Informative Data Fusion: Beyond Canonical Correlation Analysis

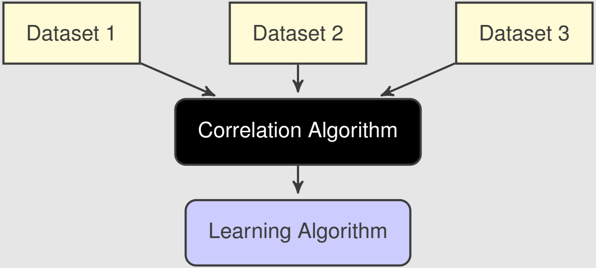
Nicholas Asendorf

asendorf@umich.edu

Department of Electrical Engineering and Computer Science
University of Michigan

Dissertation Defense

May 4, 2015



Thesis Goal

Develop theoretically justified, robust correlation algorithms for multi-dataset fusion

Thesis Outline

1. Introduction
2. Performance of Matched Subspace Detectors Using Finite Training Data
3. Extensions of Deterministic MSD to Missing Data and Useful Subspace Components
4. Detection of Correlated Signals using CCA and ICCA
5. Estimation of Canonical Vectors in CCA and ICCA
6. Top Singular Values of Cross Covariance Matrices
7. Signal Detection of Random Projections of Signal-Plus-Noise Matrices
8. Correlation Based Methods for Detection and Regression
9. Correlation Based Methods for Image Annotation
10. Detection of Correlated Signals in More than Two Datasets

Thesis Outline

1. Introduction
2. Performance of Matched Subspace Detectors Using Finite Training Data
3. Extensions of Deterministic MSD to Missing Data and Useful Subspace Components
4. Detection of Correlated Signals using CCA and ICCA
5. Estimation of Canonical Vectors in CCA and ICCA
6. Top Singular Values of Cross Covariance Matrices
7. Signal Detection of Random Projections of Signal-Plus-Noise Matrices
8. Correlation Based Methods for Detection and Regression
9. Correlation Based Methods for Image Annotation
10. Detection of Correlated Signals in More than Two Datasets

Two-Dataset Model

Linear Subspace Model

$$
\boxed{\begin{aligned} x_i &= U_x s_{x,i} + z_{x,i} \\ y_i &= U_y s_{y,i} + z_{y,i} \end{aligned}}
$$

Parameters

- $U_x^H U_x = I_{k_x}$, $U_y^H U_y = I_{k_y}$
- $z_{x,i} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_p)$, $\;z_{y,i} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_q)$
- $\mathbb{E}\left[\begin{bmatrix} s_{x,i} \\ s_{y,i} \end{bmatrix}\begin{bmatrix} s_{x,i}^H & s_{y,i}^H \end{bmatrix}\right] = \begin{bmatrix} \Theta_x & K_{xy} \\ K_{xy}^H & \Theta_y \end{bmatrix}$
- $K_{xy} = \Theta_x^{1/2} P_{xy} \Theta_y^{1/2}$
- $\Theta_x = \textbf{diag}\left(\left(\theta_1^{(x)}\right)^2, \ldots, \left(\theta_{k_x}^{(x)}\right)^2\right)$, $\;\Theta_y = \textbf{diag}\left(\left(\theta_1^{(y)}\right)^2, \ldots, \left(\theta_{k_y}^{(y)}\right)^2\right)$
- P_{xy} contains correlations ρ_{kj} between signals of x_i and y_i
- $\widetilde{K}_{xy} = \left(\Theta_x + I_{k_x}\right)^{-1/2} K_{xy} \left(\Theta_y + I_{k_y}\right)^{-1/2}$, with singular values $\kappa_1, \ldots, \kappa_{\min(k_x, k_y)}$

Canonical Correlation Analysis

What is it?

- Dimensionality reduction algorithm for exactly 2 datasets
- Correlation coefficients, linear transformations

What is it not?

- Data fusion algorithm

Covariance matrices

- $R_{xx} = \mathbb{E}\left[x_i x_i^H\right]$
- $R_{yy} = \mathbb{E}\left[y_i y_i^H\right]$
- $R_{xy} = \mathbb{E}\left[x_i y_i^H\right]$

Variable Transformation

- $f = R_{xx}^{1/2} w_x$
- $g = R_{yy}^{1/2} w_y$

Optimization problem

$$\begin{aligned} \underset{w_x, w_y}{\operatorname{argmax}} \quad & \rho = w_x^H R_{xy} w_y \\ \text{subject to} \quad & w_x^H R_{xx} w_x = 1 \\ & w_y^H R_{yy} w_y = 1 \end{aligned}$$

What is it?

- Dimensionality reduction algorithm for exactly 2 datasets
- Correlation coefficients, linear transformations

What is it not?

