BRIEF REPORTS

Brief Reports are accounts of completed research which do not warrant regular articles or the priority handling given to Rapid Communications; however, the same standards of scientific quality apply. (Addenda are included in Brief Reports.) A Brief Report may be no longer than four printed pages and must be accompanied by an abstract. The same publication schedule as for regular articles is followed, and page proofs are sent to authors.

Distributions of singular values for some random matrices

A. M. Sengupta and P. P. Mitra
Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974
(Received 20 January 1998; revised manuscript received 16 March 1999)

The singular value decomposition is a matrix decomposition technique widely used in the analysis of multivariate data, such as complex space-time images obtained in both physical and biological systems. In this paper, we examine the distribution of singular values of low-rank matrices corrupted by additive noise. Past studies have been limited to uniform uncorrelated noise. Using diagrammatic and saddle point integration techniques, we extend these results to heterogeneous and correlated noise sources. We also provide perturbative estimates of error bars on the reconstructed low-rank matrix obtained by truncating a singular value decomposition. [S1063-651X(99)09008-X]

PACS number(s): 02.50.Sk, 11.15.Pg, 05.90.+m

In analyzing large, multivariate data, certain quantities naturally arise that are in some sense ‘‘self averaging.’’ Namely, in the large size limit, a single data set can comprise a statistical ensemble for the quantity in question. One such quantity, the singular value distribution of a data matrix, is the subject of this paper. The singular value decomposition (SVD) is a representation of a general matrix of fundamental importance in linear algebra that is widely used to generate canonical representations of multivariate data. It is equivalent to principal component analysis in multivariate statistics, but, in addition, is used to generate low dimensional representations for complex multidimensional time series. One example is to generate effective low dimensional representations of high dimensional dynamical systems. Another example of current interest is to denoise and compress dynamic imaging data, in particular in the case of direct or indirect images of neuronal activity.

In this paper, we use diagrammatic and saddle point integration techniques to obtain the densities of singular values of matrices whose entries have varying degrees of randomness. In particular, we study the problem in the asymptotic limit of large matrix size; this limit is well justified in realistic cases as will be described below. The density of SV's has been obtained before, with other techniques, for matrices with each entry independently distributed normally with identical variances [1,2]. We are able to obtain distributions for some more general cases where the variances are not equal and/or correlations are present between matrix entries. Our results have implications towards isolating random components from image time series. Also, these results help in understanding the effects of truncating the SV spectrum at a given point, a technique that is widely applied to remove noise from data.

The SVD of an arbitrary (in general complex) $p\times q$ matrix $(p\geqslant q)M$ is given by $M=U\Lambda V^{\dagger}$, where the $p\times q$ matrix U has orthonormal rows, the $q\times q$ matrix Λ is diagonal with real, non-negative entries, and the $q\times q$ matrix V is unitary. Note that the matrices $MM^{\dagger}=U\Lambda^2U^{\dagger}$ and $M^{\dagger}M=V^{\dagger}\Lambda^2V$ are Hermitian, with eigenvalues corresponding to the diagonal entries of Λ^2 and U and V the corresponding matrices of eigenvectors. Consider the special case of space-time data $I(\mathbf{x},t)$. The SVD of such data is given by

$$I(\mathbf{x},t)=\sum_n \lambda_n I_n(\mathbf{x})a_n(t), \tag{1}$$

where $I_n(\mathbf{x})$ are the eigenmodes of the spatial ‘‘correlation’’ matrix $C(\mathbf{x},\mathbf{x}')=\Sigma_t I(\mathbf{x},t)I(\mathbf{x}',t)$, and similarly $a_n(t)$ are the eigenmodes of the ‘‘temporal correlation function’’ $C(t,t')=\Sigma_x I(\mathbf{x},t)I(\mathbf{x},t')$. If one considered the sequence of images as randomly chosen from an ensemble of spatial images, then $C(\mathbf{x},\mathbf{x}')$ would converge to the ensemble spatial correlation function in the limit of long times. If in addition the ensemble had space translational invariance, then the eigenmodes $I_n(\mathbf{x})$ would be plane waves $e^{i\mathbf{k}\cdot\mathbf{x}}$, the mode number n would correspond to wave vectors, and the singular values would correspond to the spatial structure factor $S(\mathbf{k})$. In general, the image ensemble in question will not have translational invariance; however, the SVD will then provide a basis set analogous to wave vectors. In physics one normally encounters the structure factors $S(\mathbf{k})$ that decay with wave vectors, and in the more general case, the singular value spectrum, organized in descending order, will show a decay indicating the structure in the data.

