

Effectuer une régression linéaire

I Méthode pour faire une régression linéaire

Nous introduisons la notion de régression linéaire à travers l'exemple du TP 5. On étudie un montage multivibrateur astable, et on s'intéresse en particulier à la période des oscillations du signal produit par le circuit.

1 – Côté théorie : quelle est la loi que l'on veut vérifier ?

L'étude théorique prédit la loi suivante pour la période T des oscillations :

$$T = \frac{4RR_2C}{R_1}. \quad (1)$$

On peut choisir de vérifier plusieurs choses. Par exemple :

- Vérifier que la période T est bien proportionnelle à C . (\rightarrow c'est ce que vous allez faire en TP)
- Vérifier que la période T est bien proportionnelle à R_2 . (\rightarrow c'est l'exemple que l'on développe ici)

►étape 1 On met la loi à vérifier sous la forme :

$$y = a_{\text{théo}}x + b_{\text{théo}}. \quad (2)$$

Ici on pose $y = T$, $x = R_2$ (ce sont les paramètres qui varient).

La loi (1) s'écrit alors $y = \frac{4RC}{R_1} \times x$.

On identifie alors $a_{\text{théo}}$ et $b_{\text{théo}}$: ici on a $a_{\text{théo}} = \frac{4RC}{R_1}$ et $b_{\text{théo}} = 0$.

►étape 2 On fait l'application numérique pour $a_{\text{théo}}$ et $b_{\text{théo}}$, avec incertitudes si besoin.

★ Ici on choisit $R = R_1 = 10 \text{ k}\Omega$, $C = 100 \text{ nF}$.

Donc $a_{\text{théo}} = 400 \text{ nF}$.

On prend une incertitude de 2% pour R et R_1 , et de même sur C , et on utilise la formule de propagation des incertitudes pour obtenir $\Delta a_{\text{théo}} = 0.2 \text{ nF}$.

On a donc finalement $a_{\text{théo}} = (4.0 \pm 0.2) \times 10^2 \text{ nF}$.

★ On a de plus $b_{\text{théo}} = 0 \text{ s}$.

2 – Côté expériences : comment vérifier la loi ?

►étape 3 On réalise des mesures expérimentales de y (ici la période T , à l'aide de l'oscilloscope) pour plusieurs valeurs de x (donc ici de R_2 qui est une résistance variable).

On entre ces valeurs sous le logiciel Regressi (par exemple) dans un tableau à deux entrées.

On peut entrer également l'incertitude sur chaque mesure expérimentale (si on la connaît), en cliquant sur "incertitude" en haut. Sous Regressi il faut entrer les *incertitudes types*.

Rappel :

Les incertitude types sont en général associées à un intervalle de confiance de 68% (la probabilité que la valeur vraie soit comprise dans l'intervalle donné par l'incertitude est de 68%), alors que les incertitudes élargies sont associées à un intervalle de confiance de 95%.

Il faut entrer les incertitudes types dans le logiciel, or dans les notices ou sur les composants

utilisés ce sont des incertitudes élargies qui sont données.

On rappelle le lien entre les deux : incertitude type = incertitude élargie / 2.

Ici on peut prendre des *incertitudes élargies* de 3% pour R_2 et de 2% pour la mesure de T à l'oscilloscope. Puis on obtient l'incertitude type en divisant l'incertitude élargie par 2. Dans notre cas elles sont donc de 1.5% sur R_2 et de 1% sur T .

→ On demande ensuite au logiciel de tracer la variable y (donc ici les valeurs de T) en fonction de la valeur de x (donc ici les valeurs de R_2).

- étape 4 On confirme d'abord *visuellement* que les points forment bien approximativement une droite.
- Si c'est le cas, on peut déjà **conclure que le modèle linéaire ou affine est le bon**, et on passe à l'étape suivante.
 - Si ce n'est pas le cas on arrête ici et on dit que le modèle linéaire ou affine n'est pas en accord avec les mesures.

- étape 5 On demande au logiciel d'effectuer une modélisation, affine dans notre cas ($y = ax + b$).

Le logiciel trouve alors les paramètres a et b qui sont tels que l'écart entre la droite $y = ax + b$ et les points expérimentaux soit minimal (cet écart étant défini comme $C(a, b) = \sum_i [y_i - (ax_i + b)]^2$, avec (x_i, y_i) les points mesurés).

Il retourne les valeurs de a et de b , assorties d'une incertitude élargie (même si on n'a pas entré d'incertitude sur les valeurs expérimentales).

Il s'agit donc des **valeurs expérimentales** de a et b , que l'on va noter $a_{\text{exp}} \pm \Delta a_{\text{exp}}$ et $b_{\text{exp}} \pm \Delta b_{\text{exp}}$.