- Data fusion algorithm

Optimization problem

$$\begin{array}{ll} \underset{f,g}{\text{argmax}} & \rho = f^H \underbrace{R_{xx}^{-1/2} R_{xy} R_{yy}^{-1/2}}_{C_{\text{cca}}} g \\ \text{subject to} & \|f\|_2 = 1 \,,\ \|g\|_2 = 1 \end{array}$$

Canonical Vectors

- $w_x = R_{xx}^{-1/2} f$
- $w_y = R_{yy}^{-1/2} g$

Insight

$$\boxed{\text{\# correlated signals} = k = \text{rank}(C_{\text{cca}})}$$

Empirical CCA

Training Datasets

- $X = [x_1, \ldots, x_n]$
- $Y = [y_1, \ldots, y_n]$

Sample Covariance Matrices

- $\widehat{R}_{xx} = \frac{1}{n} XX^H$
- $\widehat{R}_{yy} = \frac{1}{n} YY^H$
- $\widehat{R}_{xy} = \frac{1}{n} XY^H$

Estimate

$$\begin{aligned}
\widehat{C}_{\text{cca}} &= \widehat{R}_{xx}^{-1/2} \widehat{R}_{xy} \widehat{R}_{yy}^{-1/2} \\
&= \widehat{F} \widehat{K} \widehat{G}^H
\end{aligned}$$

Questions

- How to estimate k?
- When do $\widehat{\rho}_{\text{cca}}^{(i)} = \widehat{k}_i$ represent actual correlations?
- How accurate are $\widehat{w}_x^{(i)} = \widehat{R}_{xx}^{-1/2} \widehat{f}_i$ and $\widehat{w}_y^{(i)} = \widehat{R}_{yy}^{-1/2} \widehat{g}_i$?
- Can we do better?

Statistical Test for CCA Correlations

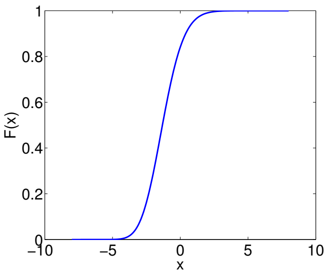
Estimate of # of Correlated Signals

$$\widehat{k}_{\text{cca}} = \sum_{i=1}^{\min(p,q)} \mathbb{1}_{\left\{ \left(\widehat{\rho}_{\text{cca}}^{(i)} \right)^2 > \tau_{\text{cca}}^{\alpha} \right\}}$$

Setting the threshold

- F_{cca} is the cdf of largest singular values of $\widehat{C}_{\text{cca}}$ in the null setting of no correlation
- $\tau_{\text{cca}}^{\alpha} = F_{\text{cca}}^{-1}(1-\alpha)$
- $\tau_{\text{cca}}^{\alpha} \approx \sigma_{n,p,q} \text{TW}_{\mathbb{C}}^{-1}(1-\alpha) + \mu_{n,p,q}$

Tracy-Widom Distribution



Empirical CCA Consistency

Theorem (Empirical CCA Consistency)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$. Given the above linear subspace data model,

$$\widehat{k}_{cca} \xrightarrow{a.s.} k \qquad \text{if } \kappa_k^2 > r_c \text{ and } n > p + q$$

where

$$r_c = \frac{c_x c_y + \sqrt{c_y c_y (1 - c_x)(1 - c_y)}}{(1 - c_x)(1 - c_y) + \sqrt{c_x c_y (1 - c_x)(1 - c_y)}}.$$

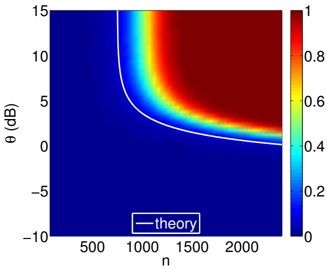
Recall

- $\widetilde{K}_{xy} = \left(\Theta_x + I_{k_x}\right)^{-1/2} \Theta_x P_{xy} \Theta_y \left(\Theta_y + I_{k_y}\right)^{-1/2}$
- Singular values $\kappa_1, \ldots, \kappa_{\min(k_x, k_y)}$

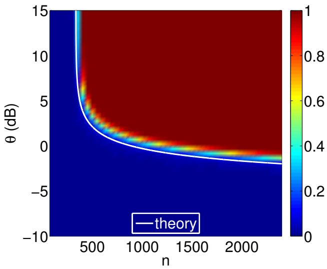
Simulation parameters

- $p = q = 150$, $k = 1$, $\theta_x = \theta_y$, $\alpha = 0.01$

$\rho = 0.5$



$\rho = 0.9$



Problems!

- Degenerate case when $n < p + q$ (Pezeshki 2004)
- Consistency boundary is dependent on correlation

Not all singular vectors are informative! (Nadakuditi, 2011)

- ✳ Trim data SVD's to only use informative components

1. Trim data SVD's: $X = \widehat{U}_x \widehat{\Sigma}_x \widehat{V}_x^H$ and $Y = \widehat{U}_y \widehat{\Sigma}_y \widehat{V}_y^H$
   - ✳ $\mathring{U}_x = \widehat{U}_x \left(:, 1 : \widehat{k}_x\right)$, $\mathring{U}_y = \widehat{U}_y \left(:, 1 : \widehat{k}_y\right)$
   - ✳ $\mathring{V}_x = \widehat{V}_x \left(:, 1 : \widehat{k}_x\right)$, $\mathring{V}_y = \widehat{V}_y \left(:, 1 : \widehat{k}_y\right)$
2. Form $\widehat{C}_{\text{icca}} = \mathring{U}_x \mathring{V}_x^H \mathring{V}_y \mathring{U}_y$
3. Take SVD: $\widehat{C}_{\text{icca}} = \widetilde{F}\widetilde{K}\widetilde{G}^H$
4. $\widehat{\rho}_{\text{icca}}^{(i)} = \widetilde{k}_i$
5. $\widetilde{w}_x^{(i)} = \widehat{R}_{xx}^{-1/2} \widetilde{f}_i$
6. $\widetilde{w}_y^{(i)} = \widehat{R}_{yy}^{-1/2} \widetilde{g}_i$

