Approche coopérative entre simulation numérique et essais

François Dubois

Professeur des Universités
Mathématiques appliquées
CNAM Paris et Université Paris Sud, Orsay

octobre 2008

Points abordés

1) Méthode des moindres carrés

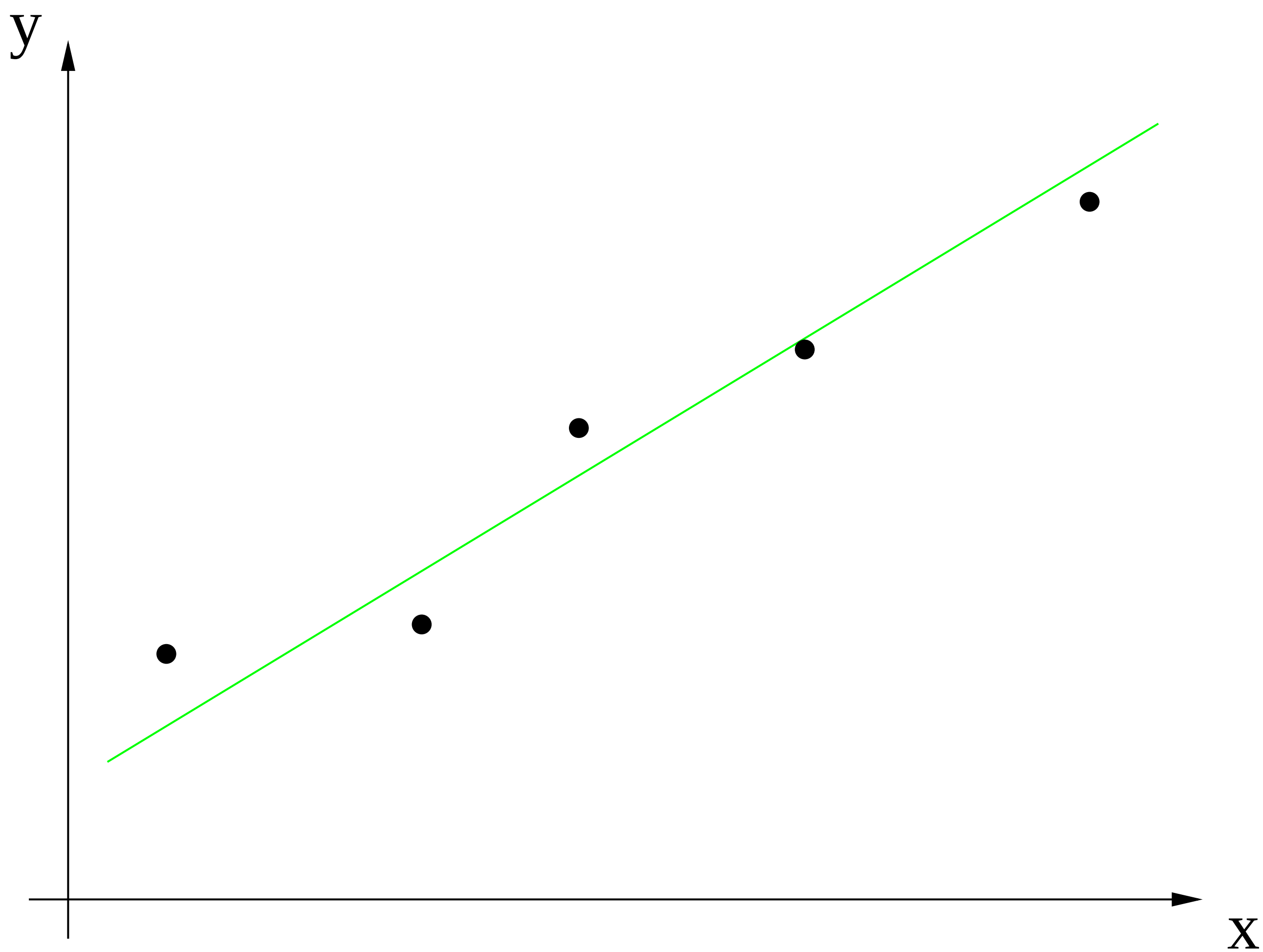
2) Descente du gradient pour les moindres carrés

3) Oscillateur harmonique

4) Une autre façon de calculer le gradient

5) Equation adjointe

6) Cas d'une équation différentielle générale



N points expérimentaux $(x_1, y_1), \ldots (x_N, y_N)$.

Modèle mathématique : $y = \alpha\, x + \beta$.

Pour trouver les paramètres α et β,

on se définit une "fonctionnelle d'erreur" J

$$J(\alpha,\,\beta) \;=\; \frac{1}{2N}\,\sum_{j=1}^{N}\bigl(y_j - \alpha\,x_j \,-\, \beta\bigr)^2$$

On minimise la fonctionnelle $J(\bullet,\,\bullet)$.

Au point de minimum éventuel $(\alpha^*,\,\beta^*)$,

$$\frac{\partial J}{\partial \alpha}(\alpha^*,\,\beta^*) \;=\; 0\,, \qquad \frac{\partial J}{\partial \beta}(\alpha^*,\,\beta^*) \;=\; 0\,.$$

Dans ce cas de la droite des moindres carrés,

on peut mener sans difficulté le calcul du gradient

$$\nabla J \;\equiv\; \Bigl(\frac{\partial J}{\partial \alpha},\,\frac{\partial J}{\partial \beta}\Bigr)^{\mathrm{t}}$$

de la fonctionnelle.

Pour $J(\alpha,\,\beta) \,\equiv\, \dfrac{1}{2N} \, \sum_{j=1}^{N} \big(y_j - \alpha \, x_j \,-\, \beta \big)^2 \,,$ on a

$$\frac{\partial J}{\partial \alpha}(\alpha,\,\beta) \,=\, \frac{1}{N} \, \sum_{j=1}^{N} \big(y_j - \alpha \, x_j \,-\, \beta \big) \, (-x_j)$$

$$\frac{\partial J}{\partial \beta}(\alpha,\,\beta) \,=\, \frac{1}{N} \, \sum_{j=1}^{N} \big(y_j - \alpha \, x_j \,-\, \beta \big) \, (-1) \,.$$

On introduit les moyennes classiques

$$< x > \equiv \tfrac{1}{N} \sum_{j=1}^{N} x_j \,, \qquad < y > \equiv \tfrac{1}{N} \sum_{j=1}^{N} y_j \,,$$

$$< x^2 > \equiv \tfrac{1}{N} \sum_{j=1}^{N} x_j^2 \,, < xy > \equiv \tfrac{1}{N} \sum_{j=1}^{N} x_j \, y_j \,, \; < y^2 > \equiv \tfrac{1}{N} \sum_{j=1}^{N} y_j^2 \,.$$

Alors

$$\frac{\partial J}{\partial \alpha} = \alpha < x^2 > + \beta < x > - < xy > , \; \frac{\partial J}{\partial \beta} \,=\, \alpha < x > + \beta - < y > \;.$$

