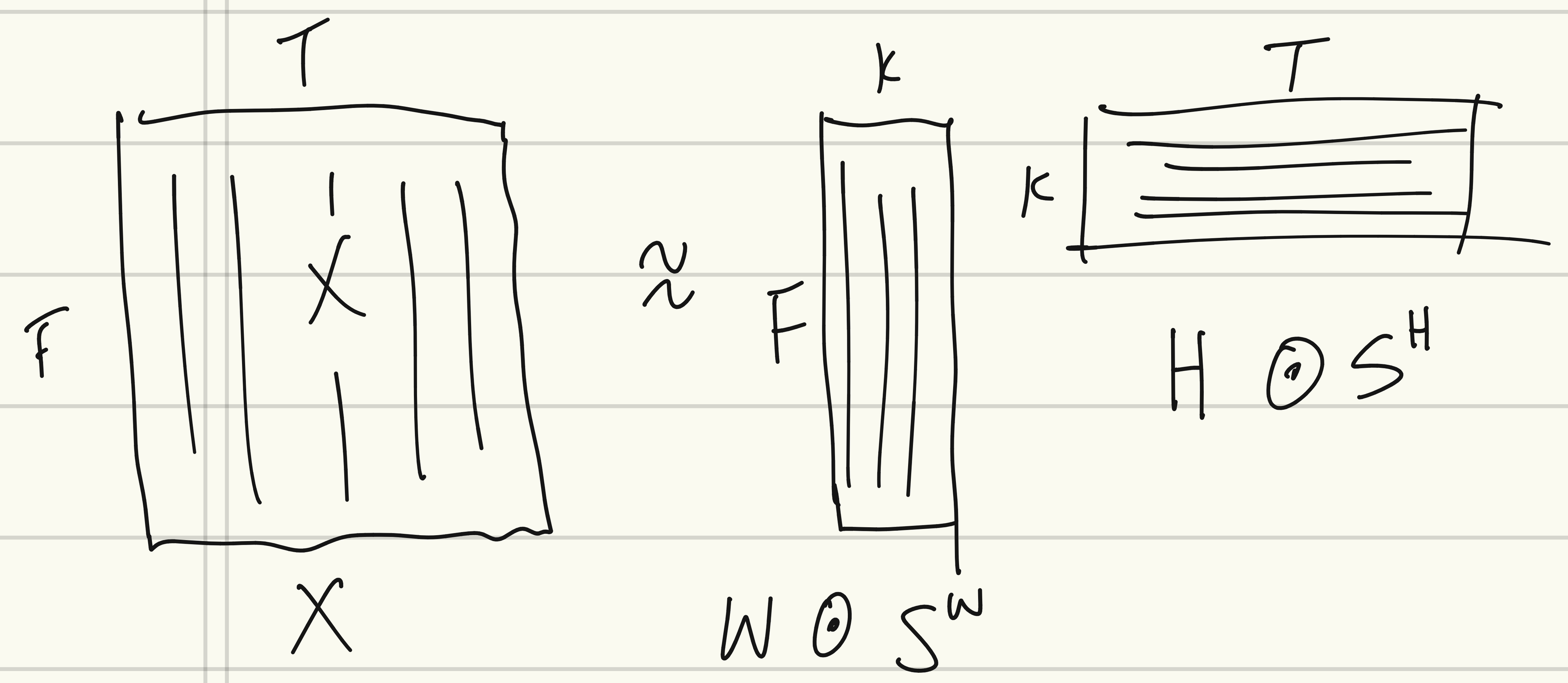
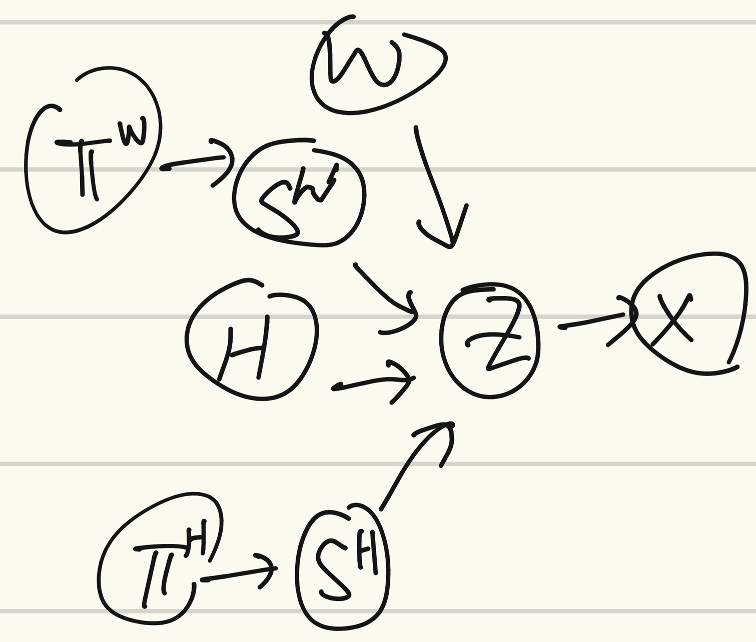
Additional Sparsity Matrix Brainstorm





