

# Modélisation dans Regressi

Jean-Michel Millet

## 1 Utilisation

### 1.1 Introduction

Regressi effectue la régression par deux méthodes

1. par défaut, la méthode des moindres carrés classique et donc sans prise en compte des incertitudes. On suppose que l'incertitude sur l'abscisse est négligeable et que l'incertitude sur l'ordonnée est la même pour tous les points. La technique de détermination des incertitudes sur les paramètres repose sur l'assimilation entre l'écart modèle-données et l'incertitude sur l'ordonnée.
2. si les incertitudes sont définies et que l'utilisateur a coché « méthode du  $\chi^2$  », la dite méthode, les incertitudes sur  $x$  et  $y$  étant prises en compte.

Le test du  $\chi^2$  suppose que les incertitudes soient des incertitudes-type, et on élargit l'incertitude par un coefficient de Student de paramètre (nombre de mesures - nombre de paramètres), ce qui suppose que la statistique sur les paramètres est bien gaussienne.

**Note :** modélisation s'entend au sens d'ajustement des paramètres d'une fonction donnée (fit en anglais). J'utilise modélisation car, pour moi, le choix de la fonction n'est pas arbitraire mais résulte de la modélisation du dispositif étudié.

#### 1.1.1 Incertitudes des données

Les incertitudes définies pour les grandeurs expérimentales devraient donc être des incertitudes-type, notées  $u$ , de manière à ce que la loi de propagation puisse s'appliquer. Les options de tracé des ellipses le supposent également.

La propagation des incertitudes (incertitudes composées) se fait par addition des variances.

Exemple :  $U = R \cdot I$ ,  $\frac{u(U)^2}{U^2} = \frac{u(R)^2}{R^2} + \frac{u(I)^2}{I^2}$

#### Rappel : expression de l'incertitude type

- Pour une mesure avec des graduations de longueur  $pas$ , l'incertitude-type est :  $pas/\sqrt{12}$
- Pour un instrument avec une *erreur* de justesse maximale donnée :  $erreur/\sqrt{3}$
- Pour un appareil de précision  $p$ , l'incertitude-type est  $p/\sqrt{3}$
- Pour un appareil type voltmètre avec une précision de ( $pc$  % de lecture +  $N \times$  chiffre le moins significatif), l'incertitude sur  $x$  est donnée par  $(x \cdot pc + N \cdot ms) / \sqrt{3}$  avec  $ms$  valeur correspondant à l'unité du dernier chiffre.

#### 1.1.2 Affichage

L'affichage des ellipses d'incertitude suppose qu'il n'y a pas corrélation entre les deux grandeurs. Si on a entré des incertitudes-type

- avec une loi normale, une ellipse de demi-axe  $u$  correspondra à un intervalle de confiance de 68%,  $2u$  de 95% et  $3u$  de 99,7%.
- pour une loi rectangulaire  $u$  correspondra à 58%,  $2u$  à 115% (sic) et  $3u$  à 173% (resic)

## 1.2 Méthode des moindres carrés

Soit une grandeur  $y$  fonction affine d'une autre grandeur  $x$  :  $y = Ax + B$ . On a  $N$  couples de mesures  $\{y_i, x_i\}$ . Pour trouver  $A$  et  $B$ , on cherche à minimiser l'écart quadratique  $\sum_{i=1}^N (y_i - Ax_i - B)^2$ .

Cela conduit, en notant  $\Delta = N \sum x^2 - \left(\sum x\right)^2$ , à :

$$B = \frac{\sum x^2 \sum y - \sum x \sum (xy)}{\Delta}, \text{ et } A = \frac{N \sum (xy) - \sum x \sum y}{\Delta}.$$

On peut alors considérer que  $(y_i - Ax_i - B)$  représente l'écart entre la valeur mesurée  $y_i$  et la valeur « vraie »  $Ax_i + B$ , et donc évaluer l'incertitude sur  $y$  par  $\sigma_y^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - Ax_i - B)^2$ ,

le 2 venant des deux paramètres  $A$  et  $B$ . L'incertitude sur  $y$  étant maintenant estimée, on peut évaluer l'incertitude sur  $A$  par propagation des incertitudes dans l'expression précédente de  $A$ .

On trouve  $\sigma_B = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$  et  $\sigma_A = \sigma_y \sqrt{\frac{N}{\Delta}}$ .

On suppose que l'incertitude sur l'abscisse est négligeable et que l'incertitude sur l'ordonnée est la même pour tous les points. La technique de détermination des incertitudes sur les paramètres repose sur l'assimilation entre écart modèle-données et l'incertitude sur l'ordonnée.

## 1.3 Prise en compte des incertitudes

On voit que dans la technique usuelle, on fait une évaluation a posteriori de l'incertitude sur  $y$ . On choisit maintenant de minimiser toujours l'écart quadratique  $y - Ax - B$ , mais on pondère chacun des éléments par l'inverse de la variance et on en profite pour prendre en compte l'incertitude sur  $x$ . On a  $\hat{y} = y + \delta y + A\delta x$ ,  $\hat{y}$  valeur mesurée,  $y$  valeur « vraie » en  $x$  et  $\delta x$ ,  $\delta y$  variables aléatoires centrées, la variance associée à  $y$  est donc  $\sigma_y^2 + (A\sigma_x)^2$ , par addition des variances. On minimise donc  $\sum_{i=1}^N \frac{(y_i - A - Bx_i)^2}{\sigma_y^2 + (A\sigma_x)^2}$ , c'est la méthode du  $\chi^2$ . Il suffit alors de faire le même calcul qu'au §??, pour déterminer  $A$  et  $B$ , on fait apparaître le coefficient de pondération

$k_i = \frac{1}{\sigma_y^2 + (A\sigma_x)^2}$  dans les sommes.

$$\Delta = \sum k \sum (kx^2) - \left(\sum kx\right)^2$$

$$A = \frac{\sum (kx^2) \sum (ky) - \sum (kx) \sum (kxy)}{\Delta}$$

$$B = \frac{\sum k \sum (kxy) - \sum (kx) \sum (ky)}{\Delta}$$

Et de même pour les incertitudes

$$\sigma_A = \sqrt{\frac{\sum (kx^2)}{\Delta}}; \sigma_B = \sqrt{\frac{\sum k}{\Delta}}.$$

Le test du  $\chi^2$  affiche un  $\chi^2$  réduit  $\frac{\chi^2}{n-p}$ , avec  $n$  nombre de données et  $p$  nombre de paramètres.

Ce  $\chi^2$  réduit devrait être proche de 1, si le modèle est correct et si les incertitudes données pour les grandeurs sont des incertitudes-type.