Le système d'équations

$$\frac{\partial J}{\partial \alpha}(\alpha^*, \beta^*) = 0 , \qquad \frac{\partial J}{\partial \beta}(\alpha^*, \beta^*) = 0$$

se traduit par un système linéaire

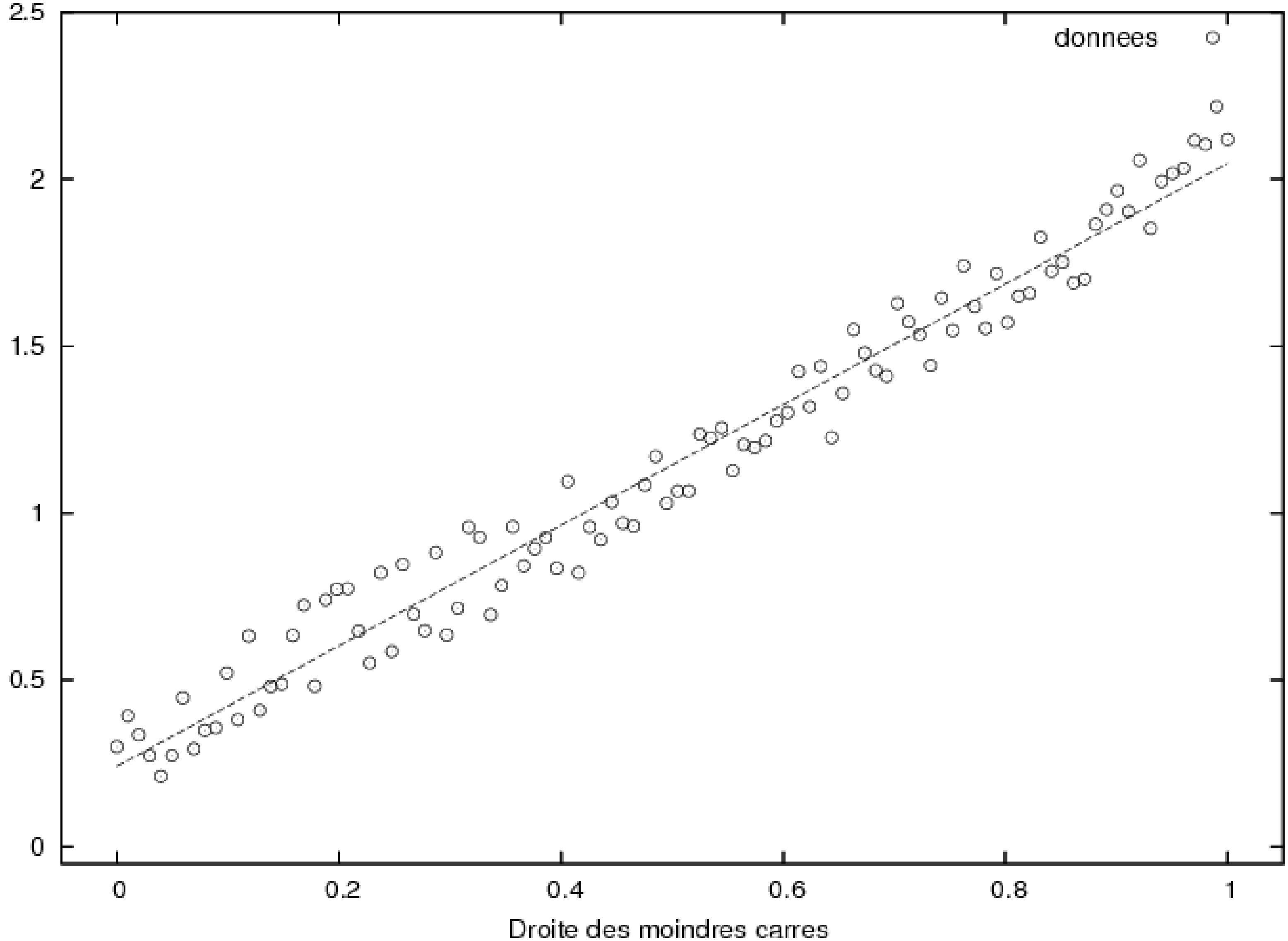
de deux équations à deux inconnues :

$$\begin{aligned} &< x^2 > \alpha^* + < x > \beta^* = < xy > \\ &< x > \ \alpha^* \ + \qquad \beta^* \quad = < y > \ . \end{aligned}$$

