Chemical Engineering 4H03

Projection of Latent Structures (Derivation)

Jake Nease
McMaster University

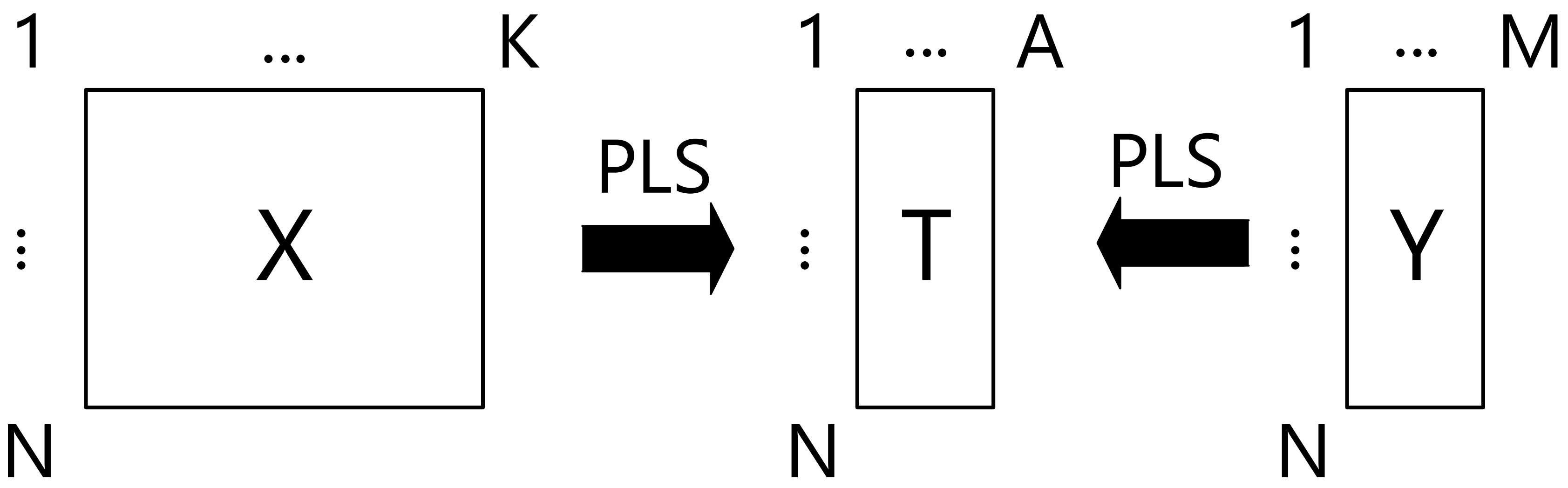
Where are We?

- MLR is good for relating dependent outcomes ($\boldsymbol{y}$) to independent inputs (X)
  - NOT appropriate if columns in X are dependent
  - Leads to singularity (ill conditioning) of $X^T X$ → inverse issues
  - 'Spinning' of model plane (unstable coefficients)
  - NOT easy to visualize or interpret if K is large

- PCR is a nice tool that dimension-reduces first
  - SOLVES dependent column issue
  - EASIER to visualize with $A < K$
  - Does NOT help that $\boldsymbol{y}$ is assumed to contain all error
  - Must be done independently for EACH column in Y

