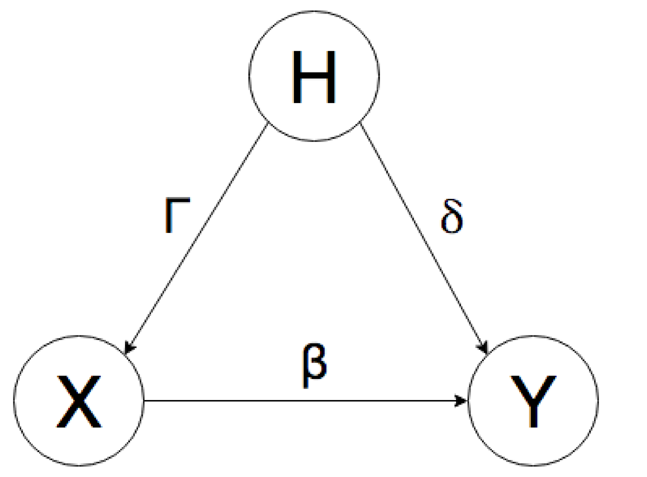
Deconfounding using Spectral transformations

Domagoj Ćevid, *joint work with P.Bühlmann and N.Meinshausen*
Seminar für Statistik, ETH Zürich, Switzerland
cevid@stat.math.ethz.ch

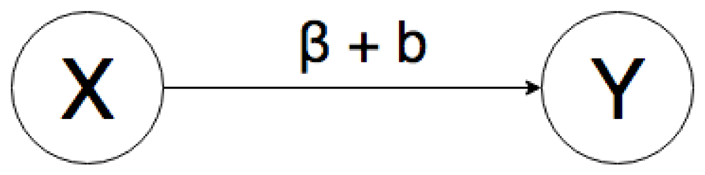
1. Confounding model

$$X = E + H\Gamma$$
$$Y = X\beta + H\delta + \eta$$



2. Perturbed linear model

$$Y = X(\beta + b) + \epsilon$$

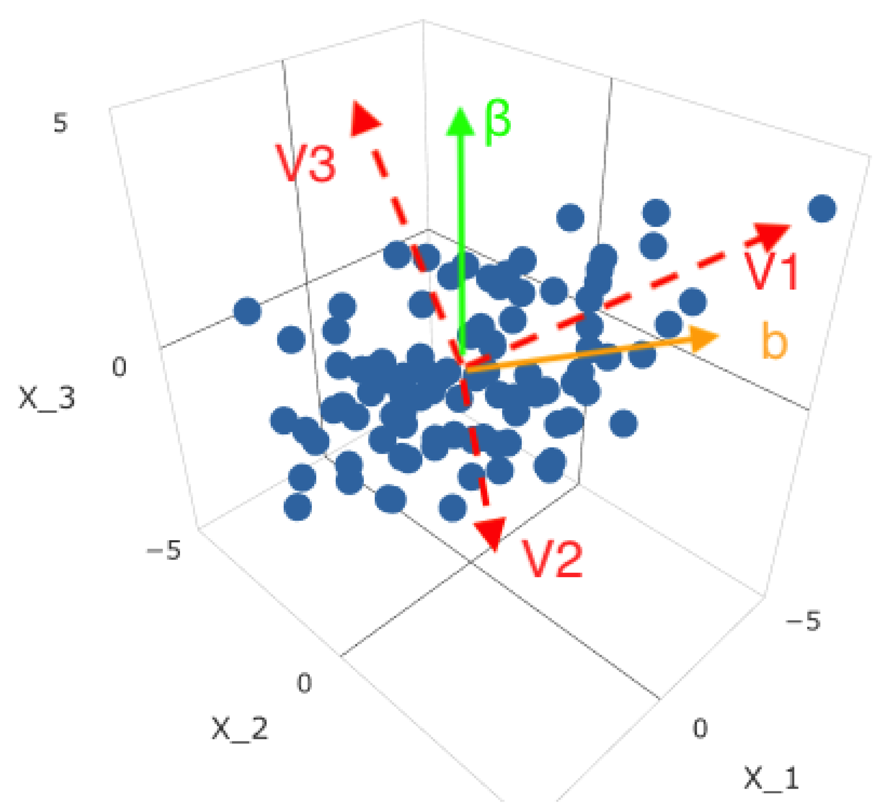


- β sparse, b dense coefficient perturbation

When E, H, η have Gaussian distribution, the confounding model and the perturbed linear model are equivalent, where

- $\Sigma = \Sigma_E + \Gamma^T\Gamma$
- $\sigma^2 = \sigma_\eta{}^2 + \delta^T(I - \Gamma\Sigma^{-1}\Gamma^T)\delta$
- $b = \Sigma^{-1}\Gamma^T\delta$

5. Motivation



- Trim transform ensures $\|\tilde{X}b\|_2$ is small regardless of direction of b
- In the confounding model b points towards large singular vectors and β does not since it is sparse and V dense, so $\|\tilde{X}b\|_2$ gets smaller compared to $\|\tilde{X}\beta\|_2$

