Online ARIMA Algorithms for Time Series Prediction (supplemental file)

Chenghao Liu[1,3], Peilin Zhao[2], Steven C.H. Hoi[3], Jianling Sun[1]
[1]School of Computer Science and Technology, Zhejiang University, China
[2]Institute for Infocomm Research, A*STAR, Singapore
[3]School of Information Systems, Singapore Management University, Singapore
twinsken@zju.edu.cn, zhaop@i2r.a-star.edu.sg, chhoi@smu.edu.sg

November 25, 2015

This document provides the materials of our detailed proofs to supplement our main submission manuscript.

Proofs of Lemma in Main Manuscript

We re-iterate some necessary assumptions in our theoretical analysis.

1. The coefficients β_i satisfy that a q-th order difference equation with coefficients $|\beta_1|, |\beta_2|, \ldots, |\beta_q|$ is a stationary process; and
2. The noise terms are stochastically and independently generated, which satisfy $\mathbb{E}[|\epsilon_t|] < M_{max} < \infty$ and $\mathbb{E}[\ell_t(X_t, X_t - \epsilon_t)] < \infty$; and
3. The loss function ℓ_t is Lipshitz continuous for some Lipshitz constant $L > 0$; and
4. The coefficients α_i satisfy $|\alpha_i| < c$ for some $c \in \mathbb{R}$.

Lemma 1. *From any time series sequence satisfies our assumption, it holds that*

$$\min_{\gamma} \sum_{t=1}^{T} \ell_t^m(\gamma) \leq \sum_{t=1}^{T} f_t^m(\alpha^\star, \beta^\star).$$

Proof. Note that if we set $\gamma_i^\star = c_i(\alpha^\star, \beta^\star)$, we immediately get that

$$\sum_{t=1}^{T} \ell_t^m(\gamma^\star) = \sum_{t=1}^{T} f_t^m(\alpha^\star, \beta^\star).$$

Trivially, it always holds that

$$\min_{\gamma} \sum_{t=1}^{T} \ell_t^m(\gamma) \leq \sum_{t=1}^{T} \ell_t^m(\gamma^\star),$$

which completes the proof. □

Lemma 2. *Given assumption 1 that a q-th order difference equation with coefficients $|\beta_1| \cdots, |\beta_q|$ and observations $\{X_t\}_{t=-(q-1)}^T$ is a stationary process, $\lambda_1, \cdots, \lambda_q$ are the q roots of this AR characteristic equation. Let we set $\lambda_{min} = \{|\lambda_1|, \cdots, |\lambda_q|\}$, it holds that*

$$X_t \leq \lambda_{min}^t (X_0 + X_1 + \cdots + X_{-(q-1)}).$$

Lemma 3. *For any time series sequence satisfies our assumption, it holds that*

$$|\sum_{t=1}^{T} \mathbb{E}[f_t^\infty(\alpha^\star, \beta^\star)] - \sum_{t=1}^{T} \mathbb{E}[f_t^m(\alpha^\star, \beta^\star)]| = O(1),$$

if we choose $m = \log_{\lambda_{min}}((TLM_{max}q)^{-1})$.

Proof. For arbitrary t, we focus on the distance between $f_t^\infty(\alpha^\star, \beta^\star)$ and $f_t^m(\alpha^\star, \beta^\star)$ in expectation. First, for any $m \in \{0, -1, \cdots, -(1-q)\}$ we have that $\nabla^d X_t^m(\alpha^\star, \beta^\star) = \nabla^d X_t$ from definition, and hence

$$\begin{aligned}
&|\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^\infty(\alpha^\star, \beta^\star)| \\
&\leq |\nabla^d X_t - \nabla^d X_t^\infty(\alpha^\star, \beta^\star)| \\
&\leq |\nabla^d X_t - \nabla^d X_t^\infty(\alpha^\star, \beta^\star) - \epsilon_t| + |\epsilon_t|.
\end{aligned}$$

From assumption 2, we know $\mathbb{E}[|\epsilon_t|] < M_{max} < \infty$ for all t and $\mathbb{E}[|\nabla^d X_t - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star) - \epsilon_t|]$ decays exponentially as proven in lemma 4, and hence we have $|\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star)| < 2M_{max}$. Next, we show that $|\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star)|$ exponentially decreases as m increases linearly.

$$\begin{aligned}
&|\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star)| \\
&= |\sum_{i=1}^{q} \beta_i^\star (\nabla^d X_{t-i} - \nabla^d X_{t-i}^{m-i}(\alpha^\star, \beta^\star)) \\
&- \sum_{i=1}^{q} (\nabla^d X_{t-i} - \nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star))| \\
&= |\sum_{i=1}^{q} \beta_i^\star (\nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star) - \nabla^d X_{t-i}^{m-i}(\alpha^\star, \beta^\star))| \\
&\le \sum_{i=1}^{q} |\beta_i^\star| |\nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star) - \nabla^d X_{t-i}^{m-i}(\alpha^\star, \beta^\star)|.
\end{aligned}$$