**Remarque** dans le cas simple où l'incertitude sur  $x$  est quasi-nulle et celle sur  $y$  constante, on devrait retrouver le même incertitude sur les paramètres que dans le cas précédent, sinon cela signifierait un  $\chi^2$  réduit relativement différent de 1, de mauvais augure.

## 1.4 Fonction quelconque

Ce qui est important dans les relations précédentes est que  $y = A + B \cdot f(x)$ , avec dans le cas usuel  $f(x) = x$ , autrement dit la linéarité qui compte dans le calcul est celle relative à  $A$  et  $B$ . Dans le cas d'une fonction quelconque  $y = f(x, A, B)$ , on suppose donc que, localement en  $A$  et  $B$ ,  $f$  est bien une fonction linéaire de  $A$  et  $B$  et on applique les relations précédentes.

### 1.4.1 Prise en compte des incertitudes

Cette fois on minimise l'écart quadratique  $y - f(x)$ . On a  $\hat{y} = y + \delta y + \frac{dy}{dx}\delta x$ ,  $\hat{y}$  valeur mesurée,  $y$  valeur « vraie » et  $\delta x$ ,  $\delta y$  variables aléatoires centrées, la variance associée à  $y$  est donc  $\sigma_y^2 + \left(\frac{df}{dx}\sigma_x\right)^2$ , par addition des variances.

Les points de la courbe  $y(x)$  sont pondérés par leur incertitude à la fois en  $x$  et en  $y$  : le coefficient de pondération est de  $\frac{1}{u(y)^2 + \left(\frac{dy}{dx}\right)^2 u(x)^2}$  si  $u(y)$  et  $u(x)$  sont les incertitudes

données dans l'onglet correspondant. Pour que les incertitudes soient actives, il faut qu'elles soient définies pour toutes les grandeurs expérimentales (double clic sur l'en-tête du tableau de valeurs pour l'éditer) et que vous ayez coché la case « méthode du  $\chi^2$  » dans la boîte de dialogue « options » obtenue à l'aide du menu local de la modélisation accessible par le clic droit. Une incertitude peut éventuellement être à zéro (cas du temps fréquemment).

## 2 Algorithmes

Remarque préliminaire : ce texte était destiné à servir de support de stage et est donc peu détaillé.

### 2.1 Introduction

#### 2.1.1 Généralités

La modélisation consiste à trouver un modèle mathématique correspondant à un phénomène physique donné. On peut élaborer un modèle à partir de théorie (modèle de connaissance : pression de vapeur saturante, Maxwell-Hertz, RLC), ou à partir de l'expérience (modèle de représentation : Galilée, Descartes). Mélange : Coulomb, interfrange,  $F = -k \cdot x$ ,  $U = R \cdot I$ .

En fait les deux sont souvent inextricablement liés. Le modèle de connaissance peut apporter la fonction représentant le phénomène et le modèle de représentation la valeur des paramètres. Le cas linéaire conduit sans problème à des modèles de représentation mais se pose le problème du domaine de validité. cf. Ampère, non linéarité (milieu matériel en élasticité, optique), interfrange, Galilée Einstein.

En automatisme, on raisonne avec des boîtes noires sur lesquelles on fait la seule hypothèse qu'elles obéissent à des équations différentielles linéaires. La modélisation dans ce cas revient à déterminer les coefficients de l'équation différentielle. Le processus étant toujours le même, on a pu élaborer des techniques particulières non traitées ici (méthodes d'identification). C'est typiquement un modèle de représentation.

On dit que l'on a une modélisation explicite lorsqu'on connaît le comportement d'une grandeur physique par une fonction. Sinon c'est une modélisation implicite (équation algébrique ou différentielle).

On dit que l'on a une modélisation linéaire si la fonction est linéaire par rapport aux paramètres.

Il faut enfin distinguer entre recherche d'une valeur et vérification d'une loi. Par exemple, pour la loi d'Ohm, pour vérifier celle-ci, il faut faire varier l'intensité du courant de manière

notable, la résistance étant alors la pente de la droite, alors que si l'on cherche la valeur d'une résistance obéissant à la loi d'Ohm, il convient plutôt de choisir l'intensité de manière à avoir une mesure précise, la répétition des mesures n'accroissant pas forcément la précision de manière significative [2].

### 2.1.2 Critère

Soit deux variables  $y$  et  $x$  reliées par une fonction  $y = f(x, a_0, b_0)$ ,  $a_0$  et  $b_0$  étant des paramètres. On appelle  $y_k, x_k$  un couple de valeurs mesurées parmi  $N$ . Le critère doit faire intervenir une somme sur  $\varepsilon_k = y_k - f(x_k, a, b)$ .  $\varepsilon_k$  n'est pas nul en général à cause des erreurs de mesures (sur  $y_k$  et  $x_k$ ) et de structure (sur la fonction  $f$ ). Cette somme ne peut être celle des  $\varepsilon_k$  à cause des compensations possibles entre valeurs positives et négatives. Pour éviter ces compensations il faut faire disparaître le signe par une valeur absolue ou un carré. On prend le carré qui se prête davantage à des études mathématiques. On va chercher à minimiser le critère

$$J = \sum \varepsilon_k^2$$

On peut remarquer que les différents points expérimentaux sont affectés du même poids, on peut envisager d'améliorer la signification de  $J$  en pondérant les points par l'inverse de l'incertitude au carré [9],[3]. Le fait d'avoir un carré va privilégier les points de forte valeur. Voir également [8] les problèmes liés à l'anamorphose. On appelle  $a_0$  et  $b_0$  les valeurs optimales des paramètres.

## 2.2 Linéaire par rapport aux paramètres (moindres carrés)

### 2.2.1 Un paramètre

La relation s'écrit  $y = a \cdot x$ .

$$J = \sum (y_k - a \cdot x_k)^2 = a^2 \sum x_k^2 + 2a \sum y_k \cdot x_k + \sum y_k^2$$

On voit que  $J(a)$  est une parabole que l'on peut écrire  $J = \alpha \cdot a^2 + \beta \cdot a + \gamma$ . Si on cherche le minimum en écrivant que  $\frac{dJ}{da} = 2\alpha a + \beta = 0$  (C'est bien un minimum puisque le coefficient de  $a^2$  est positif), on trouve :  $a_0 = \frac{\beta}{-2\alpha} = \frac{\sum y_k \cdot x_k}{\sum x_k^2}$  (à comparer avec  $a_0 = \frac{1}{N} \sum \frac{y_k}{x_k}$ ). On a d'autre part  $\frac{d^2J}{da^2} = 2\alpha$ .

