Kernel Versions of Some Orthogonal Transformations

Allan A. Nielsen
Technical University of Denmark
Applied Mathematics and Computer Science
people.compute.dtu.dk/alan
alan@dtu.dk

Symposium i anvendt Statistik, 26.-28. januar 2015

Intro

- Kernel versions of PCA and MAF/MNF analysis
- Parsimoneous representation of multi- or hypervariate data
- Interesting projections in feature space with different measures of 'interestingness': max variance, autocorrelation, signal-to-noise ratio

Data matrix

- n by p data matrix or design matrix $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$
- Each row consists of a vector of measurements x_i^T from p variables for a particular observation; X is often column centered

PCA

- R-mode or primal analysis: decompose p by p variance-covariance matrix $S = X^T X/(n-1) = 1/(n-1)\sum_{i=1}^n x_i x_i^T$

$$\frac{1}{n-1} X^T X u_i = \lambda_i u_i$$

- Projections or scores are $x^T u_i$
- Variance of scores maximized

$$u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i$$

PCA

- Q-mode or dual analysis: decompose n by n (Gram) matrix $XX^T/(n-1)$, multiply from left with X, $v_i \propto Xu_i$

$$\frac{1}{n-1} XX^T (Xu_i) = \lambda_i (Xu_i)$$

$$\frac{1}{n-1} XX^T v_i = \lambda_i v_i$$

PCA

- Multiply from left with X^T

$$\frac{1}{n-1} X^T X (X^T v_i) = \lambda_i (X^T v_i)$$

- Solutions related by (Eckart-Young, 1936)

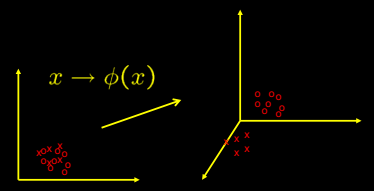
$$u_i = X^T v_i / \sqrt{(n-1)\lambda_i}$$
$$v_i = X u_i / \sqrt{(n-1)\lambda_i}$$

Gram matrix

- Consists of inner products only

$$XX^T = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & x_1^T x_n \\ x_2^T x_1 & x_2^T x_2 & \cdots & x_2^T x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n^T x_1 & x_n^T x_2 & \cdots & x_n^T x_n \end{bmatrix}$$

Nonlinear mapping



Nonlinear mapping

- Example: $x = [z_1 \ z_2]^T$ mapped into $\phi(x) = [z_1 \ z_2 \ z_1^2 \ z_2^2 \ z_1 z_2]^T$
- Maps original 2-dimensional vector into a 5-dimensional feature space so that for example a linear decision rule becomes general enough to differentiate between all linear and quadratic forms including ellipsoids

Nonlinear mapping

- Data matrix

$$X \rightarrow \quad \Phi \quad = \quad \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix}$$

- Each row consists of a q -vector of nonlinear mappings $\phi(x_i)^T$, $q \geq p$

Nonlinear mapping

- Maps Gram matrix into

$$\Phi\Phi^T \quad = \quad \begin{bmatrix} \phi(x_1)^T\phi(x_1) & \phi(x_1)^T\phi(x_2) & \cdots & \phi(x_1)^T\phi(x_n) \\ \phi(x_2)^T\phi(x_1) & \phi(x_2)^T\phi(x_2) & \cdots & \phi(x_2)^T\phi(x_n) \\ \vdots & \vdots & \cdots & \vdots \\ \phi(x_n)^T\phi(x_1) & \phi(x_n)^T\phi(x_2) & \cdots & \phi(x_n)^T\phi(x_n) \end{bmatrix}$$

- Consists of inner products of mappings only, still n by n

PCA

- Q-mode or dual analysis: decompose n by n (Gram) matrix $\Phi\Phi^T/(n-1)$

$$\frac{1}{n-1}\Phi\Phi^T v_i \quad = \quad \lambda_i v_i$$

- Solutions related by

$$u_i = \Phi^T v_i / \sqrt{(n-1)\lambda_i}$$
$$v_i = \Phi u_i / \sqrt{(n-1)\lambda_i}$$