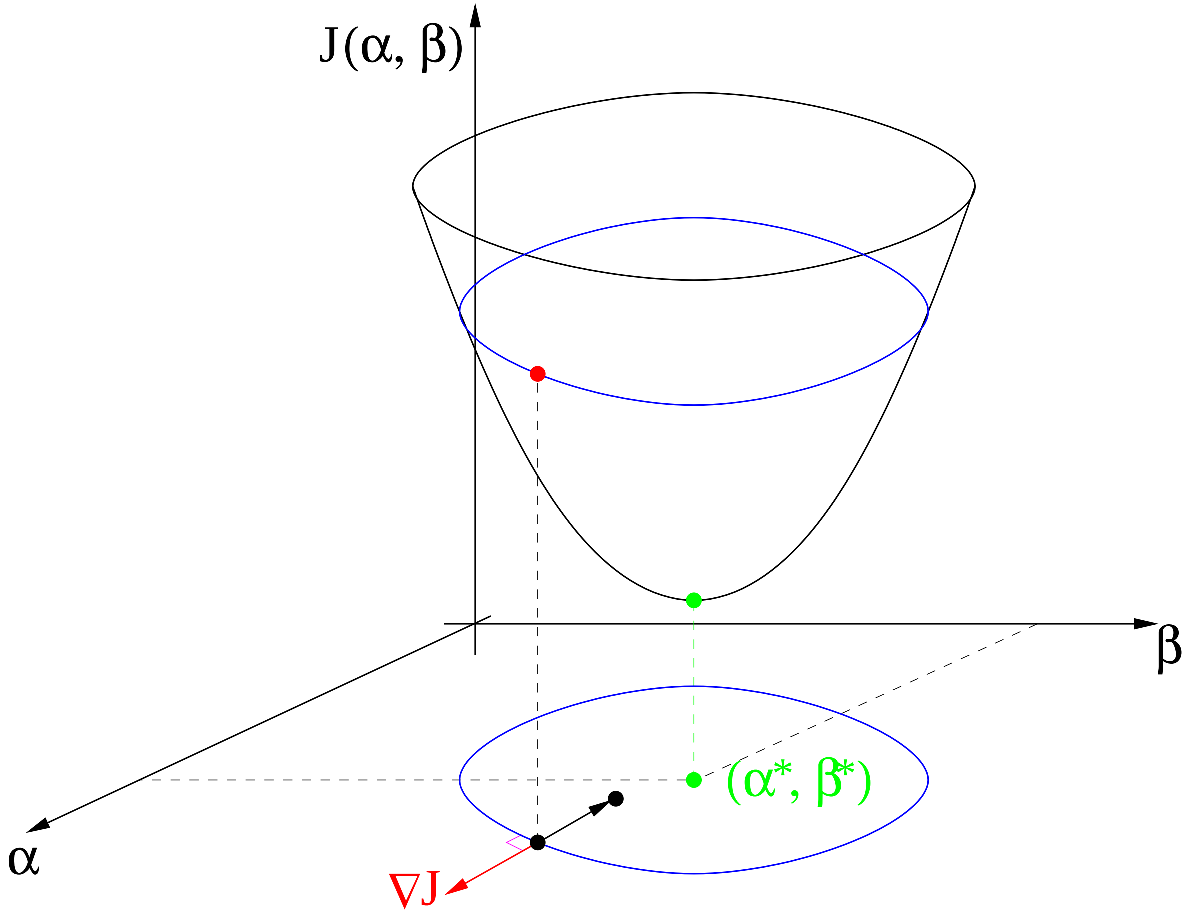
La résolution n'offre pas de difficulté

$$\alpha^* = \frac{< xy > - < x >< y >}{< x^2 > - < x >^2}$$

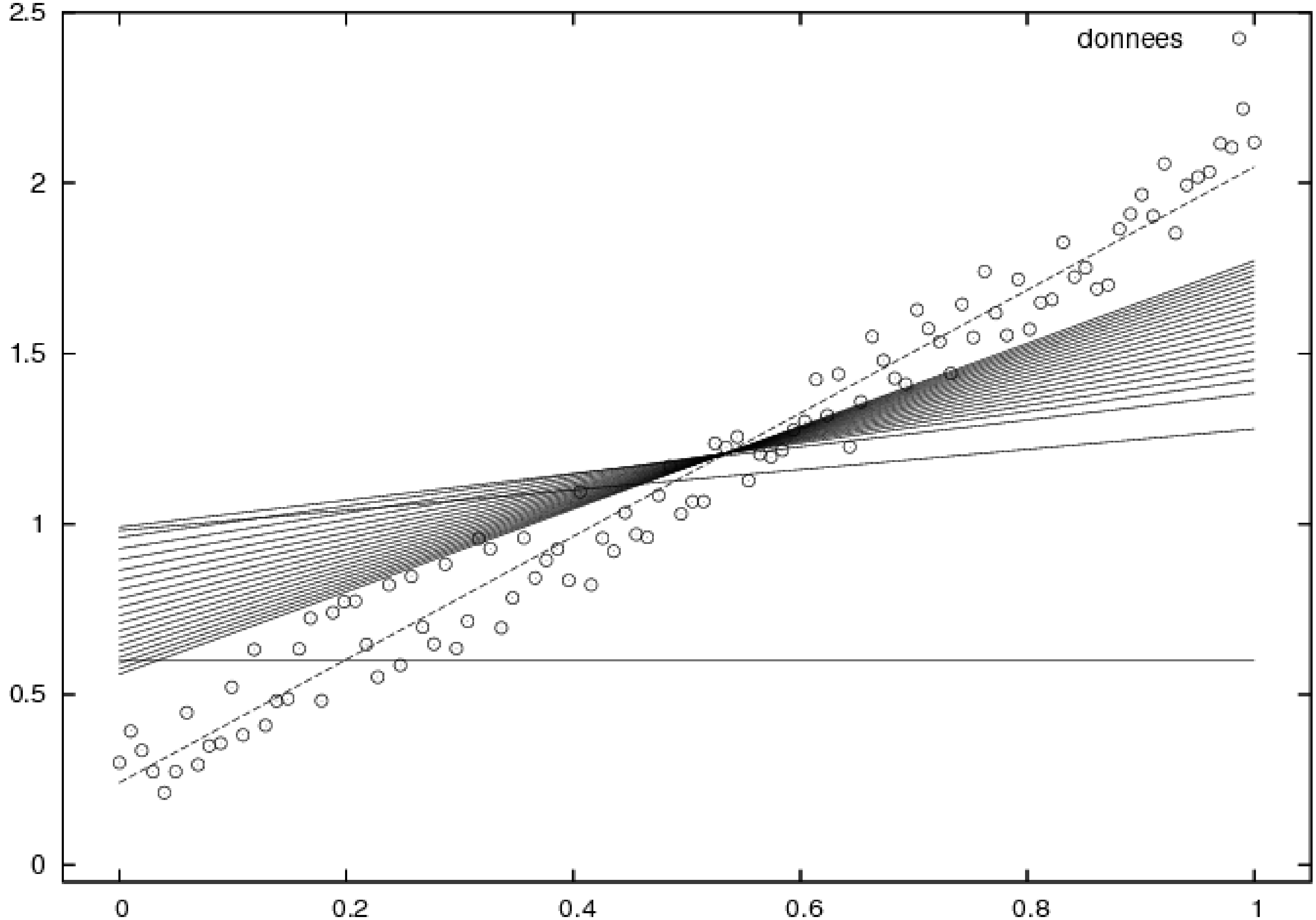
$$\beta^* = \frac{< y >< x^2 > - < x >< xy >}{< x^2 > - < x >^2} \, .$$


donnees
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Droite des moindres carres

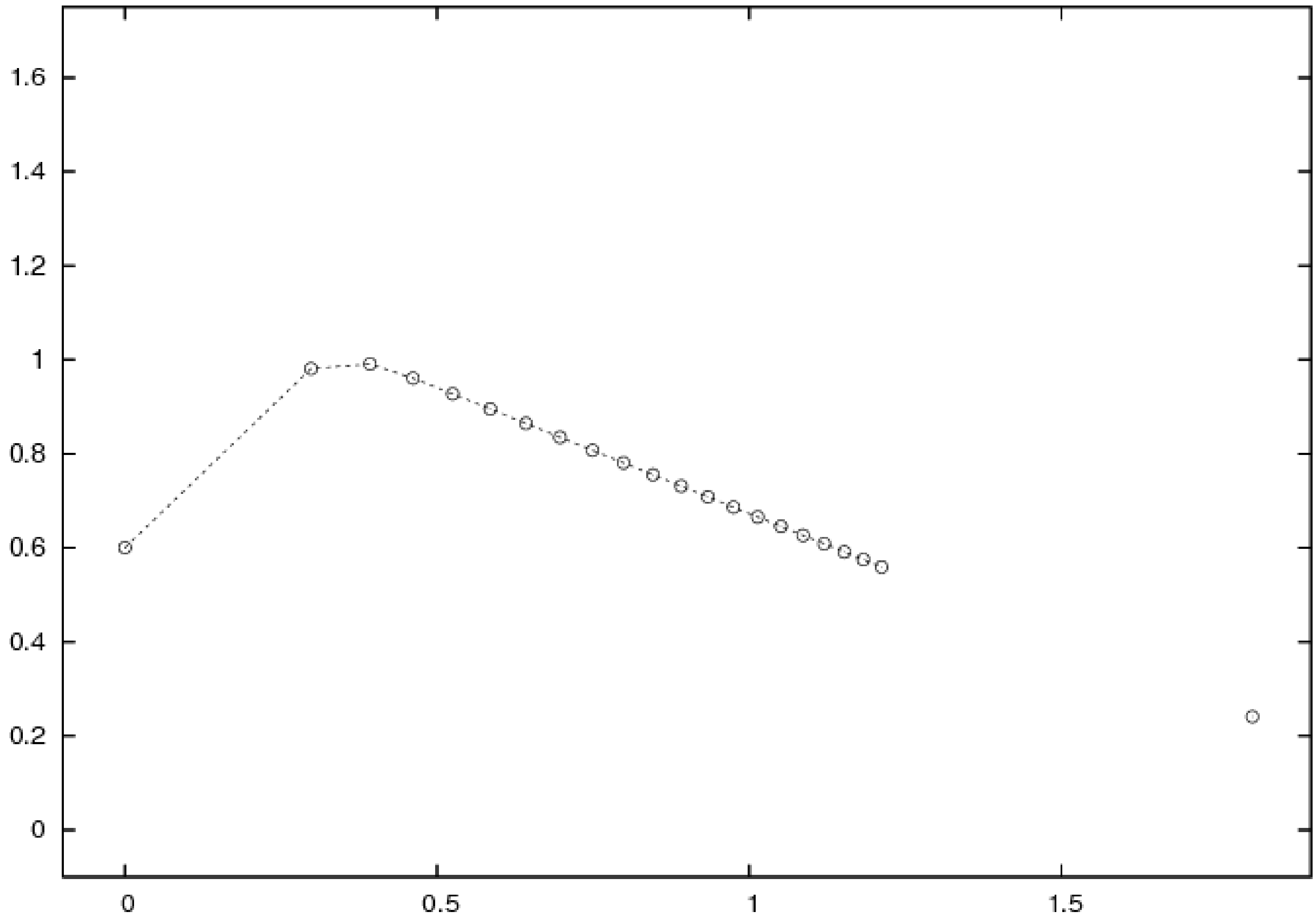
Au lieu de résoudre le systèmes d'équations $\nabla J(\alpha^*, \beta^*) = 0$,
on descend le long de la ligne de plus grande pente.



$$(\alpha_{k+1}, \beta_{k+1})^{\mathrm{t}} = (\alpha_k, \beta_k)^{\mathrm{t}} - \rho \nabla J(\alpha_k, \beta_k)$$



Droite des moindres carres. Methode de gradient a pas fixe.



