



Université de Tunis El Manar, ENIT, LAMSIN
Université de Monastir, ISIMa



École de Recherche CIMPA :

Vert Numérique : biologie mathématique et écologie théorique

Radhouane FEKIH-SALEM

**Modélisation de processus biologiques utilisant
l'identification des paramètres à partir des données réelles**

Du 24 Septembre 2022 au 03 Octobre 2022

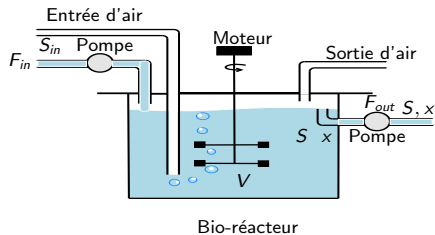
Plan

- 1 Identification des paramètres
 - Introduction
 - Droite de régression et méthode des moindres carrées
 - Importance de la structure et de la validation
- 2 Théorie de l'Identification des paramètres
- 3 Logiciel IDEAS
 - Exemple 1
 - Exemple 2

Objectif

- Identification des modèles macroscopiques des bioprocédés
- Introduire des méthodes importantes d'estimation de paramètres
 - Souligner les limites de l'approche des moindres carrés
 - Présenter l'approche du maximum de vraisemblance et quelques outils d'analyse de sensibilité paramétrique
 - Présenter les procédures d'estimation des coefficients de rendement et des paramètres cinétiques dans les modèles biologiques.
- Utilisation du logiciel IDEAS

Culture des micro-organismes dans les bioréacteurs



Structure du modèle

- Schéma réactionnel :



- Une croissance bactérienne simple



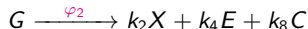
- Croissance de levure (*S. cerevisiae*) :

Les réactions sont supposées indépendantes et $N \geq M$:

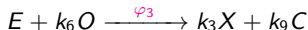
- 1 Croissance respiratoire sur le glucose :



- 2 Croissance fermentative sur le glucose :



- 3 Croissance respiratoire sur l'éthanol :



Équations du bilan de masse

- $$\begin{cases} \frac{d(V(t)\xi(t))}{dt} &= V(t)K_{\varphi}(\xi, t) - F_{out}(t)\xi(t) + F_{in}(t)\xi_{in}(t) \\ \frac{dV(t)}{dt} &= F_{in}(t) - F_{out}(t) \end{cases} \quad (1)$$

- $$\begin{cases} V(t)\frac{d\xi(t)}{dt} + \xi(t)\frac{dV(t)}{dt} &= V(t)\frac{d\xi(t)}{dt} + \xi(t)(F_{in}(t) - F_{out}(t)) \\ &= V(t)K_{\varphi}(\xi, t) - F_{out}(t)\xi(t) + F_{in}(t)\xi_{in}(t) \end{cases}$$

- $$\frac{d\xi(t)}{dt} = K_{\varphi}(\xi, t) + \frac{F_{in}(t)}{V(t)}(\xi_{in}(t) - \xi(t)) = K_{\varphi}(\xi, t) + D(t)(\xi_{in}(t) - \xi(t)).$$

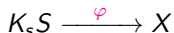
Structure du modèle

Un modèle macroscopique général :

$$\frac{d\xi(t)}{dt} = K\varphi(\xi, t) - D(t)\xi(t) + F(t) - Q(t).$$

Q : Débit massique de sortie de la composante ξ du réacteur sous forme gazeuse.

- Une croissance bactérienne simple



$$\begin{bmatrix} \dot{X} \\ \dot{S} \end{bmatrix} = \begin{bmatrix} 1 \\ -K_S \end{bmatrix} \varphi(S, X) = \begin{bmatrix} 1 \\ -K_S \end{bmatrix} \mu(S)X$$

- Croissance de levure (*S. cerevisiae*) :

$$\begin{bmatrix} \dot{X} \\ \dot{G} \\ \dot{E} \\ \dot{O} \\ \dot{P} \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 \\ -1 & -1 & 0 \\ 0 & k_4 & -1 \\ -k_5 & 0 & -k_6 \\ k_7 & k_8 & k_9 \end{bmatrix} \cdot \begin{bmatrix} \mu_1 X \\ \mu_2 X \\ \mu_3 X \end{bmatrix} - D \begin{bmatrix} X \\ G \\ E \\ O \\ P \end{bmatrix} + \begin{bmatrix} 0 \\ G_{in} \cdot D \\ E_{in} \cdot D \\ OTR \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ CTR \end{bmatrix}$$

Une variété de lois cinétiques

- Loi de Monod $\mu(S) = \mu_{max} \frac{S}{K_M + S}$
- Loi de Contois $\mu(S, X) = \mu_{max} \frac{S}{K_M X + S}$
- Loi de Haldane $\mu(S) = \mu_{max} \frac{S}{K_M + S + S^2/K_I}$

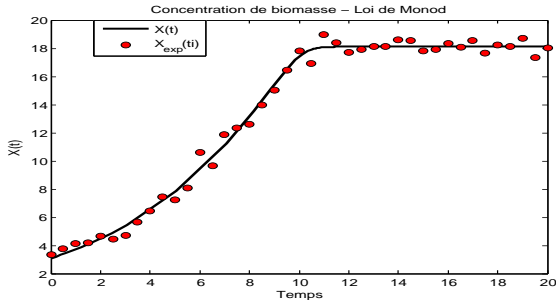
Exemple :

$$\begin{bmatrix} \dot{X} \\ \dot{S} \end{bmatrix} = \begin{bmatrix} 1 \\ -K_S \end{bmatrix} \mu_{max} \frac{S}{K_M + S} X$$

Avec des paramètres du modèle inconnus qui doivent être estimés à partir des données expérimentales... comment ?