Model

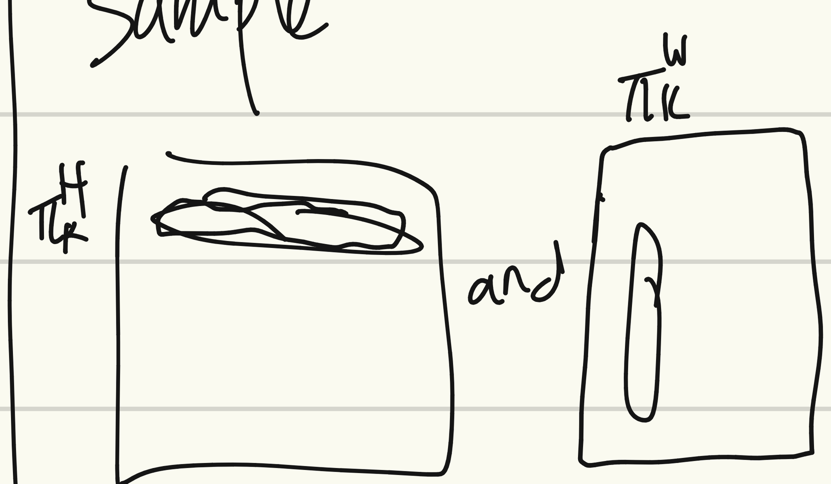
- $\underset{F\times T}{X} \approx (\underset{F\times K}{W} \odot \underset{F\times K}{S^W})(\underset{K\times T}{H} \odot \underset{K\times T}{S^H})$ $\qquad S^W$

- $W_{fk} \sim \text{Gamma}(a, b)$; $H_{kt} \sim \text{Gamma}(c, d)$
  $\pi_k^H \sim \text{Beta}\left(\frac{a_0^H}{K}, \frac{b_0^H(K-1)}{K}\right)$, $S_{kt}^H \sim \text{Bern}(\pi_k^H)$
  $\pi_k^W \sim \text{Beta}\left(\frac{a_0^W}{K}, \frac{b_0^W(K-1)}{K}\right)$, $S_{fk}^W \sim \text{Bern}(\pi_k^W)$
  $X_{ft} \sim \text{Pois}(\sum_k W_{fk} S_{fk}^W H_{kt} S_{kt}^H)$
  $\cancel{Z_{ftk}} \sim \text{Pois}(W_{fk} S_{fk}^W H_{kt} S_{kt}^H)$

- ☆ each entry as sample. ☆

  $q(\beta) \prod_n q(z_n \mid \beta)$

  Sample



  alternatively?

SSVI

$\cancel{\{Z_{ftk}\}} \; \cancel{\{S_t^H\}_t} \; \cancel{\{S_f^W\}_f}$

- local : $\{Z_{ftk}\}_{f,t,k}$, $\{S_{kt}^H\}_{k,t}$, $\{S_{fk}^W\}_{f,k}$
- global : $\{W, H, \pi^W, \pi^H\}$

$p(Z, W, H, S^W, \pi^W, \pi^H, S^H \mid X) \approx q(Z, W, H, S^W, \pi^W, \pi^H, S^H)$

$= \prod_k \Big(\underbrace{q(W_k)\, q(h_k)\, q(\pi_k^W) \cdot q(\pi_k^H)}_{q(\beta)} \Big) \cancel{\prod_{f,t,k} q(Z_{ftk}, S_{kt}^H, S_{fk}^W \mid W, H, \pi^H, \pi^W)}$ ☆

$q(Z, S^H, S^W \mid W, H, \pi^H, \pi^W)$

where $q(w_k) = \prod_f q(W_{fk})$, $q(h_k) = \prod_t q(H_{kt})$ and

$q(W_{fk}) = \Gamma(\nu_{fk}^W, \rho_{fk}^W)$, $q(H_{kt}) = \Gamma(\nu_{kt}^H, \rho_{kt}^H)$, $q(\pi_k^W) = \text{Beta}(\alpha_k^{\pi^W}, \beta_k^{\pi^W})$,

$q(\pi_k^H) = \text{Beta}(\alpha_k^{\pi^H}, \beta_k^{\pi^H})$,

max. local ELBO obtained when $q = p$(exact conditional)