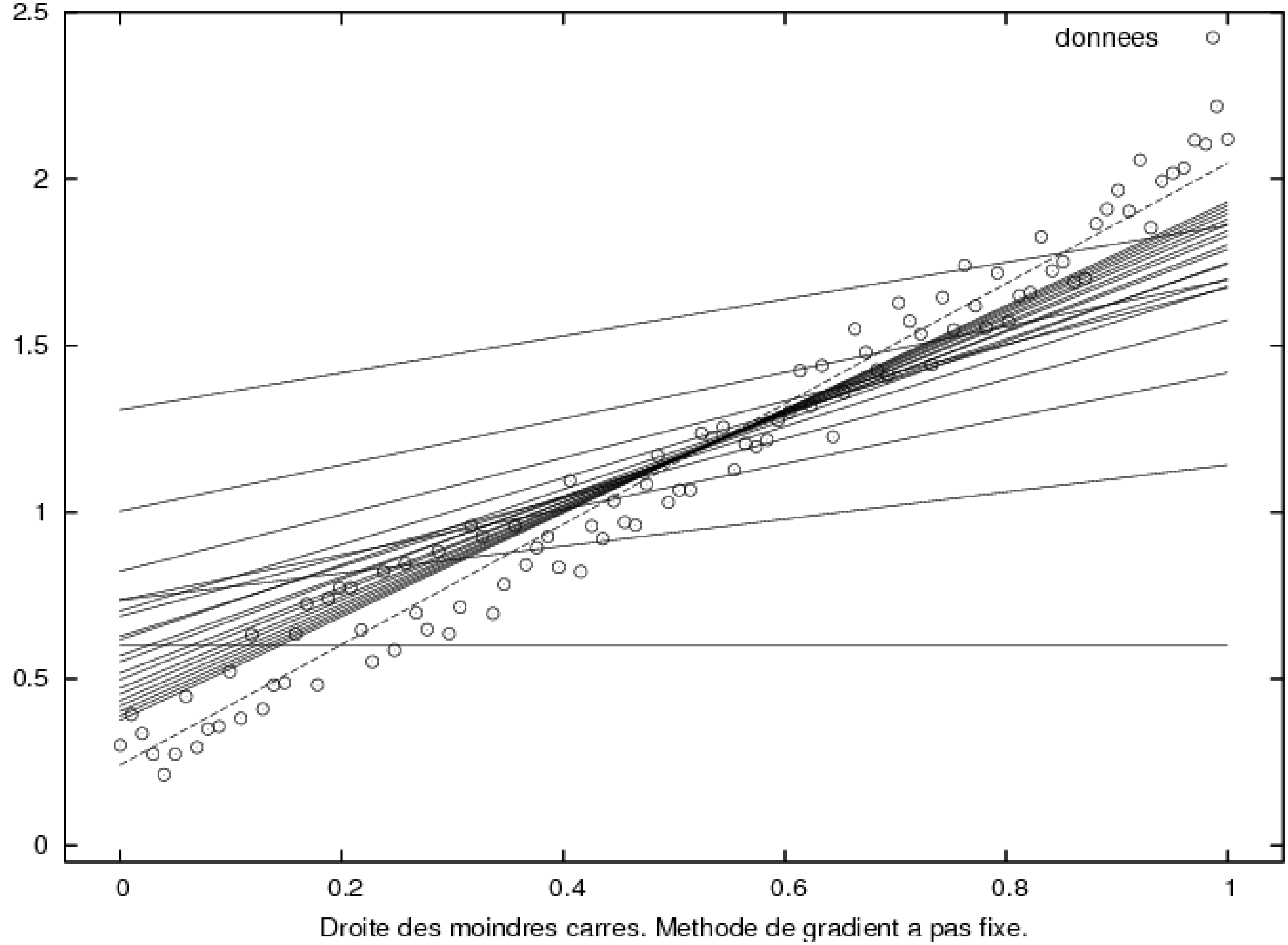
Convergence de la droite des moindres carres. Representation dans le plan des coefficients.

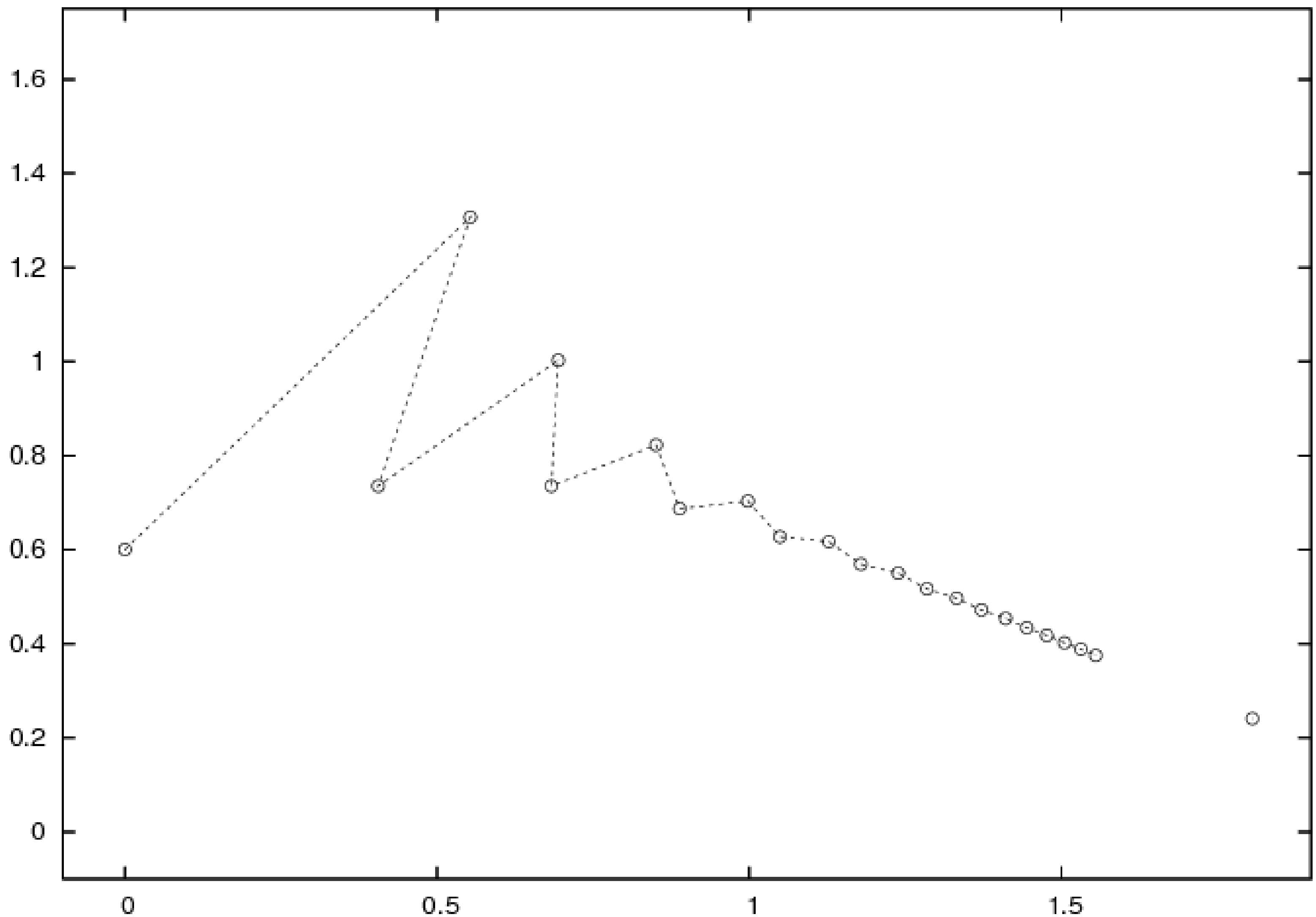
$$\lambda_{k+1} = \lambda_k - \rho \nabla J(\lambda_k)$$

Attention de ne pas choisir un coefficient de relaxation ρ trop important !