We consider the case of a $p\times q$ matrix $M=M_0+N$, where M_0 is fixed and the entries of N are normally distributed with zero mean. We consider below several cases of normal distributions for entries of N, including cases where there are correlations between entries of N. (It turns out that in the limit of large q, the results are not restricted normal distributions only.) M_0 may be thought of as the desired or underlying signal. For SVD to be useful, M_0 should effectively have a low-rank structure.

It is convenient to work in terms of the resolvent

$$\mathcal{G}(z)=\mathrm{Tr}[(z-M^{\dagger}M)^{-1}]=\sum_n \frac{1}{z-\lambda_n^2}, \tag{2}$$

where $\mathcal{G}(z)$ is a complex function. The density of SV's is given by

$$\rho(\lambda)=\sum_n \delta(\lambda-\lambda_n)=\frac{2\lambda}{\pi}\lim_{\epsilon\to 0}\mathrm{Im}[\mathcal{G}(\lambda^2-i\epsilon)]. \tag{3}$$

Let the variance of a matrix entry be σ^2. We proceed by taking the limit $p,q\to\infty$, p/q fixed, with the variances of the matrix entries tending to zero as $1/q$. $\tilde{\sigma}^2=q\sigma^2$ is kept finite. The density of states being a self-averaging quantity, we are able to apply our results, obtained by averaging over the ensemble, to the SVD of individual data matrices.

To illustrate the method, consider the simplest case, where each element of the matrix is independently and identically normally distributed with mean 0 and variance σ^2. Since

$$\mathcal{G}(z)=\partial_z \ln\det(z-M^{\dagger}M), \tag{4}$$

the average of the resolvent over the probability distribution of the matrix M can be obtained from $\langle \ln\det(z-M^{\dagger}M)\rangle$, which in turn may be computed using replicas. We introduce n replicas of q-dimensional real vectors $X_\alpha=(x_{\alpha 1},..,x_{\alpha q})$ $\times(\alpha=1,...,n)$. Consider the following identity:

$$Z_n=\int\left[\prod_{\alpha=1}^{n}\prod_{a=1}^{q}dx_{\alpha a}\right]\left\langle \exp\left(-\frac{q}{2}\sum_{\alpha=1}^{n}X_\alpha^T(z-M^{\dagger}M)X_\alpha\right)\right\rangle$$
$$=\left(\frac{2\pi}{q}\right)^{nq/2}\langle[\det(z-M^{\dagger}M)]^{-(n/2)}\rangle. \tag{5}$$

One obtains the desired quantity from the above by taking $n\to 0$. After some amount of algebraic manipulations, Z_n can be written as an integral over two $n\times n$ matrices Q and R,

$$Z_n=2^{-n^2}\left(\frac{q}{2\pi}\right)^{-(nq/2)+n^2}\int\left[\prod dR\,dQ\right]\exp\left(-\frac{q}{2}\left[\ln\det(z-\tilde{\sigma}^2Q)+\frac{p}{q}\ln\det(1-iR)+i\,\mathrm{Tr}(QR)\right]\right). \tag{6}$$

Ideally one should take the $n\to 0$ limit first and then let $q\to\infty$. In order to be able to perform analytical computations, we have to take the limit in the reverse order. That this gives the correct answer is verified later by a direct diagrammatic method.

When $p,q\to\infty$, with p/q fixed, the integral is dominated by some saddle point. We take the replica diagonal ansatz, consistent with all the symmetries, namely, $Q_{\alpha\beta}=Q(z)\delta_{\alpha\beta}$ and $R_{\alpha\beta}=-iR(z)\delta_{\alpha\beta}$.