$\hat{\eta}_n$ $\qquad \mathbb{E}_q[\eta_n(y_n, Z) \mid \beta]$

SSVI. Hoff 15 page 363 (3)

we calc. $E(z)$ but just sample S^H and S^W, to get noisy version

Collapsed Gibbs Sampling for S^H_{kt} & S^W_{fk}

- want $p(z_{ft}, S^H_{kt}, S^W_{fk} \mid X_{ft}, W, H, \pi^H, \pi^W) \propto \cancel{p(z_{ft}, X_{ft}, S^H_{kt}, \dots)}$ same
  $\propto p(S^H_{kt}, S^W_{fk} \mid X_{ft}, W, H, \pi^H, \pi^W)$
  ~~can we integrate out?~~

- initialize S^H and S^W.
  don't need z b/c we can compute $E[z]$ exactly later, just need S_{kt}, S_{fk}

S^H_{kt}

- $P(S^H_{kt} = 1 \mid S^H_{\neg k,t}, x_t, W, h_t, \pi^H, \pi^W, S^W_t)$
  $\propto P(S^H_{kt} = 1, S^H_{\neg k,t}, x_t, W, h_t, \pi^H, \pi^W, S^W_t)$ → $P(x_t) = \prod_f P(X_{ft})$
  $\propto P(S^H_{kt} = 1 \mid \pi^H)\, P(x_t \mid W, h_t, S^H_{\neg k,t}, S^H_{kt} = 1, S^W_t)$
  $\propto \pi^H_k \cdot \prod_f \Big(\underbrace{W_{fk} \cdot S^W_{fk} \cdot H_{kt} \times 1}_{S^H_{kt}=1} + \underbrace{\sum_{l \neq k} W_{fl} S^W_{fl} H_{lt} S^H_{lt}}_{S^H_{\neg k,t}} \Big)^{X_{ft}} \exp\{-W_{fk} S^W_{fk} H_{kt}\}$
  $=: p^H_1$

  $S^H_{kt} = 1$ and ignore $S^H_{\neg k,t}$
  b/c same term is multiplied in p^H_0 as well, so don't need it in $\frac{p_1 \cancel{x}}{(p_0+p_1)\cancel{x}}$

- $P(S^H_{kt} = 0 \mid S^H_{\neg k,t}, x_t, W, H, \pi^H, \pi^W, S^W)$
  $\propto (1 - \pi^H_k) \cdot \prod_f \Big(\sum_{l \neq k} W_{fl} S^W_{fl} H_{lt} \Big)^{X_{ft}}$
  $=: p^H_0$

→ generate $S^H_{kt} \sim \text{Bern}\left(\frac{p_1}{p_0 + p_1} \right)$

S^W_{fk}

- $p(S^W_{fk} = 1 \mid S_{f,\neg k}, x_f, W_f, H, \pi^H, \pi^W, S^W_f)$
  $\vdots$
  $\propto \pi^W_k \cdot \prod_t \Big(\underbrace{W_{fk} \cdot H_{kt} \cdot S^H_{kt}}_{S^W_{fk}=1} + \underbrace{\sum_{m \neq k} W_{fm} H_{mt} S^H_{mt}}_{S^H_{f,\neg k}} \Big)^{X_{ft}} \cancel{S^H_{fm}} \exp\{-W_{fk} H_{kt} S^H_{kt}\}$
  $=: p^W_1$

- $P(S^H_{kt} = 0 \mid \dots) \propto (1 - \pi^W_k) \prod_t \Big(\sum_{m \neq k} W_{fm} H_{mt} S^H_{mt} \Big)^{X_{ft}} =: p^W_0$

→ generate $S^W_{fk} \sim \text{Bern}\left(\frac{p^W_1}{p^W_0 + p^W_1} \right)$

$z_{ft} \mid X_{ft}, W_f, h_t, S^W_f, S^H_t \sim \text{Multi}(z_{ft}; X_{ft}, \phi_{ft})$ where
$\phi_{ft} \propto W_{fk} S^W_{fk} H_{kt} S^H_{kt} \Rightarrow \mathbb{E}[z_{ftk} \mid X_{ft}, \dots] = X_{ft} \phi_{ftk}$

E_q ← supposed to be?
which satisfies local ELBO.

global updates

- $p(z_n, y_n \mid \beta) \leftrightarrow p(S^{\omega}, S^{H}, Z, X \mid W, H, \pi^{W}, \pi^{H})$

$$= p(S^{\omega} \mid \pi^{\omega})\, p(S^{H} \mid \pi^{H})\, p(Z, X \mid W, H, \pi^{W}, S^{H}, \pi^{W}, \pi^{H})$$

$$= p(S^{\omega} \mid \pi^{\omega})\, p(S^{H} \mid \pi^{H})\, p(Z \mid W, H, \pi^{W}, S^{H}, \pi^{W}, \pi^{H})$$