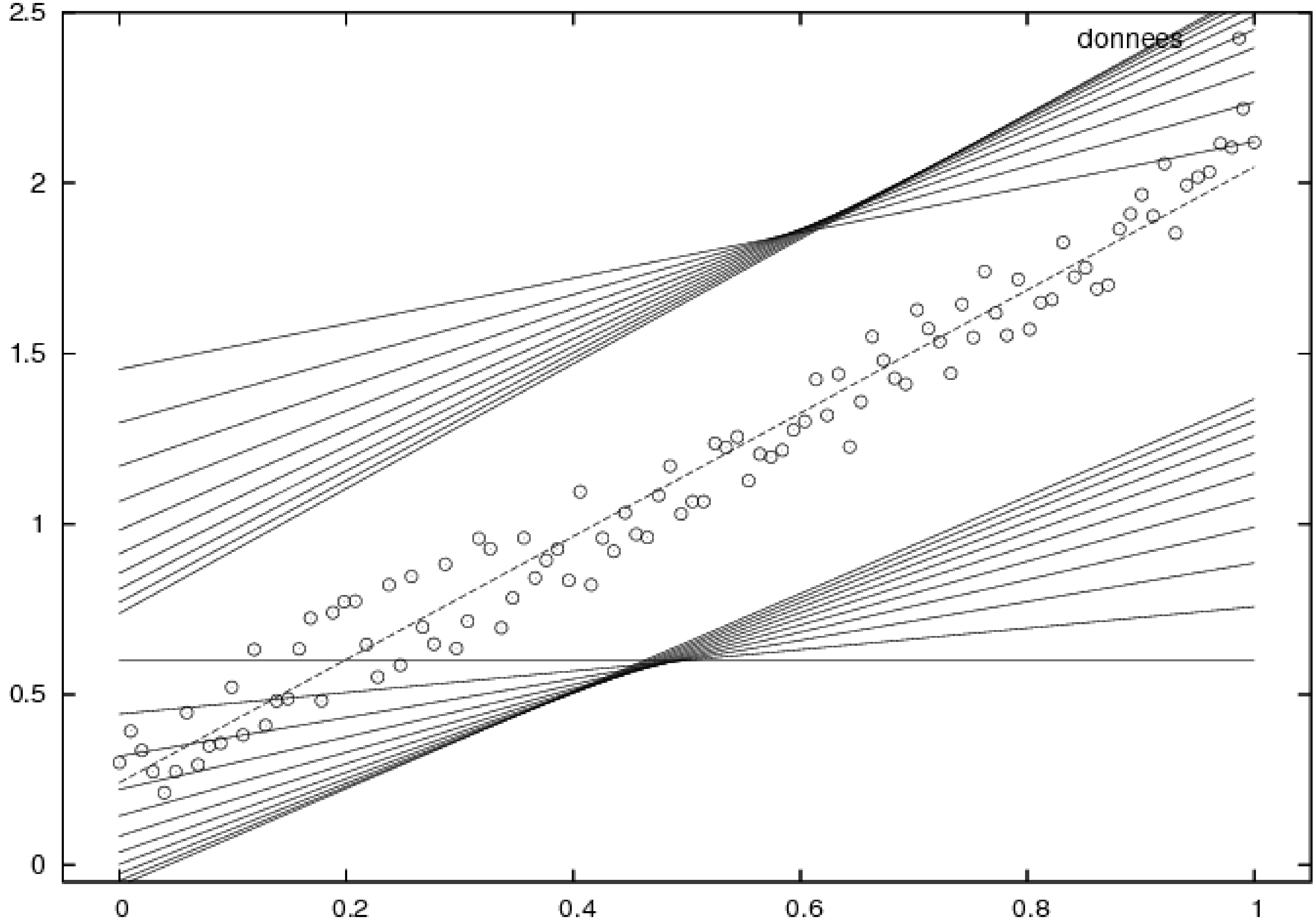
Sur les figures suivantes, on étudie l'influence sur la convergence du choix d'un paramètre de relaxation ρ de plus en plus grand.

L'algorithme finit par diverger !

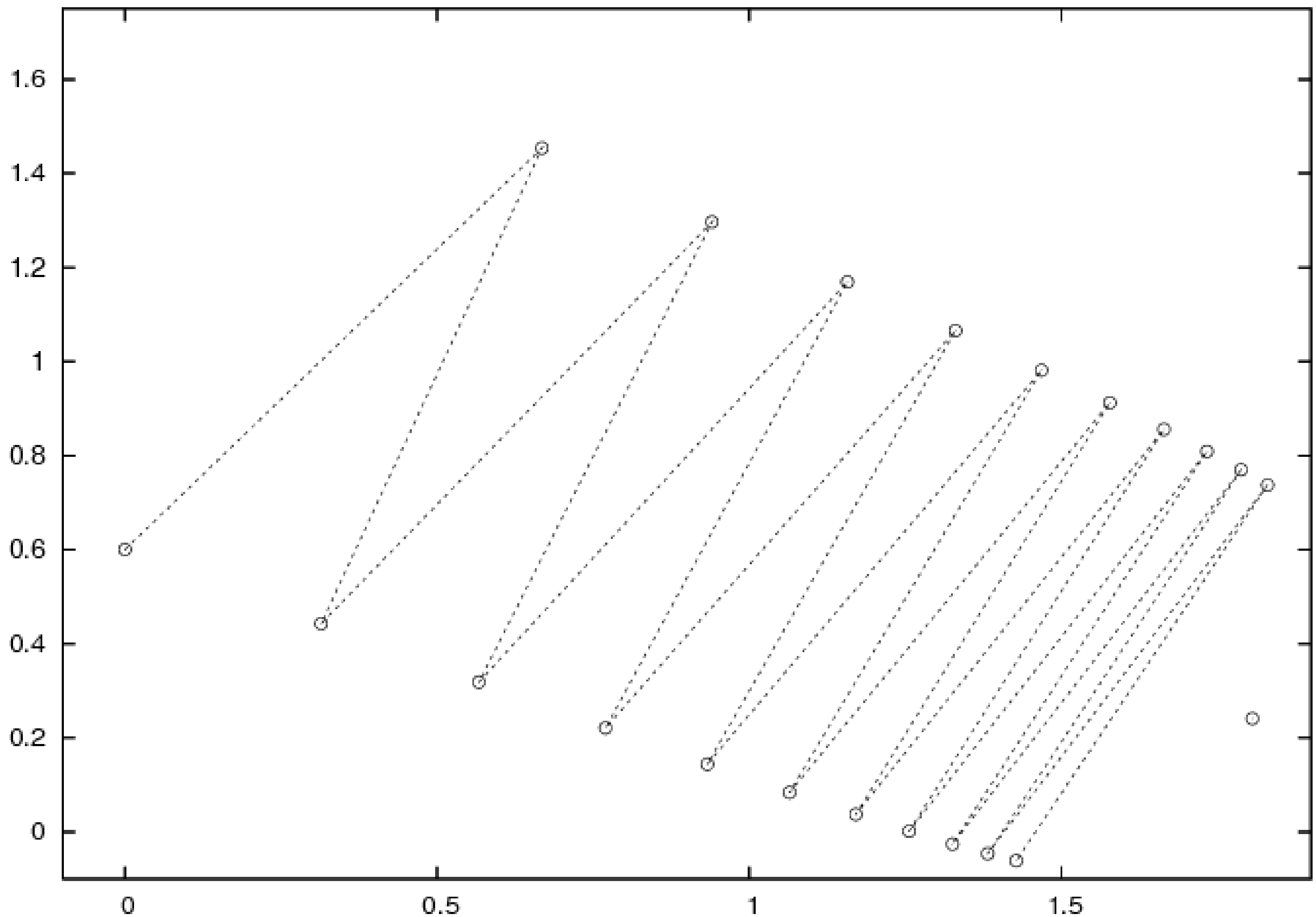

donnees
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Droite des moindres carres. Methode de gradient a pas fixe.



Convergence de la droite des moindres carres. Representation dans le plan des coefficients.



Droite des moindres carres. Methode de gradient a pas fixe.



Convergence de la droite des moindres carres. Representation dans le plan des coefficients.

