**Summary of papers**

**SPECIALIST**

*Sparse canonical correlation analysis (Hardoon, Shawe-Taylor, 2011)*

* Option 1 within the context of our goal
* Primal-dual implementation of CCA whilst implementing sparsity on the weights of both matrices, strongly resembles linear CCA formulation
* Exact algorithms for its implementation and secondary canonical combinations are provided
* Method is iterative and may be computationally intensive (3 loops)
* The formula we optimise is the following

A picture containing clock

Description automatically generated

Where X is the primal vector and K is the kernel matrix of the dual. *e* are the weights for the dual with *gamma* as sparsity parameter. *w* are the weights of the primal.

* sparsity and constraints are required for convexity of the optimisation and obtention of optimal solution

*Nonparametric canonical correlation analysis (Michaeli, Wang, Livescu, 2016)*

* Option 2 within the context of our goal
* Propose a nonparametric and **partially linear** CCA method based on Lancaster 1958 & Hannan 1961 (extends 1D method of Lancaster to multi-dim)
* Includes algorithm and matlab code for nonparametric CCA which can be modified for the partially linear case
* Does not require matrix inversion
* Computationally not intensive
* The weights are calculated as follows :



Where estimated covariance is



And

Text

Description automatically generated with medium confidence

For



Performing SVD on KPL allows us to obtain



Where U is the matrix containing the top L eigenvectors and D the diagonal matrix with eigenvalues of KPL.

*Sparse Weighted Canonical Correlation Analysis (Wenwen, Zhang, 2018)*

* Propose a sparse weighted CCA to find variables strongly correlated in specific sub-samples
* Not useful for our research

*An iterative penalized least squares approach to sparse canonical correlation analysis (Mai and Zhang, 2019)*

* Propose a sparse CCA method that closely resembles a generalization of linear CCA
* It is computationally more tractable than other sparse methods
* Not useful for our research

*A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis (Witten, Tibshirani, Hastie, 2009)*

* Propose the best-known sparse CCA method, i.e. based on penalized matrix decomposition (PMD)
* Not immediately useful, but gives a simple example as to how to deal with high-dimensional data

*A Simple and Provable Algorithm for Sparse Diagonal CCA (Asteris, Koyejo, Poldrack, 2016)*

* Propose an algorithm to improve speed of choice of sparsity parameters
* Not immediately useful, but may check again for combination with Hardoon & Shawe-Taylor 2011

*Canonical Correlation Methods for Exploring Microbe-Environment Interactions in Deep Subsurface (Uurtio, Bomberg, Nybo, Itävaara, Rousu, 2015)*

* Sample application of both KCCA and primal dual CCA (Shawe-Taylor 2011)

*KERNEL AND NONLINEAR CANONICAL CORRELATION ANALYSIS (Lai & Fyfe, 2000)*

* Original paper proposing kernel CCA
* Not particularly useful, but key reference

*Restricted kernel canonical correlation analysis (Nina Otopal, 2012)*

* Method to restrict all weights to be > 0, however, adding such constraint naturally in our minimization could also be a method?
* Paper to keep just in case this method turns out to be useful
* No algorithm or code provided

*Sparse Additive Functional and Kernel CCA (Balakrishnan, Puniyani, Lafferty, 2013)*

* Just another implementation of sparse KCCA, uses marginal thresholding to solve for the biconvex optimisation issue
* Not particularly useful

*On measure transformed canonical correlation analysis (Todros, Hero, 2012)*

* Option 3: manages to uncover non-linear relationships by applying a function to the joint distribution of the two matrices
* maintains dimensions
* computationally attractive
* implementation-wise not very intuitive
* matlab code [exists](https://herogroup.engin.umich.edu/on-measure-transformed-canonical-correlation-analysis/) but not accessible

*LARGE-SCALE APPROXIMATE KERNEL CANONICAL CORRELATION ANALYSIS (Wang, Livescu, 2016)*

* method to approximate KCCA, not immediately useful for our primal-dual goal but reasonable reference just in case

*Statistical Consistency of Kernel Canonical Correlation Analysis (Fukumizu, Bach, Gretton, 2007)*

* proves the consistency bound of KCCA and raises 2 important questions:
  + how do you choose the optimal regularization effectively?
  + How do you choose the optimal kernel?

**SUMMARIZING**

*A Tutorial on Canonical Correlation Methods* *(Uurtio, Monteiro, Shawe-Taylor, Fernandes-Reyes, Rousu, 2017)*

* Great overview of different CCA methods
* Discusses linear CCA (and the different ways of solving it), kernel CCA, sparse CCA, and Bayesian (probabilistic) CCA with numerical examples
* Provides a statistical test to check the number of statistically significant correlations (Bartlett, 1941)
* Provides a statistical test to check the statistical significance of a correlation (Fujikoshi and Veitch 1979)
* Discusses kernel CCA in detail -> **some intuition: we could use full-on kernel CCA and only augmenting one side to a higher dimensional feature space if we let p >= n (to make sure covariance matrix has same dimension as the kernel matrix)**
* Discussion of kernel CCA using Hardoon (2004) in case the data set is very large
* Discussion of sparse CCA (p>n) using Witten et al. (2009) + numerical example
* Sparse method of Hardoon et al. 2011 + example -> **method to perform partial kernel CCA !**

*Canonical Correlation Analysis: An Overview with Application to Learning Methods (Hardoon, Szedmak, Shawe-Taylor, 2004)*

* Discusses CCA and KCCA and their extensions using partial Gram-Schmidt orthogonalization (similar to partial Cholesky decomposition) and penalization to solve for dimensionality issues
* Provides an example based on image retrieval

*CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework (Chapman, Wang, 2021)*

* Discusses all the different CCA variations that have been programmed into the cca-zoo library
* Hardoon et al. 2011 is not included! Other sparse methods are -> **check if one extends Hardoon et al. 2011**
* Is an extension to other API’s CCA implementations, eg. mvlearn Perry et al. (2020)

*Convergence analysis of kernel Canonical Correlation Analysis: theory and practice (Hardoon, Shawe-Taylor, 2007)*

* Proves the convergence of KCCA
* More specifically the theorem indicates, in an indirect way through A, that the empirical value of the pattern function will be close to its expectation provided that the norms of the direction vectors are controlled, and the dimension k of the projection space is small compared with the number of features. Hence, we must trade-off between finding good correlations while not allowing the norms to become too large.

*Pyrcca: regularized kernel canonical correlation analysis in Python and its applications to neuroimaging. (Bilenko, Gallant, 2015)*

* Describes the limited options of the cca implementations in python: linear or kernel, with and without regularization
* Provides on example on neuroimaging

**RELATED**

*Avoiding Backtesting Overfitting by Covariance-Penalties: an empirical investigation of the ordinary and total least squares cases (Firoozye, Koshiyama, 2019)*

* Refers to 3 main issues in backtest overfitting
* Discusses relevant solutions to the 3 issues + refers to papers -> good for thesis tests
* Proposes covariance-penalization of sharpe ratio based on Mallow’s C and Stein’s Unbiased Regression Estimates (SURE)

*OPTIMAL DYNAMIC STRATEGIES ON GAUSSIAN RETURNS (Koshiyama, Firoozye, 2019)*

* Most importantly establishes the relation between maximizing sharpe ratio and the correlation between portfolio returns and signals
* Derives the higher order moments under specific conditions explicitly incorporating the correlation
* TLS in singular case & provides argumentative grounds for CCA in multivariate case