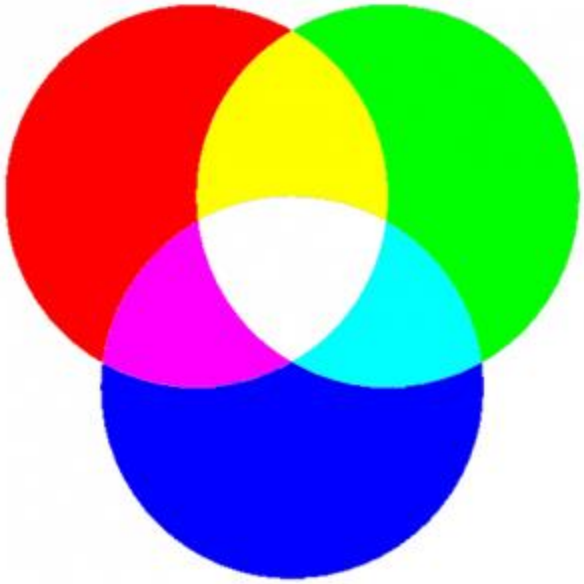
A Better Idea

- Instead, let's fit the latent variable space to have A components, where each component represents data in X **and** data in Y **simultaneously**
  - We will have TWO sets of scores and loadings



Projection of Latent Structures

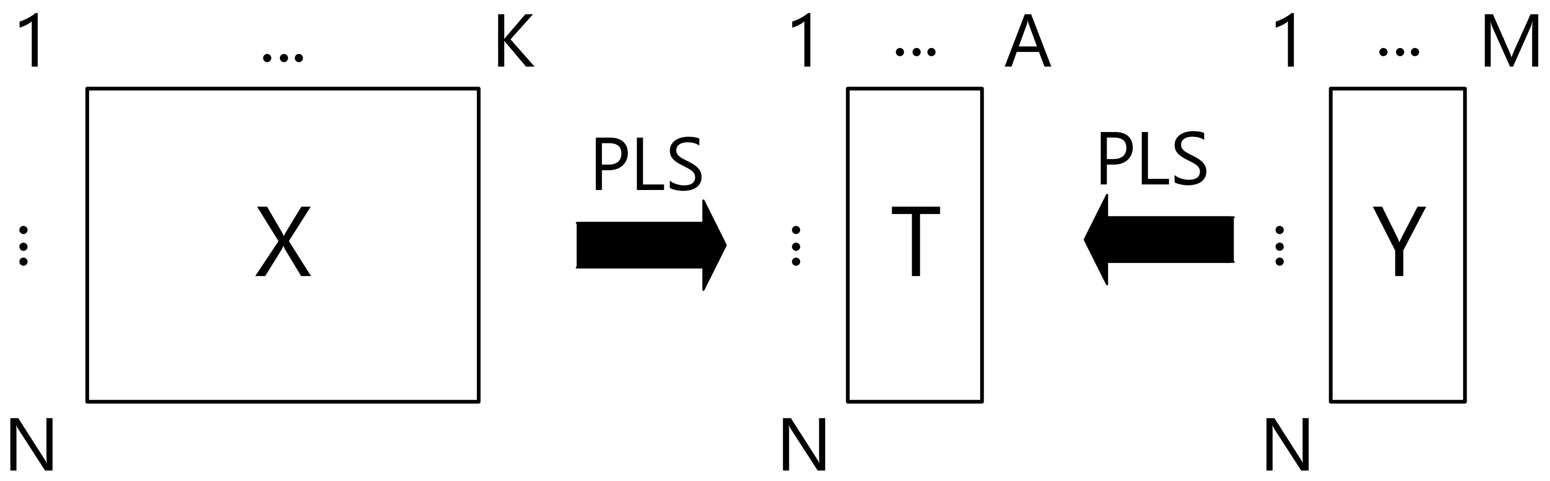
The Deluxe Linear Modeling Technique



https://drawpaintacademy.com/wp-content/uploads/2019/04/Additive-Color-300x300.png

