Dynamic Rank Factor Model for Text Streams Supplementary Material

Shaobo Han, Lin Du, Esther Salazar and Lawrence Carin Duke University, Durham, NC 27708 {shaobo.han, lin.du, esther.salazar, lcarin}@duke.edu

1 Gibbs Sampling for the Basic Model

1.1 Model

$$y_{p,t} = g\left(z_{p,t}\right) \tag{1}$$

$$\boldsymbol{z}_t = \boldsymbol{\Lambda} \boldsymbol{s}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \quad \boldsymbol{R} = \boldsymbol{I}_P$$
 (2)

$$\lambda_{p,k} \sim \text{TPBN}_{+}(a, b, \phi_k), \quad \phi_k^{1/2} \sim \mathcal{C}^{+}(0, d)$$
(3)

$$s_{k,t} = \rho_k s_{k,t-1} + \delta_{k,t}, \quad 0 < \rho_k < 1$$
 (4)

$$\delta_{k,t} \sim \text{TPBN}(e, f, \nu), \quad \nu^{1/2} \sim \mathcal{C}^+(0, h)$$
(5)

$$\rho_k \sim \text{TN}_{(0,1)}(\mu_0, \sigma_0^2), \quad s_{0,k} \sim \mathcal{N}(0, \sigma_s^2)$$
(6)

1.2 Conjugate hierarchy

$$\lambda_{p,k} \sim \mathcal{N}_+(0, l_{p,k}) \quad l_{p,k} \sim \mathcal{G}(a, u_{p,k}), \quad u_{p,k} \sim \mathcal{G}(b, \phi_k) \tag{7}$$

$$\phi_k \sim \mathcal{G}(1/2, \omega_k), \quad \omega_k \sim \mathcal{G}(1/2, d^2)$$
(8)

$$\delta_{k,t} \sim \mathcal{N}(0, \tau_{k,t}), \quad \tau_{k,t} \sim \mathcal{G}(e, \eta_{k,t}), \quad \eta_{k,t} \sim \mathcal{G}(f, \nu)$$
(9)

$$\nu \sim \mathcal{G}(1/2,\zeta), \quad \zeta \sim \mathcal{G}(1/2,h^2)$$
 (10)

1.3 Gibbs Sampling Updates

Denote $\Theta = \{\Lambda, S, L, U, \phi, \omega, \rho, \tau, \eta, \nu, \zeta\}$, we use Gibbs sampling to approximate the joint posterior distribution of (Z, Θ) ,

1. Given Θ , find $p(z_{p,t}|\Theta, Z \in R(Y), Z_{-p,-t})$, for $p = 1, \ldots, P, t = 1, \ldots, T$. 2. Given Z, find $p(\Theta|Z, Z \in R(Y))$ reduce to $p(\Theta|Z)$

Treat \boldsymbol{Z} as augmented data, the full likelihood for $(\boldsymbol{Z}, \boldsymbol{\Theta})$ is

$$p(\boldsymbol{Z}, \boldsymbol{\Theta}) = \left(\prod_{t=1}^{T} \mathcal{N}(\boldsymbol{z}_{t}; \boldsymbol{\Lambda} \boldsymbol{s}_{t}, \boldsymbol{R})\right) \times \left(\prod_{k=1}^{K} \mathcal{G}(\phi_{k}; 1/2, \omega_{k}) \mathcal{G}(\omega_{k}; 1/2, d^{2})\right)$$
$$\times \prod_{k=1}^{K} \left[\mathcal{N}(s_{0,k}; 0, \sigma_{s}^{2}) \left(\prod_{t=1}^{T} \mathcal{N}(s_{k,t}; \rho_{k} s_{k,t-1}, \tau_{k,t}) \mathcal{G}(\tau_{k,t}; e, \eta_{k,t}) \mathcal{G}(\eta_{k,t}; f, \nu)\right) \right]$$
$$\times \left(\prod_{p=1}^{P} \prod_{k=1}^{K} \mathcal{N}_{+}(\lambda_{p,k}; 0, l_{p,k}) \mathcal{G}(l_{p,k}; a, u_{p,k}) \mathcal{G}(u_{p,k}; b, \phi_{k})\right) \times \prod_{k=1}^{K} \operatorname{TN}_{(0,1)}(\rho_{k}; \mu_{0}, \sigma_{0}^{2})$$
$$\times \mathcal{G}(\nu; 1/2, \zeta) \times \mathcal{G}(\zeta; 1/2, h^{2})$$
(11)

*contributed equally

• Sampling $z_{p,t}$

$$p(z_{p,t}|\boldsymbol{\Theta}, \boldsymbol{Z} \in R(\boldsymbol{Y}), \boldsymbol{Z}_{-p,-t}) \sim \mathrm{TN}_{[\underline{z_{p,t}}, \overline{z_{p,t}}]}(\sum_{k=1}^{K} \lambda_{p,k} s_{k,t}, 1)$$
(12)

where $\underline{z_{p,t}} = \max\{z_{p',t'} : y_{p',t'} < y_{p,t}\}$ and $\overline{z_{p,t}} = \min\{z_{p',t'} : y_{p',t'} > y_{p,t}\}$

