Inversion of electromagnetic data processed by principal component analysis

M. Andy Kass
Center for Gravity, Electrical, and Magnetic Studies
Colorado School of Mines
Golden, Colorado, USA 80401
mkass@numericalgeo.com

Yaoguo Li
Center for Gravity, Electrical, and Magnetic Studies
Colorado School of Mines
Golden, Colorado, USA 80401
ygli@mines.edu

SUMMARY

Statistical de-noising and compressive inversion methods based on Principal Component Analysis can reduce random noise, separate desired signals from correlated noise, and improve the efficiency and results of airborne EM inversions. However, inversion of PCA-processed data with standard kernels produces inaccurate results due to the improper forward mapping operators used. These inversions must incorporate the PCA rotation in the inversion process for accurate results. In order to appropriately apply these operators to the inversion kernels, the statistical distribution of the noise before and after processing and its effect on the data misfit must be understood. We can then develop compressive inversion techniques utilising PCA.

In this presentation, we demonstrate the need for incorporation of rotation into the inversion kernels through linear examples and show the utility of principal component analysis in compressive inversion. We then examine the statistical distribution of TEM data and noise and show that the noise follows a multivariate t-distribution both before and after processing with PCA. We conclude by introducing a compressive inversion technique formulated in the principal component domain.

Key words: electromagnetics, PCA, statistics, compressive inversion

INTRODUCTION

Principal component analysis (PCA) is a linear transformation whereby multi-channel data are decomposed to a new basis via a rotation defined by the covariance characteristics of the data. By winnowing principal components below a threshold or removing specific components, these data can be compressed, cleaned, or separated into constituent components much like processes based on Fourier or Wavelet transforms (Kramer and Mathews, 1956; Green, 1998).

This process, while effective at removing unwanted signals from multi-channel data, alters the statistical characteristics of the desired data components. As a consequence, any numerical interpretation must account for these changes or artefacts will be introduced in the recovered models, just as in any method of noise removal. In the case of noise removal through frequency filtering, for example, the inversion must include the same type of filter to accurately represent the causative body (e.g. Lee, 2006). In the case of TEM surveys, PCA might separate the power-law decay due to layered structure from a partial exponential signal superimposed (in low susceptibility environments) due to a compact conductor. The same kernels used to invert the entire signal for conductivity distribution cannot be used to invert only the exponential signal. In fact, the physics of the system are no longer the same--how to construct and update a global objective function is unclear without an understanding of the physics that produced the anomaly. Once the global objective function is built, a clear understanding of the statistical distribution of the data is critical for proper choice of optimum misfits.

In this paper, we briefly describe two major uses of principal component analysis in the context of transient electromagnetic surveys: noise reduction and data analysis, and compressive inversion. We then show the requirement of including the rotation matrix calculated by principal component analysis in the inversion kernels for accurate results. In this context, we then investigate the statistical distributions of TEM data to understand the resulting consequences to Tikhonov-based inversion processes using both synthetic and field examples.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is a method of rotating multichannel data onto a different orthonormal basis set. The new basis is constructed via the eigenvectors of the covariance matrix which populate the columns of the rotation matrix, **R**. Therefore the data are rotated onto a new coordinate system that sorts the data by covariance. Thus signals that are pervasive along the survey tend to be in the first few principal components while uncorrelated noise tends to exist in the last few principal components. To briefly describe the rotational scheme, we introduce the Karhunen-Lo\'eve transform.

Given a covariance matrix, Γ, we can decompose this matrix into its constituent eigenvectors as:

$$\Gamma = \mathbf{R}^{\mathrm{T}}\Lambda\mathbf{R}$$

where $\mathbf{R}^{\mathrm{T}}$ is a matrix whose columns are the eigenvectors of Γ and Λ is a diagonal matrix populated by the eigenvalues. We may then rotate our data matrix, $\mathbf{X}$ (with each column representing a single observation location):

$$\Psi = \mathbf{RX}$$

and rotate back via the transpose to yield the reconstructed data:

$$\mathbf{X} = \mathbf{R}^{\mathrm{T}}\Psi.$$

We can remove particular components in the rotated domain to isolate particular signals or remove uncorrelated noise. Multiplying by an identity matrix, $\mathbf{B}$:

$$\tilde{\mathbf{X}} = \mathrm{R}^{\mathrm{T}}\mathrm{B}\Psi$$

yields an equivalent result. We can then set the diagonal elements to zero which correspond to principal components to be removed.

