Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases

Laurent Praly

Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases

Laurent Praly*

December 26, 2016

Abstract

We establish convergence to zero of the solutions of

$$\dot{\tilde{\theta}}(t) \;=\; -\phi(t)\phi(t)^\top\tilde{\theta}(t) \qquad \text{or} \qquad \tilde{\theta}(t) \;=\; \tilde{\theta}(t-1) \;-\; \phi(t)\phi(t)^\top\tilde{\theta}(t-1)$$

under a possibly "vanishing persistent" excitation condition.

1 Continuous time case

Given a continuous vector function[1] $t \in \mathbb{R} \mapsto \phi(t) \in \mathbb{R}^n$, bounded[2] in norm by $\bar{\phi}$, we let $\mathfrak{T}(t,s)$ be the transition matrix associated with the non autonomous differential equation

$$\dot{\tilde{\theta}}(t) \;=\; -\phi(t)\phi(t)^\top\tilde{\theta}(t) \tag{1}$$

i.e. satisfying

$$\partial_1\mathfrak{T}(t,s) \;=\; -\phi(t)\phi(t)^\top\mathfrak{T}(t,s) \quad , \qquad \mathfrak{T}(s,s) \;=\; I \tag{2}$$

where ∂_1 denotes the partial derivative with respect to the first argument, t here. We are interested in sufficient conditions implying

$$\lim_{t\to+\infty} |\mathfrak{T}(t,s)| \;=\; \lim_{t\to+\infty} \sup_x \frac{|\mathfrak{T}(t,s)x|}{|x|} \;=\; 0 \;.$$

Let x be an arbitrary (constant) vector. We have

$$\partial_1 \left|\mathfrak{T}(t,s)x\right|^2 \;=\; -2|\phi(t)^\top\mathfrak{T}(t,s)x|^2$$

*MINES ParisTech, PSL Research University, CAS - Centre automatique et systèmes, 35 rue St Honoré 77300 Fontainebleau, France

[1] About the smoothness of ϕ we need only the differential equation to have solutions; so, since we have boundedness, measurability is sufficient. We need also to be able to change the order of integration.

[2] To obtain a bounded function ϕ a pre-processing from a "raw" function say Φ may be required. Often encountered pre-processsings are

$$\phi(t) \;=\; \gamma(t)\frac{\Phi(t)}{\sqrt{1+|\Phi(t)|^2}} \qquad \text{or} \qquad \phi(t) \;=\; \gamma(t)\frac{\Phi(t)}{\sqrt{r(t)}} \;, \quad \dot{r}(t) \;=\; -\lambda r(t) + |\Phi(t)|^2.$$

By defining the function $\hat{f}_s$ as

$$\hat{f}_s(t) \;=\; \phi(t)^\top \mathfrak{T}(t,s)x$$

and integrating, we get

$$|\mathfrak{T}(t,s)x|^2 \;=\; |x|^2 \;-\; 2\int_s^t \left|\hat{f}_s(r)\right|^2 dr \tag{3}$$

On another hand, by integrating (2), we get

$$\mathfrak{T}(t,s) \;=\; I \;-\; \int_s^t \phi(r)\phi(r)^\top \mathfrak{T}(r,s)dr$$

Incorporating this in the definition of $\hat{f}_s$ yields

$$\begin{aligned}
\hat{f}_s(t) &= \phi(t)^\top \mathfrak{T}(t,s)x \\
&= \phi(t)^\top \left(I \;-\; \int_s^t \phi(r)\phi(r)^\top \mathfrak{T}(r,s)dr\right) x \\
&= \phi(t)^\top x \;-\; \int_s^t \left(\phi(t)^\top \phi(r)\right)\left(\phi(r)^\top \mathfrak{T}(r,s)x\right) dr \\
&= f(t) \;-\; \int_s^t \left(\phi(t)^\top \phi(r)\right) \hat{f}_s(r)dr
\end{aligned} \tag{4}$$

