# How to identify a linear model with the least square method. 

4TTV313U - introduction to simulation of dynamic process

2023

We have a physical system. We suppose the dynamic of that system follows a model that is linear to it's parameters. That means there exists a list of functions $\phi_{0}(t), \phi_{1}(t), \phi_{2}(t), \cdots, \phi_{k}(t)$ depending of the time $t$ and a list of reals $\alpha_{1}, \alpha_{2}, \cdots, \alpha_{k}$ such that we have the following equality :

$$
\begin{equation*}
\phi_{0}(t)=\alpha_{1} \times \phi_{1}(t)+\alpha_{2} \times \phi_{2}(t)+\alpha_{3} \times \phi_{3}(t)+\cdots+\alpha_{k} \times \phi_{k}(t) . \tag{1}
\end{equation*}
$$

Suppose now, we can make some experiments to obtain, for different times $t$ the values of $\phi_{0}(t)$, $\phi_{1}(t), \phi_{2}(t), \cdots, \phi_{k}(t)$.

In this document, we learn to compute, from our data, the parameters $\alpha=\left(\alpha_{1}, \alpha_{2}, \cdots, \alpha_{k}\right)$ that fit our data. We means, by fitting our data, that the error we make between our model and the experiments are the smallest possible according to some criteria we will define later.

For example, suppose that $w_{1}, w_{2}, w_{3}$ are linear velocities. A linear model according to the parameters $\alpha_{1}, \alpha_{2}$ and $\alpha_{3}$ could be :

$$
\sqrt{w_{3}} \cdot \exp t=\alpha_{1} \times t \cdot w_{1}^{2}+\alpha_{2} \times w_{1} \cdot w_{2}+\alpha_{3} \times \cos \left(3 \cdot w_{2} \cdot t\right)
$$

In that example, we can define $\phi_{0}(t)=\sqrt{w_{3}} \cdot \exp t, \phi_{1}(t)=t \cdot w_{1}^{2}, \phi_{2}(t)=w_{1} \cdot w_{2}$ and $\phi_{3}(t)=$ $\cos \left(3 . w_{2} . t\right)$ to obtain

$$
\phi_{0}(t)=\alpha_{1} \times \phi_{1}(t)+\alpha_{2} \times \phi_{2}(t)+\alpha_{3} \times \phi_{3}(t) .
$$

Suppose, we obtain the data set of the Table 1:

| id | No. exp. | date | $t$ | $w_{1}$ | $w_{2}$ | $w_{3}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | 1 | $01 / 01$ | 0 | 0.1 | 0.5 | 0.3658 |
| 1 | 1 | $01 / 01$ | 1 | 0.4 | 0.2 | 1.1444 |
| 2 | 1 | $01 / 01$ | 2 | 0.8 | 0.6 | 10.1610 |
| 3 | 1 | $01 / 01$ | 3 | 0.5 | 0.4 | 0.3676 |
| 4 | 2 | $02 / 01$ | 0 | 0.2 | 0.6 | 4.8159 |
| 5 | 2 | $02 / 01$ | 1 | 0.5 | 0.3 | 4.0255 |
| 6 | 2 | $02 / 01$ | 2 | 0.7 | 0.7 | 7.7276 |
| 7 | 2 | $02 / 01$ | 3 | 0.4 | 0.3 | 0.1518 |

Table 1: The measures of two experiments.
We can compute the different values of $\phi_{0}, \phi_{1}, \phi_{2}$ and $\phi_{3}$ for our data set :

| id | $\phi_{0}(t)=\sqrt{w_{3}} \exp t$ | $\phi_{1}(t)=t w_{1}^{2}$ | $\phi_{2}(t)=w_{1} w_{2}$ | $\phi_{3}(t)=\cos \left(3 w_{2} t\right)$ |
| :---: | :---: | :---: | :---: | :---: |
| 0 | 0.60481402 | 0. | 0.05 | 1. |
| 1 | 2.90792637 | 0.16 | 0.08 | 0.82533561 |
| 2 | 23.55359426 | 1.28 | 0.48 | -0.89675842 |
| 3 | 12.17786617 | 0.75 | 0.2 | -0.89675842 |
| 4 | 2.19451589 | 0. | 0.12 | 1. |
| 5 | 5.45386517 | 0.25 | 0.15 | 0.62160997 |
| 6 | 20.54051275 | 0.98 | 0.49 | -0.49026082 |
| 7 | 7.82563038 | 0.48 | 0.12 | -0.90407214 |

Table 2: The values of $\phi_{0}, \phi_{1}, \phi_{2}$ and $\phi_{3}$

We want now to estimate the better parameter $\alpha=\left(\begin{array}{c}\alpha_{1} \\ \alpha_{2} \\ \alpha_{3}\end{array}\right)$ in order to fit the model to the data set.