INVERSION OF PROCESSED DATA

Inversion of data processed with PCA must incorporate the rotation into the inversion kernels for accurate results (Kass et. al., 2009). Figure 1 shows a set of TEM data after processing. Removal of the first principal component completely changes the characteristics of the decay curve—clearly this change must be incorporated into inversion.

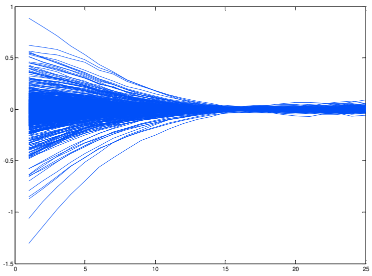

Figure 1: TEM data reconstructed with principal components 2-20. Removal of the first principal component has completely altered the standard TEM decay shape.

By applying the rotation matrix to the forward mapping operator, we can account for the changes introduced by PCA. We construct a block-diagonal matrix containing the rotation matrix calculated by PCA on the diagonal and multiply by the results from our forward mapping operator:

$$\begin{aligned}\left(\mathbf{J}_{\mathrm{r}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}\mathbf{J}_{\mathrm{r}} + \beta\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)\delta\mathbf{m} \\ = \mathbf{J}_{\mathrm{r}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}(\mathbf{d}_{obs} - \mathbf{R}\mathrm{F}[\mathbf{m}]) \\ - \beta\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\left(\mathbf{m} - \mathbf{m}_{ref}\right)\end{aligned}$$

where $\mathbf{J}$ is the calculated sensitivity matrix for the current iteration, $\mathbf{W}_d$ is a data weighting matrix, β is a tradeoff or Tikhonov parameter, $\mathbf{W}_m$ is a model weighting matrix, $\delta\mathbf{m}$ is the calculated model perturbation, $\mathbf{d}_{obs}$ is the observed data, $\mathbf{R}$ is the block-diagonal rotation matrix calculated from PCA, $\mathrm{F}[\mathbf{m}]$ is the forward mapping of the model, $\mathbf{m}$, at the current iteration, and $\mathbf{m}_{ref}$ is a reference model.

Figure 2 shows a linear example of an inversion requiring incorporation of the rotation into the sensitivity matrix. After processing with PCA, the inversion is unable to accurately construct a sensitivity which will result in a model that fits the data. Only by incorporating the rotation matrix can a suitable model be recovered.

Choice of regularization parameter via the L-curve criterion is unaffected by PCA processing. However, choice via discrepancy principle requires an estimate of the error level as well as the assumption that the noise is Gaussian. While the error level can be estimated by the amount of energy removed with the principal components, the Gaussian assumption must be validated.

STATISTICAL PROPERTIES OF TEM DATA

In a Tikhonov-based inversion process, the choice of the tradeoff parameter in the objective function can be chosen through an estimate of the optimum data misfit (discrepancy principle). If the noise comes from a Gaussian distribution, we may assume a χ^2 distribution after data normalisation, which leads to an optimum data misfit value (in a least squares sense) of the number of data. However, after processing with principal component analysis, it is not immediately clear whether or not this relationship still holds. After statistical denoising, it is intuitively clear that the noise level should be less--however the extent of the reduction and the changes in noise distribution must be quantified.

