A Regularized Linear Dynamical System Framework for Multivariate Time Series Analysis - Supplemental Materials

Zitao Liu and **Milos Hauskrecht**
Computer Science Department, University of Pittsburgh, 210 South Bouquet St., Pittsburgh, PA, 15260 USA

Notations

We use the following notations in this supplemental material. The notation is consistent with the original paper.

- $\hat{\mathbf{z}}_{t|t-1} = \mathbb{E}[\mathbf{z}_t|\{\mathbf{y}\}_1^{t-1}]$ is the *priori* estimation
- $\hat{\mathbf{z}}_{t-1|t-1} = \mathbb{E}[\mathbf{z}_{t-1}|\{\mathbf{y}\}_1^{t-1}]$ is the *posteriori* estimation.
- $P_{t|t-1} = \mathbb{E}[(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})^{'}]$ is the *priori* estimate error covariance.
- $P_{t-1|t-1} = \mathbb{E}[(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})^{'}]$ is the *posteriori* estimate error covariance.
- $\hat{\mathbf{z}}_{t|T} \equiv \mathbb{E}[\mathbf{z}_t|\mathbf{y}]$, $M_{t|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_t^{'}|\mathbf{y}]$, $M_{t,t-1|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}^{'}|\mathbf{y}]$, $P_{t|T} = \mathbb{VAR}[\mathbf{z}_t|\mathbf{y}]$, and $P_{t,t-1|T} = \mathbb{VAR}[\mathbf{z}_t\mathbf{z}_{t-1}^{'}|\mathbf{y}]$

Theorem 1 Proof

Theorem 1. *Generalized gradient descent with a fixed step size $\rho \leq 1/(||Q^{-1}||_F \cdot \| \sum_{t=1}^{T-1} \mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]\|_F + \lambda_2)$ for minimizing eq.(7) has convergence rate $O(1/k)$, where* k *is the number of iterations.*

Proof. $h(A)$ is differentiable with respect to A, and its gradient is $\bigtriangledown h(A) = Q^{-1}(A\sum_{t=1}^{T-1}\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}] - \sum_{t=2}^{T}\mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}'|\mathbf{y}]) + \lambda_2 A$. Using simple algebraic manipulation we arrive at

$$
\begin{aligned}
&|| \bigtriangledown h(X) - \bigtriangledown h(Y)||_F \\
=&||Q^{-1}(X-Y)\sum_{t=1}^{T-1}\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}] + \lambda_2(X-Y)||_F \\
\leq&||Q^{-1}||_F \cdot \| \sum_{t=1}^{T-1}\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]\|_F \cdot ||X-Y||_F \\
&+\lambda_2 \cdot ||X-Y||_F \\
=&(||Q^{-1}||_F \cdot \| \sum_{t=1}^{T-1}\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]\|_F + \lambda_2) \cdot ||X-Y||_F
\end{aligned}
$$

The inequality holds because of the sub-multiplicative property of Frobenius norm. Since we know for eq.(7), $\min_A h(A) + \lambda_2\|A\|_*$, and $h(A)$ has Lipschitz continuous gradient with constant $||Q^{-1}||_F \cdot || \sum_{t=1}^{T-1}\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]||_F + \lambda_2$, according to (Shor 1968) (Fornasier 2008), we have

$$
\begin{aligned}
&\left\| h(A^{(k)}) + \lambda_2\|A^{(k)}\|_* - h(A^{(*)}) - \lambda_2\|A^{(*)}\|_* \right\| \\
\leq&\left\| A^{(0)} - A^* \right\|_F^2 / 2tk
\end{aligned}
$$

where $A^{(0)}$ is the initial value and A^* is the optimal value for A; k is the number of iterations.

□

Backward algorithm for rLDS

Algorithm 1 EM: E-step Backward algorithm for rLDS

INPUT:
- Output from Kalman filter algorithm: $\{\hat{\mathbf{z}}_{t|t-1}\}_{t=2}^T$, $\{\hat{\mathbf{z}}_{t|t}\}_{t=1}^T$, $\{P_{t|t}\}_{t=1}^T$, $\{P_{t|t-1}\}_{t=2}^T$ and $\{K_t\}_{t=1}^T$.
- Current step LDS parameters: $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$.