Modèle dynamique classique

$$m \, \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \,+\, k \, x(t) \,=\, 0$$

$$x(0) \,=\, x_0 \,, \qquad \frac{\mathrm{d}x}{\mathrm{d}t}(0) \,=\, v_0 \,.$$

On pose $\omega \,=\, \sqrt{\dfrac{k}{m}}$; alors

$$x(t) \,=\, x_0 \, \cos(\omega \, t) \,+\, \frac{v_0}{\omega} \, \sin(\omega \, t) \,.$$

On dispose de données $\widetilde{x}(t)$ mais on ne connaît pas x_0, v_0 et ω !!

Fonction coût

$$\lambda \,=\, (x_0, \, v_0, \, \omega)^{\mathrm{t}} \,, \qquad J(\lambda) \,\equiv\, \frac{1}{2} \int_0^T |x(t) - \widetilde{x}(t)|^2 \, \mathrm{d}t$$

Erreur au sens des moindres carrés.

On introduit l'écart $\epsilon(t) \equiv x(t) - \widetilde{x}(t)$

entre le modèle et les mesures.

$$\frac{\partial J}{\partial x_0} = \int_0^T \epsilon(t) \, \cos(\omega\, t) \, \mathrm{d}t$$

$$\frac{\partial J}{\partial v_0} = \int_0^T \epsilon(t) \, \frac{\sin(\omega\, t)}{\omega} \, \mathrm{d}t$$

$$\frac{\partial J}{\partial \omega} = \int_0^T \epsilon(t) \, \big(- \omega\, x_0 \, \sin(\omega\, t) \, + \, v_0 \, \cos(\omega\, t) \big) \, \mathrm{d}t$$

test : on cherche à retrouver la fonction

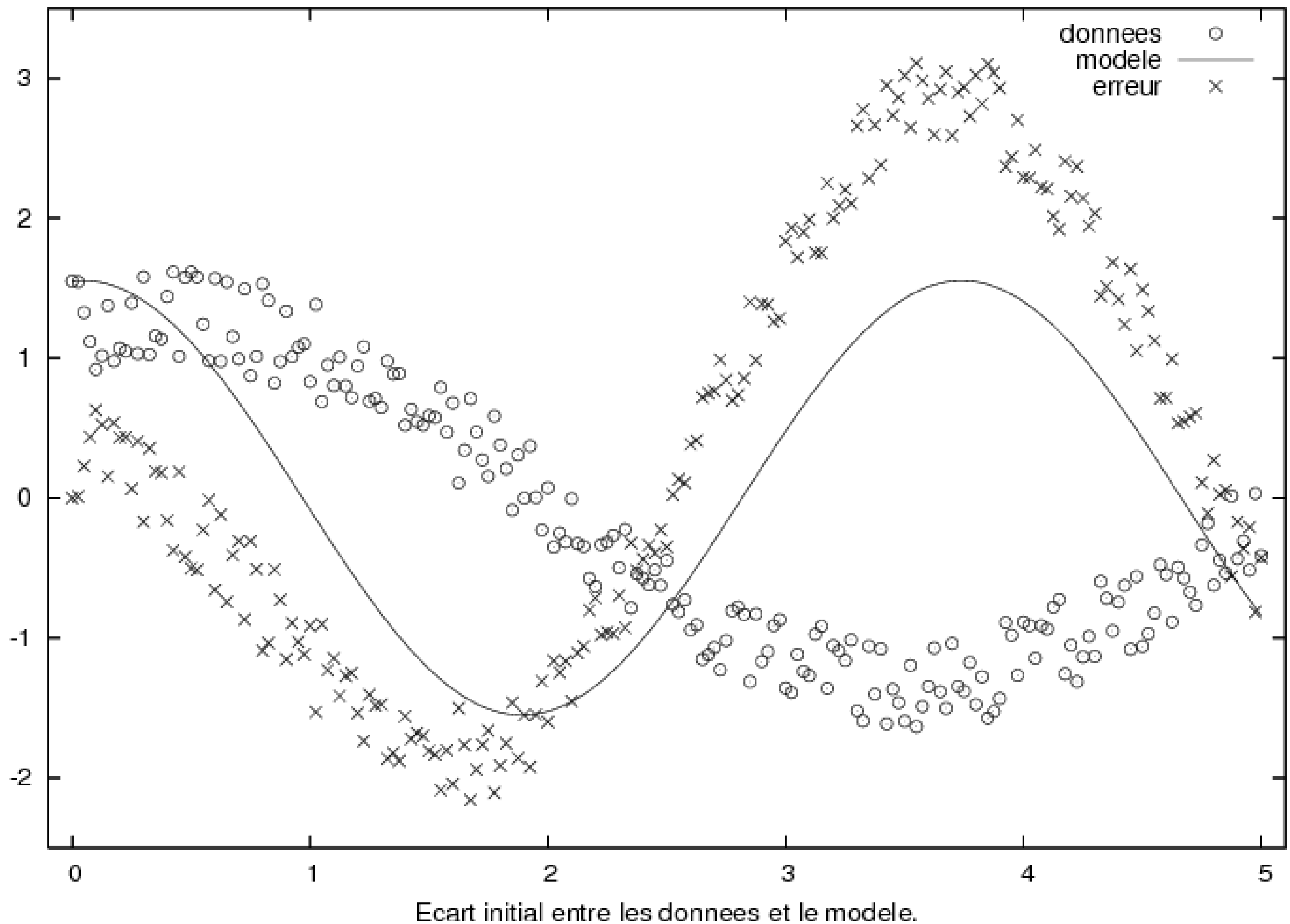
$$x(t) \, = \, 1,2 \, \cos t \, + \, 0,5 \, \sin t$$

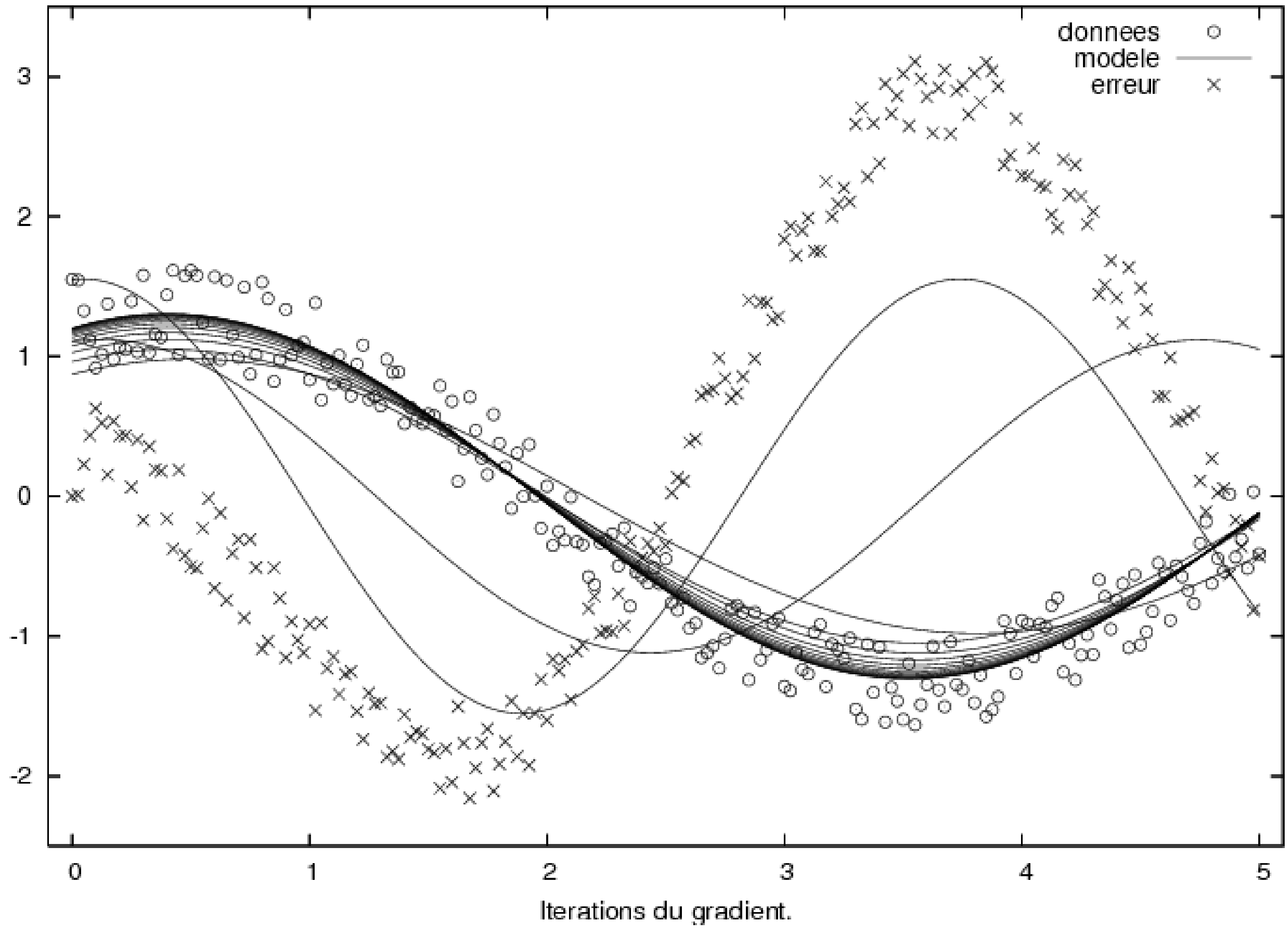
avec des données bruitées sur l'intervalle $[0, \, 5]$