The idea is to write equation 1 by using the matrix formalism :

$$
\left(\phi_{0}(t)\right)=\left(\begin{array}{cccc}
\phi_{1}(t) & \phi_{2}(t) & \cdots & \phi_{k}(t)
\end{array}\right) \times\left(\begin{array}{c}
\alpha_{1} \\
\alpha_{1} \\
\cdots \\
\alpha_{k}
\end{array}\right) .
$$

In our example, we obtain :

$$
\left(\begin{array}{lll}
\sqrt{w_{3}} \cdot \exp t
\end{array}\right)=\left(\begin{array}{lll}
t . w_{1}^{2} & w_{1} \cdot w_{2} & \cos \left(3 . w_{2} . t\right)
\end{array}\right) \times\left(\begin{array}{l}
\alpha_{1} \\
\alpha_{1} \\
\alpha_{3}
\end{array}\right)
$$

If we define by $b(t)=\left(\sqrt{w_{3}} \cdot \exp t\right)$ and by $a(t)=\left(\begin{array}{lll}t \cdot w_{1}^{2} & w_{1} \cdot w_{2} & \cos \left(3 \cdot w_{2} \cdot t\right)\end{array}\right)$, we obtain the following equation:

$$
\begin{equation*}
b(t)=a(t) \times \alpha \tag{2}
\end{equation*}
$$

In fact $[b(t), a(t)]$ is one row of Table 2 . So we can concatenate each row of our data set to obtain the following matrix definitions :

$$
B=\left(\begin{array}{c}
b\left(t_{1}\right) \\
b\left(t_{2}\right) \\
b\left(t_{3}\right) \\
\cdots \\
b\left(t_{N}\right)
\end{array}\right) \quad A=\left(\begin{array}{c}
a\left(t_{1}\right) \\
a\left(t_{2}\right) \\
a\left(t_{3}\right) \\
\cdots \\
a\left(t_{N}\right)
\end{array}\right)
$$

For example, in the previous example, the matrix $B$ is the first column of the Table 2, and the matrix $A$ contains the 3 last columns :

$$
B=\left(\begin{array}{c}
0.60481402 \\
2.90792637 \\
23.55359426 \\
12.17786617 \\
2.19451589 \\
5.45386517 \\
20.54051275 \\
7.82563038
\end{array}\right) \quad A=\left(\begin{array}{ccc}
0 . & 0.05 & 1 . \\
0.16 & 0.08 & 0.82533561 \\
1.28 & 0.48 & -0.89675842 \\
0.75 & 0.2 & -0.89675842 \\
0 . & 0.12 & 1 . \\
0.25 & 0.15 & 0.62160997 \\
0.98 & 0.49 & -0.49026082 \\
0.48 & 0.12 & -0.90407214
\end{array}\right)
$$

Now, from equation 2 we obtain the following general identity :

$$
B=A \times \alpha
$$

In fact, the model is not perfect, and we made some errors during the experiments. So we have

$$
\begin{equation*}
B=A \times \alpha+\epsilon \tag{3}
\end{equation*}
$$

where $\epsilon$ is the error vector.
So now, the problem consists in finding the parameter $\alpha$ that minimizes the error vector $\epsilon$. More precisely, we want to find the parameter $\alpha$ that minimizes the square error $\|\epsilon\|^{2}=\sum_{i} \epsilon_{i}^{2}$.

We will prove, at the end of that document, the solution of that problem is :

$$
\begin{equation*}
\alpha=\left(A^{t} \cdot A\right)^{-1} \cdot A^{t} \cdot B \tag{4}
\end{equation*}
$$

where $A^{t}$ in the transposition of the matrix $A$.
For example, in our problem :

$$
A^{T}=\left(\begin{array}{cccccccc}
0 . & 0.16 & 1.28 & 0.75 & 0 . & 0.25 & 0.98 & 0.48 \\
0.05 & 0.08 & 0.48 & 0.2 & 0.12 & 0.15 & 0.49 & 0.12 \\
1 . & 0.825 & -0.896 & -0.896 & 1 . & 0.621 & -0.490 & -0.904
\end{array}\right)
$$

$$
A^{t} \cdot A=\left(\begin{array}{ccc}
3.4798 & 1.3525 & -2.44737363 \\
1.3525 & 0.5707 & -0.62924384 \\
2.44737363 & -0.62924384 & 5.73363126
\end{array}\right)
$$

and

$$
\left(A^{t} . A\right)^{-1}=\left(\begin{array}{ccc}
10.73850093 & -23.20290054 & 2.03725027 \\
-23.20290054 & 52.12843577 & -4.18315528 \\
2.03725027 & -4.18315528 & 0.58491517
\end{array}\right)
$$