Then we have to minimize

$$S(Q(z),R(z))=\ln[z-\tilde{\sigma}^2Q(z)]+\frac{p}{q}\ln(1-R(z))+Q(z)R(z) \tag{7}$$

with respect to $Q(z)$ and $R(z)$. Hence the equations,

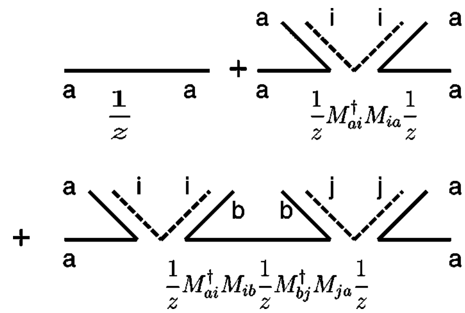

FIG. 1. Diagrammatic representation of successive terms in the resolvent.

$$\frac{1}{z-\tilde{\sigma}^2Q(z)}=\frac{1}{\tilde{\sigma}^2}R(z), \tag{8}$$

$$\frac{p/q}{1-R(z)}=Q(z). \tag{9}$$

Using the fact that $Z_n\sim\exp[-nqS(Q(z),R(z))]$ and 4, we get

$$G(z)=\langle\mathcal{G}(z)\rangle=\frac{q}{z-\tilde{\sigma}^2Q(z)} \tag{10}$$

so that $G(z)$ satisfies

$$G(z)=\frac{q}{z-\dfrac{p\tilde{\sigma}^2}{q-\tilde{\sigma}^2G(z)}}. \tag{11}$$

This equation can also be obtained from a direct diagrammatic resummation of $\langle\mathrm{Tr}\,1/(z-M^{\dagger}M)\rangle$ expanded in powers of $1/z$. We average over M with $\langle M_{ia}M^*_{jb}\rangle=\sigma^2\delta_{ij}\delta_{ab}$. The diagrammatic representation of these terms is shown in Fig. 1. In the large p,q limit, we have to consider only planar diagrams [3]. The diagrams contributing to self-energy are shown schematically in the Fig. 2. Summing the geometric series in this limit, we obtain

$$\Sigma(z)=\frac{p\tilde{\sigma}^2/q}{1-\tilde{\sigma}^2G(z)/q}, \tag{12}$$

$$G(z)=\frac{q}{z-\Sigma(z)}=\frac{q}{z-\dfrac{p\tilde{\sigma}^2}{q-\tilde{\sigma}^2G(z)}}. \tag{13}$$

The solution of this equation is

$$G(z)=\frac{q}{2\tilde{\sigma}^2z}\{-\tilde{\sigma}^2(p/q-1)+z\pm\sqrt{[z-\tilde{\sigma}^2(p/q+1)]^2-4p\tilde{\sigma}^4/q}\} \tag{14}$$

and

$$\rho(\lambda)=\frac{q}{\pi\lambda\tilde{\sigma}^2}\sqrt{(\lambda_{max}^2-\lambda^2)(\lambda^2-\lambda_{min}^2)} \tag{15}$$

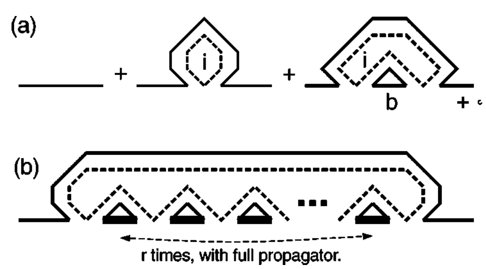

FIG. 2. Diagrams that contribute to the self-energy in the large-(p,q) limit.

for $\lambda_{min}<\lambda<\lambda_{max}$ and zero elsewhere, with $\lambda_{max,min} = \tilde{\sigma}\sqrt{(p/q+1)\pm 2\sqrt{p/q}} = \sqrt{2}\sigma\sqrt{(p+q)/2\pm\sqrt{pq}}$. Equation 15 has been derived before using different techniques [2].

Generalizing our methods, both the saddle point technique and the perturbative method, to the following cases is quite easy.

Case (i): $\langle M_{ia}\rangle = M^0_{ia}$. The matrix M^0 has singular values λ_{0a} where $a=1,...,q$.

