Figure 5: The dynamics on the C6 the manifold (cf. Fig. 4). Time is measured in snapshots, plotted on the abscissa; C6 is cyclical, and the natural unit of measure is one revolution, plotted on the ordinate. A: the maximum likelihood estimate $\tilde{\theta}(t)$ (cf. eq. 4) for the total duration of the experiment; B: same as A, but after subtraction of a globally estimated speed with a period of 8.62 snapshots/revolution (spr); C: a fraction of the record, with a local speed estimate of 8.05 spr, the duration of one stimulus condition (block) is 250 snapshots (see [4] for experimental details); D: the fine structure within one block, local speed 11.44 spr; E: same as D, uniform probability model for the fractional revolution, $\mathcal{P}[\{\theta\}]$ (cf. Fig. 6).

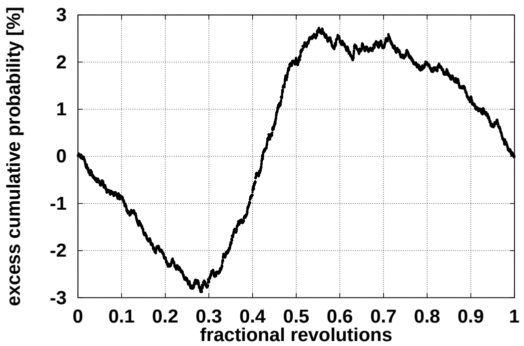

Figure 6: Departure from uniform density of the VQ-estimated probability model along the C6 manifold (cf. Fig. 4). This departure generates the oscilatory structure at the time scale of one revolution, evident in Fig. 5*D*. Its removal (see Section 4) results in the improved estimate in Fig. 5*E*.

were estimated directly from the data, as averages of the snapshots in a given Voronoi cell (cf. Fig. 4). Once an estimate $\tilde{\theta}(t)$ is available though, such as in Fig. 5*E*, $C(\theta)$ can be estimated for all θ, not just for the anchors, as an average of those snapshots with $\tilde{\theta}(t)$ in some range—of width 2Δ, centered at θ. Then, with $\bar{\theta}(t) = \theta + \bar{h}(t)$, the true parameter on the manifold (which is not known), $\bar{h}(t)$, the true, and $\tilde{h}(t)$, the estimated departure from θ,

$$\tilde{C}_\Delta(\theta) \triangleq \left\langle \phi^t \right\rangle_{|\tilde{\theta}(t)-\theta|\leq\Delta} = \left\langle C(\theta+\bar{h}(t)) + \bar{\phi}^t \right\rangle_{|\tilde{h}(t)|\leq\Delta} \quad (6)$$

where and $\bar{\phi}^t$ is the true residual (cf. eq. 4), presumably independent of θ. Its average will tend to a constant, possibly vanishing, with amplitude of the stochastic fluctuations, vanishing with $1/\sqrt{\Delta}$.

A systematic source of error in the estimate $\tilde{C}_\Delta$ is obvious when the term $C(\theta + \bar{h}(t))$ in (6) is expanded in a Taylor series. Since the averaging in (6) is symmetric around θ, the contribution of the odd derivatives will tend to vanish with $1/\sqrt{\Delta}$, and the contribution of the even ones, of order $2k$, will grow as Δ^{2k}.

These systematic errors depend on the derivatives of $C(\theta)$, whose estimates, in turn, depend on the stochastic errors. The optimal choice of Δ will be pursued elsewhere. Notably, 2Δ can never be of order smaller than the peak-to-peak error in the estimate of $\tilde{\theta}(t)$; this increases the systematic error, but does not decrease the stochastic. Since these error are coupled together, here we make the heuristic choice of $\Delta \approx$ the peak-to-peak error in $\tilde{\theta}(t)$; in Fig. 5*E* it is $\approx$ 3%.

6. CONVERGENT LEARNING OF THE MANIFOLD

With a refined estimate of the manifold, $\tilde{C}(\theta)$, for example produced by the method in Section 5, can a refined estimate be produced of the dynamics, $\tilde{\theta}(t)$? There are a variety of possibilities, some based on the search for linear subspaces in which the projection of the dynamics is even more unmixed, followed by the subsequent iteration of the VQ-based learning procedure (Section 3).

It is not necessary that the VQ subspace contains all the signal power of the manifold; rather, more useful for the estimation of the dynamics is a subspace that has a relatively small admixture of other sources. There are two causes of such mixing. On one hand, two independent manifolds could project to the same subspace; this will be addressed in Section 7. On the other, two different subspaces could get mixed because they have similar powers (second-order moments), in which case KLT cannot separate them even in principle. While, at least in principle, any ICA method—sensitive also to higher order moments [1]—could be used to find such purified linear subspaces, here we explore a result that arises naturally in the context of the factorial model from Section 3.