### 2.2.2 Régression linéaire

L'équation est  $y = a \cdot x + b$ . La technique est la même. La fonction  $J(a, b)$  est un paraboloïde. En écrivant que  $\left(\frac{\partial J}{\partial a}\right)_{a_0} = 0$  et  $\left(\frac{\partial J}{\partial b}\right)_{b_0} = 0$ , on obtient le système suivant

$$\begin{bmatrix} \sum x_k^2 & \sum x_k \\ \sum x_k & N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_k \cdot x_k \\ \sum y_k \end{bmatrix}$$

que l'on résout par la méthode de Cramer :

$$a_0 = \frac{N \sum y_k \cdot x_k \sum y_k \sum x_k}{N \sum x_k^2 (\sum x_k)^2}$$

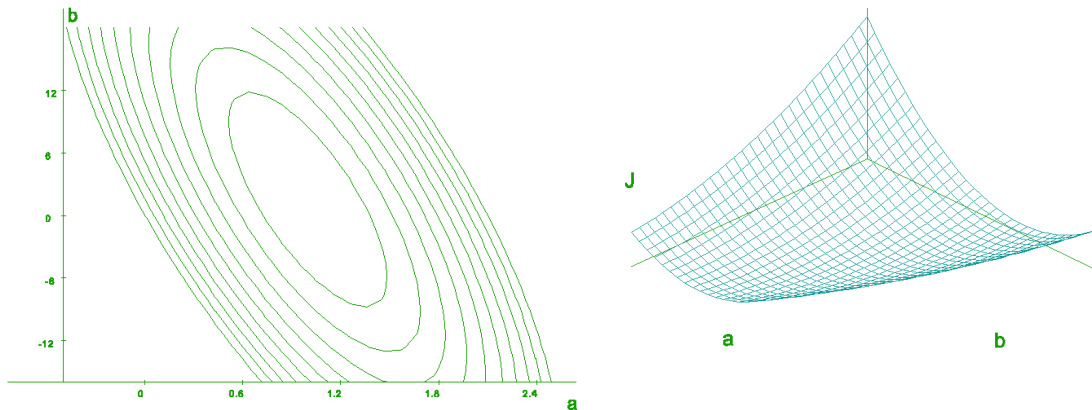
et

$$b_0 = \frac{\sum x_k^2 \sum y_k \sum x_k \sum y_k \cdot x_k}{N \sum x_k^2 (\sum x_k)^2}$$

Cela correspond à la technique de calcul des calculatrices et est le correspondant numérique de la recherche graphique d'une droite.

### 2.2.3 Isocritère

La parabolioïde  $J(a, b)$  étant plus difficilement visualisable que la courbe  $J(a)$ , on préfère représenter  $J$  par des courbes  $J = C^{te}$  dans le plan  $(a, b)$  appelées iso critères ; ce sont des courbes de niveau : ellipses centrées sur l'optimum. Ex :  $y = 1 \cdot x + 1$ .



## 2.3 Généralisation

On peut généraliser dans deux directions :

$y = ax^2 + bx + c$  c'est-à-dire une fonction de plus de deux paramètres

ou  $y = a \cdot x + b \cdot z$ , c'est-à-dire une fonction de plusieurs variables.

L'important est que la fonction soit linéaire par rapport aux paramètres. Dans ce cas on n'a pas d'équivalent graphique comme au §2. On obtient un système linéaire d'ordre égal au nombre de paramètres. Dans le cas général, il n'est pas question de donner une formule pour calculer  $a$ ,  $b$ ,  $c \dots$  on résout numériquement le système.

### 2.3.1 Linéarisation

Lorsqu'on a une formule non linéaire, on peut se ramener au cas précédent en effectuant un changement de variable appelé anamorphose. Une remarque cependant, cette transformation change le poids attribué à chaque point, et une modélisation directe ne donnera pas les mêmes valeurs pour les coefficients, sauf bien sûr si la fonction est exacte.

Ex :  $v = a \cdot \exp(-b \cdot t)$

$v = a \cdot \exp(-b \cdot t)$  conduit à  $a = 1 \pm 0,001$  et  $b = 0,996 \pm 0.001$

$\ln(v) = \ln(a) - b \cdot t$  conduit à  $a = 0,96 \pm 0,04$  et  $b = 0,97 \pm 0,01$

## 2.4 Non linéaire (Méthode du modèle)

### 2.4.1 Généralités

Si la fonction  $y(x, a, b)$  n'est plus linéaire par rapport aux paramètres, les coefficients des matrices précédentes ne sont pas des constantes et on ne peut trouver  $a$  et  $b$  en une seule étape (inversion de la matrice) comme précédemment, on va donc faire une recherche itérative. Il faut pour cela avoir un point de départ (initialiser les paramètres) et un algorithme.

Pour comprendre le principe de la méthode du modèle, on suppose d'abord le système localement linéaire (ce qui revient à dire que l'estimation proposée comme point de départ est proche de l'optimum). On a donc :

$$f(x, a, b) = f(x, a_0, b_0) + (a - a_0) \cdot \frac{\partial f}{\partial a} + (b - b_0) \cdot \frac{\partial f}{\partial b}$$

## 2.4.2 Un paramètre

S'étant donné une valeur approchée  $a$ , pour déterminer  $a_0$  il suffit de trouver la valeur de  $(a - a_0)$ , or on sait que  $a_0$  correspond au minimum de la parabole  $J(a)$ . On a

$$(a_0 - a) = \frac{\beta}{2\alpha}a = \frac{2a \cdot \alpha \beta}{2\alpha} = \frac{J'}{J''}$$

Il suffit donc de calculer  $J'$  et  $J''$ .

$$J = \sum (y_k - f(x_k, a))^2$$

soit

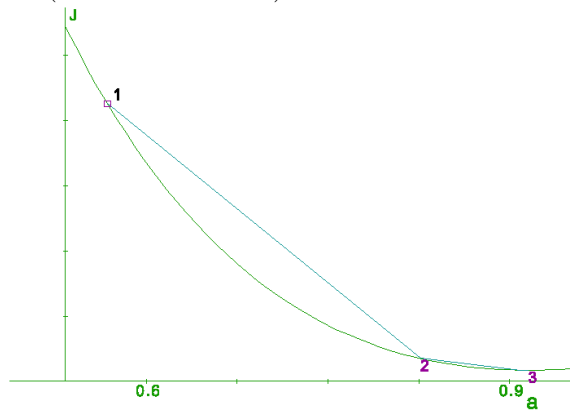
$$J' = 2 \cdot \sum (y_k - f(x_k, a)) \cdot \frac{df}{da}$$

et

$$J'' = 2 \cdot \sum (y_k - f(x_k, a)) \cdot \frac{\partial^2 f}{\partial a^2} + 2 \cdot \sum \left( \frac{\partial f}{\partial a} \right)^2$$

On néglige en général le premier terme à cause de la présence de 0, ce qui a l'avantage d'éviter le calcul de la dérivée seconde (Méthode de Gauss-Newton).