PROCEDURE:

```
// Initialize the recrusion
M_{T|T} = P_{T|T} + ẑ_{T|T} ẑ'_{T|T}
J_{T-1} = P_{T-1|T-1} A' (P_{T|T-1})^{-1}
P_{T-1|T} = P_{T-1|T-1} + J_{T-1}(P_{T|T} - P_{T|T-1}) J'_{T-1}
ẑ_{T-1|T} = ẑ_{T-1|T-1} + J_{T-1}(ẑ_{T|T} - A ẑ_{T-1|T-1})
P_{T,T-1|T} = (I - K_T C) A P_{T-1|T-1}
M_{T,T-1|T} = P_{T,T-1|T} + ẑ_{T|T} ẑ'_{T-1|T}
// Start the recrusion
for t = T-1 → 1 do
    M_{t|T} = P_{t|T} + ẑ_{t|T} ẑ'_{t|T}
    J_{t-1} = P_{t-1|t-1} A' (P_{t|t-1})^{-1}
    P_{t,t-1|T} = P_{t|t} J'_{t-1} + J_t(P_{t+1,t|T} - A P_{t|t}) J'_{t-1}
    M_{t,t-1|T} = P_{t,t-1|T} + ẑ_{t|T} ẑ'_{t-1|T}
    ẑ_{t-1|T} = ẑ_{t-1|t-1} + J_{t-1}(ẑ_{t|T} - A ẑ_{t-1|t-1})
    P_{t-1|T} = P_{t-1|t-1} + J_{t-1}(P_{t|T} - P_{t|t-1}) J'_{t-1}
end for
```

OUTPUT: $\{\hat{\mathbf{z}}_{t-1|T}\}_{t=1}^T$, $\{M_{t|T}\}_{t=1}^T$ and $\{M_{t,t-1|T}\}_{t=1}^T$.

Transformation from eq.(4) to eq.(5).

We will use the following rules to show the mathematical transformation from eq.(4) to eq.(5).

- $Q^{-1} = LL'$ (Q is a PSD matrix)
- $tr(A_{k\times l}B_{l\times m}C_{m\times n}) = \text{vec}(A')'(I_k \otimes B)\,\text{vec}(C)$
- $\text{vec}(A_{k\times l}B_{l\times m}) = (B' \otimes I_k)\,\text{vec}(A)$

By using the above rules, we can have

$$
\begin{aligned}
& arg\min_A \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[(\mathbf{z}_t - A\mathbf{z}_{t-1})'Q^{-1}(\mathbf{z}_t - A\mathbf{z}_{t-1})] + \frac{\lambda_3}{2}\|A\|_F^2 \\
\Leftrightarrow\; & arg\min_A \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_{t-1}'A'LL'A\mathbf{z}_{t-1}] - \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_{t-1}'A'LL'\mathbf{z}_t] \\
& \quad - \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_t'LL'A\mathbf{z}_{t-1}] + \frac{\lambda_3}{2}\|A\|_F^2 \\
\Leftrightarrow\; & arg\min_A \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}\Big[Tr[L'A\mathbf{z}_{t-1}\mathbf{z}_{t-1}'A'L]\Big] + \frac{\lambda_3}{2}\|A\|_F^2 \\
& \quad - \sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}\Big[Tr[L'\mathbf{z}_t\mathbf{z}_{t-1}'A'L]\Big] \\
\Leftrightarrow\; & arg\min_A \frac{1}{2}\sum_{t=2}^{T}Tr\Big[L'A(\mathbb{E}_{\mathbf{z}}[\mathbf{z}_{t-1}\mathbf{z}_{t-1}'])A'L\Big] + \frac{\lambda_3}{2}\|A\|_F^2 \\
& \quad - \sum_{t=2}^{T}Tr\Big[L'\Big(\mathbb{E}_{\mathbf{z}}[\mathbf{z}_t\mathbf{z}_{t-1}'])A'L\Big] \\
\Leftrightarrow\; & arg\min_A \frac{1}{2}Tr\Big[L'A(\underbrace{\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_{t-1}\mathbf{z}_{t-1}']}_{B})A'L\Big] + \frac{\lambda_3}{2}Tr\Big[A'A\Big] \\
& \quad - Tr\Big[L'(\underbrace{\sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_t\mathbf{z}_{t-1}']}_{D})A'L\Big] \\
\Leftrightarrow\; & arg\min_A \frac{1}{2}Tr\Big[L'ABA'L - 2L'DA'L + \lambda_3A'A\Big] \\
\Leftrightarrow\; & arg\min_A 0.5\,\text{vec}(A'L)'(I_d \otimes B)\,\text{vec}(A'L) \\
& \quad - \text{vec}(L)'(I_d \otimes D)\,\text{vec}(A'L) + 0.5\lambda_3\,\text{vec}(A')'\,\text{vec}(A') \\
\Leftrightarrow\; & arg\min_A 0.5\,\text{vec}(A')'(Q^{-1} \otimes \sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_{t-1}\mathbf{z}_{t-1}'] + \lambda_3 I_{d^2})\,\text{vec}(A') \\
& \quad - \text{vec}(L)'(L' \otimes \sum_{t=2}^{T}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_t\mathbf{z}_{t-1}'])\,\text{vec}(A')
\end{aligned}
$$