We can easily find $|\nabla^d X_t^m(\alpha^\star, \beta^\star) - X_t^{\infty}(\alpha^\star, \beta^\star)|$ satisfy a qth order difference inequality with coefficients $|\beta_1^\star|, |\beta_2^\star|, \cdots, |\beta_q^\star|$. From assumption 1 and Lemma 2, we have

$$\begin{aligned}
&|\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star)| \\
&\le \lambda_{min}^m (|\nabla^d X_{t-m}^0(\alpha^\star, \beta^\star) - \nabla^d X_{t-m}^{\infty}(\alpha^\star, \beta^\star)| \\
&+ |\nabla^d X_{t-m-1}^{-1}(\alpha^\star, \beta^\star) - \nabla^d X_{t-m-1}^{\infty}(\alpha^\star, \beta^\star)| + \cdots \\
&+ |\nabla^d X_{t-m-(q-1)}^{-(q-1)}(\alpha^\star, \beta^\star) - \nabla^d X_{t-m-(q-1)}^{\infty}(\alpha^\star, \beta^\star)|) \\
&\le 2qM_{max}\lambda_{min}^m.
\end{aligned}$$

Since all $\nabla^i X_{t-1} (i \in \{1, 2, ..., d-1\})$ are fixed and hence it follows that

$$\begin{aligned}
&|X_t^m(\alpha^\star, \beta^\star) - X_t^{\infty}(\alpha^\star, \beta^\star)| \\
&= |\nabla^d X_t^m(\alpha^\star, \beta^\star) + \sum_{i=1}^{d-1} \nabla^i X_{t-1} \\
&\quad - \nabla^d X_t^{\infty}((\alpha^\star, \beta^\star) - \sum_{i=1}^{d-1} \nabla^i X_{t-1}| \\
&= |\nabla^d X_t^m(\alpha^\star, \beta^\star) - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star)| \le 2qM_{max}\lambda_{min}^m.
\end{aligned}$$

Recall that ℓ_t is assumed to be Lipshitz continuous for some constant $L > 0$ from our assumption , and hence it follows that

$$\begin{aligned}
&|\mathbb{E}[f_t^{\infty}(\alpha^\star, \beta^\star)] - \mathbb{E}[f_t^m(\alpha^\star, \beta^\star)]| \\
&= |\mathbb{E}[\ell_t(X_t, X_t^{\infty}(\alpha^\star, \beta^\star))] - \mathbb{E}[\ell_t(X_t, X_t^m(\alpha^\star, \beta^\star))]| \\
&\le \mathbb{E}[|\ell_t(X_t, X_t^{\infty}(\alpha^\star, \beta^\star)) - \ell_t(X_t, X_t^m(\alpha^\star, \beta^\star))|] \\
&\le L \cdot \mathbb{E}[|X_t^m(\alpha^\star, \beta^\star) - X_t^{\infty}(\alpha^\star, \beta^\star)|] \\
&\le L \cdot 2qM_{max}\lambda_{min}^m,
\end{aligned}$$

where the first inequality follows from Jensen's inequality. By summing the above for all t we get that

$$\begin{aligned}
&|\sum_{t=1}^{T} \mathbb{E}[f_t^{\infty}(\alpha^\star, \beta^\star)] - \sum_{t=1}^{T} \mathbb{E}[f_t^m(\alpha^\star, \beta^\star)]| \\
&\le TL \cdot 2qM_{max}\lambda_{min}^m
\end{aligned}$$

Finally, choosing $m = \log_{\lambda_{min}}((TLq)^{-1})$ yields

$$|\sum_{t=1}^{T} \mathbb{E}[f_t^{\infty}(\alpha^\star, \beta^\star)] - \sum_{t=1}^{T} \mathbb{E}[f_t^m(\alpha^\star, \beta^\star)]| = O(1).$$

□

Lemma 4. *For any time series sequence satisfies our assumption, it holds that*

$$|\sum_{t=1}^{T} \mathbb{E}[f_t^{\infty}(\alpha^\star, \beta^\star)] - \sum_{t=1}^{T} \mathbb{E}[f_t(\alpha^\star, \beta^\star)]| = O(1).$$

Proof. First, we prove that $\mathbb{E}[|\nabla^d X_t - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star) - \epsilon_t|]$ decays exponentially as the t increases linearly.