The goal of the separation is, after all, on the basis of the estimates $\tilde{C}(\theta)$ and $\tilde{\theta}(t)$, to build a model of the residual, $\tilde{\phi}^t(\mathbf{x})$ (4), and subsequently, a model the probability of this multidimensional factor, $\mathcal{P}[\tilde{\phi}(\mathbf{x})]$ (5). As usual, a Gaussian model is a reasonable default. As illustrated in Fig. 4, subspaces in which substantial amount of the power is due to the manifold will be deprived of it after the factorization step (4), and in the residual KLT hierarchy their eigenmode indices will be higher.

A desirable situation arises when the power of such subspaces is so well accounted for by the manifold that its residual dispersion is below the white-noise level. Then, these weak subspaces are well separable by KLT, as illustrated in Fig. 7*A* for the initial estimate of the C6 source. Subsequently, the dynamics on the manifold can be re-estimated by VQ in the weak subspace. Such alternative estimation of the dynamics on, and the embedding of, the manifold can be iterated until convergence. For the C6 source, the initial spectral regime of the residual is shown in Fig. 7*B*, and the final one, in Fig. 7*A*,C. The peak-to-peak mixing errors in the final estimate of the dynamics, which seem to be dominated by the stochastic term in $\tilde{C}(\theta)$ (6), shown in Fig. 8*A*, are $\approx$ 1%.

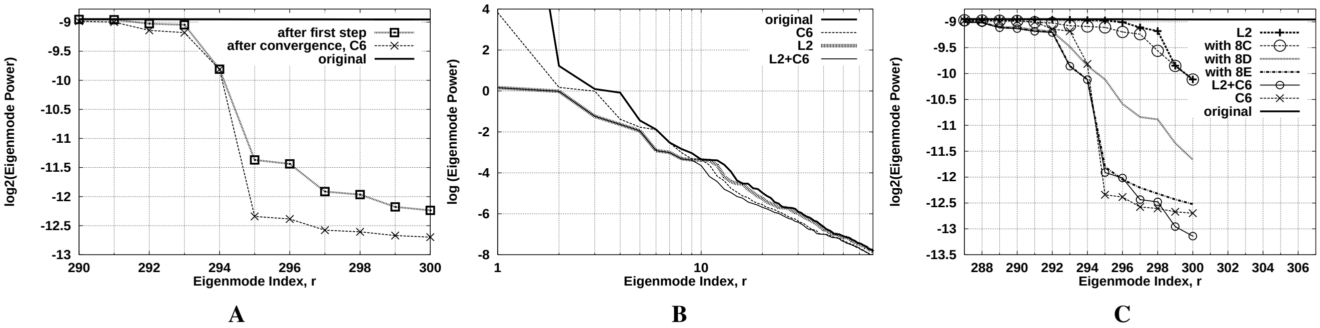

Figure 7: Eigenspectra of residuals (cf. Fig. 1). A: the eigenspectrum tails of the successive residuals as the refinement of the C6 manifold proceeds out of the original ensemble (after first step, Section 3; after convergence, Section 6). B: decrease of the unaccounted-for signal power after factorization by the indicated manifolds (original, Fig. 1; C6, Fig. 7A; L2, Section 7; L2+C6, Section 8). C: same as A, but C6 is refined out of the L2 residual this time; with 8C–E, after factorization by the successive manifold estimates $\tilde{C}(\theta)$ (6) with the dynamics estimates $\tilde{\theta}(t)$ from Fig. 8C–E, respectively.

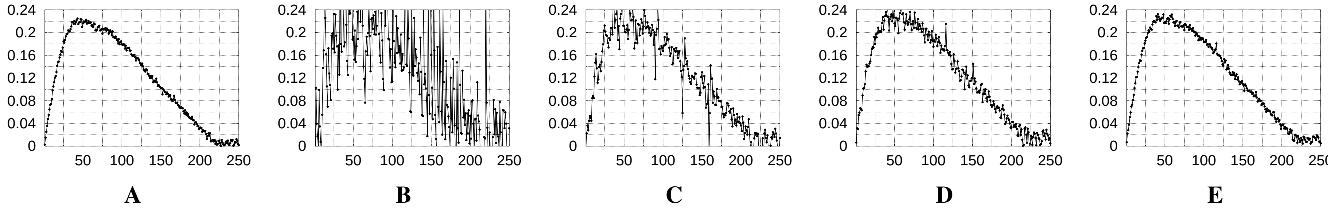

Figure 8: Estimates $\tilde{\theta}(t)$ of the dynamics on the C6 the manifold (cf. Fig. 5), with enforced uniform densities, as described in Section 4. A: convergence after one step of the iterative learning described in Section 6, all other parameters as in Fig. 5E; B: initial VQ estimate with the $Q = 6$ manifold in Fig. 10; C–E: the three subsequent estimates (cf. Fig. 7C).