On obtient donc finalement  $(a_0 - a) = \frac{\sum \varepsilon_k \sigma_k}{\sum \sigma_k^2}$ . Si le système était linéaire on obtiendrait directement  $a_0$ , en fait on recommence l'opération jusqu'à avoir  $J'$  suffisamment petit ou  $J$  stable (Minimum atteint).



$$y = 1/(1 + a \cdot t^2)$$

## 2.4.3 Méthode de Gauss - Newton

C'est la généralisation de la formule précédente à plusieurs variables. En supposant toujours le modèle localement linéaire, la recherche du minimum revient à trouver les valeurs  $a_0$  et  $b_0$  de  $a$  et  $b$  vérifiant :  $\frac{\partial J}{\partial a} = J'_a = 0$  et  $\frac{\partial J}{\partial b} = J'_b = 0$  soit

$$J'_a(a_0, b_0) = J'_a(a, b) + (a_0 - a) \cdot J''_{aa} + (b_0 - b) \cdot J''_{ab} = 0$$

et de même pour  $J'_b$ . On obtient donc le système suivant :

$$J'_a(a_0, b_0) = J'_a(a, b) + (a_0 - a) \cdot J''_{aa} + (b_0 - b) \cdot J''_{ab} = 0$$

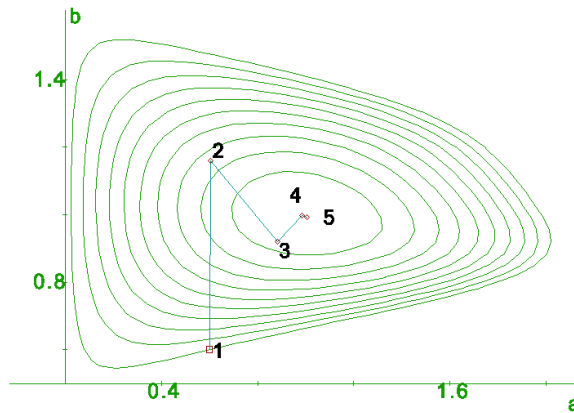
. On peut donner une interprétation géométrique simple si  $J''_{ab} = 0$  et  $J''_{aa} = J''_{bb} = \lambda$ . Alors la recherche de  $a_0$  et  $b_0$  se fait dans la direction de  $\begin{bmatrix} J'_a \\ J'_b \end{bmatrix}$ , c'est-à-dire celle de la plus grande pente (c'est d'ailleurs une méthode possible (méthode du gradient). La méthode de Gauss - Newton est plus rapide, car la ligne de plus grande pente ne décrit pas une droite dirigée vers l'optimum, sauf si les iso critères sont des cercles.

Le calcul de  $J'$  est analogue au précédent en remplaçant la dérivée par une dérivée partielle.

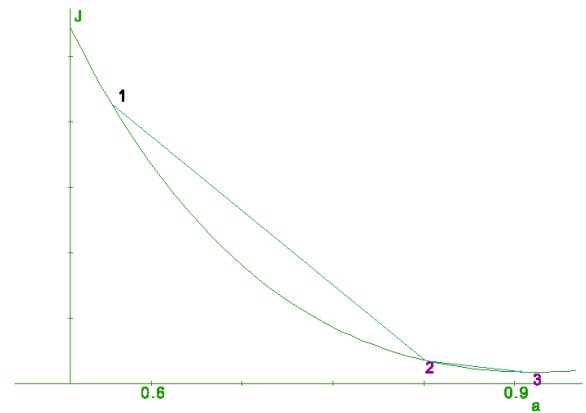
$J''$  devient  $J_{ab} = 2 \sum \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial b}$  etc.

## 2.4.4 Stabilité

Les raisonnements tenus ci-dessus supposaient une quasi linéarité. Dans la cas contraire (initialisation des paramètres avec une valeur de  $a$  trop lointaine de  $a_0$ ),  $J(a)$  n'est plus une parabole et les isocritères ne sont plus des ellipses. Si la forme est voisine (concavité), c'est simplement un problème de convergence plus lente.

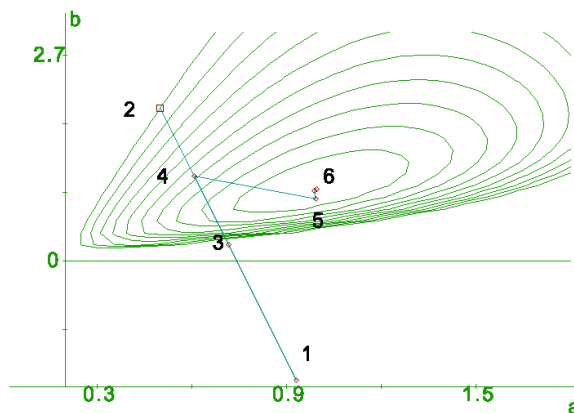


$$y = a \cdot \sin(b \cdot t) \quad a = 0,6 \quad b = 0,6$$

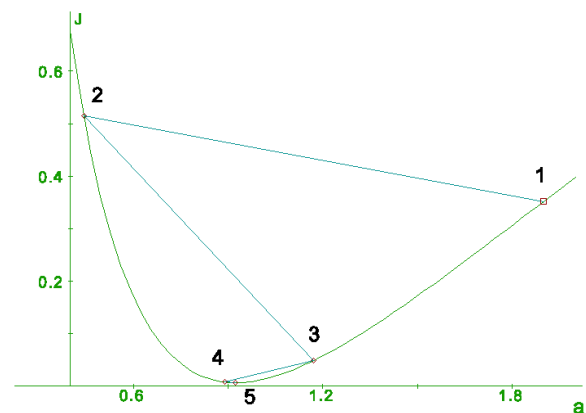


$$y = 1/(1 + a \cdot t^2) \quad a = 0,5$$

Si la concavité change,  $J''$  change de signe et la méthode diverge (on se déplace vers le maximum). Le problème de signe est évité grâce à l'approximation lors du calcul de  $J''$  assurant que celui-ci sera toujours positif. Il reste le problème des points situés au voisinage du point d'inflexion ( $J''$  tends vers 0 et donc 0 vers l' $\infty$ ). Lorsqu'on s'aperçoit d'une divergence ( $J$  augmente au lieu de diminuer), on revient en arrière par division successive de l'intervalle par 2 jusqu'à obtenir une diminution de  $J$ .



