Bayesian / Model-based TSVD

$Y = M + E \quad \in \mathbb{R}^{m_1 \times \cdots \times m_K}$

$M = S \times \mathcal{U} \qquad S \in \mathbb{R}^{r_1 \times \cdots \times r_K}$

$\qquad\qquad\qquad \mathcal{U} \in \mathcal{V}_{r_1, m_1} \times \cdots \mathcal{V}_{r_K, m_K}$

<u>Inference</u>: given Y, infer $M = S \times \mathcal{U}$

<u>OLS estimation</u>

if $\underset{\sim}{r}$ is known, $\hat{M} = \underset{\text{rank}(M) = \underset{\sim}{r}}{\arg\min} \|Y - M\|^2$

OLS estimator $\hat{M}$ can be obtained from ALS

What if 1) r isn't known?
2) we have missing data?
3) the errors aren't normal
4) we want CI's, predictions, etc.
⋮

Some of these issues can be addressed w/ model based approaches

<u>Model-based TSVD</u>

$Y = M + E \qquad E = \{e_{ijk\ldots}\} \overset{iid}{\sim} N(0, \sigma_e^2)$

$M = S \times \mathcal{U}$

Nothing is different so far. I will argue for the following "random effects" model for S:

$S = \left(\frac{\sigma_s}{\sigma_e}\right) \times Z \times \{D_1, \ldots, D_K\}, \qquad \sigma_s \geq 0$

$\qquad D_k = \text{diag}(d_{1k} \ldots d_{r_k k}), \quad \sum_1^{r_k} d_{jk}^2 = 1$

$\qquad \{Z_{ijk\ldots}\} \overset{iid}{\sim} N(0, 1)$

Invariance / Equivariance

$Y = \sigma S \times U + \sigma E$

$y = \sigma U s + \sigma e \;,\quad E[e] = 0,\; Cov(e) = I$

Parameters are $\quad (\sigma, U, s) \in \mathbb{R}^+ \times \mathcal{U} \times \mathbb{R}^{r_1 \cdots r_n} = \Theta$

$\mathcal{U} = \{ U : U = U_n \otimes U_{n-1} \otimes \cdots \otimes U_1$

Invariance: Let $\mathcal{W} = \{ W : W = W_n \otimes \cdots \otimes W_1 \,,\; W_k \in O_{m_k} \}$

Consider $\quad \mathcal{G} = \{ g : y \to \tau W y \,,\; \tau > 0 \,,\; W \in \mathcal{W} \}$

Letting $\tilde{y} = \tau W y$, we have $\tilde{y} = \tau \sigma W U s + \tau \sigma e$

$= \tilde{\sigma} \tilde{U} s' + \tilde{\sigma} e$

$(\tilde{\sigma}, \tilde{U}, s) \in \Theta$

Thus, $\mathcal{G}$ on $\mathcal{Y}$ induces a group $\bar{\mathcal{G}}$ on Θ defined by

$\bar{\mathcal{G}} = \{ \bar{g} : (\sigma, U, s) \to (\tau\sigma, WU, s) \,,\; \tau > 0 \,,\; W \in \mathcal{W} \}$

Equivariant estimation

If $\hat{\sigma}(y)$ is estimating σ, then $\hat{\sigma}(\tau W y)$ should be est. $\tau\sigma$

$\Rightarrow$ require $\quad \hat{\sigma}(\tau W y) = \tau \hat{\sigma}(y) \quad \forall\, \tau, W, y$

Similarly:

If $\hat{U}(y)$ is est. U, then $\hat{U}(\tau W y)$ should be est WU

$\rightarrow$ require $\quad \hat{U}(\tau W y) = W \hat{U}(y)$

Result: If $\mathcal{G}$ were transitive over Θ,
then optimal equivariant estimators
could be found as Bayes solns under
"the right Haar Measure"

$$R(d, \pi_{\theta|y}) = \int L(d, \theta)\, \pi_{\theta|y}(d\theta) \quad = \text{"Post risk"}$$

Let $\delta(y) = \arg\min_d R(d, \pi_{\theta|y}) \quad \forall y \in \mathcal{Y}$

$\pi_{\theta|y} = p(y|\theta) \times h(\underset{\sim}{\theta})$
right Haar measure

Problem: $\mathcal{G}$ is not transitive over Θ

$\bar{g}: (\sigma, \mu, s) \to (\sigma\tau, W\mu, s)$ leaves s unchanged $\Rightarrow$ s is a maximal invariant

Subproblem:

Consider the case that s is known

$\to \quad (\sigma, \mu) \in \mathbb{R}^+ \times \mathcal{Y} = \Theta$

$\mathcal{G} = \{g : y \to \tau W y\} \quad \bar{\mathcal{G}} = \{\bar{g} : (\sigma, \mu) \to (\tau\sigma, W\mu)\}$

$\Rightarrow \quad \bar{\mathcal{G}}$ is transitive over Θ.

an optimal equivariant estimator $\hat{\delta}$ will exist (assuming an equivarian

$\hat{\delta}(y)$ obtained as Bayes estimator under

the right Haar prior over (σ, μ).