Ici on trouve, avec le jeu de données utilisé (voir capture d'écran plus bas) : $a_{\text{exp}} = (404 \pm 20)$ S.I. et $b_{\text{exp}} = (111 \pm 117) \times 10^{-6}$ S.I.. On sait par ailleurs que l'unité de a est le farad et celle de b les secondes.

On écrira donc : $a_{\text{exp}} = (4.0 \pm 0.2 \times 10^2) \text{ nF}$ et $b_{\text{exp}} = (1 \pm 2) \times 10^{-4} \text{ s}$.

3 – Comparaison entre théorie et expérience

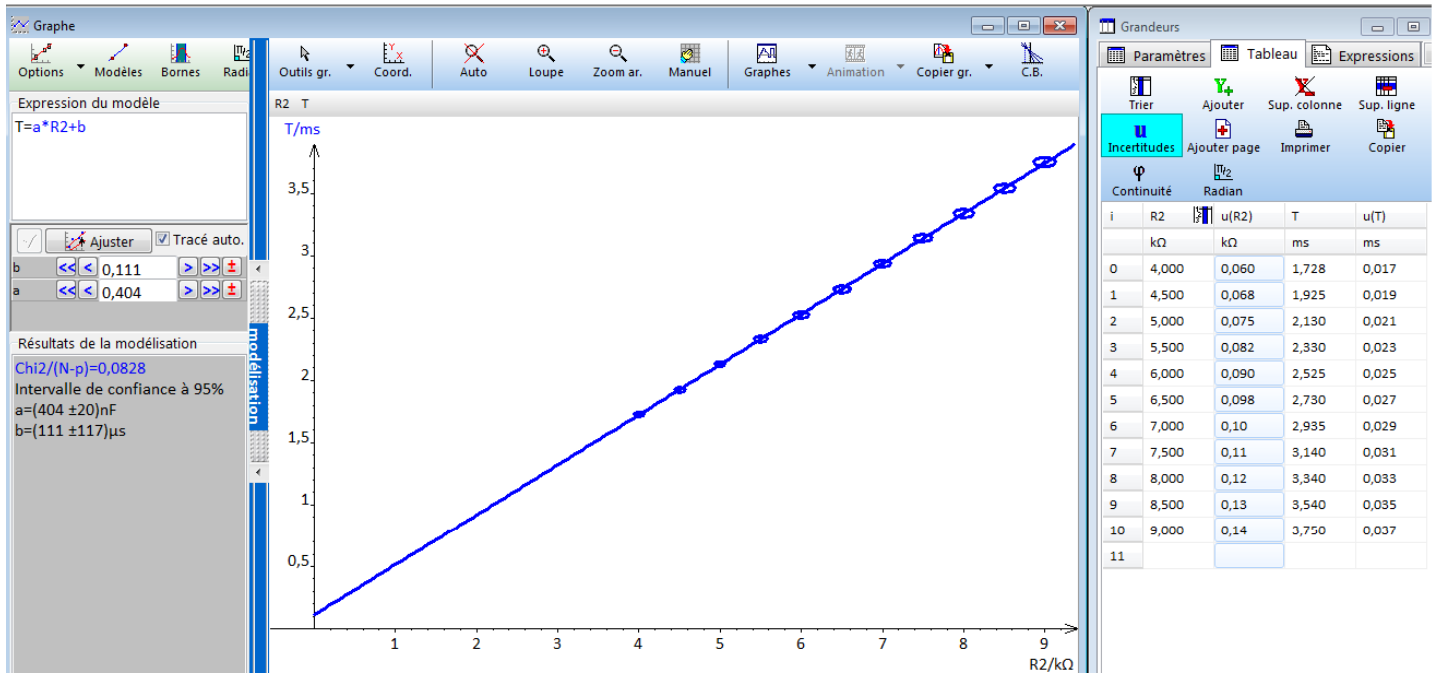
- étape 6 On conclut en comparant théorie et expérience.

On voit dans le cas présent que les intervalles $a_{\text{exp}} \pm \Delta a_{\text{exp}}$ et $a_{\text{théo}} \pm \Delta a_{\text{théo}}$ ont des valeurs en commun. De même, $b_{\text{théo}}$ est inclus dans l'intervalle $b_{\text{exp}} \pm \Delta b_{\text{exp}}$.

On peut donc conclure que *l'expérience est en accord avec la théorie*.

Remarque : Si il n'y a pas accord entre expérience et théorie *de façon grossière*, il faut dire qu'on a probablement fait une erreur quelque part, ou qu'un composant est défectueux. S'il n'y a pas accord, mais que les valeurs sont tout de même proches, il faut dire que : soit on a fait une erreur, soit le modèle théorique ne prend pas en compte tous les aspects expérimentaux. (Par exemple dans le cas présent, pour des petites périodes on voit à l'oscilloscope que la sortie de l'hystérésis met du temps à basculer entre $\pm V_{\text{sat}}$, alors que dans le modèle on suppose ceci instantané.)