Kernel trick

- We need not know the actual mappings (let λ_i subsume $n-1$)

$$\Phi\Phi^T v_i \quad = \quad \lambda_i v_i$$
$$K v_i \quad = \quad \lambda_i v_i$$

- (Mercer) kernel matrix with elements

$$K_{ij} = \kappa(x_i, x_j)$$

Kernel solution

- Solutions to $\Phi\Phi^T = \lambda_i v_i$ related by

$$u_i = \Phi^T v_i / \sqrt{\lambda_i}$$
$$v_i = \Phi u_i / \sqrt{\lambda_i}$$

- Scores are $\phi(x)^T u_i$

$$\phi(x)^T u_i \quad = \quad \begin{bmatrix} \kappa(x, x_1) & \kappa(x, x_2) & \cdots & \kappa(x, x_n) \end{bmatrix} v_i / \sqrt{\lambda_i}$$

- i.e., expressed by kernel elements (memory based, non-para.)

Kernel solution

- Several basic properties may all be expressed in terms of the kernel function without using the mapping $\phi(x)$ explicitly, including
  - the norm in kernel feature space
  - the distance between observations in feature space
  - the norm of the mean in feature space
  - ***centering to zero mean in feature space***
  - standardisation to unit variance in feature space

Kernel functions

- Popular kernels
  - Stationary, depend on vector difference, invariant under translation

$$\kappa(x_i, x_j) = \kappa(x_i - x_j)$$

  - Homogeneous, RBF

$$\kappa(x_i, x_j) = \kappa(\|x_i - x_j\|)$$

Kernel functions

- Often used RBFs, $h = \|x_i - x_j\|$
  - Multiquadric

$$\kappa(h) = (h^2 + h_0^2)^{1/2}$$

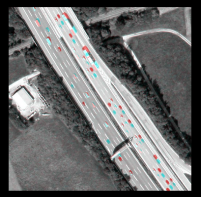
  - Inverse multiquadric

$$\kappa(h) = (h^2 + h_0^2)^{-1/2}$$

  - Gaussian

$$\kappa(h) = \exp(-\tfrac{1}{2}(h/h_0)^2)$$

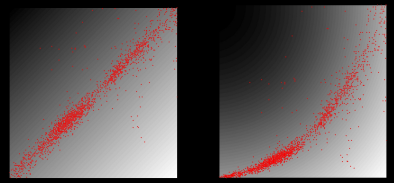
Data from DLR 3K-camera



t_1 as R, t_2 as GB

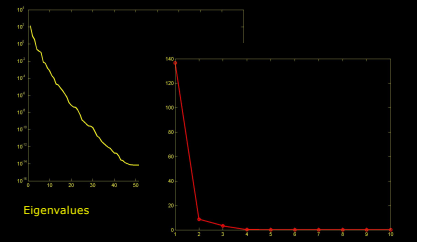
Spatial sub-sample

Data from DLR 3K-camera



PC2 PC2 artificial nonlinearity

Data from DLR 3K-camera



Eigenvalues

Data from DLR 3K-camera

kPC3 kPC3 artificial nonlinearity

MAF

- R-mode or primal analysis: max autocorrelation of linear combinations $a^T x(r)$

$$\begin{aligned}\rho &= 1 - \frac{1}{2}\frac{a^T S_\Delta a}{a^T S a} \\ &= 1 - \frac{1}{2}\frac{a^T X_\Delta^T X_\Delta a}{a^T X^T X a}\end{aligned}$$

MAF

- Q-mode or dual analysis: set $a \propto X^T b$ and kernelize

$$\begin{aligned}\rho &= 1 - \frac{1}{2}\frac{b^T \Phi\Phi_\Delta^T \Phi_\Delta \Phi^T b}{b^T \Phi\Phi^T \Phi\Phi^T b} \\ &= 1 - \frac{1}{2}\frac{b^T K_\Delta K_\Delta^T b}{b^T K^2 b}\end{aligned}$$

MAF kernel solution

- Solutions related by

$$\begin{aligned}a_i &= \Phi^T b_i \\ b_i &= \Phi a_i\end{aligned}$$

- Scores are $\phi(x)^T a_i$

$$\phi(x)^T a_i = \left[\begin{array}{cccc}\kappa(x,x_1) & \kappa(x,x_2) & \cdots & \kappa(x,x_n)\end{array}\right] b_i$$

- i.e., expressed by kernel elements (memory based, non-para.)