Estimate of # of Correlated Signals

$$\boxed{\widehat{k}_{\text{icca}} = \sum_{i=1}^{\min(\widehat{k}_x, \widehat{k}_y)} \mathbb{1}_{\left\{ \left(\widehat{\rho}_{\text{icca}}^{(i)}\right)^2 > \tau_{\text{icca}}^{\alpha} \right\}}}$$

Setting the threshold

- F_{icca} is the cdf of largest singular values of $\widehat{C}_{\text{icca}}$ in the null setting of no correlation
- $\tau_{\text{icca}}^{\alpha} = F_{\text{icca}}^{-1}(1-\alpha)$
- $\tau_{\text{icca}}^{\alpha} \approx \sigma_{n,\widehat{k}_x,\widehat{k}_y} \text{TW}_{\mathbb{C}}^{-1}(1-\alpha) + \mu_{n,\widehat{k}_x,\widehat{k}_y}$

Theorem (ICCA Consistency)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$. Given the linear subspace data model,

$$\widehat{k}_{icca} \xrightarrow{a.s.} k \quad \text{if } \min_{i=1,\ldots,k_x} \theta_i^{(x)} > c_x^{1/4} \text{ and } \min_{i=1,\ldots,k_y} \theta_i^{(y)} > c_y^{1/4}$$

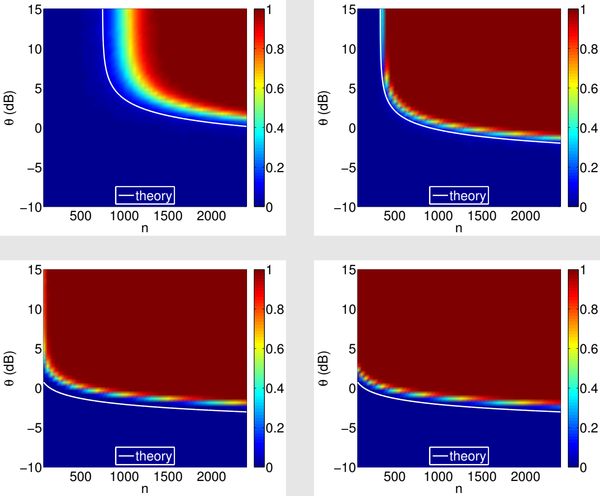
CCA and ICCA Consistency

- $p = q = 150$, $k = 1$, $\theta_x = \theta_y$, $\alpha = 0.01$

$\rho = 0.5$ $\rho = 0.9$

CCA

ICCA

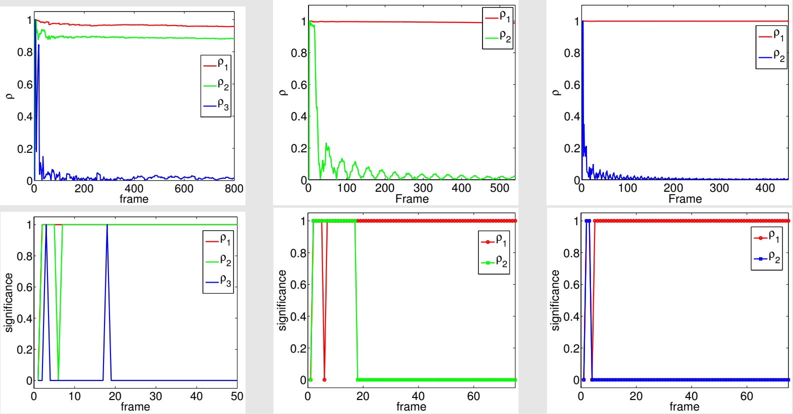


CCA and ICCA Demonstration

* Two-Camera flashing light demonstration
* Audio-Video demonstration
* Audio-Audio demonstration

CCA and ICCA Demonstration

Video-Video | **Audio-Video** | **Audio-Audio**



Missing Data Model

Matrix model

- $V_x = \left[s_{x,1}, \ldots, s_{x,n}\right]$, $V_y = \left[s_{y,1}, \ldots, s_{y,n}\right]$
- $Z_x = \left[z_{x,1}, \ldots, z_{x,n}\right]$, $Z_y = \left[z_{y,1}, \ldots, z_{y,n}\right]$

$$\boxed{\begin{aligned} X &= \left(U_x V_x^H + Z_x\right) \odot M_x \\ Y &= \left(U_y V_y^H + Z_y\right) \odot M_y \end{aligned}}$$

$$M_{ij}^x = \begin{cases} 1 & \text{with probability } \gamma_x \\ 0 & \text{with probability } 1 - \gamma_x \end{cases} \quad M_{ij}^y = \begin{cases} 1 & \text{with probability } \gamma_y \\ 0 & \text{with probability } 1 - \gamma_y \end{cases}$$

- $\odot$ denotes the Hadamard or element-wise product.

