

Learning Adaptive Forecasting Models from Irregularly Sampled Multivariate Clinical Data - Supplemental Material

Zitao Liu and Milos Hauskrecht

Computer Science Department, University of Pittsburgh, 210 South Bouquet St., Pittsburgh, PA, 15260 USA

A1. Kalman Filter Algorithm

For each patient l , we applied the DVI techniques on \mathbf{Y}^l to get the regularly sampled time series $\tilde{\mathbf{Y}}^l$. All the following operations are performed on the discretized data $\tilde{\mathbf{Y}}^l$ and we will repeat the Kalman filter algorithm on every patient sequence independently. For the sake of notational brevity, we omit the explicit sample index (l) and the tilde sign ($\tilde{\cdot}$) in the rest of Section A1.

Let $\mathbf{Y}_{i:j}$ be the multivariate sequence segment from the i th observations to the j th observations. Let denote $\hat{\mathbf{z}}_{t|T} \equiv \mathbb{E}[\mathbf{z}_t|\mathbf{Y}]$, $M_{t|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_t^\top|\mathbf{Y}]$, $M_{t,t-1|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}^\top|\mathbf{Y}]$, $P_{t|T} = \text{VAR}[\mathbf{z}_t|\mathbf{Y}]$, and $P_{t,t-1|T} = \text{VAR}[\mathbf{z}_t\mathbf{z}_{t-1}^\top|\mathbf{Y}]$. Let $\hat{\mathbf{z}}_{t|t-1}$ be the *priori* estimation of $\mathbb{E}[\mathbf{z}_t|\mathbf{Y}_{1:t-1}]$, $\hat{\mathbf{z}}_{t-1|t-1}$ be the *posteriori* estimation of $\mathbb{E}[\mathbf{z}_{t-1}|\mathbf{Y}_{1:t-1}]$, $P_{t|t-1}$ be the *priori* estimate error covariance of $\mathbb{E}[(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})^\top]$ and $P_{t-1|t-1}$ be the *posteriori* estimate error covariance of $\mathbb{E}[(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})^\top]$.

Kalman filter algorithm is used to infer the expectations at current time stamp (t) given the current observations ($\mathbf{Y}_{1:t}$), which is summarized in Algorithm 1.

Algorithm 1 Kalman filter algorithm for LDS

INPUT:

- MTS data \mathbf{Y} .
- Current step LDS parameters: $\Omega = \{A, C, Q, R, \xi, \Psi\}$.

PROCEDURE:

```

1: // Initialize the iteration
2:  $\hat{\mathbf{z}}_{1|1} = \xi$  and  $P_{1|1} = \Psi$ .
3: // Start the iteration
4: for  $t = 2 \rightarrow T$  do
5:   // Time Update:
6:    $\hat{\mathbf{z}}_{t|t-1} = A\hat{\mathbf{z}}_{t-1|t-1}$ 
7:    $P_{t|t-1} = AP_{t-1|t-1}A^\top + Q$ 
8:   // Measure Update:
9:    $K_t = P_{t|t-1}C^\top(CP_{t|t-1}C^\top + R)^{-1}$ 
10:   $\hat{\mathbf{z}}_{t|t} = \hat{\mathbf{z}}_{t|t-1} + K_t(\mathbf{y}_t - C\hat{\mathbf{z}}_{t|t-1})$ 
11:   $P_{t|t} = P_{t|t-1} - K_tCP_{t|t-1}$ 
12: 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  $\{K_t\}_{t=1}^T$ .

```

A2. Backward Algorithm in EM - E step

Here, we use the same notation defined in Section A1. Backward algorithm is used to infer the expectations at current time stamp (t) given the entire observations (\mathbf{Y}), which is summarized in Algorithm 2.

Algorithm 2 EM: E-step backward algorithm for LDS

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 $\{K_t\}_{t=1}^T$. Kalman filter algorithm is presented in Algorithm 1 in Section A1.
- Current step LDS parameters: $\Omega = \{A, C, Q, R, \xi, \Psi\}$.

PROCEDURE:

```

1: // Initialize the iteration
2:  $M_{T|T} = P_{T|T} + \hat{\mathbf{z}}_{T|T}\hat{\mathbf{z}}_{T|T}^\top$ 
3:  $J_{T-1} = P_{T-1|T-1}A^\top(P_{T|T-1})^{-1}$ 
4:  $P_{T-1|T} = P_{T-1|T-1} + J_{T-1}(P_{T|T} - P_{T|T-1})J_{T-1}^\top$ 
5:  $\hat{\mathbf{z}}_{T-1|T} = \hat{\mathbf{z}}_{T-1|T-1} + J_{T-1}(\hat{\mathbf{z}}_{T|T} - A\hat{\mathbf{z}}_{T-1|T-1})$ 
6:  $P_{T,T-1|T} = (I - K_TC)AP_{T-1|T-1}$ 
7:  $M_{T,T-1|T} = P_{T,T-1|T} + \hat{\mathbf{z}}_{T|T}\hat{\mathbf{z}}_{T-1|T}^\top$ 
8: // Start the iteration
9: for  $t = T-1 \rightarrow 1$  do
10:   $M_{t|T} = P_{t|T} + \hat{\mathbf{z}}_{t|T}\hat{\mathbf{z}}_{t|T}^\top$ 
11:   $J_{t-1} = P_{t-1|t-1}A^\top(P_{t|t-1})^{-1}$ 
12:   $P_{t,t-1|T} = P_{t|t}J_{t-1}^\top + J_t(P_{t+1,t|T} - AP_{t|t})J_{t-1}^\top$ 
13:   $M_{t,t-1|T} = P_{t,t-1|T} + \hat{\mathbf{z}}_{t|T}\hat{\mathbf{z}}_{t-1|T}^\top$ 
14:   $\hat{\mathbf{z}}_{t-1|T} = \hat{\mathbf{z}}_{t-1|t-1} + J_{t-1}(\hat{\mathbf{z}}_{t|T} - A\hat{\mathbf{z}}_{t-1|t-1})$ 
15:   $P_{t-1|T} = P_{t-1|t-1} + J_{t-1}(P_{t|T} - P_{t|t-1})J_{t-1}^\top$ 
16: end for