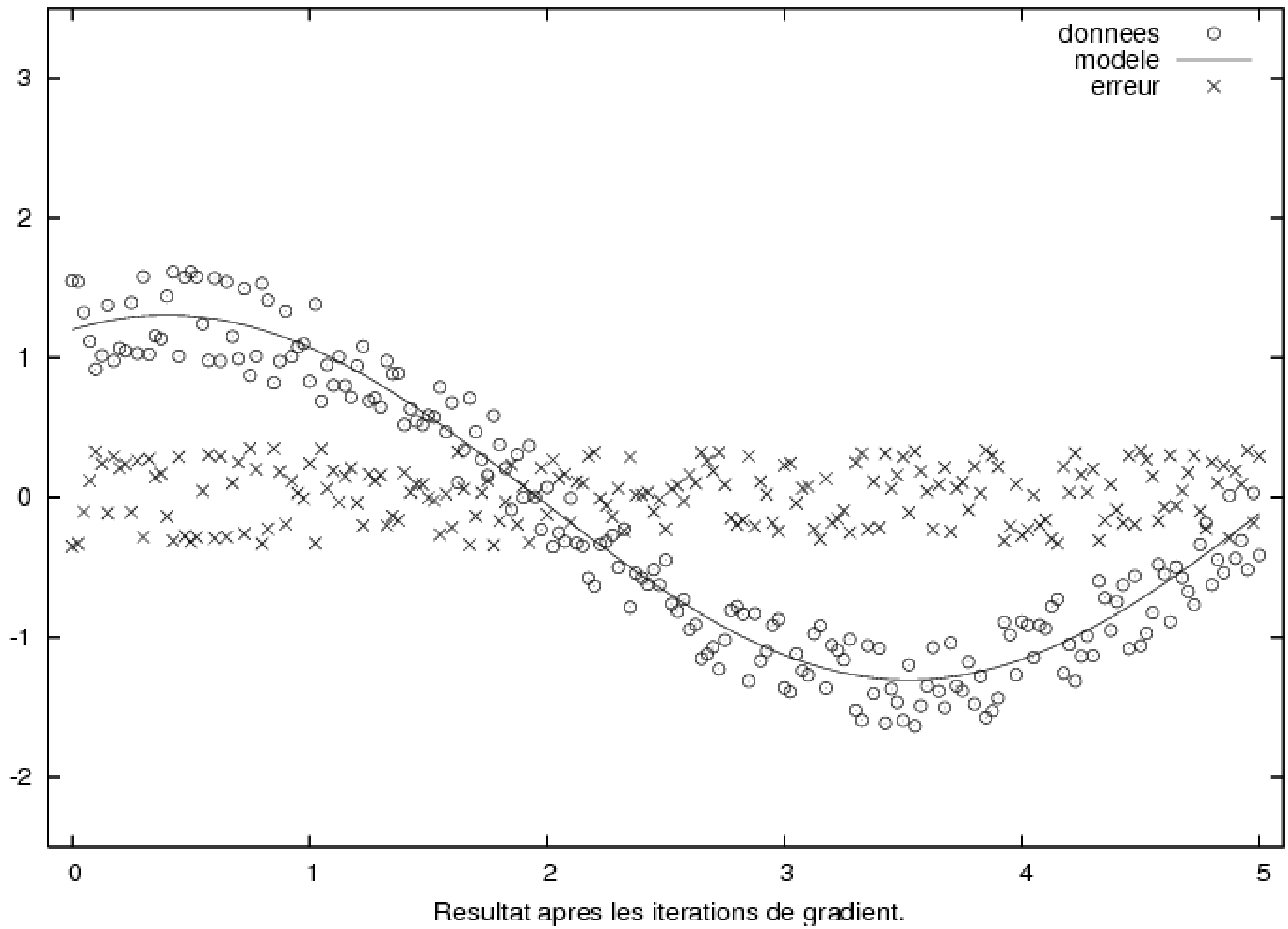
à partir de la fonction

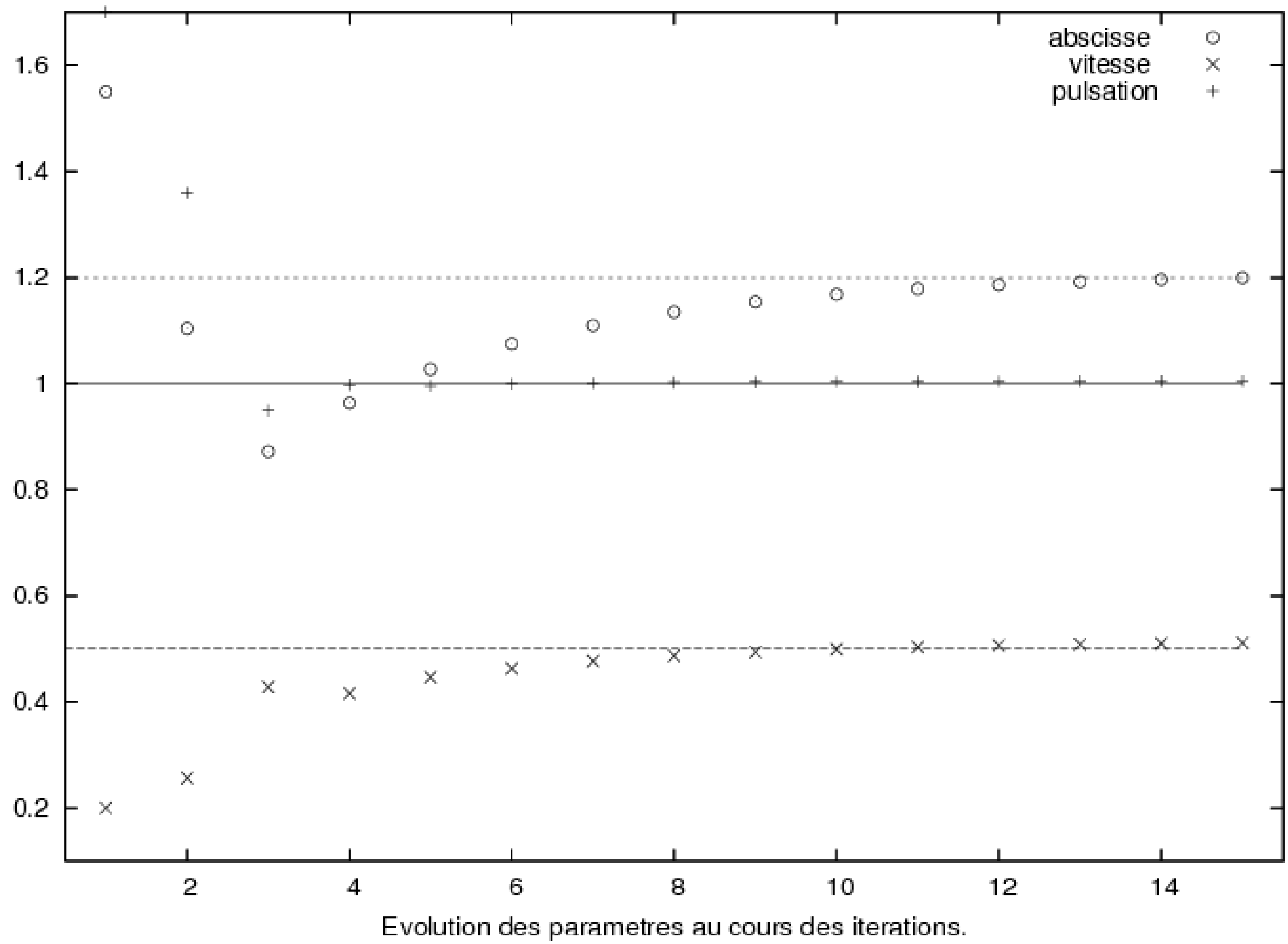
$$x_i(t) \, = \, 1,55 \, \cos(1,7\, t) \, + \, 0,2 \, \sin(1,7\, t)$$