Overview of PLS

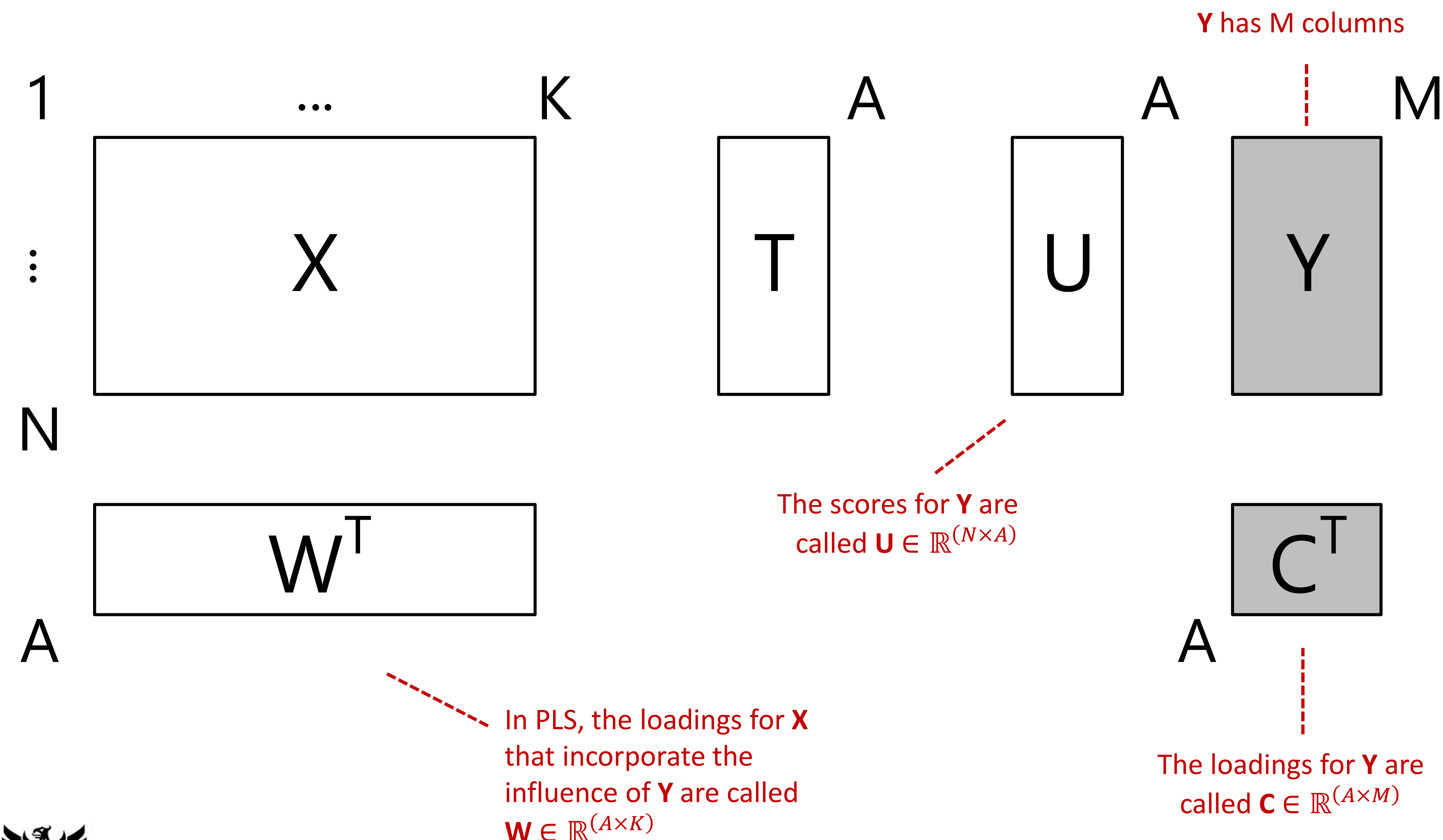
- **P**rojection of **L**atent **S**tructures:
  1. Extracts components from X and Y sequentially
  2. Uses cross-validation to fit an appropriate number of components
  3. Scores for the PLS model are computed from X and Y simultaneously
  4. Makes engineering sense: system is driven by the same underlying latent variables (in X and Y)