Principaux ingrédients d'estimation des paramètres

- Un ensemble (informatives !) de données expérimentales (avec des erreurs expérimentales)
- Une structure du modèle avec un ensemble de paramètres
- Une distance entre les données expérimentales et la prédiction du modèle (qui est fonction des valeurs des paramètres)
- Une procédure de minimisation (analytique, si possible, ou numérique)



- Un choix populaire pour la distance est la somme des carrées des erreurs

Structure du modèle et paramétrage

- Un manque d'informations dans les données (cas extrême : moins de données que les paramètres) conduit à l'incertitude et à la corrélation des paramètres (extrême cas : indétermination totale)
- Une mauvaise structure du modèle \implies des erreurs systématiques dans les paramètres
- Surparamétrisation (un trop grand nombre de degrés de liberté) \implies Incertitude des paramètres et / ou sur-ajustement (en anglais « overfitting ») (Reproduction indésirable des erreurs expérimentales)

Modèles linéaires et non linéaires

La linéarité (dans les paramètres!!!) peut être une caractéristique intéressante pour faciliter la procédure d'estimation des paramètres.

- Une droite :

$$m(t) = \theta_1 t + \theta_2$$

- Un polynôme en t :

$$m(t) = \theta_1 t^2 + \theta_2 t + \theta_3$$

- Une combinaison linéaire de fonctions :

$$m(t) = \theta_1 \exp(t) + \theta_2 \cos(t) + \theta_3 t^3 + \theta_4$$

- La linéarité peut être également exprimée par une transformation

$$\mu(S) = \mu_{\max} \frac{S}{K_M + S}$$

$$\Rightarrow \frac{1}{\mu(S)} = \frac{K_M + S}{\mu_{\max} S} = \frac{1}{\mu_{\max}} + \frac{K_M}{\mu_{\max}} \frac{1}{S} = \theta_1 + \theta_2 \frac{1}{S}$$

Distance entre les données mesurées et prédiction du modèle

La distance la plus courante entre les données mesurées $y(t_k)$ et la prédiction du modèle $m(t_k, \underline{\theta})$ est la somme des carrées des erreurs

$$J(\underline{\theta}) = \sum_{k=1}^M (y(t_k) - m(t_k, \underline{\theta}))^2$$

Mais d'autres distances peuvent également être utilisées, par exemple une somme pondérée des carrées des erreurs,

$$J(\underline{\theta}) = \sum_{k=1}^M W_k (y(t_k) - m(t_k, \underline{\theta}))^2$$

Les facteurs de pondération peuvent traduire la confiance que le modélisateur a dans les données. Quelques choix courants incluent $W_k = \frac{1}{\sigma_k^2}$ (probabiliste), $W_k = \frac{1}{t_k}$ (facteur d'oubli).

Distance entre les données mesurées et prédiction du modèle

D'autres normes (distances) peuvent également être utilisées, par exemple une somme d'erreurs absolues

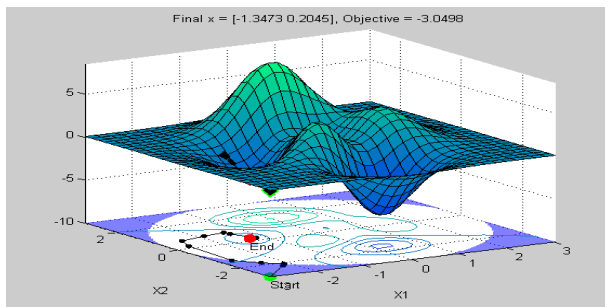
$$J(\underline{\theta}) = \sum_{k=1}^M | (y(t_k) - m(t_k, \underline{\theta})) |$$

Ce critère souligne moins les écarts les plus importants (pourrait être intéressant s'il existe des valeurs aberrantes). Cependant, la dérivée n'est pas continue, de sorte qu'il est plus difficile de minimiser (avec non-unicité de la solution).

Minimisation de la distance

La minimisation de $J(\underline{\theta})$ peut être obtenue d'une manière analytique dans les problèmes des moindres carrés linéaires. Sinon, il doit être réalisé de manière numérique. Les problèmes potentiels sont :

- Présence de minimaux locaux
- Manque de sensibilité dans certains paramètres



Emprunté de :

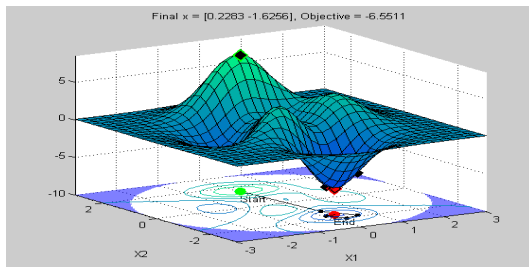
<https://www.mathworks.com/examples/global-optimization/community/19686-optimization-tips-tricks>

Minimisation de la distance

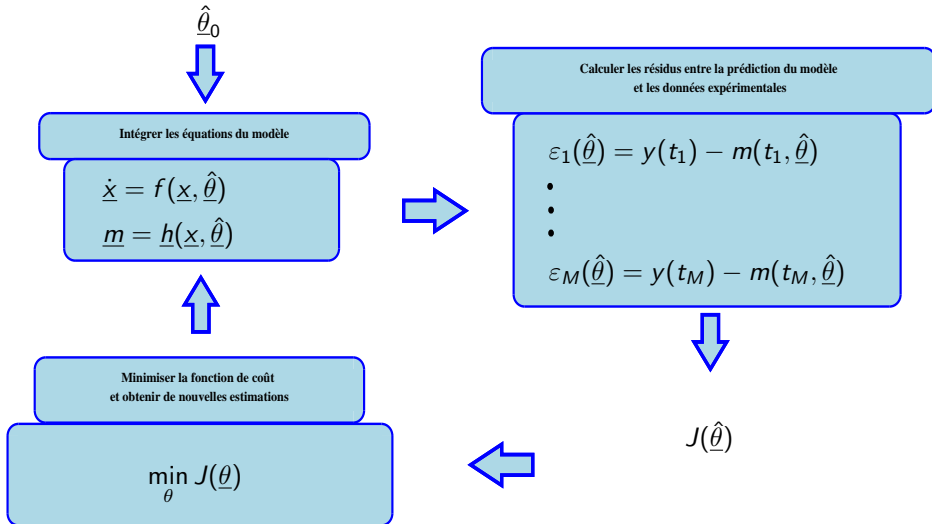
- Nous avons obtenu un minimum local, mais pas le minimum global.