where we have let

$$f(t) \;=\; \phi(t)^\top x \;.$$

It follows that we have

$$\begin{aligned}
f(u)^2 &\leq 2\hat{f}_s(u)^2 \;+\; 2\left(\int_s^u \phi(u)^\top \phi(r)\hat{f}_s(r)dr\right)^2 \\
&\leq 2\hat{f}_s(u)^2 \;+\; 2\left[\int_s^u \left(\phi(u)^\top \phi(v)\right)^2 dv\right]\left[\int_s^u \hat{f}_s(r)^2 dr\right] \\
&\leq 2\hat{f}_s(u)^2 \;+\; 2|\phi(u)|^2\left[\int_s^u |\phi(v)|^2 dv\right]\left[\int_s^u \hat{f}_s(r)^2 dr\right] \\
\int_s^t f(u)^2 du &\leq 2\int_s^t \hat{f}_s(u)^2 du \;+\; 2\int_s^t |\phi(u)|^2\left(\int_s^u |\phi(v)|^2 dv\right)\left(\int_s^u \hat{f}_s(r)^2 dr\right) du \\
&\leq 2\int_s^t \hat{f}_s(u)^2 du \;+\; 2\left(\int_s^t |\phi(u)|^2 du\right)\left(\int_s^t |\phi(v)|^2 dv\right)\left(\int_s^t \hat{f}_s(r)^2 dr\right) \\
&\leq 2\left[1 + \left(\int_s^t |\phi(u)|^2 du\right)^2\right]\left(\int_s^t \hat{f}_s(r)^2 dr\right)
\end{aligned} \tag{5}$$

So we get finally

$$
\begin{aligned}
|\mathfrak{T}(t,s)x|^2 &= |x|^2 - 2\int_s^t \left|\hat{f}_s(r)\right|^2 dr \\
&\leq |x|^2 - \frac{1}{1+\left(\int_s^t |\phi(u)|^2 du\right)^2} \int_s^t f(u)^2 du \\
&\leq |x|^2 - \frac{1}{1+\left(\int_s^t |\phi(u)|^2 du\right)^2} \int_s^t [\phi(u)^\top x]^2 du \\
&\leq x^\top \left(I - \frac{1}{1+\left(\int_s^t |\phi(u)|^2 du\right)^2} \int_s^t \phi(u)\phi(u)^\top du \right) x
\end{aligned}
$$

In other words we have simply

$$
\mathfrak{T}(t,s)^\top \mathfrak{T}(t,s) \leq I - \frac{\int_s^t \phi(u)\phi(u)^\top du}{1+\left(\int_s^t |\phi(u)|^2 du\right)^2}
$$

Remark 1

1. Up to (4), we have identities. The conservativeness we may have in this last inequality is only in the majoration obtained in (5).
2. An upperbound for $\mathfrak{T}(t,s)\mathfrak{T}(t,s)^\top$ can be obtained in a very similar way but starting from
$$
\partial_2 \mathfrak{T}(t,r) = \mathfrak{T}(t,r)\phi(r)\phi(r)^\top ,
$$
integrating in r backwards from t to s and using
$$
f_t(r) = x^\top \mathfrak{T}(t,r)\phi(r)
$$
instead of $\hat{f}_s(r)$.
3. The steps used up to (4) for the differential equation (1) can also be used for :
$$
\dot{\tilde{\theta}}(t) = -\psi(t)\phi(t)^\top \tilde{\theta}(t)
$$
as we have for example in the least square algorithm.