$$y = a \cdot \exp(-b \cdot t) \quad a = 0,5 \quad b = 2$$



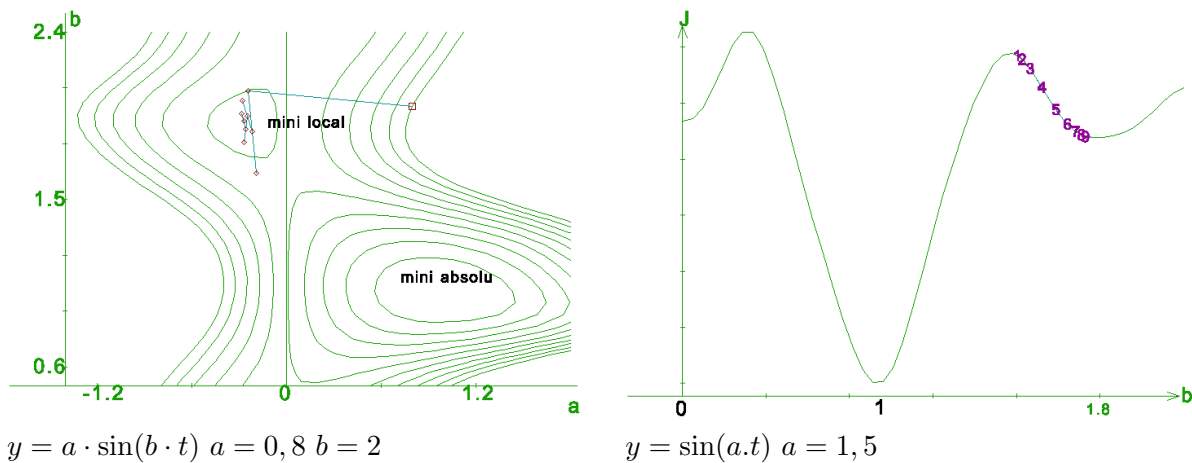
$$y = 1/(1 + a \cdot t^2) \quad a = 1,9$$

Il peut aussi se poser un problème si la nouvelle valeur du paramètre obtenu conduit à une erreur de calcul (Ex : lors de la recherche de  $v_{eq}$  par  $\log(v_{eq} - v)$ , si  $v_{eq}$  devient plus petit que l'un des points expérimentaux). On utilise dans ce cas une méthode analogue à celle de la divergence de  $J$ .

Enfin dernier cas qui ne peut être résolu de manière simple, celui où il y a des minimums secondaires. Si la valeur initiale est dans la cuvette de celui-ci, c'est cette valeur qui sera trouvée. (Ex : « harmoniques » dans le cas d'un sinus).

Ci-dessous graph de l'écart entre la fonction  $0.8 \times \sin(2 \cdot t)$  et la fonction  $a \cdot \sin(b \cdot t)$ . Les courbes bleues représentent le parcours de recherche de  $a$  et  $b$  par minimisation. Or des minimums locaux, il y en a un grand nombre (les harmoniques et même les sous harmoniques). La solution est lorsqu'on a trouvé un minimum, de réessayer ailleurs. Si on a quelques minimum locaux ça marche, mais avec les sinusoides qui en possèdent une infinité c'est plus problématique.

Autrement dit dans l'état actuel, il est difficile de trouver les paramètres d'un sinus sans une initialisation fine.



### 2.4.5 Equation différentielle

Equation d'ordre un Soit une fonction différentielle linéaire du premier ordre  $y' = f(y, t, a)$ .

La technique est la même, ce qui change est la manière de calculer  $y$  et  $\frac{\partial y}{\partial a}$ . On détermine  $y$  par une méthode de Runge - Kunta (voir annexe). Le problème est qu'on ne connaît pas explicitement  $y(t, a)$  et qu'on ne peut donc déterminer  $\frac{\partial y}{\partial a}$ . Si on dérive l'équation par rapport à  $a$ , on obtient l'équation

$$\frac{d}{dt} \left( \frac{\partial y}{\partial a} \right) = \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial a} + \frac{\partial f}{\partial a}$$

autrement dit la dérivée obéit elle-même à une équation différentielle. On obtient donc un système d'équations différentielles.

Valeurs initiales Un problème qui se pose est celui des valeurs initiales nécessaires pour démarrer l'intégration du système d'équations différentielles. Pour l'équation en  $y$ , on peut prendre la valeur expérimentale ou la considérer comme un paramètre. Pour les équations différentielles régissant les paramètres, on initialise  $\frac{dy}{da}$  à 0, puisque à l'origine  $f$  est fixée (valeur expérimentale ou autre paramètre). Si on prend  $y_0$ , valeur initiale de  $y$ , comme paramètre il faut initialiser  $\frac{\partial y}{\partial y_0}$  à 1, par définition même de  $y_0$ .

Equation d'ordre deux Le problème vient surtout de l'initialisation des dérivées premières.

Ces dernières sont connues de manière expérimentale peu précisément, on doit donc les considérer comme paramètres.

### 2.4.6 Initialisation des paramètres

Le problème posé à l'utilisateur d'une modélisation non linéaire est bien sûr la phase d'initialisation des paramètres. Il peut être aidé par une connaissance a priori de son système, sinon vu qu'un ordre de grandeur suffit pour assurer la convergence dans la plupart des cas, on rappelle ci-dessous quelques méthodes usuelles.

$$y = a \cdot \exp\left(-\frac{t}{\tau}\right); \quad y = a \cdot \sin(\omega \cdot t + \varphi) \cdot \exp\left(-\frac{t}{\tau}\right)$$

Diagramme de Bode ordre 1, 2 supérieur : rencontre des asymptotes, hauteur relative du maxi.



## 2.5 Intervalle de confiance

Dans ce paragraphe on suppose que la fonction décrit correctement le phénomène et donc que l'écart entre points expérimentaux et calculés est du aux incertitudes de mesure.

### 2.5.1 Introduction

Soit une droite  $y = a \cdot x + b$  avec des points dispersés, il est à peu près évident que  $a$  et  $b$  sont mal déterminés, mais que, de plus, les deux sont corrélés, c'est-à-dire que la probabilité de trouver  $b = b_1$  et  $a = a_1$  n'est pas égale à  $P(a_1) \cdot P(b_1)$ . Ceci limite l'interprétation des incertitudes.

### 2.5.2 Un paramètre

On dispose de  $N$  mesures  $y_k$  et  $x_k$ . On suppose  $x$  mesurée exactement et  $y$  bruitée, c'est-à-dire  $y_k = a_0 \cdot x_k + c_k$ ,  $c_k$  étant le bruit variable aléatoire indépendante de moyenne nulle et de variance  $\sigma$  (Cela revient à supposer une incertitude constante). On a vu que l'on estime  $a_0$  par  $a = a_0 + \Delta a = \frac{\sum y_k x_k}{\sum x_k^2} = a_0 + \frac{\sum c_k x_k}{\sum x_k^2}$ , on trouve que l'espérance de  $\Delta a$  c'est-à-dire en pratique la moyenne de  $\Delta a$  sur un grand nombre de mesure vaut

$$E(\Delta a) = E\left(\frac{\sum c_k x_k}{\sum x_k^2}\right) = \frac{\sum E(c_k) x_k}{\sum x_k^2} = 0$$

puisque  $x_k$  est exacte et donc que l'estimation est correcte.