Theorem (Missing data consistency)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$ and assume a low-coherence condition on the signal vectors. Given a linear subspace data model with missing data entries,

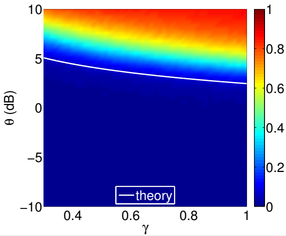
$$\widehat{k}_{cca} \xrightarrow{a.s.} k \qquad \text{if} \min_{i=1,\ldots,k} \mathring{\kappa}_i^2 > r_c \text{ and } n > p + q$$

$$\widehat{k}_{icca} \xrightarrow{a.s.} k \qquad \text{if} \min_{i=1,\ldots,\widehat{k}_x} \theta_i^{(x)} > \frac{c_x^{1/4}}{\sqrt{\gamma_x}} \text{ and } \min_{i=1,\ldots,\widehat{k}_y} \theta_i^{(y)} > \frac{c_y^{1/4}}{\sqrt{\gamma_y}}$$

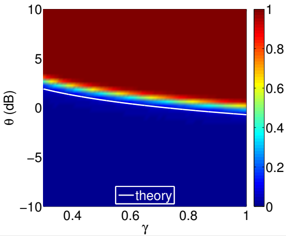
where $\mathring{\kappa}_i$ are the singular values of

$$\left(\gamma_x \Theta_x + I_{k_x}\right)^{-1/2} \left(\gamma_x \Theta_x\right)^{1/2} P_{xy} \left(\gamma_y \Theta_y\right)^{1/2} \left(\gamma_y \Theta_y + I_{k_y}\right)^{-1/2}.$$

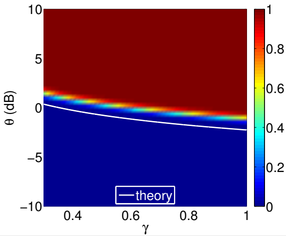
- $p = q = 150$, $k = 1$, $\theta_x = \theta_y$,$\alpha = 0.01$



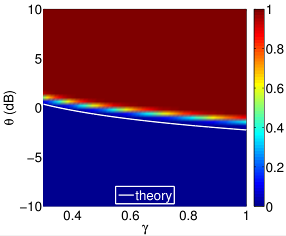
(a) CCA $\rho = 0.5$,$n = 1200$



(b) CCA $\rho = 0.9$,$n = 1200$



(c) ICCA $\rho = 0.5$, $n = 1200$



(d) ICCA $\rho = 0.9$, $n = 1200$

Regularized CCA (RCCA)

Optimization problem

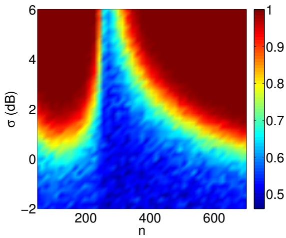
$$\begin{aligned} \underset{w_x, w_y}{\operatorname{argmax}} \quad & \rho = w_x^H R_{xy} w_y \\ \text{subject to} \quad & w_x^H R_{xx} w_x + \eta \|w_x\|_2^2 \leq 1 \\ & w_y^H R_{yy} w_y + \eta \|w_y\|_2^2 \leq 1 \end{aligned}$$

SVD Solution

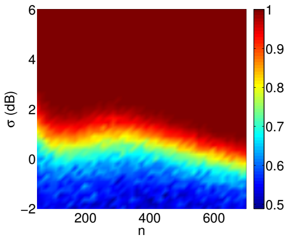
$$\widehat{C}_{\text{reg}} = \left(\widehat{R}_{xx} + \eta I_p\right)^{-1/2} \widehat{R}_{xy} \left(\widehat{R}_{yy} + \eta I_q\right)^{-1/2}$$

Limit of RCCA

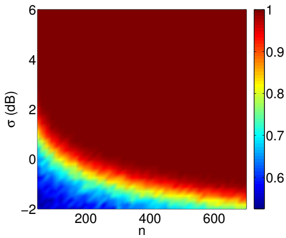
- AUC for signal vs. noise, $k = 1$, $p = 100$, $q = 150$, $\rho = 0.9$



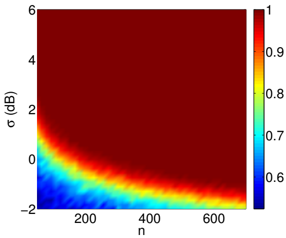
(a) $\eta = 0.0001$



(b) $\eta = 0.01$



(c) $\eta = 10$



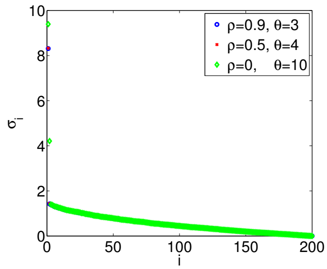
(d) $\eta = 1000$

Theorem (Limit of RCCA)

Let X and Y be the two data matrices used in CCA. As $\eta \to \infty$, the solution to the RCCA optimization problem is obtained through the SVD of $\widehat{R}_{xy} = \frac{1}{n} X Y^H$.