Est-ce un problème avec fmincon ?

- Non, ce solveur a été conçu pour être rapide et efficace pour trouver les minima locaux.
- La plupart des solveurs d'optimisation ne sont que des solveurs locaux.
- Il est très difficile de garantir une solution globale pour des problèmes d'optimisation non linéaire comme celui-ci.
- Ce problème illustre également l'importance de la condition initiale.

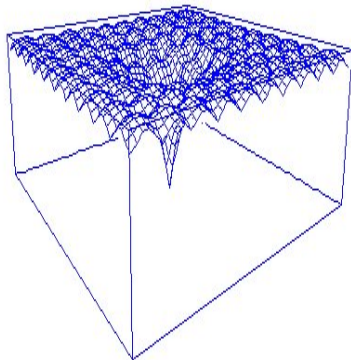
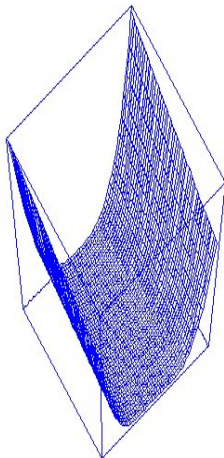
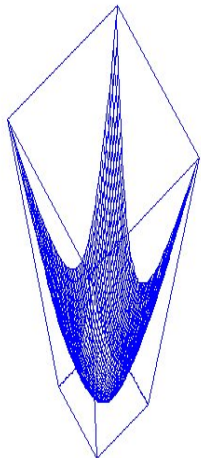


Procédure d'estimation des paramètres



Topologie de la fonction coût

Différentes formes possibles de la fonction coût (distance)



Emprunté de :

http://clerc.maurice.free.fr/psd/Semi-continuous_challenge/Semi-continuous_challenge.htm

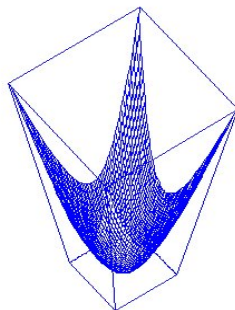
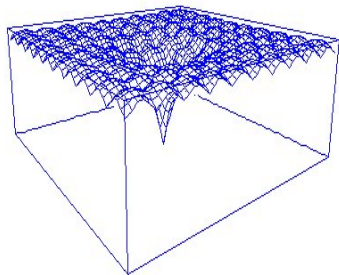
Analyse de sensibilité et conception d'expériences

Calculez les sensibilités paramétriques

$$\frac{\partial y_i}{\partial \theta_j} \Big|_{\theta_j^*}$$

Et évaluer l'influence des différents paramètres sur les sorties de mesure ainsi que les liens potentiels entre eux.

Conception d'expériences (sélection des conditions initiales et/ou des conditions de fonctionnement) pour améliorer la sensibilité



Estimation des paramètres : un art !

L'estimation des paramètres est une tâche difficile car elle combine :

- Études expérimentales informatives
- Sélection d'une bonne structure du modèle (dans ce qui suit nous allons supposer que ce choix est bien fait : nécessite certainement une bonne inspiration physique/biologique !)
- Sélection de bons outils d'estimation de paramètres (l'objectif de cette présentation !)
- Validation du modèle (validation directe et croisée)

Droite de régression

- L'idée est de transformer un nuage de point en une droite.

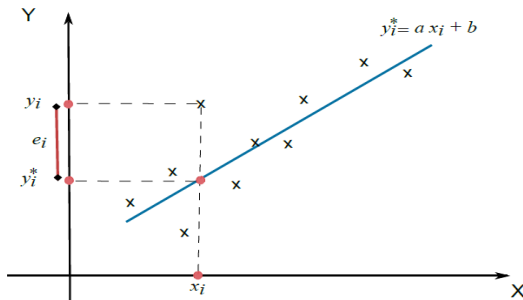


FIGURE : Ajustement linéaire : La droite la plus proche possible de chacun des points.

- On cherchera donc à minimiser les écarts entre les points et la droite

Méthode des moindres carrées

- Cette méthode vise à expliquer un nuage de points par une droite qui lie Y à X :

$$Y = aX + b,$$

telle que la distance entre le nuage de points et la droite soit minimale.

- Cette distance matérialise l'erreur : la différence entre le point réellement observé et le point prédit par la droite.
- Si la droite passe au milieu des points, cette erreur sera alternativement positive et négative, la somme des erreurs étant par définition nulle

Méthode des moindres carrées

- La méthode consiste à chercher la valeur des paramètres a et b qui minimise la somme des erreurs élevées au carré :

$$S(a, b) = \sum_{i=1}^n e_i^2$$

avec e_i est l'erreur commise sur chaque observation :

$$|e_i| = |y_i - y_i^*| = |y_i - ax_i - b|$$

- La méthode consiste donc à minimiser la fonction $S(a,b)$:

$$\frac{\partial S}{\partial a}(a, b) = \frac{\partial S}{\partial b}(a, b) = 0, \quad S(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

- L'équation $\partial S / \partial b = 0$ donne

$$-2 \sum_{i=1}^n (y_i - ax_i - b) = 0 \quad \implies \quad \bar{Y} - a\bar{X} - b = 0 \quad \iff \quad b = \bar{Y} - a\bar{X}$$

Méthode des moindres carrées

- Si $b = \bar{Y} - a\bar{X}$, on

$$\begin{aligned} S(a, b) &= \sum_{i=1}^n (y_i - ax_i - (\bar{Y} - a\bar{X}))^2 \\ &= n(\bar{Y} - a\bar{X})^2 - 2(\bar{Y} - a\bar{X}) \sum_{i=1}^n (y_i - ax_i) + \sum_{i=1}^n (y_i - ax_i)^2 \\ &= -n(\bar{Y} - a\bar{X})^2 + \sum_{i=1}^n (y_i - ax_i)^2 \\ &= a^2 \left(\sum_{i=1}^n x_i^2 - n\bar{X}^2 \right) - 2a \left(\sum_{i=1}^n x_i y_i - n\bar{X}\bar{Y} \right) + \left(\sum_{i=1}^n y_i^2 - n\bar{Y}^2 \right) \\ &= n(\text{Var}(X) a^2 - 2a \text{Cov}(X, Y) + \text{Var}(Y)) \end{aligned}$$

- Donc $S(a, b)$ est minimal pour $a = \text{Cov}(X, Y) / \text{Var}(X)$

Méthode des moindres carrés

Théorème

- La droite Δ cherchée a pour equation :

$$y - \bar{Y} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} (x - \bar{X}) \quad (3)$$

- Δ passe par le point moyen du nuage.
- Δ est appelée **droite de régression de y en x** ou droite des moindres carrés de y en x.