MNF

- R-mode or primal analysis: min NF (or max SNR) of linear combinations $a^T x(r)$

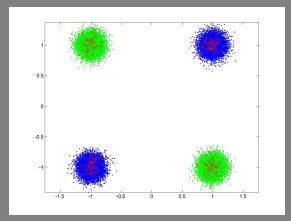
$$\begin{aligned}\frac{1}{\text{NF}} &= \frac{a^T S a}{a^T S_N a} \\ &= \frac{a^T X^T X a}{a^T X_N^T X_N a}\end{aligned}$$

MNF

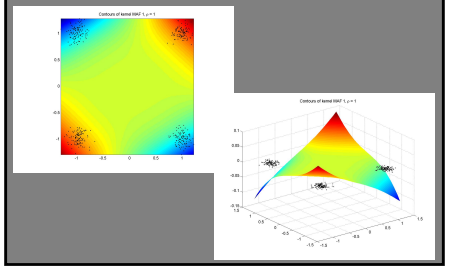
- Q-mode or dual analysis: set $a \propto X^T b$ and kernelize

$$\begin{aligned}\frac{1}{\text{NF}} &= \frac{b^T \Phi\Phi^T \Phi\Phi^T a}{b^T \Phi\Phi_N^T \Phi_N \Phi^T b} \\ &= \frac{b^T K^2 b}{b^T K_N K_N^T b}\end{aligned}$$

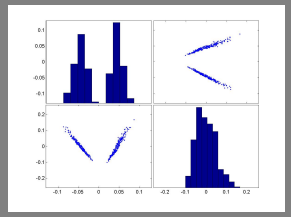
kMAF on XOR



kMAF on XOR



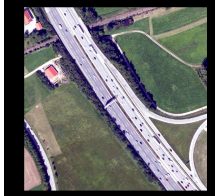
kMAF on XOR



kMAF/kMNF, DLR 3K

- kPC example used only the red spectral band, actually RGB
- (k)MAF of simple differences of the "calibrated" data
- If not calibrated data use generalised differences (e.g., MAD transformation, iterative scheme based on CCA)

kMAF/kMNF, DLR 3K



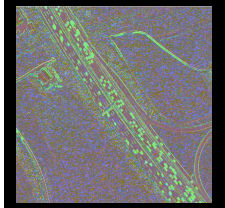
τ_1

τ_2

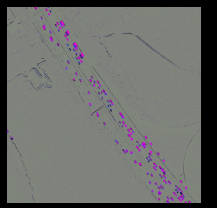
kMAF/kMNF, DLR 3K

$t_2 - t_1$

kMAF/kMNF, DLR 3K

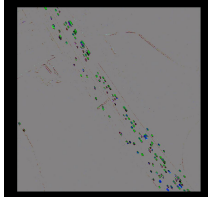


kPC 1, 2, 3

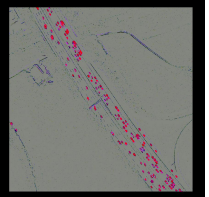


kMAF 1, 2, 3
(SNR 85.0 dB)

kMAF/kMNF, DLR 3K



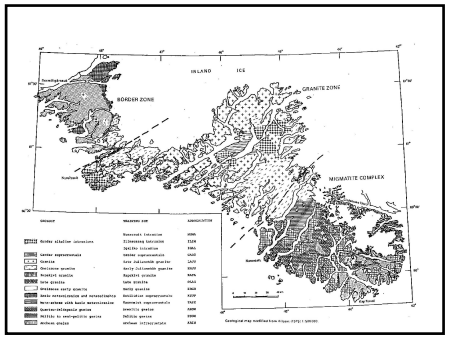
kMNFd 1, 2, 3



kMNFq 1, 2, 3
(SNR 85.5 dB; SNR 100.6 dB)