The covariance matrix is given as before by $\langle (M_{ia} - M^0_{ia})(M_{jb}-M^0_{jb})^*\rangle = \sigma^2\delta_{ij}\delta_{ab}$. In this case we obtain

$$G(z)=\sum_{a=1}^{q}\frac{1}{z-\lambda^2_{0a}-\dfrac{p\tilde{\sigma}^2}{q-\tilde{\sigma}^2G(z)}} = G_0\left(z-\frac{p\tilde{\sigma}^2}{q-\tilde{\sigma}^2G(z)}\right), \tag{16}$$

where

$$G_0(z)=\mathrm{Tr}\left(\frac{1}{z-M^{0\dagger}M^0}\right). \tag{17}$$

In case there are only a few nontrivial λ_{0a}'s $G(z)$ still satisfies a polynomial equation of order two or higher. Denby and Mallows [2] obtained similar results using a different method.

One of the simple consequences of Eq. (16) is the following. Consider a situation where there are only r nonzero singular values of M^0, each of which is much bigger than the noise. Let the nonzero SV's be $\{\lambda_{01},\ldots,\lambda_{0r}\}$. In the limit of zero noise,

$$G(z)=\frac{q-r}{z}+\sum_{a=1}^{r}\frac{1}{z-\lambda^2_{0a}}. \tag{18}$$

In presence of finite small σ, the pole at zero broadens into a branch cut that is close to the origin, and the other cuts develop around the nontrivial singular values that are far away from the origin.

Consider $G(z)$ for z close to the origin. For $z\sim\tilde{\sigma}^2$, ignoring terms of the order $(\tilde{\sigma}/\lambda_{0a})^2$ for $a\leqslant r$, we find

$$G(z)\approx\sum_{a=r+1}^{q}\frac{1}{z-\dfrac{p\tilde{\sigma}^2}{q-\tilde{\sigma}^2G(z)}}=\frac{q-r}{z-\dfrac{p\sigma^2}{1-\sigma^2G(z)}}, \tag{19}$$

which is just Eq. (13) with q replaced by $q-r,\sigma$ (but not $\tilde{\sigma}$) remaining the same. Hence, the smallest $q-r$ singular values have the same distribution as arises from a pure noise matrix of a smaller size, namely, $p\times(q-r)$. This result is useful in fitting the formula to the tail of the singular value spectrum for a real data matrix, and is used in the fit displayed in Fig. 3.

Case (ii): $\langle M_{ia}\rangle=0$ and $\langle M_{ia}M^*_{jb}\rangle=\sigma^2C_{ij}D_{ab}$, where C and D are $p\times p$ and $q\times q$ matrices, respectively. Such correlations may arise in imaging data due to filtering of an underlying uncorrelated spatial noise distribution, and/or when there is temporal filtering of data. Uncorrelated noise with spatial inhomogeneity in the variance is also a special case of the above.

We go through same manipulations as in Eqs. (5) and (6). The saddle point solution is obtained by minimizing

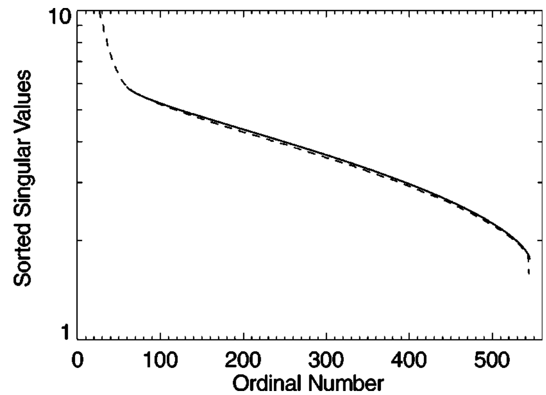

FIG. 3. Comparison of singular values from a SVD of an fMRI data set with the theoretical formula for a noise only matrix.

$$S(Q(z),R(z))=\frac{1}{q}\ln\det[z-\tilde{\sigma}^2Q(z)D] + \frac{1}{q}\ln\det[1-R(z)C]+Q(z)R(z) \tag{20}$$

with respect to $Q(z)$ and $R(z)$. Hence the equations,

$$R(z)/\tilde{\sigma}^2=\frac{1}{q}\mathrm{Tr}\frac{D}{z-\tilde{\sigma}^2Q(z)D}=\frac{1}{q}\sum_{a=1}^{q}\frac{d_a}{z-\tilde{\sigma}^2Q(z)d_a}, \tag{21}$$

$$Q(z)=\frac{1}{q}\mathrm{Tr}\frac{C}{1-R(z)C}=\frac{1}{q}\sum_{i=1}^{p}\frac{c_i}{1-R(z)c_i}. \tag{22}$$

c_i's are eigenvalues of C and d_a's are eigenvalues of D.