Insights

- From CCA, # correlated components $=$ **rank**$(C_{\text{cca}}) =$ **rank**(R_{xy})
- LRCCA examines exactly this matrix!
- Using **rank**$(\widehat{R}_{xy})$ can lead to dubious results



Theorem (Top singular values of LRCCA)

Under low rank assumptions on X and Y, the maximum singular values of $\frac{1}{n}XYHT$ *solve*

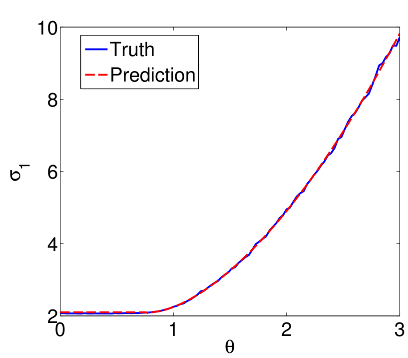
$$0 = \prod_{i=1}^{r} \left[\left(\varphi_H(\sigma_i)\varphi_F(\sigma_i) - \frac{1}{\theta_{yi}^2} \right) \left(\varphi_J(\sigma_i)\varphi_G(\sigma_i) - \frac{1}{\theta_{xi}^2} \right) \right.$$
$$\left. -\rho_i^2 \varphi_H(\sigma_i)\varphi_G(\sigma_i)\left(1 + \varphi_K(\sigma_i)\right)^2 \right]$$

where $\varphi(\cdot)$ *are functions of the Stieltjes transforms of matrix products involving* $R = X^H X$ *and* $S = Y^H Y$.

Singular Value Prediction of XY^H

Simulation parameters

- $p = 200$, $q = 400$, $n = 400$
- $r = 1$, $\rho = 1$
- $\theta = \theta_x = \theta_y$
- Numerically solve above equation using RMTool

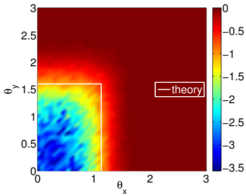


$$0 = \left(\varphi_H(b)\varphi_F(b) - \frac{1}{\theta_y^2}\right)\left(\varphi_J(b)\varphi_G(b) - \frac{1}{\theta_x^2}\right) - \rho^2\varphi_H(b)\varphi_G(b)\left(1 + \varphi_K(b)\right)^2$$

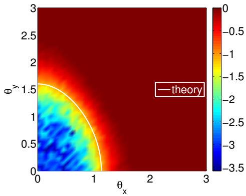
Simulation parameters

- $b = \sigma_1(XY^H)$
- $p = 200$, $q = 400$, $n = 400$, $r = 1$
- Numerically solve above equation using RMTool
- Plot log of KS-statistic

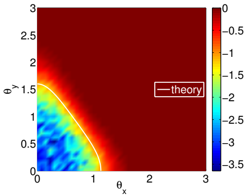
$\rho = 0$



$\rho = 0.6$

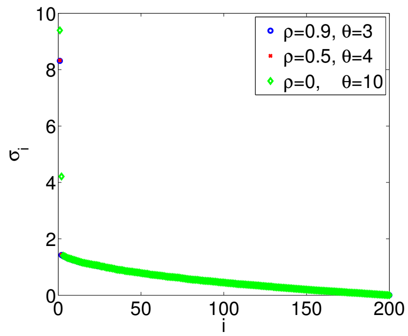


$\rho = 1$



What about using SNR estimates?

- Estimate θ_x and θ_y from individual datasets
- Whiten the singular values of $\widehat{R}_{xy}$
- This is CCA/ICCA!!



Signal-plus-noise Matrix

$$\widetilde{X}_n = \sum_{i=1}^{r} \theta_i u_i v_i^T + X_n.$$

Parameters

- $\theta_i > 0$ for $i = 1, \ldots, r$
- $u_i^H u_j = \delta_{\{i=j\}}$ and $v_i^H v_j = \delta_{\{i=j\}}$.
- $u_i \in \mathbb{C}^{n\times 1}$, $v_i \in \mathbb{C}^{N\times 1}$
- X_n is a random matrix with $\mu_{X_n} = \frac{1}{\min(n,N)} \sum_{i=1}^{\min(n,N)} \delta_{\sigma_i}$.
- $\mu_{X_n} \to \mu_X$

Signal Detection Using Random Projections

Motivation

- When n is large, SVD of $\widetilde{X}_n$ is expensive
- The detection limit of $\widetilde{X}_n$ is $\theta > \left(\frac{n}{N}\right)^{1/4}$
- Project to lower-dim space to save computation, $m < n$
- This will result in a performance loss

Projections

$$Y_n^G = G_n^H \widetilde{X}_n$$

$$Y_n^Q = Q_n^H \widetilde{X}_n$$

- $G_n \in \mathbb{C}^{n \times m}$ with independent $\mathcal{CN}(0,1)$ entries
- $Q_n \in \mathbb{C}^{n \times m}$ s.t. $Q_n^H Q_n = I_m$

Goal

Quantify the performance loss as a function of m, n, N, θ when using the SVD of Y_n^G and Y_n^Q to detect signals

Main Result

Theorem (Largest singular values)