kMAF, South Greenland

- 2,097 stream sediment samples analysed for content of 41 elements: Ca, Cu, Fe, Ga, K, Mn, Nb, Ni, Pb, Rb, Sr, Ti, Y, Zn and Zr (energy-dispersive isotope excited x-ray fluorescence) and Au, Ag, As, Ba, Br, Co, Cr, Cs, Hf, Mo, Na, Sb, Sc, Se, Ta, Th, U, W, La, Ce, Nd, Sm, Eu, Tb, Yb and Lu (instrumental neutron activation analysis)
- All obs used in Q-mode analysis



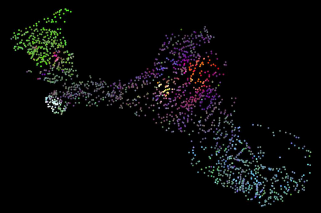
kMAF, South Greenland

The study area is underlain by a Palaeoproterozoic orogen, the Ketilidian orogen, which consists of three major tectono-stratigraphic units: (1) a northern Border zone of tectonically reworked Archaean gneissic basement overlain by Palaeoproterozoic metasediments and metavolcanics in the north-east, (2) a central zone occupied by a calc-alkaline granitic batholith, and (3) a southern migmatite complex of predominantly Palaeoproterozoic metasediments and metavolcanics intruded by post-tectonic rapakivi type granites, see Figure 1 (left) and [20]. The plate-tectonic setting of the orogen has recently been interpreted in [21]. In Mesoproterozoic times the boundary region between the border and the granite zones was subjected to rifting and intrusions of numerous dykes of basaltic to trachytic compositions as well as of felsic alkaline complexes including carbonatites. The region affected by the alkaline magmas is termed the Gardar province. [22]

kMAF, South Greenland

PC 1, 2, 3 as RGB

kMAF, South Greenland



MAF 1, 2, 3 as RGB

kMAF, South Greenland

kPC 1, 2, 3 as RGB

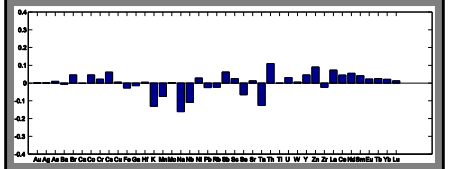
kMAF, South Greenland

kMAF 1, 2, 3 as RGB

kMAF, South Greenland

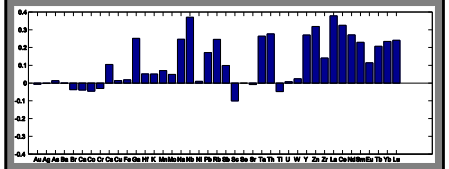
- Autocorrelations for first three MAFs 0.8379, 0.7585 and 0.7279
- Autocorrelations for first three kMAFs 0.9999, 0.9997 and 0.9996