The expression for the resolvent is

$$G(z)=\langle\mathcal{G}(z)\rangle=\mathrm{Tr}\frac{1}{z-\tilde{\sigma}^2Q(z)D}. \tag{23}$$

To see how to use this result, let us consider a special case: $\langle M_{ia}M^*_{jb}\rangle=\sigma^2C_{ij}\delta_{ab}$. In this case, we can eliminate $R(z)$ and $Q(z)$ to get

$$G(z)=\mathrm{Tr}\,\tilde{G}(z)=\frac{q}{z-\sigma^2\,\mathrm{Tr}\dfrac{C}{1-CG(z)}}. \tag{24}$$

If the eigenvalues of C are $c_i=\sigma_i^2/\sigma^2$, $i=1,...,p$, then,

$$G(z)=\frac{q}{z-\displaystyle\sum_{i=1}^{p}\frac{\sigma_i^2}{1-\sigma_i^2G(z)}}. \tag{25}$$

How do we obtain the singular value spectrum from Eq. (25)? One way is to rewrite it as

$$z=\frac{q}{G}-\sum_{i=1}^{p}\frac{1}{G-(1/\sigma_i^2)}. \tag{26}$$

We want to know G for real z. It is useful to first think of z as a function of G in the complex G plane. We now look for level sets of Im$[z(G)]$. By tracing the appropriate branch of the curve Im$[z(G)]=0$, one can solve for $G(z)$ for real z. Taking the imaginary part of the function $G(z)$ thus found gives the density of singular values. The cumulative density of states can be found by integrating the singular value density.

Qualitative insight may be gained by realizing that the real and the imaginary parts of z are the two components of the electric field in a two-dimensional electrostatic problem, with a charge q at the origin, and point charges of strength -1 placed at each of the points $(1/\sigma_i^2,0)$, $i=1,...,p$ in the complex z plane.

In addition to the density of singular values, one can try to compute the correlations between nearby singular values. It is well known in the theory of random matrices that the correlation functions of the eigenvalues of a random Hermitian matrix has interesting universal features [4]. This is true for eigenvalues chosen from a small enough region, so that the eigenvalue density in that region is more or less constant. We find that the correlations of the singular values of a matrix, having each matrix element independently and identically distributed with mean zero and variance σ^2, are in the same universality class as the Gaussian unitary ensemble. The probability density $p(\Delta\lambda)$ of level spacings $\Delta\lambda$ goes as $\Delta\lambda^2$ for $\Delta\lambda \ll \Delta^-\lambda$. The probability density of $s=\Delta\lambda/S=\Delta/\overline{\Delta}\lambda$ for the Gaussian unitary ensemble is well known in the random matrix literature [4]. It is possible that empirical level-spacing statistics can be used as a diagnostic to find out which singular values correspond mostly to ''noise'' and which correspond mostly to ''signal.''

So far, we have discussed what happens to the singular values. We would also like to estimate the errors made in reconstructing the matrix by keeping a small number terms on the left-hand side of 1, which correspond to the biggest singular values. If we keep too few terms, we lose part of the signal.

Let us go back to case (i), namely, when each element of the matrix M is independently distributed with same variance σ^2 but different means. Let $M_{ia}^0=\langle M_{ia}\rangle=\Sigma_b\lambda_{0b}u_i^{0b*}v_a^{0b}$ and $M_{ia}=\Sigma_b\lambda_b u_i^{b*}v_a^b$. We would like to calculate the mean and the variance of the variable $\hat{M}_{ia}=\Sigma_{b\in S}\lambda_b u_i^{b*}v_a^b$, where S is a subset of $\{1,...,q\}$.