3. Transformed Lasso

- We transform our data using some linear transformation F: $\tilde{X} = FX, \tilde{Y} = FY$

$$\widetilde{Y} = \widetilde{X}(\beta + b) + \widetilde{\epsilon}$$

- Afterwards we run Lasso on the transformed variables

$$\widehat{\beta} = \arg\min_{\beta} \frac{1}{n}\|\widetilde{Y} - \widetilde{X}\beta\|_2^2 + \lambda\|\beta\|_1$$

- Would like to have $\widetilde{\epsilon}$ and $\widetilde{X}b$ small, but $\widetilde{X}\beta$ large

4. Spectral Transformations

- SVD of the design matrix $X = UDV^T$, $D = \text{diag}(d_1, \ldots, d_n)$, $d_1 \geq d_2 \geq \ldots d_n \geq 0$
- We will consider **spectral transformations** F which map d_i to some $\widetilde{d}_i$, so that $\widetilde{X} = U\widetilde{D}V^T$

$$F = U \begin{bmatrix} \widetilde{d}_1/d_1 & 0 & \ldots & 0 \\ 0 & \widetilde{d}_2/d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \widetilde{d}_n/d_n \end{bmatrix} U^T$$

Examples of Spectral transformations

- **Lava** transform (Chernozhukov et al. 2017): $\widetilde{d}_i = \sqrt{\frac{n\lambda_2 d_i^2}{n\lambda_2 + d_i^2}}$
  - $(\widehat{\beta}, \widehat{b}) = \arg\min_{\beta,b} \frac{1}{n}\|Y - X(\beta + b)\|_2^2 + \lambda_2\|b\|_2^2 + \lambda_1\|\beta\|_1$
  - $\widehat{\beta} = \arg\min_\beta \frac{1}{n}\|\widetilde{Y} - \widetilde{X}\beta\|_2^2 + \lambda\|\beta\|_1$
- **Puffer** transform (Jia, Rohe 2015): $\widetilde{d}_i = 1$
- **PCA** adjustment: remove the largest r principal components of X
- **Trim** transform: $\widetilde{d}_i = \min(d_i, \text{median}(d_i))$

6. Theorem

Assume in the perturbed linear model that b is 'small' and V is 'dense'. If a spectral transformation with $\widetilde{d}_i \leq d_i$ satisfies

(B1) $\widetilde{d}_{(1)} = \mathcal{O}\left(\widetilde{d}_{(k)}\right)$ - largest transformeed singular value is 'not too big'

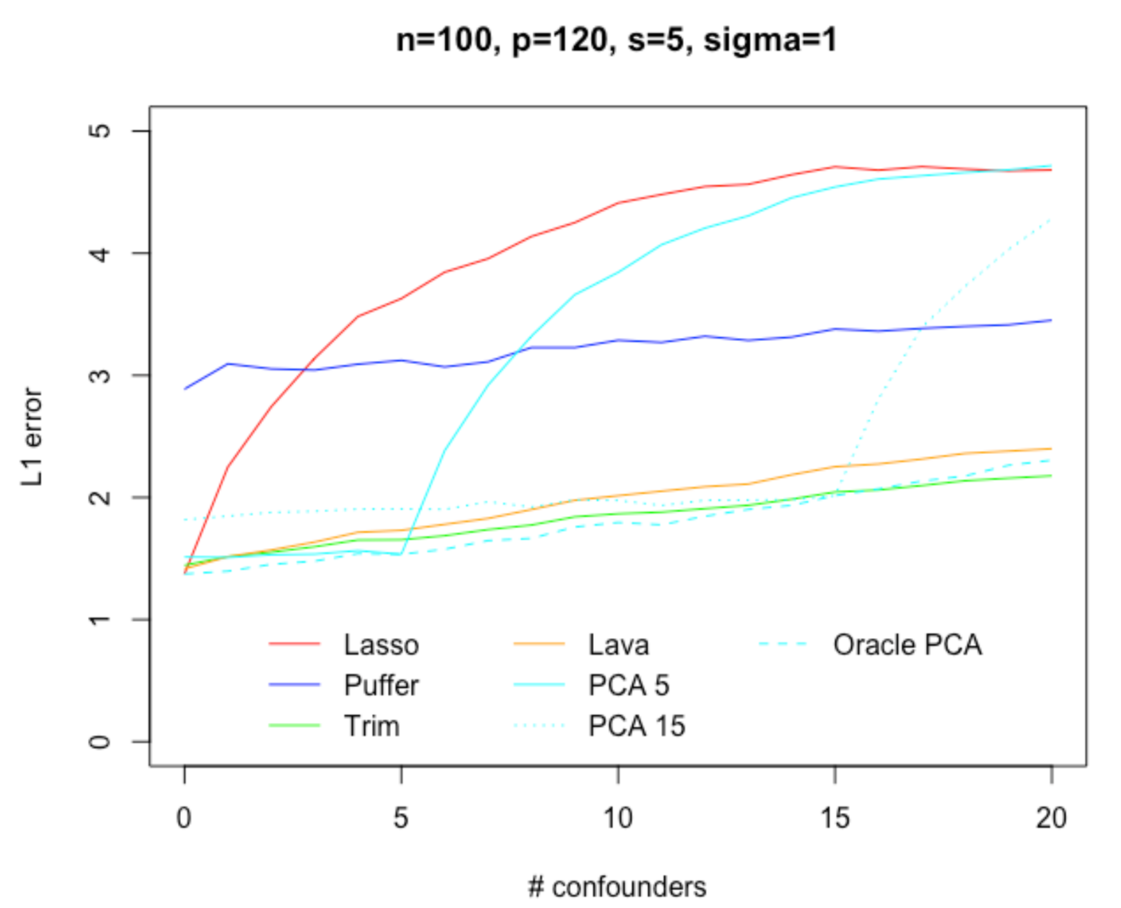
(B2) $\widetilde{d}_{(k)}^2 = \Omega\left(\lambda_{\min}(\Sigma)\, p\right)$ - enough of the transformed singular values are 'large'

