



**HAL**  
open science

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

Laurent Praly

► **To cite this version:**

Laurent Praly. Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases. [Research Report] PSL Research University; Mines ParisTech. 2017. hal-01423048v2

**HAL Id: hal-01423048**

**<https://hal.science/hal-01423048v2>**

Submitted on 15 Jan 2017 (v2), last revised 18 May 2018 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Laurent Praly\*

Started December 26, 2016, Revised January 15, 2016

Dépôt sur HAL : <https://hal.archives-ouvertes.fr/hal-01423048>  
 identifiant : hal-01423048  
 version : 2  
 date de soumission : 2016-01-15

## Abstract

We establish convergence to zero of the solutions of

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

under a possibly “vanishing persistent” excitation condition.

## 1 Continuous time case

Given a continuous vector function<sup>1</sup>  $t \in \mathbb{R} \mapsto \phi(t) \in \mathbb{R}^n$ , we let  $\mathfrak{T}(t, s)$  be the transition matrix associated with the non autonomous differential equation<sup>2</sup>

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

i.e. satisfying

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

where  $\partial_1$  denotes the partial derivative with respect to the first argument,  $t$  here. We are interested in sufficient conditions implying

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

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

<sup>1</sup>About the smoothness of  $\phi$  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.

<sup>2</sup>The function  $\phi$  is often obtained from a pre-processing from a “raw” function say  $\Phi$ . It may be

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

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

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

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

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

and integrating, we get

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

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

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

where we have let

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

It follows that we have

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

So we get finally

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

In other words we have simply

$$\mathfrak{T}(t, s)^\top \mathfrak{T}(t, s) \leq I - \frac{\int_s^t \phi(u)\phi(u)^\top du}{1 + \left( \int_s^t |\phi(u)|^2 du \right)^2} \quad (6)$$

### 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

$$\hat{f}_t(r) = x^\top \mathfrak{T}(t, r)\phi(r) \quad , \quad \hat{f}_t(t) = x^\top \phi(t)$$

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.

4. With adapting the arguments used in an Appendix of the book by Marino and Tomei, it should be possible to extend the above result to the system

$$\dot{\eta}(t) = A\eta(t) + B\phi(t)^\top \tilde{\theta}(t) \quad , \quad \dot{\tilde{\theta}}(t) = -\phi(t)C\eta(t)$$

where the triple  $(B, A, C)$  is strictly positive real.

## 2 Discrete time case

Given a sequence of vectors  $\phi(t) \in \mathbb{R}^n$  bounded<sup>3</sup> 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

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

i.e. satisfying

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

Our problem is to find sufficient conditions implying

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

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  $t \geq s+1$ ,

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

we get

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

So summation gives

$$|\mathfrak{T}(t, s)x|^2 = |x|^2 - \sum_{r=s+1}^t (2 - \phi(r)^\top \phi(r)) \hat{f}_s(r)^2 \tag{7}$$

On another hand, we have

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

So again summation gives

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

---

<sup>3</sup>See footnote 2

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

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

where we have let

$$f(t) = \phi(t)^\top x \tag{8}$$

We have also

$$\hat{f}_s(s+1) = f(s+1)$$

With the Cauchy-Schwarz inequality we obtain

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

and therefore

$$\begin{aligned}
& \sum_{u=s+1}^t f(u)^2 \\
& \leq \hat{f}_s(s+1)^2 + \sum_{u=s+2}^t \left( 1 + |\phi(u)|^2 \sum_{v=s+1}^{u-1} |\phi(v)|^2 \right) \left( \sum_{r=s+1}^u \hat{f}_s(r)^2 \right) \\
& \leq \left( 1 + \sum_{u=s+2}^t \left( 1 + |\phi(u)|^2 \sum_{v=s+1}^{u-1} |\phi(v)|^2 \right) \right) \hat{f}_s(s+1)^2 \\
& \quad + \sum_{r=s+2}^t \left( \sum_{u=r}^t \left( 1 + |\phi(u)|^2 \sum_{v=s+1}^{u-1} |\phi(v)|^2 \right) \right) \hat{f}_s(r)^2 \\
& \leq \left( (t-s) + \left( \sum_{u=s+2}^t |\phi(u)|^2 \right) \left( \sum_{v=s+1}^{t-1} |\phi(v)|^2 \right) \right) \hat{f}_s(s+1)^2 \\
& \quad + \sum_{r=s+2}^t \left( (t-r+1) + \left( \sum_{u=r}^t |\phi(u)|^2 \right) \left( \sum_{v=s+1}^{t-1} |\phi(v)|^2 \right) \right) \hat{f}_s(r)^2 \\
& \leq \left( (t-s) + \left( \sum_{u=s+2}^t |\phi(u)|^2 \right) \left( \sum_{v=s+1}^{t-1} |\phi(v)|^2 \right) \right) \left( \sum_{r=s+1}^t \hat{f}_s(r)^2 \right) \\
& \leq \left( (t-s) + \left( \sum_{u=s+1}^t |\phi(u)|^2 \right)^2 \right) \left( \sum_{r=s+1}^t \hat{f}_s(r)^2 \right) \tag{9}
\end{aligned}$$