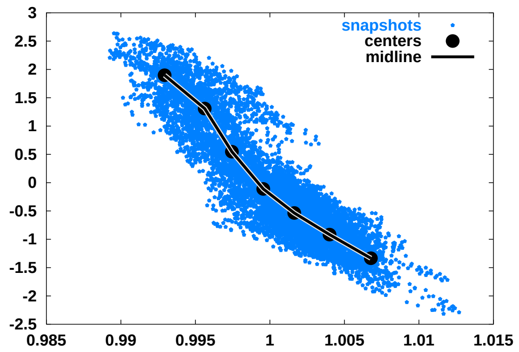

Figure 9: The 1D trajectory with $Q = 7$ anchors in the 2D 1:2 KLT subspace (cf. Fig. 3).

7. HIERARCHICAL APPLICATION

Naturally, such convergent learning can be applied to any putative independent source in the original ensemble. Here, we apply it to the L2 source (Fig. 3): its initial VQ estimate is shown in Fig. 9, the spectrum of its residual, in Fig. 7B,C, and some 2D projections of dynamics of the residual, in Fig. 10.

In this data set, we have characterized so far two independent sources of variability and have used the manifolds they generate to individually factorize the data, respectively. It would be desirable to extract the factors hierarchically—to factorize the L2 residual by the C6 source. As in Section 3, one would look for a KLT subspace in which the dynamics has a 1-dimensional regime and subsequently iterate the learning procedure until convergence.

Evidently from Fig. 10, there is a problem—the C6 manifold, while relatively unmixed in the original ensemble, happens to be much more mixed in the L2 residual. This can be understood from the eigenspectra in Fig. 7B—the 3-dimensional (10:11:12) invariant KLT subspace in the original (cf. Fig. 3) is now mixed in the L2 residual (cf. Fig. 10) and corresponds to the (8:9:10:11) subspace. It happens that the (9:10:11) projection is a 3D cylinder, and no 2D KLT projection exposes the circle at its base.

In this case, since estimate of the C6 dynamics is already available (Fig. 8A), it could be used as a starting point for manifold learning. Nevertheless, the situation in Fig. 10 is rather generic, and it would be desirable to have a systematic way to deal with such heavy mixtures. We proceed to present one.

8. A LARGE DOMAIN OF FAST CONVERGENCE

We start with the (9:10:11) subspace of the L2 residual (Fig. 10). As in Section 3, we apply VQ and choose $Q = 6$, which is at the end of the putative 1D regime of its rate-distortion curve [16]. As shown in Fig. 10, this initial estimate is very far from the true manifold; also, the respective estimate of the dynamics, shown in Fig. 8B, is very unreliable. Nevertheless, the factorization (4) with these crude estimates leads to rearrangement of the KLT spectrum that is sufficient to continue the iteration.

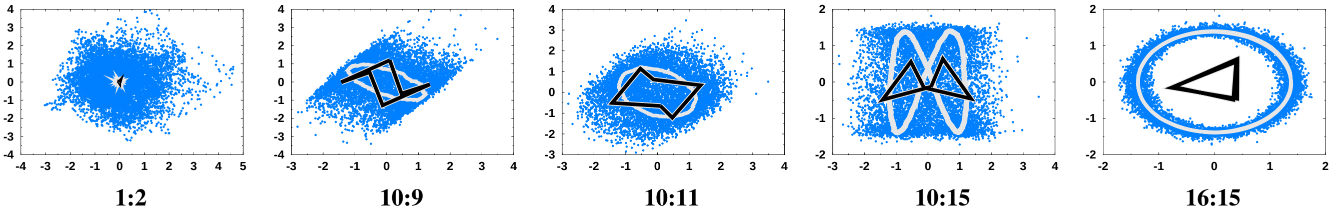

Figure 10: Phase diagrams (cf. Fig. 3) of the dynamics of the L2 residual (cf. Fig. 9). Points: all snapshots; dark lines: the initial VQ estimate of the C6 manifold, $Q = 6$; light lines: the estimate after convergence.

The second step leads to the estimate in Fig. 8*C*; this estimate already captures enough structure of the dynamics, which is sufficient for the the residual dispersion of a small subspace to drop below the noise level, shown in Fig. 7*C*. Two more steps (Fig. 8*D*,*E*) lead to an estimate, comparable to that in Fig. 8*A*, and one more, to convergence.