Kalman filter algorithm for rLDS

Algorithm 2 Kalman filter algorithm for rLDS

INPUT: Current step LDS parameters: $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$.
PROCEDURE:

```
// Initialize the recrusion
ẑ_{1|1} = ξ and P_{1|1} = Ψ.
// Start the recrusion
for t = 2 → T do
    // Time Update:
    ẑ_{t|t-1} = A ẑ_{t-1|t-1}
    P_{t|t-1} = A P_{t-1|t-1} A' + Q
    // Measure Update:
    K_t = P_{t|t-1} C' (C P_{t|t-1} C' + R)^{-1}
    ẑ_{t|t} = ẑ_{t|t-1} + K_t (y_t - C ẑ_{t|t-1})
    P_{t|t} = P_{t|t-1} - K_t C P_{t|t-1}
end for
```

OUTPUT: $\{\hat{\mathbf{z}}_{t|t-1}\}_{t=2}^{T}$, $\{\hat{\mathbf{z}}_{t|t}\}_{t=1}^{T}$, $\{P_{t|t}\}_{t=1}^{T}$, $\{P_{t|t-1}\}_{t=2}^{T}$ and $\{K_t\}_{t=1}^{T}$.

Production and Billing Data Results

In *Production and Billing* data, we increase the training size to 90 and use the rest 10 observations for tesing. The prediction results are shown in Table 1.

Table 1: Average-MAPE results on *Production and Billing* dataset with 90 training and 10 testing.

# of hidden states	2	3	4	5	6	7	8
EM	4.8707	5.9996	5.2047	4.7732	4.4478	5.3245	4.6278
SubspaceID	4.7880	5.5328	6.9020	6.0295	4.5284	5.5099	5.2205
StableLDS	4.9024	5.4111	6.4988	6.0295	4.5323	5.5099	5.2205
rLDS$_{\mathcal{G}}$	4.7880	5.5328	6.9020	6.0295	4.5284	5.5099	5.2205
rLDS$_{\mathcal{R}}$	**4.5521**	**4.9543**	**5.1984**	**4.5455**	**4.4150**	**4.2961**	**3.8780**

# of hidden states	10	12	14	16	18	20	30
EM	5.2038	5.1908	3.5408	4.2521	4.0159	4.1269	3.8743
SubspaceID	5.2182	5.1707	5.7206	5.0139	4.1461	4.7801	3.7662
StableLDS	5.2182	5.1707	5.7206	5.0139	4.1461	4.7801	3.7662
rLDS$_{\mathcal{G}}$	5.2182	5.1707	5.7206	5.0139	4.1461	4.7800	3.7662
rLDS$_{\mathcal{R}}$	**3.9218**	**4.0061**	**3.4484**	**3.7843**	**3.2868**	**3.8535**	**3.5554**