2 Discrete time case

Given a sequence of vectors $\phi(t) \in \mathbb{R}^n$ bounded[3] in norm by $\bar{\phi} \leq \sqrt{2}$, we let $\mathfrak{T}(t,s)$ be the transition matrix associated with the non autonomous discrete time system

$$
\tilde{\theta}(t) = \left(I - \phi(t)\phi(t)^\top\right)\tilde{\theta}(t-1)
$$

[3]See footnote 2

i.e. satisfying

$$\mathfrak{T}(t,s) \;=\; \left(I - \phi(t)\phi(t)^\top\right)\mathfrak{T}(t-1,s) \quad , \qquad \mathfrak{T}(s,s) \;=\; I$$

Our problem is to find sufficient conditions implying

$$\lim_{t\to+\infty} |\mathfrak{T}(t,s)| \;=\; \lim_{t\to+\infty} \sup_x \frac{|\mathfrak{T}(t,s)x|}{|x|} \;=\; 0\ .$$

Let x be an arbitrary unit vector. We have

$$\mathfrak{T}(t,s)x = \mathfrak{T}(t-1,s)x \;-\; \phi(t)\phi(t)^\top\mathfrak{T}(t-1,s)x$$

With denoting, for $t \geq s+1$,

$$\hat{f}_s(t) \;=\; \phi(t)^\top\mathfrak{T}(t-1,s)x$$

we get

$$\begin{aligned}
|\mathfrak{T}(t,s)x|^2 &= \left(\mathfrak{T}(t-1,s)x - \phi(t)\hat{f}_s(t)\right)^\top\left(\mathfrak{T}(t-1,s)x - \phi(t)\hat{f}_s(t)\right)\\
&= |\mathfrak{T}(t-1,s)x|^2 - 2x^\top\mathfrak{T}(t-1,s)^\top\phi(t)\hat{f}_s(t)\phi(t)^\top\phi(t)\hat{f}_s(t)^2\\
&= |\mathfrak{T}(t-1,s)x|^2 - \left(2-\phi(t)^\top\phi(t)\right)\hat{f}_s(t)^2\\
&\vdots\\
|\mathfrak{T}(s+1,s)x|^2 &= |x|^2 - \left(2-\phi(s+1)^\top\phi(s+1)\right)\hat{f}_s(s+1)^2
\end{aligned}$$

So summation gives

$$|\mathfrak{T}(t,s)x|^2 \;=\; |x|^2 \;-\; \sum_{r=s+1}^{t}\left(2-\phi(r)^\top\phi(r)\right)\hat{f}_s(r)^2 \tag{6}$$

On another hand, we have

$$\begin{aligned}
\mathfrak{T}(t,s) &= \mathfrak{T}(t-1,s) \;-\; \phi(t)\phi(t)^\top\mathfrak{T}(t-1,s)\\
\mathfrak{T}(t-1,s) &= \mathfrak{T}(t-2,s) \;-\; \phi(t-1)\phi(t-1)^\top\mathfrak{T}(t-2,s)\\
&\vdots\\
\mathfrak{T}(s+1,s) &= I \;-\; \phi(s+1)\phi(s+1)^\top\mathfrak{T}(s,s)
\end{aligned}$$

So again summation gives

$$\mathfrak{T}(t,s) \;=\; I \;-\; \sum_{r=s+1}^{t}\phi(r)\phi(r)^\top\mathfrak{T}(r-1,s)$$

Incorporating this in the expression of $\hat{f}_s$ yields, for $t \geq s+2$,

$$\begin{aligned}
\hat{f}_s(t) &= \phi(t)^\top\mathfrak{T}(t-1,s)x\\
&= \phi(t)^\top\left(I \;-\; \sum_{r=s+1}^{t-1}\phi(r)\phi(r)^\top\mathfrak{T}(r-1,s)\right)x\\
&= \phi(t)^\top x \;-\; \sum_{r=s+1}^{t-1}\left(\phi(t)^\top\phi(r)\right)\left(\phi(r)^\top\mathfrak{T}(r-1,s)x\right)\\
&= f(t) \;-\; \sum_{r=s+1}^{t-1}\phi(t)^\top\phi(r)\hat{f}_s(r) \qquad \forall t \geq s+2\ ,
\end{aligned}$$