• Sampling $\lambda_{p,k}$

$$p(\lambda_{p,k}|-) \propto \left(\prod_{t=1}^{T} \mathcal{N}(z_{p,t};\lambda_{p,k}s_{k,t} + \sum_{j \neq k} s_{j,t}\lambda_{p,j}, 1)\right) \mathcal{N}_{+}(\lambda_{p,k};0,l_{p,k})$$
$$= \mathcal{N}_{+} \left(\lambda_{p,k};v_{\lambda_{p,k}}\sum_{t=1}^{T} \left[s_{k,t}z_{p,t} - s_{k,t}\sum_{j \neq k} s_{j,t}\lambda_{p,j}\right], v_{\lambda_{p,k}}\right)$$
$$v_{\lambda_{p,k}} = (l_{p,k}^{-1} + \sum_{t=1}^{T} s_{k,t}^{2})^{-1}$$
(13)

• Sampling $l_{p,k}$, $u_{p,k}$

 $p(l_{p,k}|-) = \text{GIG}(a - 1/2, 2u_{p,k}, (\lambda_{p,k})^2), \quad p(u_{p,k}|-) = \mathcal{G}(a + b, u_{p,k} + \phi_k) \quad (14)$ The Generalized Inverse Gaussian (GIG) distribution can be expressed as

$$\operatorname{GIG}(x; p, a, b) = \frac{(a/b)^{\frac{p}{2}}}{2K_p(\sqrt{ab})} x^{P-1} \exp\left(-\frac{1}{2}(ax + \frac{b}{x})\right) \quad (x > 0)$$

where $K_p(\theta)$ is the modified Bessel function of the second kind

$$K_{p}(\theta) = \int_{0}^{\infty} \frac{1}{2} \theta^{-p} t^{P-1} \exp\left(-\frac{1}{2}(t+\frac{\theta^{2}}{t})\right) dt$$

with property $K_{-\frac{1}{2}}(\theta) = \frac{1}{2}\sqrt{2\pi}\theta^{-\frac{1}{2}}\exp(-\theta)$ and $K_{p+1}(\theta) = K_{P-1}(\theta) + \frac{2p}{\theta}K_p(\theta)^{-1}$.

• Sampling ϕ_k, ω_k

$$p(\phi_k|-) = \mathcal{G}(1/2 + bP, \omega_k + \sum_{p=1}^{P} u_{p,k}), \quad p(\omega_k|-) = \mathcal{G}(1, \phi_k + d^2)$$
(15)

- Sampling $\tau_{k,t}, \eta_{k,t}$ $p(\tau_{k,t}|-) = \operatorname{GIG}(e-1/2, 2\eta_{k,t}, (s_{k,t}-\rho_k s_{k,t-1})^2), \quad p(\eta_{k,t}|-) = \mathcal{G}(e+f, \tau_{k,t}+\nu)$ (16)
- Sampling ν, ζ

$$p(\nu|-) = \mathcal{G}(1/2 + fTK, \zeta + \sum_{k=1}^{K} \sum_{t=1}^{T} \eta_{k,t}), \quad p(\zeta|-) = \mathcal{G}(1, \nu + h^2)$$
(17)

• Sampling ρ_k

$$p(\rho_k|-) = \operatorname{TN}_{(0,1)}\left(\sigma_{\rho_k}^2(\sigma_0^{-2}\mu_0 + \sum_{t=1}^T \tau_{k,t}^{-1} s_{k,t-1} s_{k,t}), \sigma_{\rho_k}^2\right)$$
(18)

where $\sigma_{\rho_k}^2 = 1/(\sigma_0^{-2} + \sum_{t=1}^T \tau_{k,t}^{-1} s_{k,t-1}^2).$

• Sampling $s_{k,t}$: we have the state model and the observation model²

$$s_t | s_{t-1} \sim \mathcal{N}(\boldsymbol{A} s_{t-1}, \boldsymbol{Q}_t), \quad \boldsymbol{A} = \operatorname{diag}(\boldsymbol{\rho}), \quad \boldsymbol{Q}_t = \operatorname{diag}(\boldsymbol{\tau}_t),$$
(19)

$$\boldsymbol{z}_t | \boldsymbol{s}_t \sim \mathcal{N}(\boldsymbol{\Lambda} \boldsymbol{s}_t, \boldsymbol{R}), \quad \boldsymbol{R} = \boldsymbol{I}_P$$
 (20)

for t = 1, ..., T

¹Code for simulating GIG distribution is available at: http://jonaswallin.github.io/articles/2013/07/simulation-of-gig-distribution/

²For brevity, we omit the dependencies on Θ in notation

1. Forward Filtering: beginning at t = 0 with $s_0 \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_K)$, we have, for all $t = 1, \ldots, T$, the on-line posteriors $p(s_t | \mathbf{z}_{1:t}) = \mathcal{N}(\mathbf{m}_t, \mathbf{V}_t)$. Start from

$$p(s_{t-1}|z_{1:(t-1)}) = \mathcal{N}(m_{t-1}, V_{t-1})$$
(21)