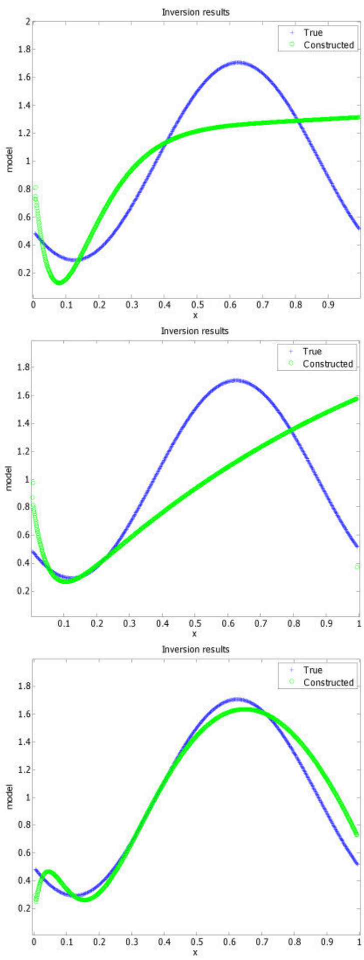

Figure 2: Inversion with PCA. (top) Inversion of noisy data. (middle) Inversion of data processed with PCA. (bottom) Inversion of data processed with PCA and incorporating the rotation matrix into the inversion kernels.

A common method of noise reduction in PCA is to truncate the reconstruction—that is use fewer principal components than are available, setting to zero the components corresponding to the smallest eigenvalues. The number of zeroed components usually is computed by adding up the value of the small eigenvalues until the estimated noise threshold is reached. For example, a 5% error level would lead to removing 5% of the sum of the eigenvalues. If the estimation is exactly correct and all the error comes from independent and identically distributed variables drawn from a Gaussian distribution, then the optimum data misfit would be zero. Because the data errors do not always draw from a Gaussian distribution and because the process does not perfectly separate signals, an understanding of the statistical changes of the noise before and after PCA is required.

Synthetic Test

To understand the effects of PCA rotation on mixtures of distributions, we processed synthetic datasets containing both Gaussian mixtures and Gaussian/Poisson mixtures. When a small number of variables drawn from Gaussians are added together, the aggregate distribution is a multivariate T-distribution. As the number of independent additive distributions increases, the T-distribution approaches a Gaussian, as described by the Central Limit Theorem.

Figure 3 shows the aggregate distribution taken from four Gaussian distributions. Figure 4 shows the resulting histogram after processing with PCA. Figure 4(top) shows the histogram of the data reconstructed with the first principal component. The mean of the recovered distribution is equal to the mean of the original distribution. Subsequent principal components have zero mean (Figure 4(bottom)).

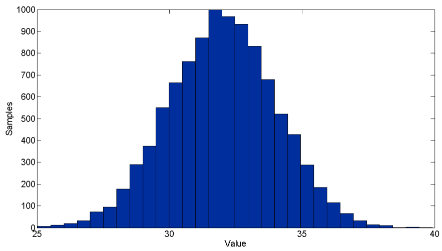

Figure 3: Aggregate distribution of four Gaussian variables.

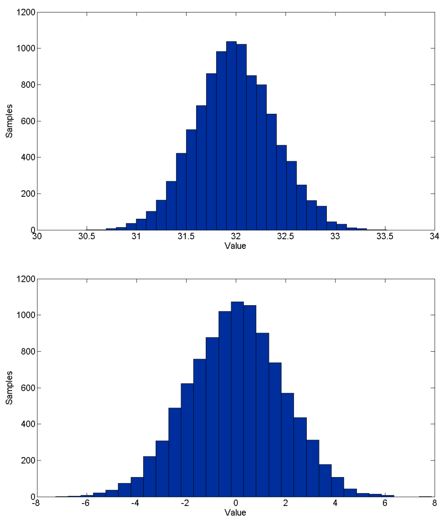

Figure 4: (top) Histogram of reconstructed data with the first principal component only. (bottom) Histogram of reconstructed data from the fifth principal component.