where we have let

$$f(t) \;=\; \phi(t)^\top x \tag{7}$$

We have also

$$\hat{f}_s(s+1) \;=\; \phi(s+1)^\top \mathfrak{T}(s,s)x \;=\; f(s+1)$$

With the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} f(u)^2 &\leq \left(1+\sum_{r=s+1}^{u-1}[\phi(u)^\top\phi(r)]^2\right)\left(\sum_{r=s+1}^{u}\hat{f}_s(r)^2\right) \\ &\leq \left(1+|\phi(u)|^2\sum_{v=s+1}^{u-1}|\phi(v)|^2\right)\left(\sum_{r=s+1}^{u}\hat{f}_s(r)^2\right) \qquad \forall u \geq s+2 \\ f(s+1)^2 &= \hat{f}_s(s+1)^2 \end{aligned}$$

and therefore

$$\begin{aligned} &\sum_{u=s+1}^{t} f(u)^2 \\ &\quad\leq \hat{f}_s(s+1)^2 + \sum_{u=s+2}^{t}\left(1+|\phi(u)|^2\sum_{v=s+1}^{u-1}|\phi(v)|^2\right)\left(\sum_{r=s+1}^{u}\hat{f}_s(r)^2\right) \\ &\quad\leq \left(1+\sum_{u=s+2}^{t}\left(1+|\phi(u)|^2\sum_{v=s+1}^{u-1}|\phi(v)|^2\right)\right)\hat{f}_s(s+1)^2 \\ &\qquad\qquad + \sum_{r=s+2}^{t}\left(\sum_{u=r}^{t}\left(1+|\phi(u)|^2\sum_{v=s+1}^{u-1}|\phi(v)|^2\right)\right)\hat{f}_s(r)^2 \\ &\quad\leq \left((t-s)+\left(\sum_{u=s+2}^{t}|\phi(u)|^2\right)\left(\sum_{v=s+1}^{t-1}|\phi(v)|^2\right)\right)\hat{f}_s(s+1)^2 \\ &\qquad\qquad + \sum_{r=s+2}^{t}\left((t-r+1)+\left(\sum_{u=r}^{t}|\phi(u)|^2\right)\left(\sum_{v=s+1}^{t-1}|\phi(v)|^2\right)\right)\hat{f}_s(r)^2 \\ &\quad\leq \left((t-s)+\left(\sum_{u=s+2}^{t}|\phi(u)|^2\right)\left(\sum_{v=s+1}^{t-1}|\phi(v)|^2\right)\right)\left(\sum_{r=s+1}^{t}\hat{f}_s(r)^2\right) \\ &\quad\leq \left((t-s)+\left(\sum_{u=s+1}^{t}|\phi(u)|^2\right)^2\right)\left(\sum_{r=s+1}^{t}\hat{f}_s(r)^2\right) \end{aligned} \tag{8}$$

With the definition (7) of $f(u)$ this inequality is

$$x^\top\left(\sum_{u=s+1}^{t}\phi(u)\phi(u)^\top\right)x \;\leq\; \left((t-s)+\left(\sum_{u=s+1}^{t}|\phi(u)|^2\right)^2\right)\left(\sum_{r=s+1}^{t}\hat{f}_s(r)^2\right)$$

With (6), it allows us to obtain the following upperbound for $|\mathfrak{T}(t,s)x|^2$

$$\begin{aligned}|\mathfrak{T}(t,s)x|^2 &\leq |x|^2 - \min_{r\in\{s+1,\dots,t\}} \left\{2-\phi(r)^\top\phi(r)\right\} \left(\sum_{r=s+1}^{t} \hat{f}_s(r)^2\right) \\ &\leq |x|^2 - \frac{\min\limits_{r\in\{s+1,\dots,t\}} \left\{2-\phi(r)^\top\phi(r)\right\}}{\left((t-s)+\left(\sum\limits_{u=s+1}^{t}|\phi(u)|^2\right)^2\right)} x^\top \left(\sum_{u=s+1}^{t}\phi(u)\phi(u)^\top\right) x\end{aligned}$$