Let Y_n be the projection of $\widetilde{X}_n$ onto either G_n or Q_n. The largest r singular values of the $m \times N$ matrix Y_n exhibit the following behavior as $n, m, N \to \infty$ with $n/N \to c_1$ and $m/N \to c_2$. We have that for each fixed $1 \leq i \leq r$, $\sigma_i(Y_n)$ solves

$$\sigma_i^2 \varphi_F(\sigma_i)\varphi_H(\sigma_i) = \frac{1}{\theta_i^2}, \tag{1}$$

where

$$\varphi_F(\sigma_i) \xrightarrow{a.s.} -\mathbb{E}\left[x m_{\mu_{RS|R}}\left(\sigma_i^2, x\right)\right]_{\mu_R}$$

$$\varphi_H(\sigma_i) \xrightarrow{a.s.} -\frac{n}{N} m_{M_3}(\sigma_i^2) - \frac{1}{\sigma_i^2}\frac{n-N}{N}$$

where m_{μ_M} is the Stieltjes transform of a matrix M and μ_R is the limiting eigenvalue density of either $G_n G_n^H$ or $Q_n Q_n^H$, μ_S is the limiting eigenvalue density of $X_n X_n^H$ and M_3 is either $G_n G_n^H X_n X_n^H$ or $Q_n Q_n^H X_n X_n^H$.

Corollaries to Main Result

Corollary (Phase transition)

When

$$\theta_i \leq \theta_{crit} = \frac{1}{b\sqrt{\varphi_F(b)\varphi_H(b)}}$$

Then

$$\sigma_i \xrightarrow{a.s.} b,$$

where b is the supremum of the support of the limiting density of $G_n^H X_n$ or $Q_n^H X_n$.

Corollary (Unitary closed form)

When Y_n is generated using a unitary matrix Q_n,

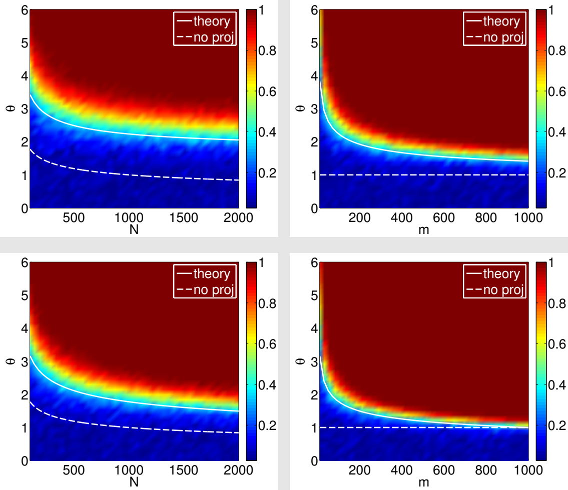
$$\sigma_i \xrightarrow{a.s.} \begin{cases} \sqrt{\frac{c_1}{\theta_i^2} + c_2\theta_i^2 + 1 + c_1 c_2} & \text{if } \theta_i \geq \left(\frac{c_1}{c_2}\right)^{1/4} \\ \sqrt{c_1 c_2} + 1 & \text{if } \theta_i < \left(\frac{c_1}{c_2}\right)^{1/4} \end{cases}$$

Numerical Verification

$m = 100$ | $N = 1000$

Gaussian

Unitary



Content Based Image Retrieval

Training Data

Caption 1

Caption 2

⋮ ⋮

Caption *n*

Text Query

snowy mountains

$\Downarrow$

Automatic Image Annotation

Training Data

→ Caption 1

→ Caption 2

⋮ ⋮

→ Caption *n*

Image Query

⇓

snowy mountains

Correlation Based Approaches

Intuition

- Image features are correlated to words
- CCA/ICCA seem like natural algorithms to identify correlations

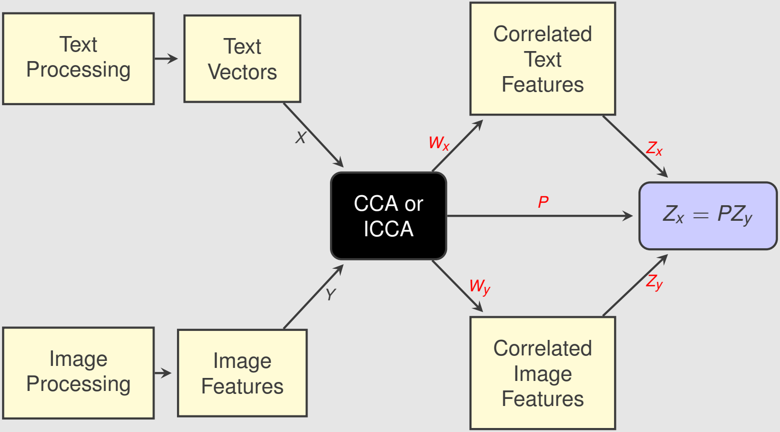
Text Processing - *tf*idf* feature vectors**

- term frequency (*tf*): frequent words are important
- inverse document frequency (*idf*): unique document words are important
- optional Porter stemming and stopword removal