Evaluons maintenant l'incertitude ou plus précisément l'écart type de  $\Delta a$ . En effet  $\Delta a$  est nul en moyenne sur un grand nombre d'expérience mais si on en fait une seule ... La variance vaut par définition  $\text{var}(x) = \langle (x - \langle x \rangle)^2 \rangle$  donc ici  $\text{var}(a) = \langle \Delta a^2 \rangle$  ce qui donne après calcul

$$\text{var}(a) = \frac{\sum E(c_k^2) x_k^2}{(\sum x_k^2)^2} = \frac{\sum x_k^2 \cdot E(c_k^2)}{(\sum x_k^2)^2}$$

soit  $\frac{\sigma^2}{\sum x_k^2}$  et on évalue  $\sigma^2$  par  $\frac{J_{\min}}{N-1}$ , ce qui suppose entre autre que les écarts sont dus uniquement aux incertitudes sur  $y$  et non à la fonction elle-même (Ex. 1 : dans l'étude de la chute libre on néglige les frottements fluides ; Ex. 2 : utilisation des formules approchées des courbes de dosage) ou aux incertitudes sur  $x$ . En effet

$$J = \sum (y_k - a_0 x_k)^2 \simeq \sum (c_k)^2$$

et  $\sigma = \sqrt{\sum (c_k)^2 / N}$ , le calcul exact conduisant à remplacer  $N$  par  $N - 1$ . L'écart type est la racine de la variance.

### 2.5.3 Deux paramètres

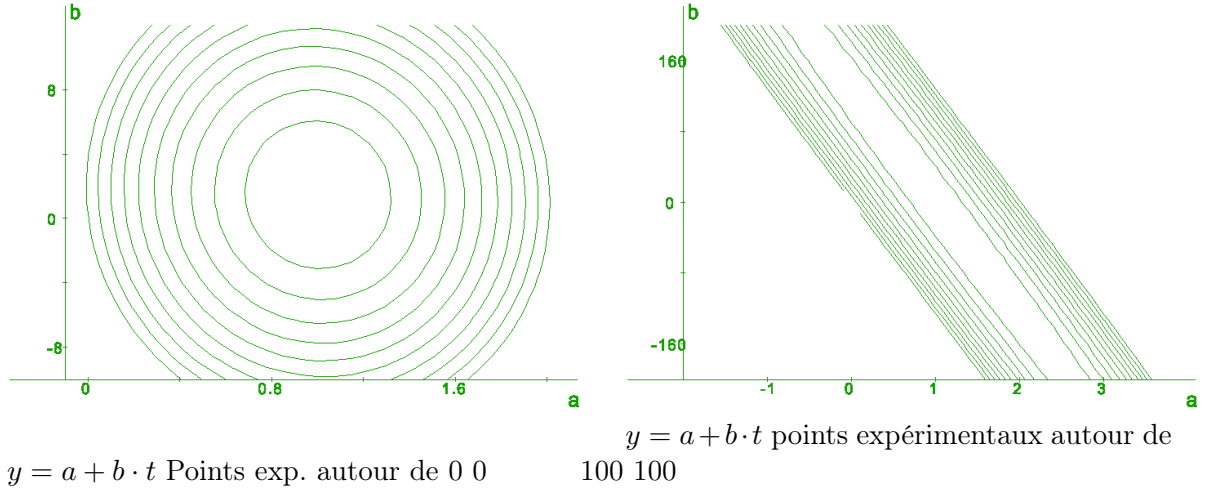
On dispose de  $N$  mesures  $y_k$  et  $x_k$ . On suppose  $x$  mesurée exactement et  $y$  bruitée, c'est-à-dire  $y_k = f(x_k) + c_k$ ,  $c_k$  étant le bruit variable aléatoire indépendante de moyenne nulle et de variance  $\sigma$ ,  $f$  étant une fonction de  $x$  et de paramètres  $a_i$ . Rappel (voir § III-C) : on linéarise au voisinage du minimum soit  $y_k = \sum_i a_{i0} f_i(x_k) + c_k = \vec{a}_0 \cdot \vec{f}(\vec{x}_k) + \vec{c}_k$ , on estime  $a_{i0}$  par  $a_i = a_{i0} + \Delta a_i$  et on obtient  $\Delta a_i$  en résolvant (écriture matricielle)  $J''_{lm} \cdot \Delta a_m = J'_l$ . On se place au minimum de  $J$  ( $J' = 0$ ), on a donc  $J'(\vec{a}_0) = J'(\vec{a})_{\min} + J''(-\vec{\Delta a}) = J''(-\Delta a)$  soit  $\Delta a = -(J'')^{-1} \cdot J'$  avec

$$J'_i = 2 \cdot \sum (y_k - f(x_k, a)) \cdot \frac{\partial f}{\partial a_i}(x_k) = 2 \cdot \sum_k c_k \cdot \frac{\partial f}{\partial a_i}(x_k)$$

On montre alors que l'espérance de  $\overrightarrow{\Delta a}$  est nulle et donc que l'estimation est correcte. On obtient la matrice de covariance de  $\Delta a$  par  $2(J'')^{-1}\sigma^2$  et on évalue  $\sigma^2$  par  $J_{\min}/(N-n)$ ,  $n$  étant le nombre de paramètres. La variance de  $a_i$  est donc  $\frac{J_{\min}}{N-n}(2J''^{-1})_{ii}$  et la covariance entre  $a_i$  et  $a_j$   $\frac{J_{\min}}{N-n}(2J''^{-1})_{ij}$ .

#### 2.5.4 Isocritère

Les courbes d'isoprobabilités et d'isocritères se confondent. En particulier, la corrélation entre les deux paramètres est liée à l'orientation de l'isocritère. Si les axes de l'ellipse sont confondus avec les axes  $a$  et  $b$ , les deux valeurs  $a$  et  $b$  ne sont pas corrélées.



#### 2.5.5 Biais

On reprend l'étude du §2 avec  $v_k$  mesuré exactement et  $x_k$  bruité :  $x_k = y_k/a + c_k$ . En développant l'estimation de  $a$  on trouve :

$$\Delta a = \frac{Na\sigma^2}{\sum x_k^2 + N\sigma^2}$$

de valeur moyenne non nulle. La valeur de  $a$  est donc « fausse » en ce sens que même en moyenne le  $a$  estimé sera différent du  $a$  optimum (réel?). Pour évaluer l'ordre de grandeur de ce biais, supposons la mesure relativement précise de manière à négliger  $\sigma^2$  au dénominateur et appelons  $\langle x \rangle$  la valeur quadratique moyenne de  $x$ , on obtient :  $\frac{\Delta a}{a} = \frac{\sigma^2}{\langle x \rangle^2}$  donc un biais relatif d'ordre l'incertitude relative au carré.