9. DISCUSSION

Here we have confirmed the result [16] that the loci of maximum likelihood of optical imaging data are smooth, curved, low-dimensional manifolds. Further, we have described a method for the iterative refinement of, alternatively, the parameters of the embedding of, and the dynamics along, those manifolds. We have shown rapid convergence after a good initial guess, from a relatively unmixed KLT subspace. In a more typical scenario, we have shown fast convergence from a heavily mixed KLT subspace. Hierarchically, we have factorized the initial data set by these curved manifolds, each time exposing successively simpler residuals, with less signal power and potentially smaller number of sources of variability.

This procedure utilizes curved *independent manifolds*, embedded in larger-dimensional linear subspaces; these manifolds generalize the 1-dimensional linear subspaces of ICA.

The fact that fast convergent learning is possible of curved manifolds, the dynamics along which is independent, could lead to the robust estimation of the parameters of a full nonlinear factorial model of the probability density of optical imaging and, possibly, other biomedical data. In such models, the spatial action of the various independent sources of variability would be separated, and their temporal dynamics, estimated robustly.

The current method can be refined in a number of directions. Most interestingly, the approach described here is completely oblivious to the short-term temporal correlations that exist among the KLT coefficients and the parameters of manifold dynamics (not shown); these are attractive targets for further research—from both probabilistic, and neuroscientific points of view.

10. REFERENCES

[1] J.-F. Cardoso. High-order contrasts for Independent Component Analysis. *Neural Computation*, 11(1):157–192, 1999.

[2] D. W. Dong, J. A. S. Kelso, W. D. Wilke, and F. Steinberg. Spatiotemporal decorrelated activity patterns in functional MRI data during real and imagery motor tasks. In *Society for Neuroscience Abstracts*, volume 25, page 786. Society for Neuroscience, 1999.

[3] R. Everson, B. Knight, and L. Sirovich. Separating spatially distributed response to stimulation from background. I. Optical imaging. *Biological Cybernetics*, 77(6):407–417, 1997.

[4] R. Everson, A. K. Prashanth, M. Gabbay, B. W. Knight, L. Sirovich, and E. Kaplan. Representation of spatial frequency and orientation in the visual cortex. *Proc. Natl. Acad. Sci. USA*, 95:8334–8338, July 1998.

[5] R. M. Everson and S. J. Roberts. Inferring the eigenvalues of covariance matrices from limited, noisy data. *IEEE Trans. Sig. Proc.*, 2000. in press.

[6] R. Fisher. The use of multiple measurments in taxonomic problems. *Ann. Eugenics*, 7(II):179–188, 1936.

[7] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999. http://www.icsi.berkeley.edu/~jagota/NCS/.

[8] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York Berlin Heidelberg Tokio, 1986.

[9] K. Karhunen. Zur Spektraltheorie Stochasticher. *Prozesse Ann. Acad. Sci. Fennicae*, 37, 1946.

[10] Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *IEEE Trans. Commun.*, 28:84–95, Jan. 1980.

[11] R. R. Llinas, U. Ribary, D. Jeanmonod, E. Kronberg, and P. P. Mitra. Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proc. Natl. Acad. Sci. USA*, 96(26):15222–15227, Dec. 1999.

[12] M. Loève. *Probability Theory*. Van Nostrand, Princeton, N.J., 1955.

[13] E. Mandel and P. S. Penev. Facial feature tracking and pose estimation in video sequences by factorial coding of the low-dimensional entropy manifolds due to the partial symmetries of faces. In *Proc. 25th IEEE Int'l Conf. Acoust., Speech, Sig. Proc. (ICASSP2000)*, Istanbul, Turkey, 2000. IEEE. in press.

[14] P. P. Mitra and B. Pesaran. Analysis of dynamic brain imaging data. *Biophysical J.*, 76(2):691–708, Feb. 1999.

[15] P. S. Penev. *Local Feature Analysis: A Statistical Theory for Information Representation and Transmission*. PhD thesis, The Rockefeller University, New York, NY, May 1998. available at http://venezia.rockefeller.edu/penev/thesis/.

[16] P. S. Penev, M. Gegiu, and E. Kaplan. Using vector quantization to build nonlinear factorial models of the low-dimensional independent manifolds in optical imaging data. In *Proc. 2000 Int'l Conf. Image Proc. (ICIP-2000)*, Vancouver, Canada, 2000. IEEE CS. in press.

[17] A. M. Sengupta and P. P. Mitra. Distributions of singular values for some random matrices. *Phys. Rev. E*, 60(3):3389–3392, Sept. 1999.

[18] J. W. Silverstein. Eigenvalues and eigenvectors of large-dimensional sample covariance matrices. *Contemporary Mathematics*, 50:153–159, 1986.

[19] R. R. Sokal and F. J. Rohlf. *Biometry*. W. H. Freeman and Co., San Francisco, CA, 1969.

[20] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70:1055–1096, 1982.