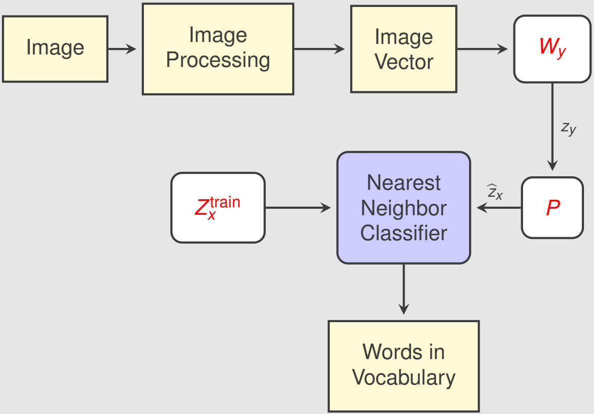
Image Processing - Visual Words

1. create SIFT features for all training images
2. k-mean clustering of all SIFT features to "visual words"
3. assign each SIFT keypoint to closest visual word
4. bag of words count the occurrences of visual word in each image

Training Pipeline



Automatic Caption Annotation Pipeline



Pascal Dataset

- 1000 images each with 5 captions
- Vocabulary: 2393 words

- A D-ERFW-6 in flight.
- An army green plane flying in the sky.
- An old fighter plane flying with German military markings.
- A small green and yellow plane in the sky.
- A WWII fighter plane with its landing gear down.

Text Query: airplane

(a) CCA Results

(b) ICCA Results

Image Query

CCA Annotation

1. hairless
2. buddi
3. swan
4. leaf-less
5. bnsf
6. desert
7. fluffi
8. salad
9. majest
10. memorabilia

ICCA Annotation

1. plane
2. ship
3. cruis
4. fly
5. blue
6. jet
7. airplan
8. dock
9. fighter
10. through

Application to Image Annotation

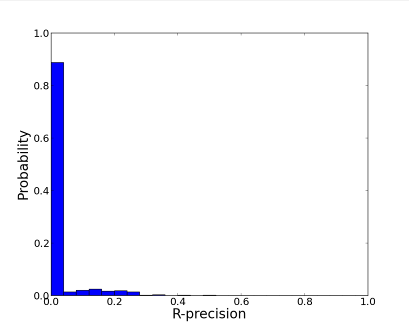
University of Washington Ground Truth Dataset

- 1109 image-label pairs
- Vocabulary 346 unique words

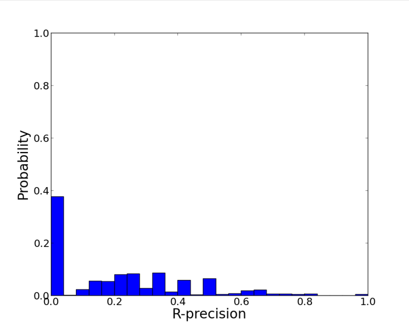
clear, sky, building, ground, statue, people, bush, grass, truck

Clear, sky, snow, Mountains, Rockes

CCA
Mean Accuracy: 0.0184



ICCA
Mean Accuracy: 0.2067



What if we have m datasets $Y_1, \ldots, Y_m$?

Notation

- covariance matrices: $R_{ij} = \mathbb{E}\left[y_i y_j^H\right]$
- canonical vectors: $x_1, \ldots, x_m$
- canonical variates: $w_1, \ldots, w_m$, $w_i = x_i^H y_i$

$$\Phi(x) = E[ww^H] = \begin{bmatrix} x_1^H R_{11} x_1 & \ldots & x_1^H R_{1m} x_m \\ \vdots & \ddots & \vdots \\ x_m^H R_{m1} x_1 & \ldots & x_m^H R_{mm} x_m \end{bmatrix}$$

Optimization Problem

$$\begin{aligned} \underset{x}{\text{optimize}} \quad & J(\Phi(x)) \\ \text{subject to} \quad & h(x, R) \end{aligned}$$

Ambiguity in Objective Function

SUMCORR $\underset{x_1,\dots,x_m}{\operatorname{argmax}} \sum_{i=1}^{m}\sum_{j=1}^{m} x_i^H R_{ij} x_j$

SSQCORR $\underset{x_1,\dots,x_m}{\operatorname{argmax}} \sum_{i=1}^{m}\sum_{j=1}^{m} \left(x_i^H R_{ij} x_j\right)^2$

MAXVAR $\underset{x_1,\dots,x_m}{\operatorname{argmax}} \lambda_1 \left(\Phi(x)\right)$

MINVAR $\underset{x_1,\dots,x_m}{\operatorname{argmin}} \lambda_m \left(\Phi(x)\right)$

GENVAR $\underset{x_1,\dots,x_m}{\operatorname{argmin}} \prod_{i=1}^{m} \lambda_i \left(\Phi(x)\right)$

Ambiguity in Constraint Function

NORM $\|x_i\|_2^2 = 1$

AVGNORM $\sum_{i=1}^{m} \|x_i\|_2^2 = m$

VAR $x_i^H R_{ii} x_i = 1$

AVGVAR $\sum_{i=1}^{m} x_i^H R_{ii} x_i = m$

Objective Function

$$\underset{x_1,\ldots,x_m}{\operatorname{argmax}} \quad \lambda_1\left(\Phi(x)\right)$$
$$\text{subject to} \quad x_i^H R_{ii} x_i = 1, \forall i$$