Droite de régression

Exemple : Le tableau suivant donne le nombre x_i de voitures vendues par un concessionnaire officiel, ainsi que le montant y_i correspondant des frais de publicité au cours de 10 trimestres successifs.

x_i	480	450	480	540	570	420	390	520	470	480
y_i	22	18	20	24	24	22	14	22	18	16

$$\bar{X}=480, \bar{Y}=20, \text{Var}(X)=2600, \text{Var}(Y)=10.4, \text{Cov}(X,Y)=118, r(X,Y)=0.72$$

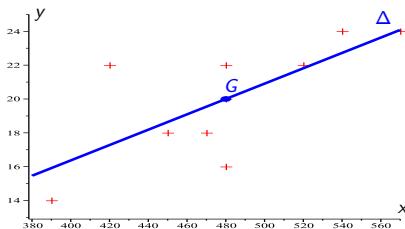


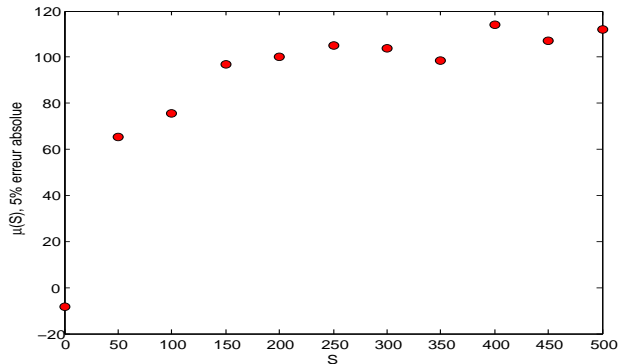
FIGURE : Droite de régression de y en x

$$y = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} (x - \bar{X}) + \bar{Y} = 0.045x + 1.785$$

Un exemple de l'importance de la structure et de la validation

Identification d'un modèle cinétique basé sur 11 mesures de μ pour S de 0 à 500 (par étapes de 50) avec un bruit blanc additif avec distribution normale, moyenne nulle et écart type de 5

$$\mu = 120 \frac{S}{40 + S}$$



Un exemple de l'importance de la structure et de la validation

Différents candidats de structures de modèles

- 1 Structure du modèle exact (loi de Monod)

$$\mu_1 = \theta_1 \frac{S}{\theta_2 + S}$$

- 2 Même niveau de complexité (loi de Tessier)

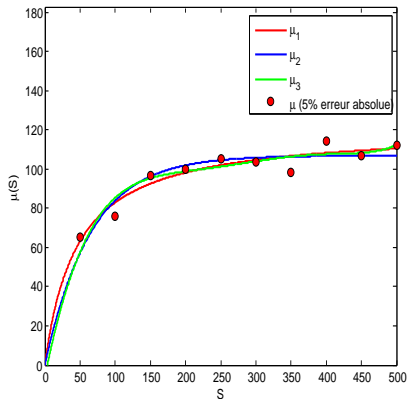
$$\mu_2 = \theta_1 \left(1 - e^{-S/\theta_2}\right)$$

- 3 Modèle sur-paramétré (modèle "boîte noire" polynomial)

$$\mu_3 = \theta_1 + \theta_2 S + \theta_3 S^2 + \theta_4 S^3 + \theta_5 S^4 + \theta_6 S^5$$

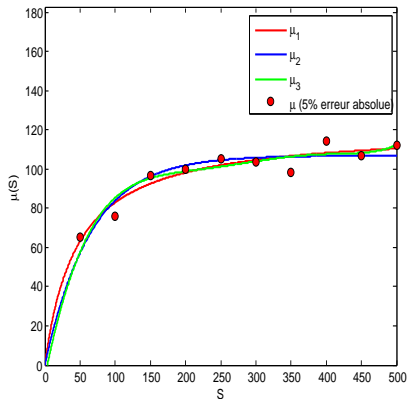
Un exemple de l'importance de la structure et de la validation

Validation directe

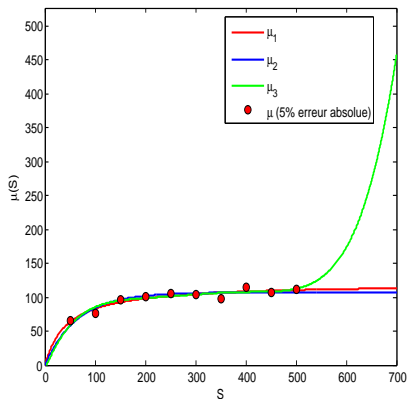


Un exemple de l'importance de la structure et de la validation

Validation directe



Validation croisée



Procédure d'estimation des paramètres

**Objectifs de modélisation
+
Connaissance à priori**

Collecte de données expérimentales

Sélection d'une structure de modèle

Sélection d'un critère

Minimisation du critère

Validation de modèle

Utilisation du logiciel IDEAS

- ① Qu'est-ce que c'est le logiciel IDEAS
- ② Contexte
 - ① Estimation des paramètres
 - ② Intervalles de confiance des estimations
- ③ Utiliser IDEAS
 - ① Exemple
 - ② Quelques conseils

Qu'est-ce que c'est le logiciel IDEAS

- IDEAS est l'acronyme de "IDEntification and Analysis of Sensitivity"
- Une boîte à outils d'identification des paramètres avec analyse de l'incertitude symbolique
- Ce logiciel gratuit est une boîte à outils Matlab pour l'identification des paramètres des modèles d'équation différentielle ordinaire (ODE).
- L'estimation des paramètres est effectuée dans le sens du maximum de vraisemblance (ML).
- Toutes les fonctions générées sont accessibles et peuvent être utilisées dans d'autres routines et modifié si nécessaire.