Combine (19) with (21), integrate out s_{t-1} , we have the predictive density at t,

$$p(\boldsymbol{s}_t | \boldsymbol{z}_{1:(t-1)}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{m}_{t-1}, \boldsymbol{Q}_t + \boldsymbol{A}\boldsymbol{V}_{t-1}\boldsymbol{A}^T)$$
(22)

Further combine (20) with (22), we have the on-line posteriors at t, $p(\mathbf{s}_t | \mathbf{z}_{1:t}) = \mathcal{N}(\mathbf{m}_t, \mathbf{V}_t)$, where $\mathbf{m}_t = \mathbf{V}_t \{ \mathbf{\Lambda}^T \mathbf{R}^{-1} \mathbf{z}_t + \mathbf{H}_t^{-1} \mathbf{A} \mathbf{m}_{t-1} \}$, $\mathbf{V}_t = [\mathbf{H}_t^{-1} + \mathbf{\Lambda}^T \mathbf{R}^{-1} \mathbf{\Lambda}]^{-1}$, and $\mathbf{H}_t = \mathbf{Q}_t + \mathbf{A} \mathbf{V}_{t-1} \mathbf{A}^T$.

2. Backward Sampling: define the backward smoothing density

$$p(\boldsymbol{s}_t | \boldsymbol{z}_{1:T}) = \mathcal{N}(\widetilde{\boldsymbol{m}}_t, \boldsymbol{V}_t)$$
(23)

At t = T, we have the initialization condition $p(s_T | z_{1:T}) = \mathcal{N}(\widetilde{m}_T, \widetilde{V}_T) = \mathcal{N}(m_T, V_T)$. Combine (19) with (21), we have the conditional distribution of s_{t-1} given s_t ,

$$p(\boldsymbol{s}_{t-1}|\boldsymbol{s}_t, \boldsymbol{z}_{1:(t-1)}) = \mathcal{N}(\widetilde{\boldsymbol{\mu}}_{t-1}, \widetilde{\boldsymbol{\Sigma}}_{t-1})$$
(24)

where $\widetilde{\mu}_{t-1} = \widetilde{\Sigma}_{t-1} \{ A^T Q_t^{-1} s_t + V_{t-1}^{-1} m_{t-1} \}, \widetilde{\Sigma}_{t-1} = (V_{t-1}^{-1} + A^T Q_t^{-1} A)^{-1}.$

3. Backward recursion for the posterior covariance: For each t = T - 1, T - 2, ..., 0, start from (23), we are able to find $p(s_{t-1}|z_{1:T}) = \mathcal{N}(\widetilde{m}_{t-1}, \widetilde{V}_{t-1})$ via backward recursion. According to the Markov property,

$$p(s_{t-1}|s_{t:T}, z_{1:T}) \equiv p(s_{t-1}|s_t, z_{1:T}) \equiv p(s_{t-1}|s_t, z_{1:(t-1)})$$
(25)

using the second equality in (25), we obtain

$$p(\boldsymbol{s}_{t-1}|\boldsymbol{s}_t, \boldsymbol{z}_{1:T}) = \mathcal{N}(\widetilde{\boldsymbol{\mu}}_{t-1}, \widetilde{\boldsymbol{\Sigma}}_{t-1})$$
(26)

Combine (23) with (26), integrated out s_t , we have the backward smoothing density at t-1,

$$p(\mathbf{s}_{t-1}|\mathbf{z}_{1:T}) = \mathcal{N}(\widetilde{\mathbf{m}}_{t-1}, \mathbf{V}_{t-1})$$
$$\widetilde{\mathbf{m}}_{t-1} = \widetilde{\mathbf{\Sigma}}_{t-1}(\mathbf{A}^T \mathbf{Q}_t^{-1} \widetilde{\mathbf{m}}_t + \mathbf{V}_{t-1}^{-1} \mathbf{m}_{t-1})$$
$$\widetilde{\mathbf{V}}_{t-1} = \widetilde{\mathbf{\Sigma}}_{t-1} + \widetilde{\mathbf{\Sigma}}_{t-1} \mathbf{A}^T \mathbf{Q}_t^{-1} \widetilde{\mathbf{V}}_t \mathbf{Q}_t^{-1} \mathbf{A} \widetilde{\mathbf{\Sigma}}_{t-1}$$
(27)

2 Gibbs Sampling for the Extended Model

2.1 Dealing with Multiple Documents

At each time point t, for each document n_t , the likelihood is

$$y_{p,t}^{n_t} = g\left(z_{p,t}^{n_t}\right)$$
(28)

To consider N_t documents per time point, add additional layer,

$$\boldsymbol{z}_{t}^{n_{t}} = \boldsymbol{\Lambda} \boldsymbol{b}_{t}^{n_{t}} + \boldsymbol{\epsilon}_{t}^{n_{t}}, \quad \boldsymbol{\epsilon}_{t}^{n_{t}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \quad \boldsymbol{R} = \boldsymbol{I}_{P}, \quad n_{t} = 1, \dots, N_{t}$$
$$\boldsymbol{b}_{t}^{n_{t}} \sim \mathcal{N}(\boldsymbol{s}_{t}, \boldsymbol{\Gamma}), \quad \boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma}), \quad \boldsymbol{\gamma}_{k}^{-1} \sim \mathcal{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad k = 1, \dots, K$$
(29)