$$\begin{aligned}
&\mathbb{E}[|\nabla^d X_t - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star) - \epsilon_t|] \\
&= \mathbb{E}[|\sum_{i=1}^{k} \alpha_i^\star \nabla^d X_{t-i} + \sum_{i=1}^{q} \beta^\star \epsilon_{t-i} + \epsilon_t - \sum_{i=1}^{k} \alpha^\star \nabla^d X_{t-i} \\
&- \sum_{i=1}^{q} \beta^\star (\nabla^d X_{t-i} - \nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star)) - \epsilon_t|] \\
&= \mathbb{E}[|\sum_{i=1}^{q} \beta_i^\star (\nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star) - \nabla^d X_{t-i} + \epsilon_{t-i})|] \\
&\le \sum_{i=1}^{q} |\beta_i^\star| \mathbb{E}[|\nabla^d X_{t-i} - \nabla^d X_{t-i}^{\infty}(\alpha^\star, \beta^\star) - \epsilon_{t-i}|].
\end{aligned}$$

We could regard $\mathbb{E}[|\nabla^d X_t - \nabla^d X_t^{\infty}(\alpha^\star, \beta^\star) - \epsilon_t|]$ as a qth order difference inequality with coefficients

$|\beta_1^\star|, |\beta_2^\star|, \cdots, |\beta_q^\star|$. From assumption 1 and Lemma 2, we have

$$\begin{aligned}
&\mathbb{E}[|\nabla^d X_t - \nabla^d X_t^\infty(\alpha^\star, \beta^\star) - \epsilon_t|] \\
&\leq \lambda_{min}^t(\mathbb{E}[|\nabla^d X_0 - \nabla^d X_0^\infty(\alpha^\star, \beta^\star) - \epsilon_{-1}|] \\
&+\mathbb{E}[|\nabla^d X_{-1} - \nabla^d X_{-1}^\infty(\alpha^\star, \beta^\star) - \epsilon_{-1}|] + \cdots \\
&+\mathbb{E}[|\nabla^d X_{-(q-1)} - \nabla^d X_{-(q-1)}^\infty(\alpha^\star, \beta^\star) - \epsilon_{-q})|]) \\
&= \lambda_{min}^t \rho,
\end{aligned}$$

where ρ represents the right-hand side of λ_{min}^t for simplicity. According to Lemma 2, we know $|\lambda_{min}| < 1$, which ends the proof. Since all $\nabla^i X_{t-1} (i \in \{1, 2, ..., d-1\})$ are fixed and hence it follows that

$$\begin{aligned}
&\mathbb{E}[|X_t - X_t^\infty(\alpha^\star, \beta^\star) - \epsilon_t|] \\
&= \mathbb{E}[|\nabla^d X_t + \sum_{i=1}^{d-1} \nabla^i X_{t-1} - \nabla^d X_t^\infty((\alpha^\star, \beta^\star) \\
&- \sum_{i=1}^{d-1} \nabla^i X_{t-1} - \epsilon_t|] \\
&= \mathbb{E}[|\nabla^d X_t - \nabla^d X_t^\infty(\alpha^\star, \beta^\star) - \epsilon_t|] \leq \lambda_{min}^t \rho.
\end{aligned}$$

From Assumption 2, ϵ_t is stochastic and independent of $\epsilon_1, \ldots, \epsilon_{t-1}$ and hence the best prediction available at time t will cause a loss of at lest $\ell_t(X_t, X_t - \epsilon_t)$. The best ARIMA model coefficients (α', β') in hindsight are those have generated signal, which follows that

$$\sum_{t=1}^{T} f_t(\alpha', \beta') = \sum_{t=1}^{T} \ell_t(X_t, X_t - \epsilon_t).$$

Recall that ℓ_t is assumed to be Lipshitz continuous for some constant $L > 0$, and hence it follows that

$$\begin{aligned}
&|\mathbb{E}[f_t^\infty(\alpha^\star, \beta^\star)] - \mathbb{E}[f_t(\alpha^\star, \beta^\star)]| \\
&= |\mathbb{E}[\ell_t(X_t, X_t^\infty(\alpha^\star, \beta^\star))] - \mathbb{E}[\ell_t(X_t, X - t - \epsilon_t)]| \\
&= |\mathbb{E}[\ell_t(X_t, X_t^\infty(\alpha^\star, \beta^\star)) - \ell_t(X_t, X - t - \epsilon_t)]| \\
&\leq \mathbb{E}[|\ell_t(X_t, X_t^\infty(\alpha^\star, \beta^\star)) - \ell_t(X_t, X - t - \epsilon_t)]|] \\
&\leq L \cdot \mathbb{E}[|X_t - X_t^\infty(\alpha^\star, \beta^\star) - \epsilon_t|] \leq \rho \lambda_{min}^t.
\end{aligned}$$

Finally, summing over all iterations yields

$$|\sum_{t=1}^{T} \mathbb{E}[f_t^\infty(\alpha^\star, \beta^\star)] - \sum_{t=1}^{T} \mathbb{E}[f_t(\alpha^\star, \beta^\star)]| = O(1).$$

□