Haar Measure for (σ, U)

$h(\sigma, U) = h_\sigma(\sigma) \times h_u(U)$

$\propto \frac{1}{\sigma} \cdot 1$

uniform distribution on $\mathcal{V}_{r_1 \times m_1} \times \cdots \times$

Jeffreys prior for a scale param

This suggests $h(\sigma, U)$ as a prior over σ, U even when S is unknown. We just now need to specify a prior over S.

Model-Based TSVD

$Y = M + E \in \mathbb{R}^{m_1 \times \cdots \times m_K}$ $\quad E \sim N_{m_1 \cdots m_K}(0, \sigma_\epsilon^2 I)$, rank$(M) = (r_1 \ldots$

$= X \times U + E$ $\quad X \in \mathbb{R}^{r_1 \cdots r_K}, \; U_k \in \mathcal{V}_{r_k m_k}$

Random Effects / Factor Model Representation

$M = X \times U$, $\quad X = S \times \{D_1 \ldots D_K\}$, $\quad D_k = \text{diag}(d_{k1} \ldots d_{kr_k})$ $\quad \text{tr}(D_k^2)$

$S \sim N(0, I, \ldots, I\sigma_s^2)$

$\Rightarrow$ $X \sim N(0, D_1^2, \ldots, D_K^2 \sigma_s^2)$

$\mathrm{Cov}(\mathrm{vec}(X)) = \sigma_s^2 D_K^2 \otimes \cdots \otimes D_1^2$

Under this model/prior: treating only X and E as random

$E[Y_{(1)} Y_{(1)}^T] = E[U_1 X_{(1)} U_*^T U_* X_{(1)}^T U_1^T] + E[E_{(1)} E_{(1)}^T] + 0$

$= U_1 E[D_1 S_{(1)} S_{(1)}^T D_1] U_1^T + M_{-1} \sigma_\epsilon^2 I$

$= \sigma_s^2 \, U_1 D_1^2 U_1^T + M_{-1} \sigma_\epsilon^2 I$

under this model,

$$E[Y_{(1)} Y_{(1)}^T] = E[\mathcal{U}_1 S_{(1)} \mathcal{U}_{-1}^T \mathcal{U}_{-1} S_{(1)}^T \mathcal{U}_1^T] + E[E_{(1)} E_{(1)}^T]$$

$$= \mathcal{U}_1 E[S_{(1)} S_{(1)}^T] \mathcal{U}_1^T + m_2 \cdots m_n \sigma_e^2 I$$

$$= \mathcal{U}_1 E[D_1 Z_{(1)} Z_{(1)}^T D_1 \sigma_s^2] \mathcal{U}_1^T + m_{-1} \sigma_e^2 I$$

$$= \mathcal{U}_1 D_1^2 \mathcal{U}_1^T (\sigma_s^2 \cdot r_{-1}) + (\sigma_e^2 m_{-1}) I$$

$$= \underbrace{\mathcal{U}_1 \Lambda_1 \mathcal{U}_1^T}_{\text{rank } r_1} (r_{-1} \sigma_s^2) + \underbrace{I}_{\text{full rank}} (m_{-1} \sigma_e^2)$$

$\simeq$ factor analysis type of model

Slight relabeling ?: $Y = S \times \Lambda^{1/2} \times U + E$, $S \sim iid\ N(0, \sigma_s^2)$

$= S \times D \times U + E$ $\quad E \sim iid\ N(0, \sigma_e^2)$

Parameter Space

$\sigma_e^2 \in \mathbb{R}^+$

$\sigma_s^2 \in \mathbb{R}^+$

$= D_k^2$, $\quad \Lambda_1 \ldots \Lambda_K$, $\quad tr(\Lambda_k) = 1 \quad \longleftarrow (\lambda_{1k} \ldots \lambda_{r_k k}) \in \text{Simplex}(r_k)$ comp...