kMAF, South Greenland



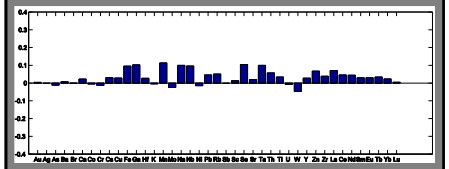
Corr kMAF 1 with all

kMAF, South Greenland



Corr kMAF 2 with all

kMAF, South Greenland

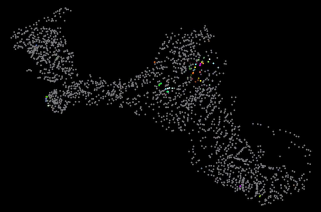


Corr kMAF 3 with all

kMAF, South Greenland

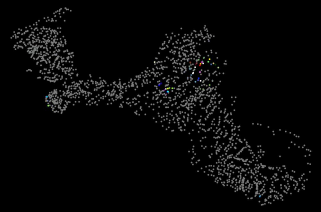
kMAF 1, 2, 3 as RGB

kMAF, South Greenland



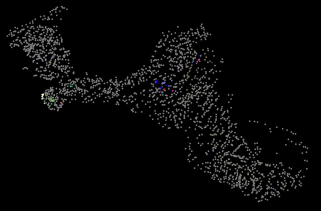
kMAF 1, 2, 3 as RGB, mq

kMAF, South Greenland



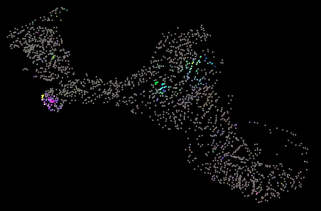
kMAF 1, 2, 3 as RGB, imq

kMAF, South Greenland



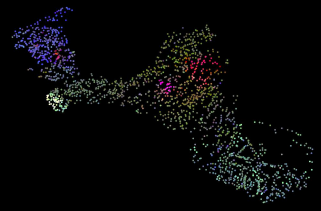
kMAF 1, 2, 3 as RGB, 10x width

kMAF, South Greenland



kMAF 1, 2, 3 as RGB, 100x width

kMAF, South Greenland



kMAF 1, 2, 3 as RGB, 1000x width

kMAF, South Greenland

- Partial reconstruction of simplified geological map: nice, makes one trust the technique
- Potentially more important: where does the analysis not agree with conventional wisdom?

Future work

- Kernel function
- Scale parameter, kernel width
- Dependence on particular training samples
- Parallel implementation (test data part), or run on graphics board GPU
- Pre-image problem
- …

Conclusions

- Kernel orthogonalisation is based on Q-mode analysis
- Data are replaced by nonlinear mappings
- Inner products of nonlinear mappings in turn replaced by kernel function, 'kernel trick'
- kPCA unlike linear PCA successfully finds change observations in a DLR 3K case where nonlinearities are introduced artificially

Conclusions

- kMAF/kMNF unlike kPCA successfully finds change observations in DLR 3K colour image data
- (k)MAF unlike (k)PCA successfully finds relevant geological structures
- kMAF variates with increasing kernel width find geologically relevant regions of increasing scale; focus on extreme observations – smooth background esp. for small kernel widths

References

- H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology* **24**, 417-441 and 498-520 (1933).
- P. Switzer and A. A. Green, "Min/max autocorrelation factor analysis", Technical Report 6, Department of Statistics, Stanford University (1984).
- A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing* **26**(1), 65-74 (1988).
- B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation* **10**(5), 1299-1319 (1998).
- J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press (2004).
- C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).

References

- H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology* **24**, 417-441 and 498-520, 1933.
- P. Switzer and A. A. Green, "Min/max autocorrelation factor analysis", Technical Report 6, Department of Statistics, Stanford University, 1984.
- A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing* **26**(1), 65-74, 1988.
- A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, "Spatial Factor Analysis of Stream Sediment Geochemistry Data from South Greenland," in Vera Pawlowsky-Glahn (Ed.) Proceedings of the Third Annual Conference of the International Association for Mathematical Geology, IAMG'97, Barcelona, Spain, 22-27 September 1997, pp. 955-960.
- B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation* **10**(5), 1299-1319, 1998.
- A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, "Maximum Autocorrelation Factorial Kriging," in W.J. Kleingeld and D.G. Krige (editors), Proceedings of the 6th International Geostatistics Congress, Geostats 2000, pp. 538-547, Cape Town, South Africa, 10-14 April 2000.
- J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- A. A. Nielsen and M. J. Canty, "Kernel principal component and maximum autocorrelation factor analyses for change detection," SPIE Europe Remote Sensing Conference, Berlin, Germany, 31 August-3 September 2009.
- A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing* **20**(3), 612-624, 2011.