#### 2.5.6 Student

Jusqu'à présent on a confondu écart type et incertitude. Si on connaît parfaitement la variable aléatoire (moyenne et écart type) on peut trouver la probabilité de trouver  $a = a_0$ , on trouve que la probabilité de trouver  $a$  dans l'intervalle  $a_0 \pm 3\sigma$  est de 95%. Le problème est ici que l'on a seulement une estimation de ces grandeurs et donc à l'imprécision liée au s'ajoute le fait que le  $a$  estimé est lui-même aléatoire. La formule de Student permet de résoudre ce problème : celle-ci donne le coefficient par lequel il faut multiplier  $\sigma$  pour avoir un intervalle de confiance à  $x\%$  (par ex 95 %) avec  $N$  expériences. Voir bibliographie [1],[2].

### 2.5.7 Influence des bornes

Pour vérifier si le comportement d'un système obéit à une équation donnée, il faut bien sûr faire l'étude sur une large gamme de variation. Toute fonction est localement assimilable à une droite. Si tel est le cas on ne peut trouver que deux coefficients : l'équivalent de la pente et l'ordonnée à l'origine. Mathématiquement cela se traduit par des matrices mal conditionnées, i.e. dont le calcul de l'inverse conduit à des erreurs d'arrondis.

Exemple : charge de condensateur  $E = E_0 (1 - e^{-t/\tau})$ , détermination de  $E_0$  et  $\tau$ . Pour une mesure sur  $2\tau$ , on obtient des précisions de 0,6 et 1,3 %, pour une mesure sur  $2\tau$  11 et 22 %.

Exemple 2 [6] : l'étude de la fonction  $\log(P) = A - B/T - C \cdot \log(T)$  ( $P$  pression de vapeur saturante à  $T$  (en K)) sur un intervalle restreint à 80 °C to 120 °C conduit à un terme  $C \cdot \log(T)$  de l'ordre de 10 variant de 0,2, un terme  $B/T$  de l'ordre de 7,5 variant de 0,8 alors que l'incertitude sur  $\log(P)$  est de l'ordre de 0,05. Le terme  $C$  est donc quasi inaccessible.

### 2.5.8 Influence du nombre de points

Dans [2], on montre que l'on n'augmente pas sensiblement la précision d'une mesure en prenant un grand nombre de mesures. Le cas ici est légèrement différent en ce sens que pour vérifier une loi, il faut le faire sur un intervalle assez grand de la variable explicative (), mais le résultat est le même : la détermination d'un paramètre de la loi ne nécessite pas forcément un grand nombre de points. La précision est proportionnelle à  $1/\sqrt{N}$ . Le passage de 25 à 250 points augmente donc la précision d'un facteur 3 environ mais multiplie le temps de calcul par 10.

## 2.6 Justesse de la fonction

### 2.6.1 Présentation

On a jusqu'à présent supposé que la fonction était correcte. Pour vérifier ceci, on peut utiliser des méthodes simples.

Visualisation : les points expérimentaux doivent être répartis aléatoirement autour de la courbe théorique.

Comparaison : on compare plusieurs fonctions et on prend celle dont l'écart est la moins grand. Il ne faut pas pousser ce raisonnement trop loin, car on peut toujours améliorer la précision en ajoutant un paramètre supplémentaire (augmentation du degré du polynôme). On s'aperçoit que le paramètre supplémentaire est non significatif à l'ordre de grandeur de l'incertitude sur celui-ci.

### 2.6.2 Méthode du $\chi^2$

Soit une variable aléatoire  $x$  d'espérance  $m$  et d'écart-type  $\sigma$ , alors

$$\sum_{i=1}^N \frac{(x_i - m)^2}{\sigma^2}$$

suit une loi du  $\chi^2$  à  $m = N - 1$  degrés de liberté. Pour  $\chi^2(m)$  la moyenne est  $m$ , le maximum en  $(m - 2)$  et la variance  $2m$ , (soit un écart-type de  $\sqrt{2m}$ ). La valeur de la moyenne est quasi évidente d'après la définition de la variance. Ceci implique que pour des valeurs assez grandes de  $m$ , on a une valeur du  $\frac{c^2}{m}$  qui doit être proche de 1, avec un écart de l'ordre de  $1/\sqrt{m}$ .

### 2.6.3 Nouvelle expression du critère

À l'aide du  $\chi^2$ , on peut vérifier si une fonction décrit correctement une loi à condition de connaître les incertitudes de mesure.

On généralise l'expression précédente pour une fonction  $y = f(x)$  sous la forme 0 où  $\sigma_i$  représente la variance de  $y_i$  et  $a$  les  $n$  paramètres de  $f$ .

La variable aléatoire  $J$  suit une loi du  $\chi^2$  à  $m = N - n - 1$  degrés de liberté. Si la valeur de  $J/m$  est plus grande que 1, c'est que la fonction est mal adaptée, les incertitudes de mesure n'arrivant pas à justifier l'écart entre grandeurs expérimentales et calculées. Si la fonction  $J/m$  est plus petite que 1, ça marche trop bien ! :

- soit les incertitudes ont été surévaluées
- soit la fonction a un nombre de paramètres trop important (équation polynomiale d'ordre important par exemple).
- soit les résultats ont été modifiés (élimination ou « correction » des valeurs « aberrantes »)

## 2.7 Runge-Kunta

On opère avec un pas de  $\delta t$  sur  $x' = f(x, t)$

$$k0 = \delta t \cdot f(x_k, t_k)$$

$$k1 = \delta t \cdot f(x_k + k0/2, t_k + \delta t/2)$$

$$k2 = \delta t \cdot f(x_k + k1/2, t_k + \delta t/2)$$

$$k3 = \delta t \cdot f(x_k + k2, t_k + \delta t)$$

$$t_{k+1} = t_k$$

$$x_{k+1} = x_k + (k0 + 2 \cdot k1 + 2 \cdot k2 + k3)/6$$

## 3 Dérivation

La dérivation numérique est une opération posant des problèmes de précision. Si on définit  $y'_k = \frac{y_k - y_{k+1}}{x_k - x_{k+1}}$ , l'incertitude même raisonnable sur  $y$  risque de devenir importante sur la différence. On a donc intérêt pour dériver de lisser d'abord, soit mathématiquement (polynôme) soit physiquement (modèle adapté) et à dériver ensuite.