Objective of PLS

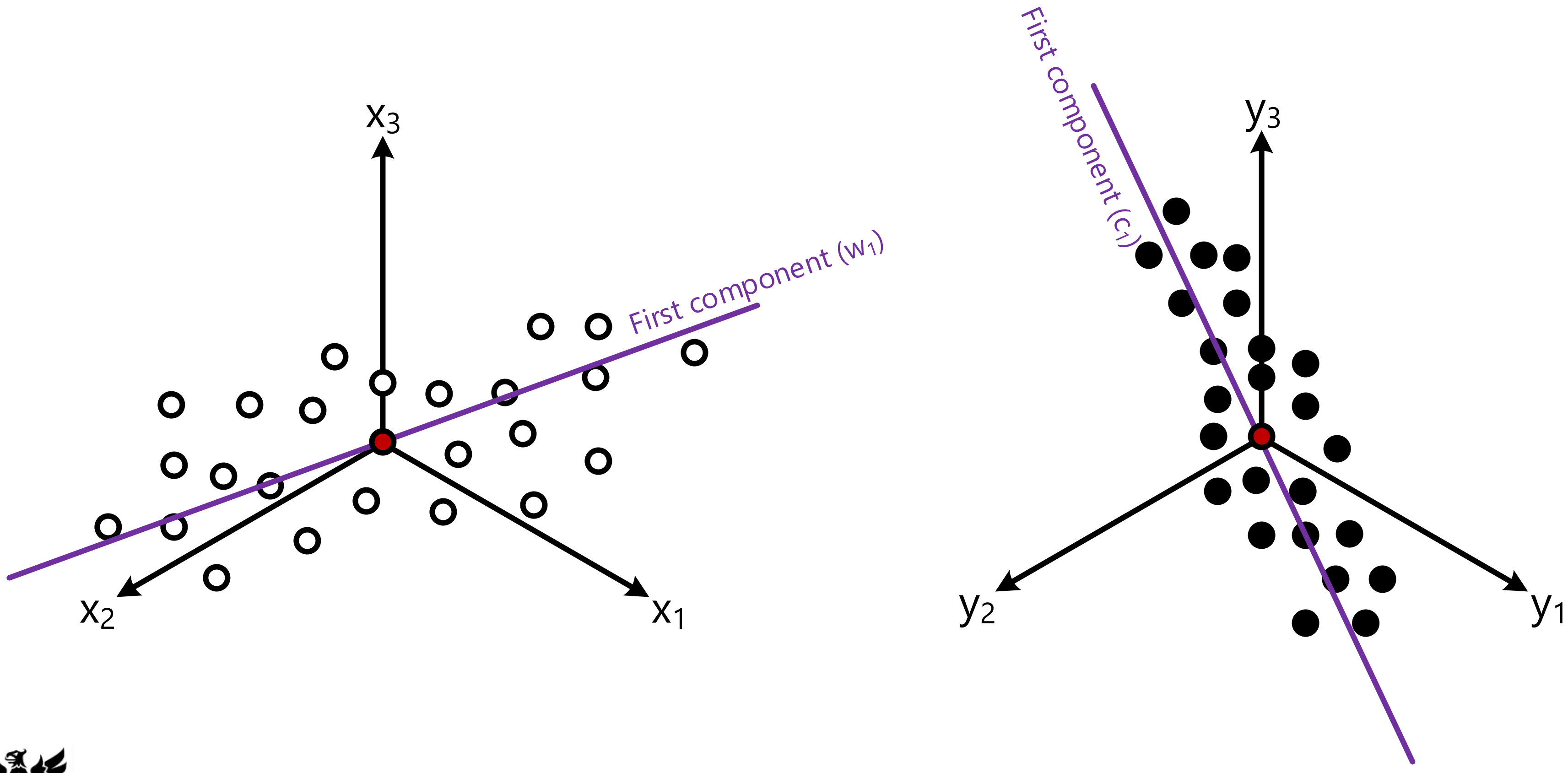
- Objective of **PCA** is to explain X
- What do we want from **PLS**?
  - Best explanation of X
  - Best explanation of Y
  - Maximize the relationship between X and Y
- Covariance is going to come up again here