Qu'est-ce que c'est le logiciel IDEAS

- IDEAS est développée au sein des équipes :
 - ① Mathématiques et Informatique Appliquées and Unité d'Écologie et Physiologie du Système Digestif (INRA)
 - ② Laboratoire des Signaux et Systèmes (L2S), CNRS, SUPELEC et Univ Paris-Sud de France.

Contexte et Estimation des paramètres

- La théorie de l'identification des paramètres et de l'analyse de l'incertitude utilisée dans le logiciel est rappelée pour favoriser la compréhension du logiciel.

Considérons le modèle d'espace d'état :

$$\dot{x} = f(x, \theta, t), \quad (4)$$

où

- $x(t, \theta)$ est le vecteur d'état ($x : \mathbb{R}^+ \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$)
- θ est le vecteur de paramètres ($\theta \in \mathbb{R}^{n_p}$)
- f est une fonction vectorielle de classe C^1 (continue avec dérivées partielles continues du premier ordre) de l'état et des paramètres ($f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \times \mathbb{R}^+ \rightarrow \mathbb{R}^{n_x}$),
- Dans le cas particulier des observations synchrones, la sortie du modèle, représentée par le vecteur y_m de dimension \mathbb{R}^{n_y} , satisfait

$$y_m(t, \theta) = h(x, \theta, t)$$

Estimation des paramètres

- Le vecteur de données collectées à l'instant t_i est modélisé par :

$$y(t_i) = y_m(t_i, \theta^*) + \varepsilon_i, \quad i = 1, \dots, n_t,$$

- n_t : le nombre de temps d'observation
- θ^* : la vraie valeur du vecteur de paramètre.
- Les erreurs de mesure $\varepsilon_i (i = 1, \dots, n_t)$ sont supposées indépendantes, homoscédastiques, moyenne nulle et gaussienne, ce qui signifie que

$$\varepsilon_i \sim N(0, \Sigma).$$

- Σ : la matrice de covariance

Estimation des paramètres

Sous ces hypothèses, l'estimateur ML est :

$$(\hat{\theta}, \hat{\Sigma}) = \arg \min_{\theta, \Sigma} L(\theta, \Sigma)$$

où

$$L(\theta, \Sigma) = \frac{n_t}{2} \ln \det \Sigma + \frac{1}{2} \sum_{i=1}^{n_t} [y(t_i) - y_m(t_i, \theta)]^T \Sigma^{-1} [y(t_i) - y_m(t_i, \theta)]$$

Moindres carrés non pondérés.

IDEAS offre à l'utilisateur quatre alternatives :

- **Moindres carrés non pondérés.** Si Σ est supposé être proportionnel à la matrice identité, l'estimateur ML pour θ est l'estimateur des moindres carrés non pondérés, ce qui minimise la fonction coût

$$J_1(\theta) = \sum_{i=1}^{n_t} [y(t_i) - y_m(t_i, \theta)]^T [y(t_i) - y_m(t_i, \theta)]$$

et l'estimation ML de la matrice de covariance est

$$\hat{\Sigma} = \frac{J_1(\hat{\theta})}{n_t} I.$$

Maximum de vraisemblance avec Σ inconnue et diagonale.

- **Maximum de vraisemblance avec Σ inconnue et diagonale.** L'estimateur ML pour θ minimise la fonction coût

$$J_2(\theta) = \sum_{k=1}^{n_y} \frac{n_t}{2} \ln \left[\sum_{i=1}^{n_t} [y_k(t_i) - y_{m_k}(t_i, \theta)]^2 \right]$$

et l'estimation ML de la matrice de covariance est

$$\hat{\Sigma} = \text{diag} \left(\hat{\sigma}_1^2, \dots, \hat{\sigma}_{n_y}^2 \right)$$

avec

$$\hat{\sigma}_k^2 = \frac{1}{n_t} \sum_{i=1}^{n_t} [y_k(t_i) - y_{m_k}(t_i, \hat{\theta})]^2$$

Maximum de vraisemblance avec Σ inconnue.

- **Maximum de vraisemblance avec Σ inconnue.** L'estimateur ML pour θ minimise la fonction coût :

$$J_3(\theta) = \ln \left(\det \sum_{i=1}^{n_t} [y(t_i) - y_m(t_i, \theta)][y(t_i) - y_m(t_i, \theta)]^T \right),$$

et l'estimation ML de la matrice de covariance est

$$\hat{\Sigma} = \frac{1}{n_t} \sum_{i=1}^{n_t} [y(t_i) - y_m(t_i, \hat{\theta})][y(t_i) - y_m(t_i, \hat{\theta})]^T$$

Maximum de vraisemblance avec Σ connue.

- **Maximum de vraisemblance avec Σ connue** (fourni par l'utilisateur)
L'estimateur ML correspond à l'estimateur de Gauss-Markov, qui minimise la fonction coût

$$J_4(\theta) = \sum_{i=1}^{n_t} [y(t_i) - y_m(t_i, \theta)]^T \Sigma^{-1} [y(t_i) - y_m(t_i, \theta)].$$

Intervalles de confiance des estimations

- La matrice de covariance P de l'estimateur ML est approchée par l'inverse du FIM (F) calculé en $\hat{\theta}$. L'estimation de P est donnée par

$$\hat{P} = F^{-1}(\hat{\theta}, \Sigma_0),$$

où Σ_0 est une valeur nominale pour la covariance de bruit. Pour les cas où Σ est inconnue, elle est approximée par son estimation $\Sigma_0 = \hat{\Sigma}$ et le FIM s'écrit

$$F(\hat{\theta}) = \sum_{i=1}^{n_t} \left[\frac{\partial y_m}{\partial \theta} \right]_{(t_i, \hat{\theta})}^T \Sigma_0^{-1} \left[\frac{\partial y_m}{\partial \theta} \right]_{(t_i, \hat{\theta})} \quad (5)$$

- Lorsque Σ est inconnu, une approche largement utilisée consiste à l'approximer par son estimation $\hat{\Sigma}$ du ML. Le FIM est alors calculé en prenant $\Sigma_0 = \hat{\Sigma}$ dans (5).
- La racine carrée η_j du j ème élément de la diagonale de \hat{P} est une estimation de l'écart type de $\hat{\theta}_j$, qui est utilisé pour obtenir une approximation de l'intervalle de confiance à 95% pour θ_j c'est à dire : $[\hat{\theta}_j \pm 2\eta_j]$.