2.2 Gibbs Sampler

Denote $\Theta = \{\Lambda, S, B_{1:T}, L, U, \Gamma, \phi, \omega, \rho, \tau, \eta, \nu, \zeta\}$, we use Gibbs sampling to approximate the joint posterior distribution of (Z, Θ) ,

- 1. Given Θ , find $p(z_{p,t}^{n_t}|\Theta, Z \in R(Y), Z_{-p,-t,-n_t})$, for $p = 1, \ldots, P, t = 1, \ldots, T, n_t = 1, \ldots, N_t$
- 2. Given Z, find $p(\Theta|Z, Z \in R(Y))$ reduce to $p(\Theta|Z)$

Treat $oldsymbol{Z}$ as augmented data, the full likelihood for $(oldsymbol{Z}, \Theta)$ is

$$p(\boldsymbol{Z},\boldsymbol{\Theta}) = \left(\prod_{t=1}^{T}\prod_{n_t=1}^{N_t}\mathcal{N}(\boldsymbol{z}_t^{n_t};\boldsymbol{\Lambda}\boldsymbol{b}_t^{n_t},\boldsymbol{R})\right) \times \left(\prod_{k=1}^{K}\mathcal{G}(\phi_k;1/2,\omega_k)\mathcal{G}(\omega_k;1/2,d^2)\right)$$
$$\times \left(\prod_{t=1}^{T}\prod_{n_t=1}^{N_t}\mathcal{N}(\boldsymbol{b}_t^{n_t},\boldsymbol{s}_t,\boldsymbol{\Gamma})\right) \times \prod_{k=1}^{K}\mathcal{G}(\gamma_k^{-1};\alpha,\beta)$$
$$\times \prod_{k=1}^{K}\left[\mathcal{N}(s_{0,k};0,\sigma_s^2)\left(\prod_{t=1}^{T}\mathcal{N}(s_{k,t};\rho_k s_{k,t-1},\tau_{k,t})\mathcal{G}(\tau_{k,t};e,\eta_{k,t})\mathcal{G}(\eta_{k,t};f,\nu)\right)\right]$$
$$\times \left(\prod_{p=1}^{P}\prod_{k=1}^{K}\mathcal{N}_{+}(\lambda_{p,k};0,l_{p,k})\mathcal{G}(l_{p,k};a,u_{p,k})\mathcal{G}(u_{p,k};b,\phi_k)\right) \times \prod_{k=1}^{K}\mathrm{TN}_{(0,1)}(\rho_k;\mu_0,\sigma_0^2)$$
$$\times \mathcal{G}(\nu;1/2,\zeta) \times \mathcal{G}(\zeta;1/2,h^2) \tag{30}$$

• Sampling $z_{p,t}^{n_t}$

$$p(z_{p,t}^{n_t}|\boldsymbol{\Theta}, \boldsymbol{Z} \in R(\boldsymbol{Y}), \boldsymbol{Z}_{-p,-t,-n_t}) \sim \mathrm{TN}_{[\underline{z_{p,t}^{n_t}}, \overline{z_{p,t}^{n_t}}]}(\sum_{k=1}^{K} \lambda_{p,k} b_{k,t}^{n_t}, 1)$$
(31)

where $\underline{z_{p,t}^{n_t}} = \max\{z_{p',t'}^{n_t} : y_{p',t'}^{n_t} < y_{p,t}^{n_t}\}$ and $\overline{z_{p,t}^{n_t}} = \min\{z_{p',t'}^{n_t} : y_{p',t'}^{n_t} > y_{p,t}^{n_t}\}$

• Sampling $\lambda_{p,k}$

$$p(\lambda_{p,k}|-) \propto \left(\prod_{t=1}^{T} \prod_{n_t=1}^{N_t} \mathcal{N}(z_{p,t}^{n_t}; \lambda_{p,k} b_{k,t}^{n_t} + \sum_{j \neq k} b_{j,t}^{n_t} \lambda_{p,j}, 1)\right) \mathcal{N}_+(\lambda_{p,k}; 0, l_{p,k})$$

= $\mathcal{N}_+ \left(\lambda_{p,k}; v_{\lambda_{p,k}} \sum_{t=1}^{T} \sum_{n_t=1}^{N_t} b_{k,t}^{n_t} \left[z_{p,t}^{n_t} - \lambda_p \mathbf{b}_t^{n_t} + b_{k,t}^{n_t} \lambda_{p,k} \right], v_{\lambda_{p,k}} \right)$
 $v_{\lambda_{p,k}} = (l_{p,k}^{-1} + \sum_{t=1}^{T} \sum_{n_t=1}^{N_t} (b_{k,t}^{n_t})^2)^{-1}$ (32)