Some New Notation



PLS Geometrically

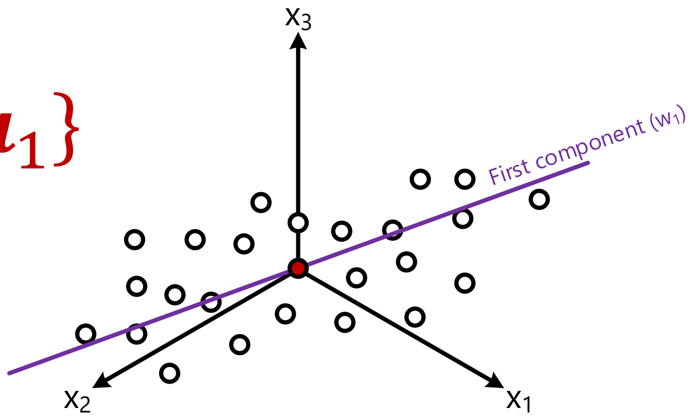
- Consider a data set with $K = M = 3$ (three variables)
  - The first component explains X and Y "well"
  - Why not "best?"

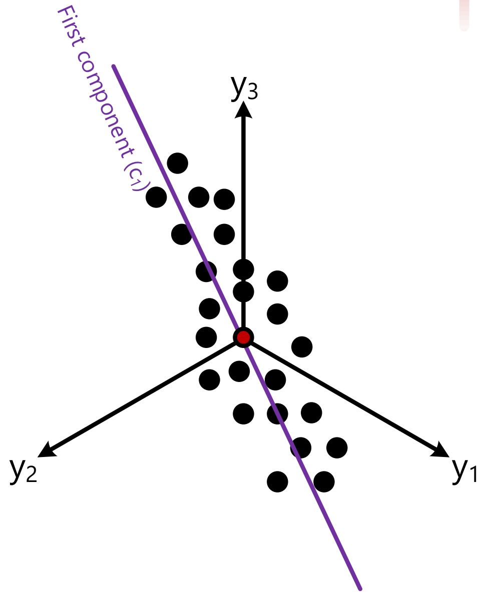


PLS Geometrically

- Compute the SCORES t_1 and u_1 for both model spaces
  - IN PCA, we would want to maximize the variance of t_1
  - If we were magically fitting one component to Y, we would want to maximize the variance of u_1
- Surprising absolutely no one, our objective in PLS is to maximize the **COVARIANCE** of t_1 and u_1

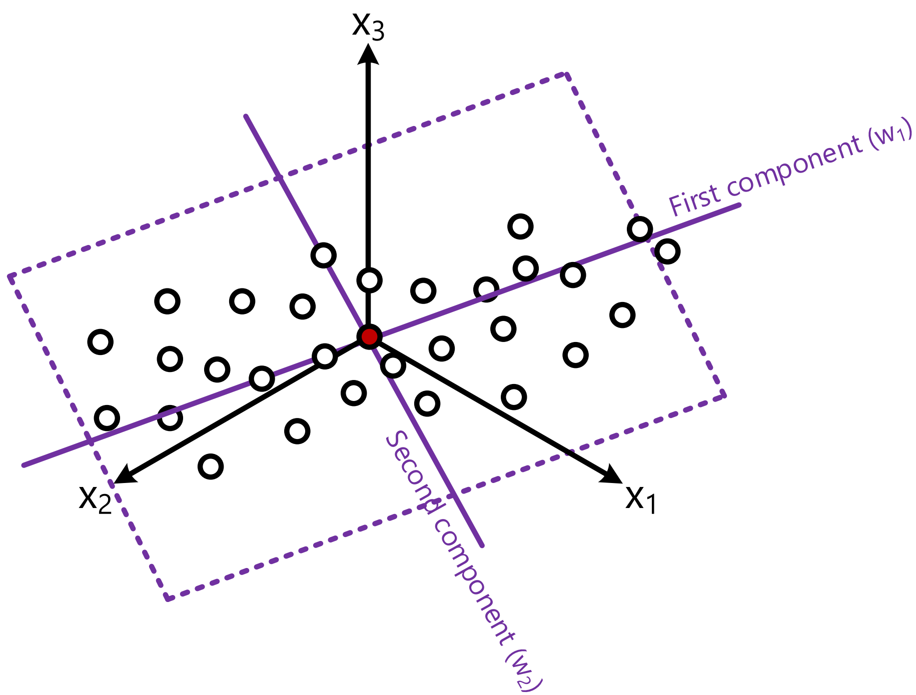
$$\max_{\boldsymbol{w}_1, \boldsymbol{c}_1} \phi = \mathcal{C}\{\boldsymbol{t}_1, \boldsymbol{u}_1\}$$