By constructing QQ plots for the data before and after processing, we can see if the distribution shape has changed. If the QQ plot is linear, we may infer that the data still come from the same distribution type as before processing (e.g. Gaussian); however we may not assume the parameters (e.g. mean or variance) have not changed.

Figure 5 shows the QQ plots calculated from the original and processed data. Both for the data constructed from the first principal component as well as for the data constructed from PCs 5-25, the QQ plots indicate that the data come from the same distribution (in this case, a multivariate T). The tails of the distributions however, indicate that the degrees of freedom have changed between the original and processed data.

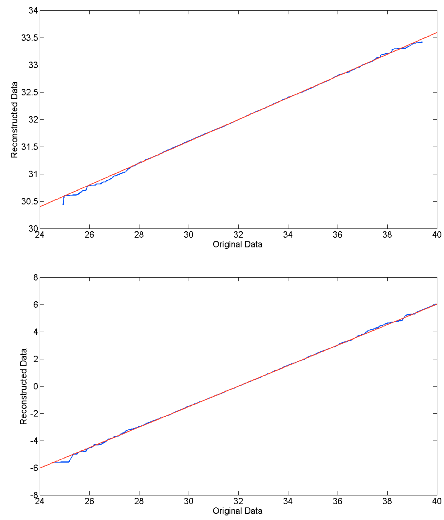

Figure 5: QQ plots between original data and (top) data reconstructed with the first principal component, (bottom) data constructed from PCs 5-25. The blue line represents the QQ plot while the red line is a linear fit for visual reference.

We performed the same test by adding random data sampled from a Poisson distribution (not shown) with the same results. Again, the distribution shape did not change except for the tails of the distribution.

Field Example

We compared the distributions of the noise in a field survey before and after processing in the same manner as the previous section. These data were acquired as part of an unexploded ordnance survey from Kaho'olawe, Hawaii. Figure 6 shows the histogram constructed from the first time-gate. Unlike the synthetic examples, however, the histogram indicates the data come from a bimodal distribution.

To investigate whether the statistical distribution of the noise has changed, we first inverted the raw data. The residuals of this inversion were considered as the original or true noise. We then processed the data with PCA to obtain the PCA-estimated noise. By constructing QQ plots, we see that the distribution shape has remained largely unchanged by PCA (Figure 7). The tails of the distribution have changed indicating a change in the degrees of freedom.

The results of these tests show that the distribution of the data errors remains largely unchanged before and after processing. However, the data errors are not exactly represented by a Gaussian distribution, but rather a multivariate T-distribution with differing degrees of freedom. If a relatively small number of principal components are removed, then any

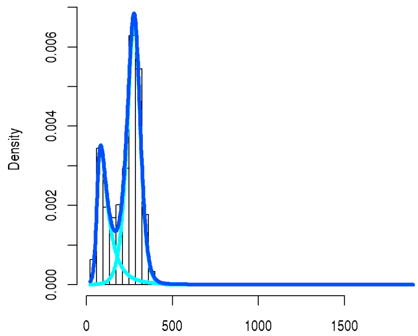

Figure 6: Bimodal distribution from a TEM survey. The blue line indicates two multivariate T-distribution fits to the data.

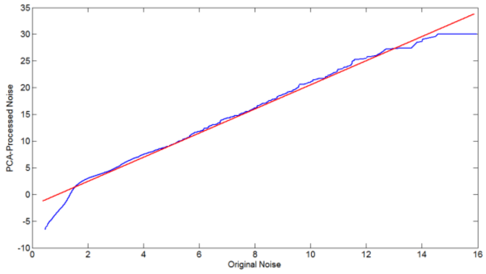

Figure 7: QQ plot between inversion residuals and PCA residuals. The blue line represents the QQ plot while the red line is a linear fit for visual reference.

methods used for selection of tradeoff parameter are no more or less valid than before processing.