W_{fk}:

only consider terms containing W_{fk},

$$\propto p(Z_f \mid W_f, H, S_f^{W}, S^{H})$$

$$= \prod_t \prod_K p(Z_{ftk} \mid W_f, H_t, S_f^{W}, S_t^{H})$$

$$\propto \prod_t \Big[(W_{fk} S_{fk}^{W} H_{kt} S_{kt}^{H})^{Z_{ftk}} \cdot \exp\{-W_{fk} S_{fk}^{W} H_{kt} S_{kt}^{H}\}$$

$$\times \cancel{\prod_{l \neq k} (W_{fl} S_{fl}^{W} H_{lt} S_{lt}^{H})^{Z_{ftl}} \cdot \exp\{-W_{fl} S_{fl}^{W} H_{lt} S_{lt}^{H}\}} \Big]$$

no W_{fk}

$$\propto \prod_t \Big[(W_{fk} S_{fk}^{W} H_{kt} S_{kt}^{H})^{Z_{ftk}} \cdot \exp\{-W_{fk} \cdot S_{fk}^{W} H_{kt} S_{kt}^{H}\}$$

$$= (W_{fk} \cdot S_{fk}^{W} H_{kt} S_{kt}^{H})^{\sum_t Z_{ftk}} \cdot \exp\{-W_{fk} \cdot S_{fk}^{W} \textstyle\sum_t H_{kt} S_{kt}^{H}\}$$

$$\approx \sum_t X_{ft} \phi_{ftk}$$

$$\rightarrow \nu_{fk}^{W} \leftarrow (1-\eta^{(i)}) \nu_{fk}^{W} + \eta^{(i)} (a + \textstyle\sum_t X_{ft} \phi_{ftk}^{(i)})$$

$$\rho_{fk}^{W} \leftarrow (1-\eta^{(i)}) \rho_{fk}^{W} + \eta^{(i)} (b + S_{fk}^{W} \textstyle\sum_t H_{kt}^{(i)} S_{kt}^{H(i)})$$

H_{kt}:

(H_{kt})

$$\propto p(Z_t \mid H_t, W, S^{W}, S_t^{H})$$

$$= \prod_f \prod_K p(Z_{ftk} \mid H_t, W_f, S_k^{W}, S_t^{H})$$

$$\propto \prod_f \Big[(H_{kt} \cdots)^{Z_{ftk}} \cdot \exp\{-H_{kt} \cdots\} \times \prod_{l \neq k} (\ )^{Z_{ftk}} \exp(-\ ) \Big]$$

$$\propto (H_{kt} W_{fk} S_{fk}^{W} S_{kt}^{H})^{\sum_f Z_{ftk}} \cdot \exp\{-H_{kt} S_{kt}^{H} \textstyle\sum_f W_{fk} S_{fk}^{\omega}\}$$

π_k^{H}:

$$\propto p(S_k^{H} \mid \pi_k^{H}) = \prod_t p(S_{kt}^{H} \mid \pi_k^{H}) = \prod_t \Big[(\pi_k^{H})^{S_{kt}^{H}} (1-\pi_k^{H})^{1-S_{kt}^{H}} \Big]$$

$$= (\pi_k^{H})^{\sum_t S_{kt}^{H}} \cdot (1-\pi_k^{H})^{T-\sum_t S_{kt}^{H}}$$

π_k^{W}:

$$\propto p(S_k^{W} \mid \pi_k^{W}) = \cdots = (\pi_w^{H})^{\sum_f S_{fk}^{W}} \quad (1-\pi_k^{\omega})^{F-\sum_f S_{fk}^{W}}$$

H_{kt} updates: $\nu_{kt}^{H} \leftarrow (1-\eta^{(i)}) \nu_{kt}^{H} + \eta^{(i)} (c + \sum_f X_{ft} \phi_{ftk}^{(i)})$

$\rho_{kt}^{H} \leftarrow (1-\eta^{(i)}) \rho_{kt}^{H} + \eta^{(i)} (d + S_{kt}^{H(i)} \sum_f W_{fk}^{(i)} S_{fk}^{W(i)})$

π updates: $\alpha_k \leftarrow (1-\eta^{(i)}) \alpha_k + \eta^{(i)} \left(a_0^{H \text{ or } W}/K + \sum_t S_{kt}^{H(i)} \text{ or } \sum_f S_{fk}^{W(i)} \right)$

$\beta_k \leftarrow (1-\eta^{(i)}) \beta_k + \eta^{(i)} \left(\frac{b_0^{H \text{ or } W}(K-1)}{K} + (T-\sum_t S_{kt}^{H(i)}) \text{ or } (F-\sum_f S_{fk}^{W(i)}) \right)$