We deduce that, the best parameter of our model is given by

$$
\alpha=\left(\begin{array}{lll}
9.90646281 & 21.72126782 & -0.46834528
\end{array}\right)^{t} .
$$

The error me make with the model is :

$$
\epsilon^{t}=(A \cdot \alpha-B)^{t}=\left(\begin{array}{llllllll}
0.012 & 0.028 & -0.027 & 0.016 & -0.056 & -0.010 & 0.040 & -0.040
\end{array}\right)
$$

and the square root error is : $\|\epsilon\|=0.0924$.

We will prove now that, the solution of our problem is :

$$
\alpha=\left(A^{t} \cdot A\right)^{-1} \cdot A^{t} \cdot B
$$

We want to minimize the squared error :

$$
L=\|\epsilon\|^{2}=\sum_{i} \epsilon_{i}^{2} .
$$

As $L \geq 0$ there exists a minimum. This minimum is reached when, the derivative of $L$ according to the parameters $\alpha_{i}$ are equal to 0 :

$$
\frac{\partial L}{\partial \alpha_{i}}=0 .
$$

Let $\nabla_{\alpha} L$ be the gradient of $L$ defined by :

$$
\nabla_{\alpha} L=\left(\begin{array}{c}
\frac{\partial L}{\partial \alpha_{1}} \\
\frac{\partial L}{\partial \alpha_{2}} \\
\cdots \\
\frac{\partial L}{\partial \alpha_{k}}
\end{array}\right)
$$

If $L$ is minimum then its gradient is null :

$$
\exists \alpha_{\min }, \forall \alpha, L(\alpha) \geq L\left(\alpha_{\min }\right) \quad \Longrightarrow \quad \nabla_{\alpha} L\left(\alpha_{\min }\right)=0 .
$$

We will now determine $\alpha_{\text {min }}$ the solution of our problem by computing $\nabla_{\alpha} L$ :

$$
\begin{aligned}
\nabla_{\alpha} L & =\nabla_{\alpha} \epsilon^{t} \epsilon=\nabla_{\alpha}(B-A \alpha)^{t}(B-A \alpha)=\nabla_{\alpha}\left(B^{t}-\alpha^{t} A^{t}\right)(B-A \alpha) \\
& =\nabla_{\alpha} B^{t} B-\nabla_{\alpha} B^{t} A \alpha-\nabla_{\alpha} \alpha^{t} A^{t} B+\nabla_{\alpha} \alpha^{t} A^{t} A \alpha
\end{aligned}
$$

As $B^{t} B$ is a constant, we have $\nabla_{\alpha} B^{t} B=0$.
Let's define the vector $e_{i}$ in $\mathbb{R}^{k}$ as a vector where each row contains a zero, except for row $i$, which contains a 1. Now, we can compute $\frac{\partial B^{t} A \alpha}{\partial \alpha_{i}}=B^{t} A e_{i}$ and deduce that: $\nabla_{\alpha} B^{t} A \alpha=\left(B^{t} A\right)^{t}=A^{t} B$. Similarly, we can express, $\frac{\partial \alpha^{t} A^{t} B}{\partial \alpha_{i}}=e_{i}^{t} A^{t} B$ and obtain : $\nabla_{\alpha} \alpha^{t} A^{t} B=A^{t} B$.
Finally, by using the Leibniz rule,

$$
\nabla_{\alpha} \alpha^{t} A^{t} A \alpha=\nabla_{\beta} \beta^{t} A^{t} A \alpha+\left.\nabla_{\beta} \alpha^{t} A^{t} A \beta\right|_{\beta \rightarrow \alpha}=A^{t} A \alpha+\left(\alpha^{t} A^{t} A\right)^{t}=2 A^{t} A \alpha,
$$

and we obtain that

$$
\nabla_{\alpha} L\left(\alpha_{\min }\right)=-2 A^{t} B+2 A^{t} A \alpha_{\min }=0
$$

We conclude that :

$$
\alpha_{\min }=\left(A^{t} A\right)^{-1} A^{t} B
$$