Intervalles de confiance des estimations

- Pour évaluer (5), il faut calculer la matrice de sensibilité de la sortie. Soit $s_j = \frac{\partial x}{\partial \theta_j}$ dénotent la sensibilité de l'état au paramètre θ_j . Le vecteur s_j est la solution de

$$\dot{s}_j = \left[\frac{\partial f}{\partial x} \right]_{(x,\theta,t)} s_j + \left[\frac{\partial f}{\partial \theta_j} \right]_{(x,\theta,t)}, \quad s_j(0) = 0. \quad (6)$$

- La sensibilité de la sortie au paramètre θ_j est évaluée par

$$\frac{\partial y_m}{\partial \theta_j} = \left[\frac{\partial h}{\partial x} \right]_{(x,\theta,t)} s_j + \left[\frac{\partial h}{\partial \theta_j} \right]_{(x,\theta,t)}, \quad \frac{\partial y_m}{\partial \theta_j}(0) = 0. \quad (7)$$

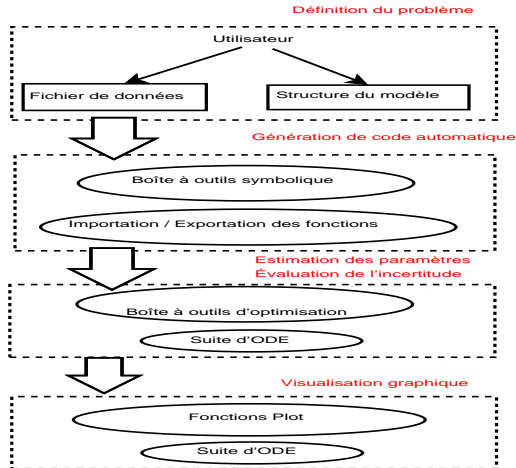
- Le FIM est donc calculé après la solution du système augmenté donné par (4,6) en $\hat{\theta}$, et en faisant les substitutions sur (7). Les parties à droite de (6-7) sont évalués à l'aide de la boîte à outils symbolique de Matlab.

IDEAS

- IDEAS se compose de cinq fichiers. L'utilisateur doit créer un dossier et copier les fichiers source.
- Le logiciel fonctionne sous Matlab v7.0 ou versions ultérieures.
- Il nécessite l'optimisation et des boîtes à outils symbolique à exécuter.
- IDEAS est exploité via une interface graphique et des boîtes de dialogue dans la commande Fenêtre.
- Il est composé de trois panneaux correspondant aux étapes de la résolution du problème.

Étapes suivies par IDEAS

- Les panneaux sont appelés "Root", "Optimisation" et "Visualisation".
- Une boîte de dialogue indique à l'utilisateur quelle est l'étape en cours.



Étapes suivies par IDEAS

- La Figure montre l'intégration des étapes suivies dans une exécution typique.
- IDEAS génère automatiquement des fonctions `m` qui sont appelées dans les différentes étapes de l'exécution.
- Toutes ces fonctions partagent le nom défini par l'utilisateur (problème Nom).
- La fonction sans suffixe est le modèle ODE.

Explication des suffixes des fonctions générées par IDEAS

Suffixe	Signification
cost	Évalue le critère d'optimisation
load	Télécharge le fichier de données
optim	Effectue l'étape d'optimisation
ploty	Trace l'ajustement des données
plotys	Trace les trajectoires de sensibilité
se	État du modèle augmenté + équations de sensibilité à l'état
sy	Calcule les sensibilités des sorties
out	Calcule les sorties du modèle
uncert	Calcule la FIM

- Ces fonctions sont ouvertes à l'utilisateur et peuvent être modifiées ou utilisées dans d'autres routines définies par l'utilisateur.

Exemple 1

Considérons le modèle classique du chémostat

$$\begin{cases} \dot{s} &= D(s^{in} - s) - \frac{1}{Y}\mu(s)x, \\ \dot{x} &= (\mu(s) - D_1)x, \end{cases} \quad (8)$$

et le vecteur de sortie est

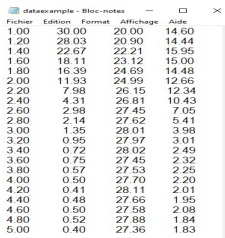
$$y_m = \left(s, x, \frac{1-Y}{Y}\mu(s)x + D_1x \right)^T \quad (9)$$

avec

$$\mu(s) = \frac{\mu_{max}s}{K+s}, \quad s(0) = 30, \quad x(0) = 20.$$

Utiliser IDEAS

- Les paramètres à estimer sont K , μ_{max} , Y , D_1 . Soit $D = 0.041$ et $s^{in} = 40$.
- Le fichier de données doit être fourni au format .txt avec le temps d'échantillonnage dans la première colonne.
- Les sorties (mesures) doivent suivre le même ordre du vecteur de sortie y_m (9).
- Le fichier de données ("dataexample") doit être dans le même dossier avec les fichiers sources.



Fichier	Edition	Format	Affichage	Aide
1.00	30.00	20.00	14.60	
1.20	28.03	20.90	14.44	
1.40	22.67	22.21	15.95	
1.60	18.11	23.12	15.00	
1.80	16.39	24.69	14.48	
2.00	11.93	24.99	12.66	
2.20	7.98	26.15	12.34	
2.40	4.31	26.81	10.43	
2.60	2.98	27.45	7.05	
2.80	2.14	27.62	5.41	
3.00	1.35	28.01	3.98	
3.20	0.95	27.97	3.01	
3.40	0.72	28.02	2.49	
3.60	0.75	27.45	2.32	
3.80	0.57	27.53	2.25	
4.00	0.50	27.70	2.20	
4.20	0.41	28.11	2.01	
4.40	0.48	27.66	1.95	
4.60	0.50	27.58	2.08	
4.80	0.52	27.88	1.84	
5.00	0.40	27.36	1.83	

Étapes IDEAS

Pour utiliser IDEAS, suivez les étapes suivantes :

- ➊ Dans Matlab, sélectionnez le répertoire où se trouvent les fichiers sources et les données.
- ➋ Tapez IDEAS dans la fenêtre de commande. Les informations de licence de la boîte à outils apparaissent.
- ➌ Cliquez sur le bouton OK et l'interface s'affiche.
- ➍ Suivez les instructions données dans la boîte de dialogue. Pour l'illustration, nous utilisons "ChemClass1" comme nom du problème et le fichier de données est nommé dataexample.txt.