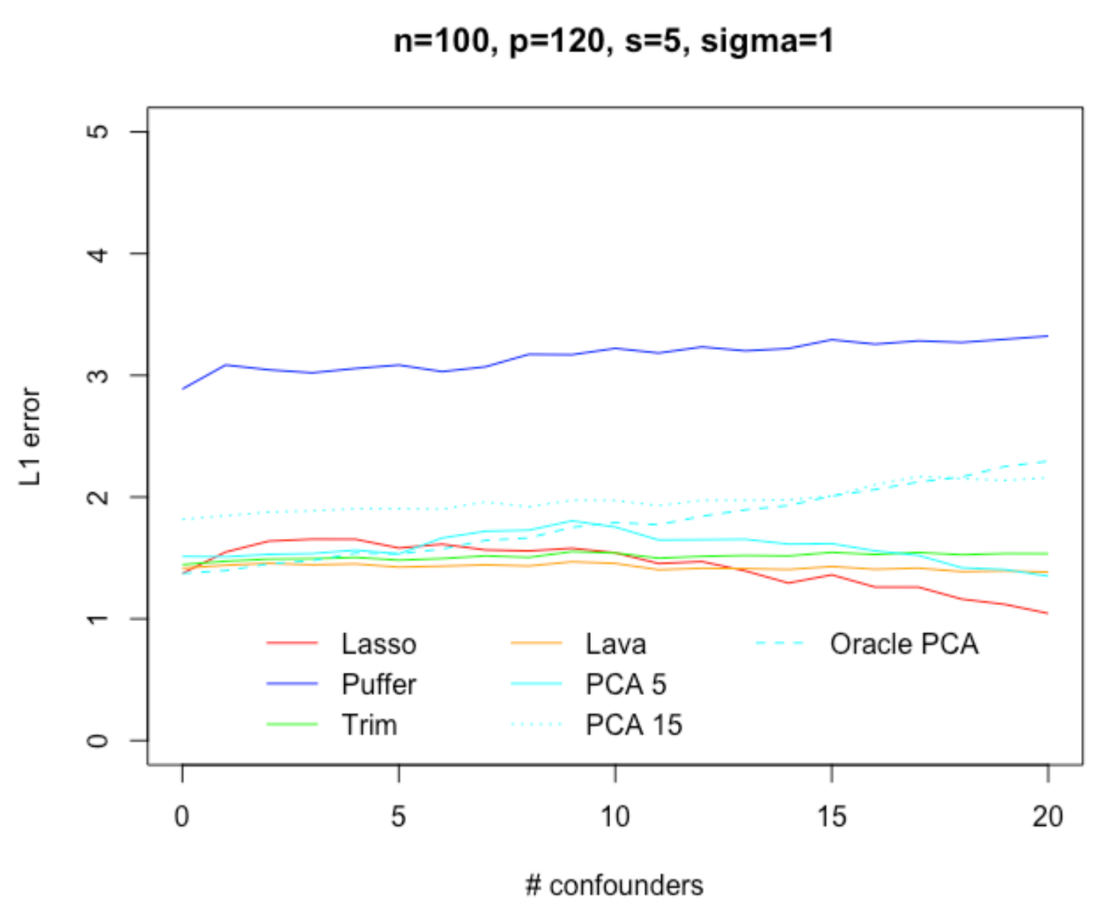
for $k = \Omega(n)$, then we can choose λ so that

$$\|\widehat{\beta} - \beta\|_1 = \mathcal{O}_p\left(\frac{\sigma s}{\lambda_{\min}(\Sigma)}\sqrt{\frac{\log p}{n}}\right)$$

If in addition 'enough singular values d_i are large', then the Trim transform with $\tau = d_{\lfloor tn \rfloor}$, $t \in (0, 1)$, satisfies the conditions **(B1)** and **(B2)**.

7. Simulations: Confounders vs no confounders





8.Application: Gene expression