Remarque : Si $b_{\text{théo}} = 0$ et que cette valeur nulle est comprise dans l'intervalle $b_{\text{exp}} \pm \Delta b_{\text{exp}}$, alors il est possible de faire une régression avec une loi linéaire (du type $y = ax$). Il faut alors vérifier visuellement qu'il n'y a pas d'ordonnée à l'origine. Puis voir si la nouvelle valeur de a est compatible avec la valeur théorique.



Capture d'écran du logiciel Regressi, montrant les valeurs utilisées.

Bien penser à :

- dans les options de modélisation, sélectionner "prendre en compte les incertitudes",
- entrer des incertitudes types pour les données (donc des incertitudes élargies divisées par 2).

Le logiciel donne alors a_{exp} et b_{exp} , ainsi que les incertitudes élargies (à 95%) Δa_{exp} et Δb_{exp} .

Remarque sur l'exemple choisi : Si on prend encore plus de points expérimentaux (tous les 0.5 kΩ de 1.5 à 9 kΩ), ceci permet de réduire les incertitudes. On obtient alors $a_{\text{exp}} = (4.06 \pm 0.08) \text{ nF}$, ce qui est plus précis que tout à l'heure et encore compatible avec la valeur théorique. En revanche, on a $b_{\text{exp}} = (9 \pm 3) \times 10 \mu\text{s}$, ce qui n'est pas compatible avec la valeur théorique nulle. Et on constate bien visuellement qu'il y a une ordonnée à l'origine non nulle !

En fait, on se rend compte en zoomant à l'oscilloscope que la sortie de l'hystérésis mets un certain temps à basculer de $+V_{\text{sat}}$ à $-V_{\text{sat}}$, temps de l'ordre de $40 \mu\text{s}$. On comprend donc que même si les paliers sont de longueur nulle, les temps de basculement vont impliquer une période non nulle de l'ordre de $2 \times 40 \mu\text{s}$. Ceci explique la présence de l'ordonnée à l'origine.

On voit donc ici que cette ordonnée à l'origine qui peut paraître embêtante est en réalité bien expliquée par un défaut de l'ALI (son temps de monté non infini) non pris en compte dans le modèle théorique.

II Cas où la loi à vérifier n'est pas linéaire

Il est possible que la loi à vérifier ne soit pas linéaire.

Par exemple, supposons que l'on veuille vérifier que la vitesse du son dans l'air est bien proportionnelle à la racine carrée de la température : $c_s = \alpha\sqrt{T}$, avec α une constante :

- On effectue des mesures de la vitesse du son c_s pour différentes températures T . On les rentre dans Régressi ou dans la calculatrice.
- On essaie de se ramener à une loi linéaire. La raison est que le cerveau humain est doué pour voir si des points sont alignés, mais pas du tout pour estimer s'ils suivent une loi en racine carrée, en $1/x$, en exponentiel ou autre.

Ici par exemple on peut décider de tracer c_s en fonction de \sqrt{T} . En effet, si on pose $y = c_s$ et $x = \sqrt{T}$, alors la relation à vérifiée devient $y = \alpha x$, et on peut employer la méthode précédente.

Concrètement, sous Régressi on crée une nouvelle variable dans l'onglet expression : $x = T^{0.5}$. Le logiciel calcule alors x pour toutes les valeurs de T .

On trace alors c_s en fonction de x , et on fait la régression linéaire sur ce graphique.

Il se peut dans certains cas qu'on ne puisse pas se ramener à une loi affine ou linéaire. Dans ce cas on utilise les autres modèles disponibles dans le logiciel, tout en regardant à l'œil si le modèle est vraisemblable.

Exemples :

- La loi théorique à vérifier est celle donnant la période T en fonction de R_1 : $T = \frac{4RR_2C}{R_1}$.
On mesure T pour différentes valeurs de R_1 .
Que faut-il poser pour mettre la loi à vérifier sous la forme $y = a_{\text{théo}}x + b_{\text{théo}}$?
- On souhaite vérifier la loi de Descartes de la réfraction : $n_1 \sin i_1 = n_2 \sin i_2$.
Expérimentalement, pour plusieurs angles d'incidence i_1 , on mesure l'angle réfracté i_2 .
Que faut-il poser pour mettre la loi à vérifier sous la forme $y = a_{\text{théo}}x + b_{\text{théo}}$?

Exemple :

On reprend l'exemple précédent de la température. On a le tableau de valeurs expérimentales suivant :

c_s (m/s)	310	325	331	345	355
T (K)	240	260	280	300	320

À l'aide de votre calculatrice, vérifier si ces données expérimentales sont en accord avec la loi théorique $c_s = 20.05\sqrt{T}$ (T en K, c_s en m/s).