With the definition (8) 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 (7), it allows us to obtain the following upperbound for  $|\mathfrak{I}(t, s)x|^2$

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

In other words we have

$$\mathfrak{I}(t, s)^\top \mathfrak{I}(t, s) \leq I - \frac{\min_{r \in \{s+1, \dots, t\}} \{2 - |\phi(r)|^2\}}{\left( (t-s) + \left( \sum_{u=s+1}^t |\phi(u)|^2 \right)^2 \right)} \left( \sum_{u=s+1}^t \phi(u)\phi(u)^\top \right). \tag{10}$$

**Remark 2** The same final remarks as for the continuous time case can be done here. In particular about the majoration (9), 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

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

and the above analysis is 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  $t_i$  be strictly positive real numbers going to  $+\infty$  with  $t_0 = 0$ . 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  $\pi_i$  in  $[0, 1]$  satisfying

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

Specifically, since  $\sqrt{1-a} \leq 1 - \frac{a}{2}$

– in the continuous time case, (6) gives

$$\pi_i \geq \frac{1}{2} \frac{\lambda_{\min} \left( \int_{t_{i-1}}^{t_i} \phi(u) \phi(u)^\top du \right)}{1 + \left( \int_{t_{i-1}}^{t_i} |\phi(u)|^2 du \right)^2} \quad (11)$$

With denoting  $\bar{\phi}_i = \text{esssup}_{t \in [t_{i-1}, t_i]} |\phi(t)|$ , a more conservative lowerbound for  $\pi_i$  is

$$\pi_i \geq \frac{1}{2} \frac{\lambda_{\min} \left( \int_{t_{i-1}}^{t_i} \phi(u) \phi(u)^\top du \right)}{1 + (t_i - t_{i-1})^2 \bar{\phi}_i^4} \quad (12)$$

– in the discrete time case, with  $\bar{\phi}^2$  smaller than 2, (10) gives

$$\pi_i \geq \frac{\min_{r \in \{1+t_{i-1}, \dots, t_i\}} \{2 - |\phi(r)|^2\}}{2 \left( (t_i - t_{i-1} - 1) + \left( \sum_{u=1+t_{i-1}}^{t_i} |\phi(u)|^2 \right)^2 \right)} \lambda_{\min} \left( \sum_{1+t_{i-1}}^{t_i} \phi(u) \phi(u)^\top \right) \quad (13)$$



or the more conservative lower bound

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

So we have

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

where to obtain the last inequality we have used the property

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

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

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

**Discussion :** To show the interest of this result, we compare it with the persistent excitation (spanning) condition. We do this here for the continuous time case only, but the same holds for the discrete time case.

The vector function  $\phi$  is said persistently exciting or spanning if there exist two strictly positive real numbers  $\varepsilon$  and  $T$  such that, for any  $t$ , the Gram matrix  $\int_t^{t+T} \phi(s)\phi(s)^\top ds$  on a time window with width  $T$  is above the level  $\varepsilon$ , i.e.

$$\lambda_{\min} \left( \int_t^{t+T} \phi(s)\phi(s)^\top ds \right) \geq \varepsilon \quad \forall t \geq 0 .$$

The condition of non summability of  $\pi_i$ , says that the above property can be relaxed in two ways: the level  $\varepsilon$  may decrease with  $t$ , the width  $T$  of the time window may increase with  $t$ .

Specifically, let  $T$  be fixed and let  $\varepsilon_i$  be the level reached by the Gram matrix on the  $i$ th time window  $[(i-1)T, iT]$ , i.e.

$$\varepsilon_i = \lambda_{\min} \left( \int_{(i-1)T}^{iT} \phi(s)\phi(s)^\top ds \right)$$

then, with (12),  $\pi_i$  is not summable if  $\varepsilon_i$  is not, i.e.

$$\sum_{i=1}^{\infty} \lambda_{\min} \left( \int_{(i-1)T}^{iT} \phi(s)\phi(s)^\top ds \right) = +\infty .$$

Now, let  $\varepsilon$  be fixed and, with  $t_0 = 0$ , let  $t_i$  be the smallest time such that the Gram matrix on the time window  $[t_{i-1}, t]$  is larger than the level  $\varepsilon$ , i.e.

$$t_i = \min_{t : \lambda_{\min} \left( \int_{t_{i-1}}^t \phi(s)\phi(s)^\top ds \right) \geq \varepsilon} t$$

then, with (12),  $\pi_i$  is not summable if

$$\sum_{i=1}^{\infty} \frac{1}{1 + (t_i - t_{i-1})^2} = +\infty .$$

## 4 History of the versions

### Modifications on January 15, 2017

Addition of the discussion on the relation between non summability of  $\pi_i$  and the persistent excitation (spanning) condition.

### Modifications on January 13, 2017

The lower bounds (11) and (13) have been changed to follow a suggestion of Romeo Ortega of giving less conservative lower bound for  $\pi_i$ . Before they were, with  $t_{i+1} - t_i \leq T$ ,

– for the continuous time case,

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

– for the discrete time case

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