In other words we have

$$\mathfrak{T}(t,s)^\top\mathfrak{T}(t,s) \leq I - \frac{\min\limits_{r\in\{s+1,\dots,t\}} \left\{2-\phi(r)^\top\phi(r)\right\}}{\left((t-s)+\left(\sum\limits_{u=s+1}^{t}|\phi(u)|^2\right)^2\right)} \left(\sum_{u=s+1}^{t}\phi(u)\phi(u)^\top\right) .$$

Remark 2 The same final remarks as for the continuous time case can be done here. In particular about the majoration (8), a less conservative bound is obtained in the proof of [1, Theorem 4.5] (or of [2, Theorem 2.2, 1st column, p. 2052], for the case where the preprocessing mentioned at the beginning of this section is

$$\phi(t) = \frac{\Phi(t)}{\sqrt{r(t)}} \quad , \qquad r(t) = r(t-1) + |\Phi(t)|^2$$

and the above analysis in carried out exploiting the assumption that the sequence r goes to infinity in some specific way.

References

[1] H.-F. Chen, L. Guo. Identification and stochastic adaptive control. Springer Science+Business, LLC 1991.

[2] W. Ren, P. Kumar. Stochastic adaptive prediction and model reference control. IEEE Transactions on Automatic Control, Vol. 39. No. 10. october 1994 2047

3 Convergence for the discrete and continuous time case

Let T and t_i be strictly positive real numbers satisfying

$$t_{i+1} - t_i \leq T$$

and let $t_0 = 0$. For any t, there exists $\tau(t)$ such that t is between $t_{\tau(t)}$ and $t_{1+\tau(t)}$. We have seen that both in the discrete and continuous time case, there exist a real number π_i in $[0,1]$ satisfying

$$|\mathfrak{T}(t_i,t_{i-1})| \leq 1-\pi_i$$

Specifically, assuming $t \mapsto |\phi(t)|$ is bounded by $\bar{\phi}$, and since $\sqrt{1-a} \leq 1 - \frac{a}{2}$

– in the continuous time case, we have

$$\pi_i \;=\; \frac{1}{2(1+T^2\bar{\phi}^4)}\lambda_{\min}\left(\int_{t_{i-1}}^{t_i}\phi(u)\phi(u)^\top du\right)$$

– in the discrete time case, with $\bar{\phi}^2$ smaller than 2, we have

$$\pi_i \;=\; \frac{2-\bar{\phi}^2}{2T(1+T\bar{\phi}^4)}\lambda_{\min}\left(\sum_{1+t_{i-1}}^{t_i}\phi(u)\phi(u)^\top\right)$$

So we have

$$\begin{aligned}
|\mathfrak{T}(t,0)| &= \left|\mathfrak{T}(t,t_{\tau(t)})\prod_{i=1}^{\tau(t)}\mathfrak{T}(t_i,t_{i-1})\right| \\
&\leq \left|\mathfrak{T}(t,t_{\tau(t)})\right|\prod_{i=1}^{\tau(t)}|\mathfrak{T}(t_i,t_{i-1})| \\
&\leq \prod_{i=1}^{\tau(t)}(1-\pi_i) \\
&\leq \exp\left(-\sum_{i=1}^{\tau(t)}\pi_i\right)
\end{aligned}$$

where to obtain the last inequality we have used the property

$$(1-x) \;\leq\; \exp(-x) \qquad \forall x\in\mathbb{R}\ .$$

We conclude that $|\mathfrak{T}(t,0)|$ tends to 0 if we can find T and the t_i's such that we get

$$\sum_{i=1}^{\infty}\pi_i \;=\; +\infty\ .$$