Welcome to IDEAS

Écrire le nom de votre problème et du fichier de données. Cliquez ensuite sur le bouton "Model definition".

ideastoolbox

IDEAS: a toolbox for parameter identification

Copyright(c) INRA - Univ Paris-Sud - CNRS 2010
Reference number from APP:
IDDN.FR.001.110021.000.R.C.2010.000.20700.

Root

Problem name
ChemClass1

Name of data file
dataexample

Model definition

Code generation

Optimization

Unweighted Least Squares

Parameter estimation

Visualization

Data fit

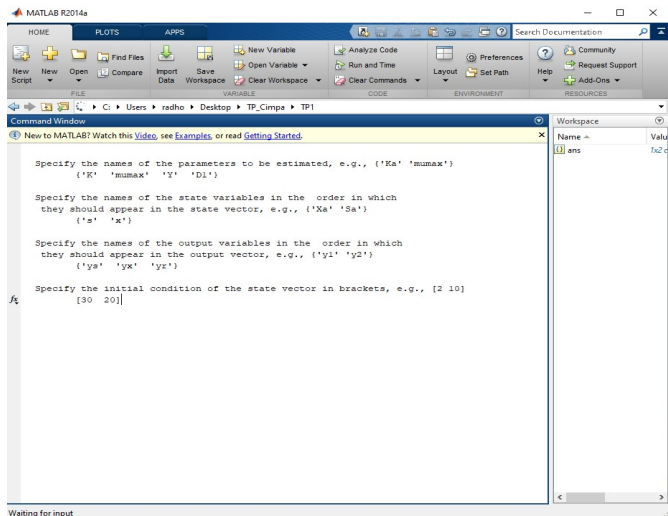
Sensitivities

Root

Start

Welcome to IDEAS. Please click on the Start button

Définition des variables du modèle



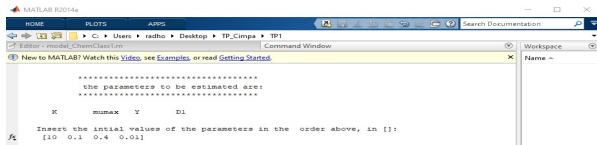
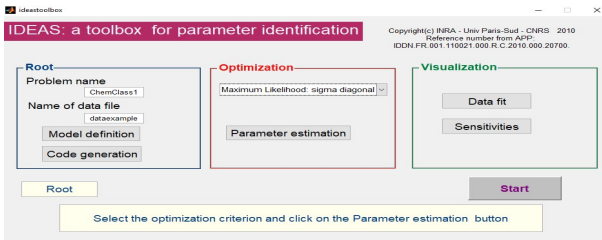
Définition du modèle mathématique

```

1 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2 % Program to define the mathematical model
3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
5 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
6 % This part was automatically generated. Do not change
7 % Symbolic definition of the parameters
8
9 syma K mumax Y D1
10 parameters = [ K mumax Y D1 ];
11 % Symbolic definition of the state variables
12 syma s x
13 Xx = [ s x ];
14 % Symbolic definition of the outputs
15 syma ys yx yr
16 Yy = [ ys yx yr ];
17 % Initial conditions of the state variables
18 Cinit = [ 30 20 ];
19 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
20 % In this part the user defines the model.
21 % Follow the structure below
22 % Do not change the names of the items:
23 % parameters % F % H
24 % Step 1 : Write the constant parameters of the model, e.g.
25 % Km = 10;
26 D = 0.042; % 1/h, dilution rate
27 Sin = 40; % Concentration du substrat à l'entrée mg COD/l
28 % Step 2 : Write the functions of the model, e.g.
29 % ro = mumax * Xx(1)*Xx(2) / (K+Xx(1));
30 mu = mumax*s/(K+s); % Taux de croissance
31 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
32 % The mathematical model is represented as follows :
33 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
34 % dx/dt = F(x,p)
35 % y = H(x,p)
36 % x has a dimension of n
37 % F has a dimension of n
38 % y has a dimension of q
39 % H has a dimension of q
40 % Step 3 : Write the elements of the vector F, e.g. % F(1) = -ro/Y - Km*Xx(1);
41 F(1) = D*(Sin-s) - mu*x/Y;
42 F(2) = (mu-D1)*x;
43 % Step 4 : Write the elements of the vector H, e.g. % H(1) = ro;
44 H(1) = s;
45 H(2) = x;
46 H(3) = (1-Y)/Y*mu*x + D1*x;
47 % Step 5 : Save the file

```

- Ce fichier est généré automatiquement
- L'utilisateur doit spécifier le modèle selon la structure décrite ci dessous