For small σ,

$$\langle\hat{M}_{ia}\rangle=\sum_{b\in S}\Bigg[\lambda_{0b}u_i^{0b*}v_a^{0b}+2\lambda_{0b}^2\sigma^2\sum_{c\neq b}\frac{\lambda_{0c}}{(\lambda_{0b}^2-\lambda_{0c}^2)}u_i^{0c*}v_a^{0c} -2\sigma^2\lambda_{0b}\sum_{c\neq b}\frac{\lambda_{0c}^2}{(\lambda_{0b}^2-\lambda_{0c}^2)^2}u_i^{0b*}v_a^{0b}\Bigg]+o(\sigma^4), \quad (27)$$

$$\mathrm{var}(\hat{M}_{ia})=\sigma^2\sum_{b\in S}\Bigg[|u_i^{0b*}v_a^{0b}|^2 +\lambda_{0b}^2\sum_{j\neq b}\frac{(\lambda_{0b}^2+\lambda_{0j}^2)}{(\lambda_{0b}^2-\lambda_{0j}^2)^2}|u_i^{0j*}v_a^{0b}|^2 +\lambda_{0b}^2\sum_{c\neq b}\frac{(\lambda_{0b}^2+\lambda_{0c}^2)}{(\lambda_{0b}^2-\lambda_{0c}^2)^2}|u_i^{0b*}v_a^{0c}|^2 -4\sum_{c\in S,c\neq b}\frac{(\lambda_{0b}^2\lambda_{0c}^2)}{(\lambda_{0b}^2-\lambda_{0c}^2)^2}|u_i^{0c*}v_a^{0b}|^2\Bigg]+o(\sigma^4). \quad (28)$$

In this expression j runs from 1 to p with $\lambda_{0j}=0$, for $j>q$.

To illustrate the utility of these results, we consider the SV distribution obtained from a space-time data set consisting of functional magnetic resonance images (FMRI). The experimental details regarding the chosen data set can be found in [5]. For our purposes, the data constitutes a 1724×550 matrix. The longer dimension corresponds to a subset of the pixels in a 64×64 spatial image obtaining by discarding pixels which have intensity below a selected threshold. In Fig. 3, the tail of the SV distribution from this data is displayed along with a fit to a theoretical curve obtained from Eq. (15). The distribution has two adjustable parameters. One of them is the variance σ. A second adjustable parameter in the fit is the rank of the original matrix, which in this case has been assumed to be 60. We fit the tail to the singular values of a $1724\times(550-60)$ pure-noise matrix. In fact, in the present case the uncorrelated noise can be estimated independently, and is therefore not really a free fitting parameter. We found that the fitted value of σ is in close correspondence with the independently estimated value of the noise variance.

In the example above, the good fit obtained between the theoretical curve and the tail in the SV distribution indicates that the noise entries in the original data were uniform and uncorrelated. It is easy to find experimental data where these conditions are violated, for example optical measurements of electrical activity in brain tissue [6,7] where the illumination is not fully uniform and the shot noise varies from point to point in space. Alternatively, the data may be spatially filtered and correlations may be introduced in space but not in time. Both of these cases produces SV distributions that cannot be fit by the procedure described above, but may be understood using Eq. (25).

We acknowledge useful discussions with C. L. Mallows. This work was supported by internal funding at Lucent Technologies. One of us (A.M.S.) was supported by NSF Grant No. DMR 96-32294 while working at Rutgers University, where part of the work was done.

[1] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains* (American Mathematical Society, Providence, RI, 1963).

[2] L. Denby and C. L. Mallows, *Computing Sciences and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas, (Interface Foundation, Fairfax Station, VA, 1991), pp. 54–57.

[3] G. 't Hooft, Nucl. Phys. B **72**, 461 (1974).

[4] M. L. Mehta, *Random Matrices* (Academic Press, New York, 1991).

[5] P. P. Mitra, S. Ogawa, X. Hu, and K. Ugurbil, Magn. Reson. Med. **37**, 511 (1997).

[6] J. C. Prechtl, L. B. Cohen, B. Pesaran, P. P. Mitra, and D. Kleinfeld, Proc. Natl. Acad. Sci. USA **94**, 7621 (1997).

[7] D. S. Broomhead and G. P. King, Physica D **20**, 217 (1986); A. I. Mess, P. E. Rapp, and L. S. Jennings, Phys. Rev. E **37**, 340 (1987).