- Let V_i be the right singular vectors of dataset Y_i

$$C_{\text{mcca}} = \begin{bmatrix} I_{d_1} & V_1^H V_2 & \cdots & V_1^H V_m \\ V_2^H V_1 & I_{d_2} & \cdots & V_2^H V_m \\ \vdots & \vdots & \ddots & \vdots \\ V_m^H V_1 & V_m^H V_2 & \cdots & I_{d_m} \end{bmatrix}$$

- Let $C_{\text{mcca}} = FKF^H$ be eigenvalue decomposition
- $\rho^{(j)} = k_j$
- $x_i = R_{ii}^{-1/2} f_i$

MAXVAR for CCA

Two dataset setting

$$
\begin{aligned}
C_{\text{cca}} &= \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix} \\
&= \begin{bmatrix} I_{d_1} & V_1^H V_2 \\ V_2^H V_1 & I_{d_2} \end{bmatrix} \\
&= \begin{bmatrix} F & -F \\ G & G \end{bmatrix} \begin{bmatrix} I + \left(KK^H\right)^{-1/2} & 0 \\ 0 & I - \left(K^H K\right)^{-1/2} \end{bmatrix} \begin{bmatrix} F & -F \\ G & G \end{bmatrix}^H
\end{aligned}
$$

Insights

- Eigenvalues come in pairs $\{1 + k_i, 1 - k_i,\}$
- Perfect correlation implies $\lambda_{\text{max}} = 2$, $\lambda_{\text{min}} = 0$
- Eigenvectors have specific structure
- MINVAR solves the same problem!
- Correlation implies *ALL* datasets correlated!

MCCA - Ambiguity Example

3 Dataset Example

$$C_{\text{mcca}}^{(1)} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C_{\text{mcca}}^{(2)} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

- Both have $\lambda_1 = 2$
- $u^{(1)} = \frac{1}{\sqrt{2}} [1, 1, 0]^T$
- $u^{(2)} = \frac{1}{\sqrt{3}} [1, 1, 1]^T$

Insights

- Eigenvalue detects correlation
- Eigenvector reveals structure

MCCA Results

Theorem

If $2n < \min_{i \neq j \neq k}(d_i + d_j + d_k)$ *then the largest eigenvalue of* C_{mcca} *is equal to* m*.*

Theorem

If $n < \sum_{i=1}^{m} d_i$ *then the smallest eigenvalue of* C_{mcca} *is zero.*

Claim

We conjecture that when $n < \sum_{i=1}^{m} d_i$*, the largest eigenvalue of* C_{mcca} *is determined entirely based on* n *and* $\sum_{i=1}^{m} d_i$ *and not on the underlying correlation.*

Proposed correlation statistic

$$\boxed{\widehat{\rho}^{(i)} = \lambda_i \left(C_{\text{mcca}} - I\right)}$$

Informative MCCA

Idea - Trim right singular vectors

- $\mathring{U}_j = \widehat{U}_j\left(:, 1 : \widehat{k}_j\right)$
- $\mathring{V}_j = \widehat{V}_j\left(:, 1 : \widehat{k}_j\right)$
- $\mathring{U} = \textbf{blkdiag}(\mathring{U}_1, \ldots, \mathring{U}_m)$
- $\mathring{V} = \left[\mathring{V}_1, \ldots \mathring{V}_m\right]$

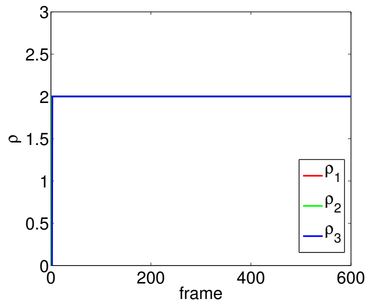
$$\boxed{\widetilde{C}_{\text{mcca}} = \mathring{U}\mathring{V}^H\mathring{V}\mathring{U}^H.}$$

Estimate the number of correlations

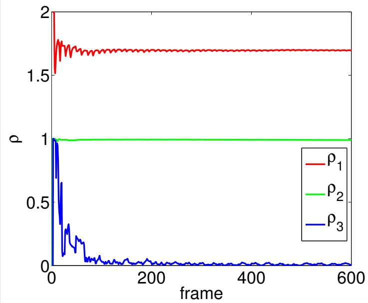
$$\boxed{\widehat{t}_{\text{imcca}} = \sum_{i=1}^{\widehat{k}} \mathbb{1}_{\{\widehat{\kappa}_i > \tau^{\alpha}_{\text{imcca}}\}},}$$

Multi-Camera Flashing Light Demonstration

MCCA and IMCCA Demonstration



(a) MCCA Correlations



(b) IMCCA Correlations

Future Work

Kernel CCA

- analysis for nonlinear low-rank signal-plus-noise model
- analysis of choice of kernel
- analysis of choice of regularization parameter

MCCA

- MAXVAR theorem for deterministic correlation when $n < \sum_i d_i$
- null distribution for statistical test for number of correlated signals
- consistency analysis for estimate of number of correlated signals

Image Annotation and Retrieval

- clever feature engineering with NLP techniques
- extension to nonlinear algorithms

Acknowledgments

* people i want to thank