- Sampling $\boldsymbol{b}_{t}^{n_{t}}$ $p(\boldsymbol{b}_{t}^{n_{t}}|-) = \mathcal{N}(\boldsymbol{\Sigma}_{\boldsymbol{b}_{t}^{n_{t}}}(\boldsymbol{\Lambda}^{T}\boldsymbol{R}^{-1}\boldsymbol{z}_{t}^{n_{t}}+\boldsymbol{\Gamma}^{-1}\boldsymbol{s}_{t}),\boldsymbol{\Sigma}_{\boldsymbol{b}_{t}^{n_{t}}}), \quad \boldsymbol{\Sigma}_{\boldsymbol{b}_{t}^{n_{t}}} = (\boldsymbol{\Gamma}^{-1}+\boldsymbol{\Lambda}^{T}\boldsymbol{R}^{-1}\boldsymbol{\Lambda})^{-1}$ (33)
- Sampling γ_k^{-1}

$$p(\gamma_k^{-1}|-) \sim \mathcal{G}\left(\alpha + \frac{1}{2}\sum_{t=1}^T N_t, \beta + \frac{1}{2}\sum_{t=1}^T \sum_{n_t=1}^{N_t} (b_{k,t}^{n_t} - s_{k,t})^2\right)$$
(34)

- Sampling $l_{p,k}$, $u_{p,k}$ $p(l_{p,k}|-) = \operatorname{GIG}(a - 1/2, 2u_{p,k}, (\lambda_{p,k})^2), \quad p(u_{p,k}|-) = \mathcal{G}(a + b, l_{p,k} + \phi_k)$ (35)
- Sampling ϕ_k, ω_k

$$p(\phi_k|-) = \mathcal{G}(1/2 + bP, \omega_k + \sum_{p=1}^{P} u_{p,k}), \quad p(\omega_k|-) = \mathcal{G}(1, \phi_k + d^2)$$
(36)

- Sampling $\tau_{k,t}, \eta_{k,t}$ $p(\tau_{k,t}|-) = \operatorname{GIG}(e-1/2, 2\eta_{k,t}, (s_{k,t}-\rho_k s_{k,t-1})^2), \quad p(\eta_{k,t}|-) = \mathcal{G}(e+f, \tau_{k,t}+\nu)$ (37)
- Sampling ν, ζ

$$p(\nu|-) = \mathcal{G}(1/2 + fTK, \zeta + \sum_{k=1}^{K} \sum_{t=1}^{T} \eta_{k,t}), \quad p(\zeta|-) = \mathcal{G}(1, \nu + h^2)$$
(38)

Sampling ρ_k

$$p(\rho_k|-) = \operatorname{TN}_{(0,1)}\left(\sigma_{\rho_k}^2(\sigma_0^{-2}\mu_0 + \sum_{t=1}^T \tau_{k,t}^{-1} s_{k,t-1} s_{k,t}), \sigma_{\rho_k}^2\right)$$
(39)

where $\sigma_{\rho_k}^2 = 1/(\sigma_0^{-2} + \sum_{t=1}^T \tau_{k,t}^{-1} s_{k,t-1}^2).$

We have the state model and the observation model³

$$\boldsymbol{s}_t | \boldsymbol{s}_{t-1} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{s}_{t-1}, \boldsymbol{Q}_t), \quad \boldsymbol{A} = \operatorname{diag}(\boldsymbol{\rho}), \quad \boldsymbol{Q}_t = \operatorname{diag}(\boldsymbol{\tau}_t),$$
(40)

$$\boldsymbol{b}_{t}^{n_{t}}|\boldsymbol{s}_{t} \sim \mathcal{N}(\boldsymbol{s}_{t}, \boldsymbol{\Gamma}), \quad \boldsymbol{\Gamma} = \operatorname{diag}(\gamma_{k})$$
 (41)

for $n_t = 1, ..., N_t, t = 1, ..., T$

1. Forward Filtering: beginning at t = 0 with $s_0 \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_K)$, we have, for all $t = 1, \ldots, T$, the on-line posteriors $p(s_t | \mathbf{B}_{1:t}) = \mathcal{N}(\mathbf{m}_t, \mathbf{V}_t)$. Start from

$$p(s_{t-1}|B_{1:(t-1)}) = \mathcal{N}(m_{t-1}, V_{t-1})$$
 (42)

Combine (40) with (42), integrate out s_{t-1} , we have the predictive density at t,

$$p(\boldsymbol{s}_t | \boldsymbol{B}_{1:(t-1)}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{m}_{t-1}, \boldsymbol{Q}_t + \boldsymbol{A}\boldsymbol{V}_{t-1}\boldsymbol{A}^T)$$
(43)