## 4 Intersection

### 4.1 Notations, calcul de l'intersection

$$y_1 = a_1 * x + b_1$$

$$y_2 = a_2 * x + b_2$$

$$\text{Intersection } y_1 = y_2, x = \frac{b_2 - b_1}{a_1 - a_2}; y = \frac{a_1 b_2 - b_1 a_2}{a_1 - a_2}$$

### 4.2 Incertitude sur $x$

$$\frac{dx}{x} = \frac{db_2 - db_1}{b_2 - b_1} - \frac{da_1 - da_2}{a_1 - a_2}$$

$$dx = \frac{db_2 - db_1}{a_1 - a_2} - \frac{da_1 - da_2}{(a_1 - a_2)^2} (b_2 - b_1), \text{ soit}$$

$$\frac{dx}{db_1} = -\frac{1}{a_1 - a_2}; \frac{\partial x}{\partial b_2} = \frac{1}{a_1 - a_2}; \frac{\partial x}{\partial a_1} = -\frac{b_2 - b_1}{(a_1 - a_2)^2}; \frac{\partial x}{\partial a_2} = \frac{b_2 - b_1}{(a_1 - a_2)^2}$$

$$u(x)^2 = \frac{u(b_2)^2 + u(b_1)^2}{(a_1 - a_2)^2} + (u(a_2)^2 + u(a_1)^2) \frac{(b_2 - b_1)^2}{(a_2 - a_1)^4}$$

### 4.3 Vérification sur un exemple

Les deux régressions sur les droites donnent (Regressi, incertitude élargie à 95% U, et Reg-Lin-MC, incertitude-type) :

	$a_1$	$b_1$	$\text{cov}(a_1, b_1)$	$a_2$	$b_2$	$\text{cov}(a_1, b_1)$
GUM	$-0,0590 \pm 0,0012$	$1,327 \pm 0,008$	$-823 \cdot 10^{-8}$	$0,1470 \pm 0,0034$	$-1,40 \pm 0,06$	$-204 \cdot 10^{-6}$
Reg	$-0,059 \pm 0,003$	$1,327 \pm 0,022$		$0,147 \pm 0,011$	$-1,4 \pm 0,2$	

Pour la droite 2, 5 points, soit 3 degrés de liberté, élargissement de Student à 95% de 4,3.

Pour la droite 1, 7 points, soit 5 degrés de liberté, soit élargissement de Student à 95% de 2,78.

**Calcul avec  $U(x)$**  : on obtient  $V_{eq} = 13,21$  mL et  $U(V_{eq}) = 1,2$  mL, soit  $V_{eq} = (13,2 \pm 1,2)$  mL

**Calcul avec GUM** : à 95%  $V_{eq} = (13,2 \pm 1,4)$  mL, cohérent, l'écart vient du fait que j'ai donné à GUM comme demi-largeur d'une distribution rectangulaire la valeur de  $U_{95\%}$ .

**Mesure avec le réticule** à partir de l'intervalle de confiance à 95%,  $12,78 < V_{eq} < 13,61$ , on obtient  $V_{eq} = (13,14 \pm 0,46)$  mL.

De même, pour l'intervalle de prédiction,  $12,57 < V_{eq} < 13,82$  soit  $V_{eq} = (13,21 \pm 0,64)$  mL. L'écart est cette fois ci important ;

Pour rendre compatible les mesures graphiques et les calculs, il faut prendre en compte la covariance entre  $a$  et  $b$ .

#### 4.4 Prise en compte de la covariance

$$u(x)^2 = \frac{u(b_2)^2 + u(b_1)^2}{(a_2 - a_1)^2} + (u(a_2)^2 + u(a_1)^2) \frac{(b_2 - b_1)^2}{(a_2 - a_1)^4} + 2(\text{cov}(a_1, b_1) + \text{cov}(a_2, b_2)) \frac{b_2 - b_1}{(a_1 - a_2)^3}$$

qui donne, avec un facteur d'élargissement de 1,96,  $V_{eq} = (13,14 \pm 0,53)$  mL, quelque part entre l'intervalle de prédiction et de confiance avec en plus les erreurs d'arrondis, d'élargissement ... on va dire que c'est bon.

#### 4.5 Incertitude sur y

$$y = \frac{a_1 b_2 - b_1 a_2}{a_1 - a_2}$$

$$\frac{dy}{y} = \frac{a_1 db_2 + b_2 da_1 - b_1 da_2 - a_2 db_1}{a_1 b_2 - b_1 a_2} - \frac{da_1 - da_2}{a_1 - a_2}$$

$$dy = \frac{a_1 db_2 - a_2 db_1}{a_1 - a_2} + \frac{(a_1 da_2 - a_2 da_1)(b_2 - b_1)}{(a_1 - a_2)^2}$$

$$\frac{dy}{db_1} = -\frac{a_2}{a_1 - a_2}; \frac{\partial y}{\partial b_2} = \frac{a_1}{a_1 - a_2}; \frac{\partial y}{\partial a_1} = -\frac{a_2(b_2 - b_1)}{(a_1 - a_2)^2}; \frac{\partial y}{\partial a_2} = \frac{a_1(b_2 - b_1)}{(a_1 - a_2)^2}$$

$$u(y)^2 = \frac{a_1^2 u(b_2)^2 + a_2^2 u(b_1)^2}{(a_1 - a_2)^2} + \frac{(a_2^2 u(a_1)^2 + a_1^2 u(a_2)^2)(b_2 - b_1)^2}{(a_1 - a_2)^4}$$

$$+ 2(a_2^2 \text{cov}(a_1, b_1) + a_1^2 \text{cov}(a_2, b_2)) \frac{b_2 - b_1}{(a_1 - a_2)^3}$$

#### 4.6 Conclusion

Le tout est implémenté dans Regressi à ce jour.

Pour ce qui est de ce que doivent faire/savoir les élèves ?!

Et ça, c'est pour un dosage conductimétrique ...

## 5 Bibliographie

[1] Théorie des probabilités Ventsel MIR p. 318.

[2] BUP 691 Février 1987 p. 159-208 Le calcul des incertitudes H. Gié R. Moreau

- [3] BUP 725 Juillet 1990 p. 769-791 A propos de la méthode des moindres carrés Y. Cortial
- [4] BUP 596 Juillet 1977 p. 1249-1303 Exploitation d'une série de mesures R. Moreau
- [5] BUP 596 Juillet 1977 p. 1305-1314 Données expérimentales J.C. Legrand
- [6] BUP 750 Janvier 1993 p. 63-80 Température et pression dans une cocotte-minute P.

Drouillon

- [7] Recherche de modèles expérimentaux J.C. Trigeassou Tec & Doc 1988.
- [8] BUP 702 Mars 1988 p. 309-313 L'anamorphose C. Ruhla
- [9] BUP 752 Mars 1993 p. 353-369 La régression linéaire et ses conditions d'applications R.

Journeaux

- [10] Identification de modèles paramétriques Walter Pronzato Masson 1994