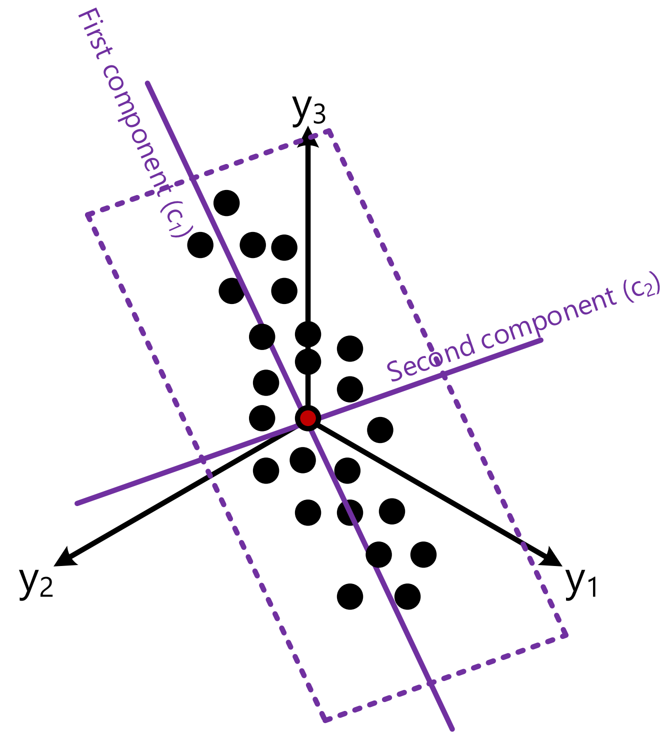




PLS Geometrically

- Now add the second component
  - w_2 is orthogonal to w_1
  - c_2 is orthogonal to c_1





PLS Analytically (Basic Version)

- PLS scores explain variance in X (recall $\boldsymbol{w}$ are loadings)
  - $\boldsymbol{t}_a = X_a \boldsymbol{w}_a$ Variance in X for in a^{th} component
  - $\max \phi = \boldsymbol{t}_a^T \boldsymbol{t}_a$ Subject to $\boldsymbol{w}_a^T \boldsymbol{w}_a = 1.0$

- PLS scores ALSO explain variance in Y
  - $\boldsymbol{u}_a = Y_a \boldsymbol{c}_a$ Variance in Y for in a^{th} component
  - $\max \nu = \boldsymbol{u}_a^T \boldsymbol{u}_a$ Subject to $\boldsymbol{c}_a^T \boldsymbol{c}_a = 1.0$

- It does this by maximizing the **relationship** between **explained variance** in X and Y
  - $\max \psi = f(\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\})$

PLS Maximizes Covariance of $\{X, Y\}$

- Recall the calculation for covariance

$$\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} = \mathcal{E}\{(\boldsymbol{t}_a - \bar{\boldsymbol{t}}_a)(\boldsymbol{u}_a - \bar{\boldsymbol{u}}_a)\}$$

$$\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} \equiv \frac{1}{N}\, \boldsymbol{t}_a^T \boldsymbol{u}_a$$ ---------- WHY!!!???

- Although this is our objective, covariance is hard to interpret and is dependent on units

- Instead, let's maximize **correlation**, which is a value between -1 and 1!