Further combine (41) with (43), we have the on-line posteriors at t,

$$p(\boldsymbol{s}_t | \boldsymbol{B}_{1:t}) = \mathcal{N}(\boldsymbol{m}_t, \boldsymbol{V}_t)$$
$$\boldsymbol{m}_t = \boldsymbol{V}_t \{ N_t \boldsymbol{\Gamma}^{-1} \overline{\boldsymbol{b}_t} + (\boldsymbol{Q}_t + \boldsymbol{A} \boldsymbol{V}_{t-1} \boldsymbol{A}^T)^{-1} \boldsymbol{A} \boldsymbol{m}_{t-1} \},$$
$$\boldsymbol{V}_t = [(\boldsymbol{Q}_t + \boldsymbol{A} \boldsymbol{V}_{t-1} \boldsymbol{A}^T)^{-1} + N_t \boldsymbol{\Gamma}^{-1}]^{-1}, \quad \overline{\boldsymbol{b}_t} = \frac{1}{N_t} \sum_{n_t=1}^{N_t} \boldsymbol{b}_t^{n_t} \quad (44)$$

Define $\widetilde{\Omega}_t = (\boldsymbol{Q}_t + \boldsymbol{A} \boldsymbol{V}_{t-1} \boldsymbol{A}^T)$, according to the Woodbury lemma,

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$
(45)

we have

$$(\widetilde{\Omega}_t^{-1} + N_t \Gamma^{-1})^{-1} = \widetilde{\Omega}_t - \widetilde{\Omega}_t (N_t^{-1} \Gamma + \widetilde{\Omega}_t)^{-1} \widetilde{\Omega}_t$$
(46)

2. Backward Sampling: define the backward smoothing density

$$p(\boldsymbol{s}_t | \boldsymbol{B}_{1:T}) = \mathcal{N}(\widetilde{\boldsymbol{m}}_t, \widetilde{\boldsymbol{V}}_t)$$
(47)

At t = T, we have the initialization condition $p(s_T | B_{1:T}) = \mathcal{N}(\widetilde{m}_T, \widetilde{V}_T) = \mathcal{N}(m_T, V_T)$. Combine (40) with (42), we have the conditional distribution of s_{t-1} given s_t ,

$$p(\mathbf{s}_{t-1}|\mathbf{s}_{t}, \mathbf{B}_{1:(t-1)}) = \mathcal{N}(\widetilde{\boldsymbol{\mu}}_{t-1}, \boldsymbol{\Sigma}_{t-1})$$
$$\widetilde{\boldsymbol{\mu}}_{t-1} = \widetilde{\boldsymbol{\Sigma}}_{t-1} \{ \mathbf{A}^{T} \mathbf{Q}_{t}^{-1} \mathbf{s}_{t} + \mathbf{V}_{t-1}^{-1} \mathbf{m}_{t-1} \}$$
$$\widetilde{\boldsymbol{\Sigma}}_{t-1} = (\mathbf{V}_{t-1}^{-1} + \mathbf{A}^{T} \mathbf{Q}_{t}^{-1} \mathbf{A})^{-1}$$
(48)

Similarly, apply Woodbury matrix inversion lemma we have

$$\widetilde{\boldsymbol{\Sigma}}_{t-1} = \boldsymbol{V}_{t-1} - \boldsymbol{V}_{t-1} \boldsymbol{A}^T \widetilde{\boldsymbol{\Omega}}_t^{-1} \boldsymbol{A} \boldsymbol{V}_{t-1}$$
(49)

• Sampling $\widetilde{V}_{0:(T-1)}$

Integrated out B_t , we have the observation model

$$\boldsymbol{z}_{t}^{n_{t}}|\boldsymbol{s}_{t} \sim \mathcal{N}(\boldsymbol{\Lambda}\boldsymbol{s}_{t}, \widetilde{\boldsymbol{R}}), \quad \widetilde{\boldsymbol{R}} = \boldsymbol{I}_{P} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{T}, \quad \widetilde{\boldsymbol{R}}^{-1} = \boldsymbol{I}_{P} - \boldsymbol{\Lambda}(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Lambda}^{T}\boldsymbol{\Lambda})^{T}\boldsymbol{\Lambda}^{T}$$
(50)

for $n_t = 1, \ldots, N_t, t = 1, \ldots, T$. We have the on-line posteriors at t,

$$p(\boldsymbol{s}_{t}|\boldsymbol{Z}_{1:t}) = \mathcal{N}(\boldsymbol{m}_{t}, \boldsymbol{V}_{t})$$
$$\boldsymbol{m}_{t} = \boldsymbol{V}_{t} \{ N_{t} \boldsymbol{\Lambda}^{T} \widetilde{\boldsymbol{R}}^{-1} \overline{\boldsymbol{z}}_{t} + \widetilde{\boldsymbol{\Omega}}_{t}^{-1} \boldsymbol{A} \boldsymbol{m}_{t-1} \},$$
$$\boldsymbol{V}_{t} = [\widetilde{\boldsymbol{\Omega}}_{t}^{-1} + N_{t} \boldsymbol{\Lambda}^{T} \widetilde{\boldsymbol{R}}^{-1} \boldsymbol{\Lambda}]^{-1}, \ \overline{\boldsymbol{z}}_{t} = \frac{1}{N_{t}} \sum_{n_{t}=1}^{N_{t}} \boldsymbol{z}_{t}^{n_{t}}$$
(51)