MATLAB R2014a

HOME PLOTS APPS

C:\Users\radho\Desktop\TP_Cimpa\TP1

Editor - model_ChemClass1.m Command Window

New to MATLAB? Watch this Video, see Examples, or read Getting Started.

```

*****
Finding the estimates
*****

Diagnostic Information

Number of variables: 4

Functions
Objective: ChemClassicost
Gradient: finite-differencing
Hessian: finite-differencing (or Quasi-Newton)

Algorithm selected
medium-scale: Quasi-Newton line search

End diagnostic information

Iteration Func-count f(x) Step-size First-order optimality
0 5 225.095 0.000771482 313
1 15 216.391 0.145677 280
2 30 192.752 0.145677 289
3 35 180.694 1 512
4 60 87.8121 0.207788 991

24 220 32.8439 1 2.03
25 240 32.8439 1.09 1.98
26 250 32.8439 0.1 1.87
27 265 32.8439 0.01 1.85

Local minimum possible.

fminunc stopped because the size of the current step is less than
the default value of the step-size tolerance.

<stopping criteria details>

Computing finite-difference Hessian using user-supplied objective function.

*****
The estimation has finished
*****
K : 0.4673 +/- 1.3656
mumax : 0.37694 +/- 0.023773
Y : 0.29653 +/- 0.010031
D1 : 0.027648 +/- 0.0045679

The correlation matrix of the parameters is:

ans =
0.84051 0.045939 -0.024031
0.045939 0.52382 0.45538
-0.024031 0.45538 0.5394 1

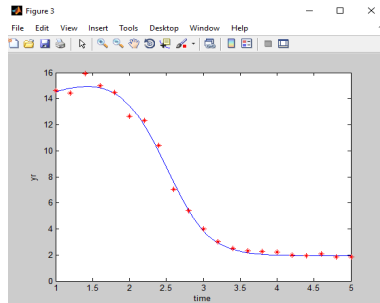
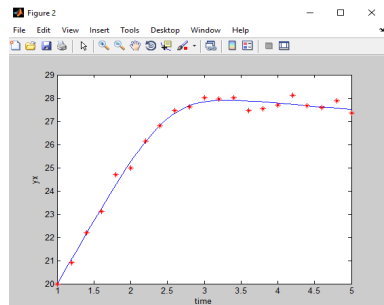
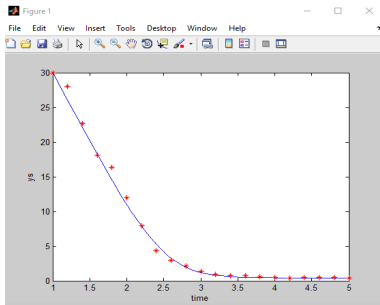
The Fisher matrix condition number is:

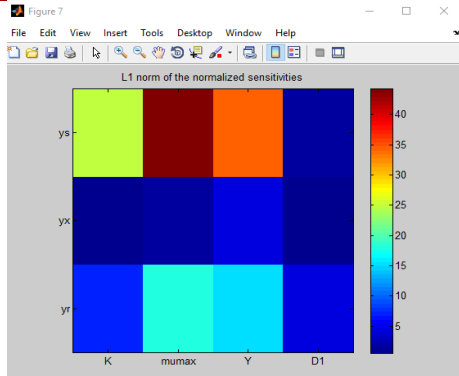
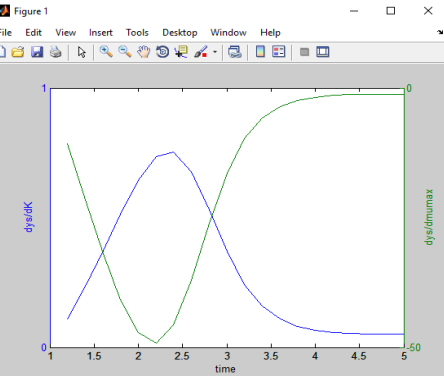
ans =
954107.2691

*****
Export your Results
*****

Type 1 to export in xls format or 2 for txt format 1

```





- Appuyez sur le bouton "Sensitivities" pour afficher la trajectoire des sensibilités de la sortie. La dernière figure est une représentation matricielle de la norme L1 des sensibilités normalisées. L'élément k, j est donné

$$\sum_{i=1}^{n_t} \left| \frac{\hat{\theta}_j}{y_{m_k}(t_i, \hat{\theta})} \left[\frac{\partial y_{m_k}}{\partial \theta_j} \right]_{(t_i, \hat{\theta})} \right|$$

Quelques conseils

- L'étape d'optimisation peut être sensible aux problèmes numériques.
- Si votre problème d'estimation ne peut pas être résolu et que vous avez en retour des messages d'avertissement, cela peut être dû à l'initialisation du vecteur de paramètres ou à des erreurs dans l'intégration des équations de l'ODE.
- Vous pouvez essayer de faire un reparamétrage afin de travailler dans la base logarithmique.
- Vous pouvez entrer dans les fonctions avec les suffixes `optim` et `cost` pour décommenter les instructions et travailler avec cette base.
- L'avantage de cette option est que la positivité des paramètres est garantie.
- Gardez à l'esprit que l'optimisation est locale, vous devez donc tester l'estimation avec différentes valeurs de paramètre pour l'initialisation.

Exemple 2

Considérons le modèle classique du chémostat de deux espèces en compétition sur un seul nutriment.

$$\begin{cases} \dot{s} &= D (s^{in} - s) - \frac{1}{Y_1} \mu_1(s) x_1 - \frac{1}{Y_2} \mu_2(s) x_2, \\ \dot{x}_1 &= (\mu_1(s) - D) x_1, \\ \dot{x}_2 &= (\mu_2(s) - D) x_2, \end{cases} \quad (10)$$

et le vecteur de sortie est

$$y_m = (s, x_1, x_2)^T \quad (11)$$

avec

$$\mu_1(s) = \frac{\mu_{1max} S}{k_1 + s}, \quad \mu_2(s) = \frac{\mu_{2max} S}{k_2 + s}, \quad s(0) = 1.5, \quad x_1(0) = 3.5, \quad x_2(0) = 4.5$$

- Générer des données des variables S , x_1 et x_2 dans un fichier appelé data.txt entre l'instant 0 à 100 avec un bruit blanc additif et une distribution normale, de moyenne nulle et d'écart type de 5.
- Déterminer l'estimation des paramètres et l'analyse de sensibilité en utilisant Ideas.

Références

- ① Goodwin, G., and Payne, R. Dynamic System Identification : Experiment Design and Data Analysis. Academic Press, New York, 1977.
- ② Munoz-Tamayo, R., Laroche, B., Leclerc, M., and Walter, E. IDEAS : a parameter identification toolbox with symbolic analysis of uncertainty and its application to biological modelling. In 15th Symposium on System Identification (2009).
- ③ Walter, E., and Pronzato, L. Identification of Parametric Models from Experimental Data. Springer, London, 1997.