COMPRESSIVE INVERSION

With an understanding of the effect of PCA on noise, we can formulate a compressive inversion scheme to dramatically increase the efficiency of constrained 1D inversions, such as those used commonly in airborne electromagnetic surveys. By performing the inversion in the principal component domain, we can reduce the number of inversions required to recover the same-sized model region as the original, stitched inversions (Kass and Li, 2011). Compressive inversion has a similar formulation to that of nonlinear inversion of processed data:

$$\left(\mathbf{J}_r^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J}_r + \beta\mathbf{W}_m^T\mathbf{W}_m\right)\delta\mathbf{m} = \mathbf{J}_r^T\mathbf{W}_d^T\mathbf{W}_d\left(\mathbf{d}_{obs} - \mathbf{R}F[\mathbf{m}]\right) - \beta\mathbf{W}_m^T\mathbf{W}_m\left(\mathbf{m} - \mathbf{m}_{ref}\right)$$

but the rotation matrix, **R**, does not include the inverse rotation. Thus the model perturbations are calculated in the principal component domain. The number of model elements that must be calculated in the principal component domain is much less than what must be calculated normally.

Figure 8 shows a simple linear example with PCA-domain compression. By thresholding the principal component reconstruction at seven (Figure 9), only seven real-valued inversions were required to recover the model rather than the 128 required with standard approaches.

CONCLUSIONS

We have shown that processing with principal component analysis requires incorporation of the rotation matrix into the inversion kernels for accurate results. PCA does not change the distribution of the noise, however the parameters of the distribution are dependent on the processing. The noise follows a multivariate T-distribution, the degrees of freedom of which is a function of the principal components kept and rejected. Since it is the degrees of freedom which controls how close the T-distribution is to Gaussian, this has implications on the validity of the Gaussian assumption. When eliminating a relatively large number of principal components during processing, the choice of the tradeoff parameter should depend on methods such as the L-curve or Generalized Cross Validation. When few PCs are removed, the validity of the Gaussian assumption remains unchanged. With an understanding of the noise characteristics, we have constructed a compressive inversion technique tailored to airborne EM surveys which significantly reduces the computational cost of constrained 1D inversions while maintaining a high recovered-model quality.

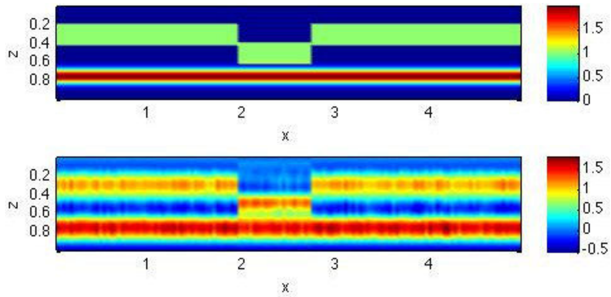

Figure 8: Compressive inversion results. (top) True model. (bottom) Inversion results calculated in the principal component domain.

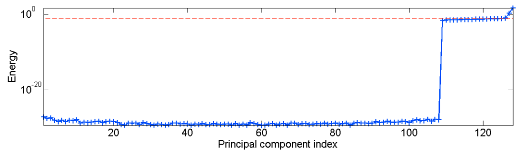

Figure 9: Principal component energy as a function of index.

REFERENCES

Green, A., 1998, The use of multivariate statistical techniques for analysis and display of AEM data: Exploration Geophysics, 29, 77-82.

Kass, M.A., Irons, T., and Li, Y., 2009, Inversion of multichannel data with rotated kernels: 79th SEG International Meeting, Houston, Texas, USA, Expanded Abstracts 28.

Kass, M.A., and Li, Y., 2011, Data-adaptive compressive inversion of multichannel geophysical data: 81st SEG International Meeting, San Antonio, Texas, USA, Expanded Abstracts 30.

Kramer, H., and Mathews, M., 1956, A linear coding for transmitting a set of correlated signals: IRE Transactions on Information Theory, IT-2, 41-46.

Lee, J., 2006, Effects of low pass filtering on inversion of airborne gravity gradiometry data: Masters Thesis, Colorado School of Mines.