³For brevity, we omit the dependencies on Θ in notation

The conditional distribution of s_{t-1} given s_t ,

$$p(s_{t-1}|s_t, Z_{1:(t-1)}) = \mathcal{N}(\widetilde{\mu}_{t-1}, \Sigma_{t-1})$$
$$\widetilde{\mu}_{t-1} = \widetilde{\Sigma}_{t-1} \{ A^T Q_t^{-1} s_t + V_{t-1}^{-1} m_{t-1} \}$$
$$\widetilde{\Sigma}_{t-1} = (V_{t-1}^{-1} + A^T Q_t^{-1} A)^{-1}$$
(52)

Similarly, apply Woodbury matrix inversion lemma we have

$$\widetilde{\boldsymbol{\Sigma}}_{t-1} = \boldsymbol{V}_{t-1} - \boldsymbol{V}_{t-1} \boldsymbol{A}^T \widetilde{\boldsymbol{\Omega}}_t^{-1} \boldsymbol{A} \boldsymbol{V}_{t-1}$$
Further, the backward smoothing density at $t-1$, (53)

$$p(\boldsymbol{s}_{t-1}|\boldsymbol{Z}_{1:T}) = \mathcal{N}(\widetilde{\boldsymbol{m}}_{t-1}, \widetilde{\boldsymbol{V}}_{t-1})$$
$$\widetilde{\boldsymbol{m}}_{t-1} = \widetilde{\boldsymbol{\Sigma}}_{t-1}(\boldsymbol{A}^{T}\boldsymbol{Q}_{t}^{-1}\widetilde{\boldsymbol{m}}_{t} + \boldsymbol{V}_{t-1}^{-1}\boldsymbol{m}_{t-1})$$
$$\widetilde{\boldsymbol{V}}_{t-1} = \widetilde{\boldsymbol{\Sigma}}_{t-1} + \widetilde{\boldsymbol{\Sigma}}_{t-1}\boldsymbol{A}^{T}\boldsymbol{Q}_{t}^{-1}\widetilde{\boldsymbol{V}}_{t}\boldsymbol{Q}_{t}^{-1}\boldsymbol{A}\widetilde{\boldsymbol{\Sigma}}_{t-1}$$
(54)

3 Experimental results

3.1 Simulation Study: DRFM with different innovations

We conducted a simulation study to assess the performance of our proposed approach. We first generate the latent continuous variable Z from the augmented model with e = f = 0.5, $\nu = 1$, $\rho_k = 0.5$, $l_{p,k} = 1/P$ for $k = 1, \ldots, K$, $p = 1, \ldots, P$ and then round it to integer value. Three different approaches are considered here: Gaussian innovation with fixed variance $\delta \sim \mathcal{N}(0, 1)$, Gaussian innovation with unknown variance $\delta \sim \mathcal{N}(0, \tau)$, and $\tau^{-1} \sim \mathcal{G}(0.01, 0.01)$, and heavy-tailed innovation $\delta \sim \text{TPBN}(0.5, 0.5, \phi)$ with $\phi^{1/2} \sim C^+(0, 1)$. The results are shown in Figure 1.



Figure 1: Estimated posterior mean of factor score s with 95% confidence interval for P = 10, K = 2 and T = 150. Left column: Gaussian innovation with fixed variance $\tau = 1$; middle column: Gaussian innovation with unknown variance $\tau^{-1} \sim \mathcal{G}(0.01, 0.01)$; right column: heavy-tailed innovation.

Note that here we are dealing with a simple two factor dynamic model, and we are lucky to recover the ground truth of the trajectory of factor score. In contrast to other models that involve nonlinear transformation, the smooth transition and sudden jumps can be well preserved under the proposed DFRM framework, using heavy-tailed innovation.

Figure 2 shows the monotone relationship between observation y and the latent variable z inferred by extended rank likelihood in our DFRM model. It can be seen that the rank likelihood approach maintains the order information of y in z and provides a flexible link between y and z.



Figure 2: Estimated posterior mean of latent variable z vs. the observed data y for P = 10, K = 2and T = 150. Left: Gaussian innovation with fixed variance $\tau = 1$; middle: Gaussian innovation with unknown variance $\tau^{-1} \sim \mathcal{G}(0.01, 0.01)$; right: heavy-tailed innovation.

3.2 Case Study I: State of the Union dataset

The State of the Union dataset contains the transcripts of 225 US State of the Union addresses, from 1790 to 2014. We take each transcript as a document, *i.e.*, there is one document per year. We have 7518 unique words in total. Table 1 shows all 25 learned topics and and the top 12 most probable words associated with each of them. Figure 3 presents the learned trajectory for each topic.