$\mathcal{U}_1 \ldots \mathcal{U}_n \quad \mathcal{U}_n^T \mathcal{U}_n = I_{r_n \times r_n} \quad \longleftarrow \mathcal{U}_n \in \mathcal{V}_{r_n, m_n}$

$\sigma_e^2 \sim$ Jeffreys / "diffuse inv-gamma"

$\sigma_s^2 \sim$ " "

$diag(\Lambda_n) = (\lambda_1 \ldots \lambda_{r_n}) \sim Dir(\alpha)$ or uniform

$\mathcal{U}_n \sim$ Uniform

Posterior calculations: Unknowns: $(\mathcal{U}, \Lambda, S, \sigma_s^2, \sigma_e^2)$

$\pi(\mathcal{U}, \Lambda, S, \sigma_s^2, \sigma_e^2 \mid y)$ can be approximated w/ 1) Gibbs sampling
2) Variational Bayes
3) ..?

Gibbs Sampling

$$\pi(\sigma_e^2 \mid y, S, \Lambda \dots) \propto \pi(\sigma_e^2) \times p(y \mid S, \sigma_e^2, \Lambda, \mathcal{U} \dots)$$

$$\pi(\sigma_e^2) \propto 1/\sigma_e^2$$

$p(y \mid \dots)$: model is $(y - \mathcal{U}Ds) \sim N(0, \sigma_e^2)$

$$p(y \mid \dots) \underset{\sigma_e^2}{\propto} (\sigma_e^2)^{-M/2} \exp\{\tfrac{1}{2}(\tfrac{1}{\sigma_e^2}) \|y - \mathcal{U}Ds\|^2\}$$

$$\pi(\sigma_e^2 \mid \dots) \underset{\sigma_e^2}{\propto} (\sigma_e^2)^{-\frac{M+1}{2}} \exp\{-\tfrac{1}{\sigma_e^2} \|y - \mathcal{U}Ds\|^2/2\}$$

$$\propto \text{inv-gamma}\left(\frac{1+M}{2}, \|y - \mathcal{U}Ds\|^2/2\right)$$

Similarly,

$$\pi(\sigma_s^2 \mid \dots) \propto \text{inv-gam}\left(\frac{1+r}{2}, \frac{\|S\|^2}{2}\right)$$

More interestingly:

Full cond of S :

We can write $y = AS + e$, $A = (\underbrace{\mathcal{U}_n D_n}_{\Lambda_n^{1/2}} \otimes \dots \otimes \underbrace{\mathcal{U}_1 D_1}_{\Lambda_1^{1/2}})$

$$\pi(S \mid \dots) \propto e^{-\frac{1}{2\sigma_s^2} S^T S} \cdot e^{-\frac{S^T A^T A S}{2\sigma_e^2} + S^T A^T y / 2\sigma_e^2}$$

$$V = (A^T A / \sigma_e^2 + I / \sigma_s^2)^{-1}$$

$\propto$ MVN(m, V) where $m = V A^T y / \sigma_e^2$

Full cond of U

$Y = S \times D \times U + E$ $\qquad \Lambda_n^{1/2} \qquad \Lambda_2^{1/2}$

$Y_{(1)} = U_1 D_1 S_{(1)} B_1^T + E_{(1)} \qquad B_1 = (U_n D_n \otimes \cdots \otimes U_2 D_2)$

$\pi(U_1 | Y, \dots) \underset{U_1}{\propto} p(Y | U_1 \dots) \times \cancel{\pi(U_1)}$ unif

$\underset{U_1}{\propto} p(Y | U_1 \dots)$

$Y G U^T$

$\propto_{U_1} \exp\{-\frac{1}{2} \| Y_{(1)} - U_1 D_1 \overbrace{S_{(1)} B_1^T}^{G^T} \| / \sigma_e^2\}$

$\| Y_{(1)} - UG \|^2 / \sigma_e^2 = \| Y_{(1)} \|^2 / \sigma_e^2 - 2\, tr(U G^T Y^T) + tr(G \cancel{U^T U} G^T)/\sigma_e^2$

$= \| Y_{(1)} \|^2 / \sigma_e^2 + \| G \|^2 / \sigma_e^2 - 2\, tr(U^T Y G)$

$\pi(U | \cdots) \underset{U}{\propto} etr(U^T H) \qquad H = YG$

This is a "exp. fam model" for random matrices on $\mathcal{V}_{r_n, m_n}$

called the von Mises / Fisher / Langevin distribution over $\mathcal{V}_{r_n m_n}$

You can simulate from it using the rstiefel package in R.

Demo:

① estimation under correct rank: what is the estimate of M?

② estimation under incorrect rank

③ "Big data"