```

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

A3. Update Rules in EM - M step

The updated rules of $\Omega = \{A, C, Q, R, \xi, \Psi\}$ are shown in eq.(1) - eq.(6).

$$A^{(k+1)} = \left(\sum_{l=1}^N \sum_{t=2}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_t^l (\mathbf{z}_{t-1}^l)^\top | \tilde{\mathbf{Y}}^l] \right) \cdot \left(\sum_{l=1}^N \sum_{t=2}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_{t-1}^l (\mathbf{z}_{t-1}^l)^\top | \tilde{\mathbf{Y}}^l] \right)^{-1} \quad (1)$$

$$C^{(k+1)} = \left(\sum_{l=1}^N \sum_{t=1}^{\tilde{T}_l} \tilde{\mathbf{y}}_t^l (\mathbb{E}[\mathbf{z}_t^l | \tilde{\mathbf{Y}}^l])^\top \right) \cdot \left(\sum_{l=1}^N \sum_{t=1}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_t^l (\mathbf{z}_t^l)^\top | \tilde{\mathbf{Y}}^l] \right)^{-1} \quad (2)$$

$$R^{(k+1)} = \frac{1}{\sum_{l=1}^N \tilde{T}_l} \left(\sum_{l=1}^N \sum_{t=1}^{\tilde{T}_l} \tilde{\mathbf{y}}_t^l (\tilde{\mathbf{y}}_t^l)^\top - C^{(k+1)} \sum_{l=1}^N \sum_{t=1}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_t^l | \tilde{\mathbf{Y}}^l] (\tilde{\mathbf{y}}_t^l)^\top \right) \quad (3)$$

$$Q^{(k+1)} = \frac{1}{\sum_{l=1}^N \tilde{T}_l - N} \left(\sum_{l=1}^N \sum_{t=2}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_t^l (\mathbf{z}_t^l)^\top | \tilde{\mathbf{Y}}^l] - A^{(k+1)} \sum_{l=1}^N \sum_{t=2}^{\tilde{T}_l} \mathbb{E}[\mathbf{z}_{t-1}^l (\mathbf{z}_t^l)^\top | \tilde{\mathbf{Y}}^l] \right) \quad (4)$$

$$\xi^{(k+1)} = \sum_{l=1}^N \mathbb{E}[\mathbf{z}_1^l | \tilde{\mathbf{Y}}^l] \quad (5)$$

$$\Psi^{(k+1)} = \sum_{l=1}^N \mathbb{E}[\mathbf{z}_1^l (\mathbf{z}_1^l)^\top | \tilde{\mathbf{Y}}^l] - \sum_{l=1}^N \mathbb{E}[\mathbf{z}_1^l | \tilde{\mathbf{Y}}^l] (\mathbb{E}[\mathbf{z}_1^l | \tilde{\mathbf{Y}}^l])^\top \quad (6)$$

A4. Avg-MAPE Results of Different Initial Observation Lengths

Table 1: Avg-MAPE results of different initial observation lengths on a real-world clinical dataset.

Initial lengths	1	2	3	4	5	6	7	8	9	10
P_Mean	6.80	6.76	6.72	6.68	6.66	6.67	6.67	6.70	6.77	6.82
L_Mean	4.28	4.32	4.37	4.38	4.42	4.46	4.47	4.51	4.55	4.57
GP	3.87	3.68	3.63	3.58	3.57	3.57	3.54	3.54	3.54	3.55
MTGP	3.94	3.76	3.68	3.59	3.57	3.54	3.49	3.46	3.42	3.38
LDS	3.02	3.01	2.98	2.95	2.95	2.94	2.92	2.91	2.91	2.90
LDS+reGP	3.05	3.02	3.00	2.96	2.95	2.94	2.92	2.90	2.90	2.89
LDS+reMTGP	3.05	3.01	2.99	2.96	2.94	2.93	2.91	2.89	2.88	2.88
Initial lengths	11	12	13	14	15	16	17	18	19	20
P_Mean	6.91	6.97	7.11	7.23	7.39	7.50	7.58	7.70	7.74	7.78
L_Mean	4.58	4.58	4.63	4.70	4.76	4.79	4.80	4.81	4.80	4.79
GP	3.57	3.58	3.61	3.65	3.70	3.72	3.74	3.77	3.79	3.82
MTGP	3.34	3.32	3.36	3.37	3.36	3.33	3.30	3.29	3.29	3.27
LDS	2.91	2.91	2.94	2.97	2.99	3.02	3.00	3.04	3.09	3.12
LDS+reGP	2.90	2.90	2.93	2.95	2.96	3.00	2.99	3.01	3.05	3.07
LDS+reMTGP	2.88	2.87	2.90	2.92	2.92	2.95	2.92	2.94	2.99	3.01