PLS Maximizes Covariance of $\{X, Y\}$

- Correlation for x and y is

$$r\{x,y\} = \frac{\mathcal{E}\{(x-\bar{x})(y-\bar{y})\}}{\sqrt{\mathcal{V}(x)\mathcal{V}(y)}} = \frac{\mathcal{C}\{x,y\}}{\sqrt{\mathcal{V}(x)\mathcal{V}(y)}} = \frac{\mathcal{C}\{x,y\}}{\sqrt{\mathcal{V}(x)}\sqrt{\mathcal{V}(y)}}$$

$$\mathcal{C}\{x,y\} \therefore= r\{x,y\} \cdot \sqrt{\mathcal{V}(x)} \cdot \sqrt{\mathcal{V}(y)}$$

For PLS scores:

$$\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} = r\{\boldsymbol{t}_a, \boldsymbol{u}_a\} \cdot \sqrt{\mathcal{V}(\boldsymbol{t}_a)} \cdot \sqrt{\mathcal{V}(\boldsymbol{u}_a)}$$

$$\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} = r\{\boldsymbol{t}_a, \boldsymbol{u}_a\} \cdot \sqrt{\boldsymbol{t}_a^T\,\boldsymbol{t}_a} \cdot \sqrt{\boldsymbol{u}_a^T\,\boldsymbol{u}_a}$$

$\max \phi = \boldsymbol{t}_a^T\,\boldsymbol{t}_a$

$\max \nu = \boldsymbol{u}_a^T\,\boldsymbol{u}_a$

PLS Maximizes Covariance of $\{X, Y\}$

$$\mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} = r\{\boldsymbol{t}_a, \boldsymbol{u}_a\} \cdot \sqrt{\boldsymbol{t}_a^T\, \boldsymbol{t}_a} \cdot \sqrt{\boldsymbol{u}_a^T\, \boldsymbol{u}_a}$$

$\max \phi = \boldsymbol{t}_a^T\, \boldsymbol{t}_a$ $\max \nu = \boldsymbol{u}_a^T\, \boldsymbol{u}_a$

Maximizing covariance simultaneously maximizes the RELATIONSHIP between t_a and u_a as well as their own variance (which was the point of PCA!)

I consider this important

PLS Summarized

- The OPTIMIZATION objective for PLS is to maximize the covariance between $\boldsymbol{t}_a$ and $\boldsymbol{u}_a$ for each component
  - $\max \psi = \mathcal{C}\{\boldsymbol{t}_a, \boldsymbol{u}_a\} = r\{\boldsymbol{t}_a, \boldsymbol{u}_a\} \cdot \sqrt{\boldsymbol{t}_a^T \boldsymbol{t}_a} \cdot \sqrt{\boldsymbol{u}_a^T \boldsymbol{u}_a}$
  - Explains X via $\boldsymbol{t}_a^T \boldsymbol{t}_a$
  - Explains Y via $\boldsymbol{u}_a^T \boldsymbol{u}_a$
  - But we also want a high **CORRELATION** between $\boldsymbol{t}_a$ and $\boldsymbol{u}_a$
    - Unsurprisingly, a high $r\{\boldsymbol{t}_a, \boldsymbol{u}_a\}$
  - Decision variables are $\boldsymbol{w}_a$ and $\boldsymbol{c}_a$

- Some notes
  - Requires formal constrained optimization (outside 4H scope)
  - Can also be obtained through NIPALS (described next)

Final Remarks

- PLS elegantly exploits the covariance between $\boldsymbol{t}$ and $\boldsymbol{u}$ scores to simultaneously maximize correlation AND the individual variances of each
  - **Thus we explain all dimensions as well as possible**
- This is **fundamental** to the understanding of PLS
  - DISCUSSION: differences between PCR and PLS?
- Next up: **Computing PLS components using NIPALS**
  - Good news: the algorithm is similar to PCA
  - Bad news: it takes more steps and the interpretation is (a teensy bit)messier