			1						1			
Topic#1	Topic#2		Topic#3		Topic#4		Topic#5		Topic#6		Topic#7	
UNITED	Dollars		ADMINISTRATION		GOVERNMENT		GOVERNMENT		LAW		GOVERNMENT	
ACT	WAR		FEDERAL		AMERICAN		SERVICE		COUNTRY		UNITED	
PUBLIC	MILLION		PROGRAM		UNITED		PUBLIC		NATIONAL		DEPARTMENT	
TREATY	FISCAL		POLICY		FOREIGN		DEPARTMENT		PUBLIC		PUBLIC	
Duties	EXPENDITURES		ENERGY		DEPARTMENT		Report		BUSINESS		LAW	
Present	resent GOVERNMENT		Programs		NATIONAL	L	SECRETARY		GOVERNMENT		COURT	
NATIONS	NATIONS Billion		ECONOMIC		CANAL		DISTRICT		ACTION		SERVICE	
TREASURY	ASURY PROGRAM		DEVELO	OPMENT	POLICY		Attention		CONTROL		FEDERAL	
Session	UNITED		SECURITY		REPUBLIC		Present		UNITED		CANAL	
COMMERCE	RCE FEDERAL		Nation		Order		FISCAL		INTERSTATE		TARIFF	
Citizens	Estimated		Major		ADMINISTRATION		LAWS		LABOR		DISTRICT	
WAR	LEGISLATION		ACT		BANKS		COURT		CORPORATIONS		LANDS	
Topic#8	Topic#9		Topic#10		Topic#1		1 Topic		#12 To		ic#13	
GOVERNMENT		Constitution		MEXICO		INCREA	SE GOV		ERNMENT GO		VERNMENT	
GENERAL	GENERAL COUN			GOVERN	MENT	UNITEI)	PUBI	PUBLIC		ITED	
PUBLIC	PUBLIC WAR		Texas		Cent			Natio	Nation		ISLANDS	
CHARACTER PRESIDENT		UNITED			LAW		AMERICAN		COMMISSION			
Interests POWER		WAR			LEGISL	ATION	LAW		Island			
Subject MEXICO		MEXICAN		V	SECRE	TARY	POW	POWER		Cuba		
COUNTRY PUBLIC		ARMY			Free		CONDITIO		Spain			
POWER	POWER UNION		Territory			INCREA	ASED	BUSINESS		ACT		
Duty California		COUNTRY		Y	FISCAL		ISLANDS		GENERAL			
Attention SERVICE		PEACE		AMERIC		CAN SERVICE		ICE	MILITARY			
FEDERAL HOUSE		HOUSE	POLICY		TARIFF		WAR		IN		ERNATIONAL	
Means Period		LANDS			Products		LAWS		OFFICERS			
Topic#14 Topic#15		Topic#15	Topic#16			Topic#1	7	Topic#18		Тор	ic#19	
Free		GOVERNMENT		STATUTE		Jobs		CHILDREN		AM	IERICA	
NATIONS FE		FEDERAL		LAW		COUNTRY		AMERICA		GOVERNMENT		
FREEDOM PUBL		PUBLIC	3LIC		BUSINESS		TAX		AMERICANS		Nation	
ECONOMIC NATIO		NATIONAL	ATIONAL		GENERAL		AMERICAN		CARE		AMERICAN	
MILITARY COUNTR		COUNTRY	AMERIC		AN ECO		CONOMY		Tonight		FEDERAL	
DEFENSE ECONOMI		ECONOMIC	PURPOSE		6	DEFICT	I Supp		ort		ight	
DEACE AGRICULT		JRE COURT			AMERI		Centu	ITU P		ACE		
PEACE BANKS		MEXICO			ENERG		HEA	ALIH		WAR		
STRENGTH Present			SERVICE		Business	u C		Vorking		AMEDICANS		
DECURITY AMERICA		т	FEDERAL	SION DI AN		п	Chall	UDITY E		IEKICANS		
Nation Construction		N COMMISS Present		CARE		Fami		JKIII FU		IUKE		
Topio#20		Tradiction		Tonio#2		Tanin D Tanic		#24 T		i=#25		
Gold	Gold GOVERNM		ENT	NT WAR		GOVERN		MENT UNIT		GO	VERNMENT	
GOVERNMENT Constitution		Constitution	MEXICO		UNITE				REATY		PUBLIC	
Notes UNIT		UNITED	TED PEAC		EACE		Spain		Isthmus		BANKS	
TREASURY POWER		ARMY		Cuba			PUBI	PUBLIC		BANK		
Silver UNION		ENEMY		Spanish			PANA	PANAMA		RRENCY		
UNITED FEDERAL		FORCES		WAR		L		W		Money		
Bonds Duty		Duty	MILITAR		Y	Island	Island		Territory		UNITED	
CURRENCY AI		AMERICAN	AMERICAN		MEXICAN		SECRETARY		AMERICA		FEDERAL	
RESERVE Kansas		Kansas		Production		June		CAN	NAL A		IERICAN	
Circulation		Question	JAPANES		E Duty		SE		VICE N		TIONAL	
Issued LAW		LAW	FIGHTING		G DEPAR		IMENT BA		NKS D		У	
Large Present			AMERICAI		FISCAL	Col		mbia Ins		itutions		

Table 1: Top 12 words associated with the State of the Union Topics



Figure 3: (*State of the Union* dataset) Time evolving topics from 1790 to 2014. Left up panel: Topics 1 to 7. Right up panel: Topics 8 to 13. Left bottom panel: Topics 14 to 19. Right bottom panel: Topics 20 to 25. The plotted values represent the posterior means.