On utilise encore l'algorithme du gradient : $\lambda_{k+1} \, = \, \lambda_k \, - \, \rho \, \nabla J(\lambda_k) \, .$


donnees
modele
erreur
3
2
1
0
-1
-2
0
1
2
3
4
5
Ecart initial entre les donnees et le modele.


donnees
modele
erreur
3
2
1
0
-1
-2
0
1
2
3
4
5
Iterations du gradient.


donnees
modele
erreur
3
2
1
0
-1
-2
0
1
2
3
4
5
Resultat apres les iterations de gradient.


abscisse
vitesse
pulsation
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
2
4
6
8
10
12
14
Evolution des parametres au cours des iterations.

Idée de Pontryaguine (1950) :
écrire l'équation d'évolution comme une **contrainte**

Introduire un Lagrangien :

$$\mathcal{L}(x,\, \lambda,\, p) \,\equiv\, J(\lambda) \,+\, \int_0^T p(t) \Big(m \, \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \,+\, k\, x(t) \Big) \, \mathrm{d}t$$

On remarque que $\mathcal{L}$ est identique à $J(\lambda)$
si $x(t)$ est un oscillateur harmonique

Mais on traite $x(t)$ comme une variable indépendante

et on remarque que $J \,\equiv\, \dfrac{1}{2} \displaystyle\int_0^T \lvert x(t) - \widetilde{x}(t) \rvert^2 \, \mathrm{d}t$

est en fait une fonction de $x(t)$!

Dans une variation arbitraire $(\delta x,\, \delta\lambda,\, \delta p)$ des paramètres $(x,\, \lambda,\, p)$,

on a tout simplement $\qquad \delta J \,=\, \displaystyle\int_0^T \big(x(t) - \widetilde{x}(t) \big) \, \delta x(t) \, \mathrm{d}t$

On a $$\delta\bigg(\int_0^T p(t) \Big(m \, \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \, + \, k \, x(t) \Big) \, \mathrm{d}t \bigg) \, =$$

$$\int_0^T \delta p(t) \Big(m \, \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \, + \, k \, x(t) \Big) \, \mathrm{d}t \, + \int_0^T p(t) \, m \, \delta \Big(\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \Big) \, \mathrm{d}t \, + \int_0^T p(t) \, \delta \Big(k \, x(t) \Big) \, \mathrm{d}t$$

Trois termes très différents à traiter

- le premier est identiquement nul
  si $x(t)$ est un oscillateur harmonique
- on garde le second pour plus tard !
- le troisième s'écrit (puisque $k \, = \, m \, \omega^2$)

$$\int_0^T p(t) \, \delta \Big(k \, x(t) \Big) \, \mathrm{d}t \, =$$

$$= \int_0^T 2 \, m \, \omega \, p(t) \, x(t) \, \delta\omega \, \mathrm{d}t \, + \int_0^T p(t) \, k \, \delta x(t) \, \mathrm{d}t$$

Que vaut $\displaystyle \int_0^T p(t)\, m\, \delta\Big(\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}\Big)\, \mathrm{d}t$?

On remarque d'abord que si $y(t) \,=\, x(t) + \delta x(t)$

on a par dérivation par rapport au temps

$$\frac{\mathrm{d}y}{\mathrm{d}t} \,=\, \frac{\mathrm{d}x}{\mathrm{d}t} \,+\, \frac{\mathrm{d}}{\mathrm{d}t}\Big(\delta x(t)\Big)\,;$$

donc l'écart $\displaystyle \delta\Big(\frac{\mathrm{d}x}{\mathrm{d}t}\Big) \equiv \frac{\mathrm{d}y}{\mathrm{d}t} - \frac{\mathrm{d}x}{\mathrm{d}t}$ entre les deux dérivées

vaut $\displaystyle \frac{\mathrm{d}}{\mathrm{d}t}\Big(\delta x(t)\Big)\,.$

En résumé, $\displaystyle \delta\Big(\frac{\mathrm{d}x}{\mathrm{d}t}\Big) \,=\, \frac{\mathrm{d}}{\mathrm{d}t}\Big(\delta x(t)\Big)\,.$

La difficulté est que cet écart $\displaystyle \delta\Big(\frac{\mathrm{d}x}{\mathrm{d}t}\Big)$ n'est **pas** indépendant de $\delta x(t)$!

On remarque qu'on a toute liberté de **choisir** le multiplicateur $p(t)$

On a $$\int_0^T p(t)\, m\, \delta\Big(\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}\Big)\, \mathrm{d}t \,=\, \int_0^T p(t)\, m\, \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big(\delta x(t)\Big)\, \mathrm{d}t$$

$$= \Big[p(t)\, m\, \frac{\mathrm{d}}{\mathrm{d}t}\Big(\delta x(t)\Big)\Big]_0^T \,-\, \int_0^T \frac{\mathrm{d}p}{\mathrm{d}t}\, m\, \delta\Big(\frac{\mathrm{d}x}{\mathrm{d}t}\Big)\, \mathrm{d}t$$

$$= - p(0)\, m\, \delta v_0 \,-\, \int_0^T \frac{\mathrm{d}p}{\mathrm{d}t}\, m\, \delta\Big(\frac{\mathrm{d}x}{\mathrm{d}t}\Big)\, \mathrm{d}t \qquad \text{si on } \textbf{choisit} \quad p(T) \,=\, 0$$

$$= - p(0)\, m\, \delta v_0 \,-\, \Big[\frac{\mathrm{d}p}{\mathrm{d}t}\, m\, \delta x(t)\Big]_0^T \,+\, \int_0^T \frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\, m\, \delta x(t)\, \mathrm{d}t$$

$$= - p(0)\, m\, \delta v_0 \,+\, \frac{\mathrm{d}p}{\mathrm{d}t}(0)\, m\, \delta x_0 \,+\, \int_0^T \frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\, m\, \delta x(t)\, \mathrm{d}t$$

$$\text{si on } \textbf{choisit} \quad \frac{\mathrm{d}p}{\mathrm{d}t}(T) \,=\, 0$$

On a dans par ordre d'apparition, lorsque $p(T) = 0$ et $\dfrac{\mathrm{d}p}{\mathrm{d}t}(T) = 0$:

$$\begin{aligned}\delta\mathcal{L} = &\int_0^T \big(x(t) - \widetilde{x}(t)\big)\,\delta x(t)\,\mathrm{d}t + \int_0^T \delta p(t)\left(m\,\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + k\,x(t)\right)\mathrm{d}t \\ &- p(0)\,m\,\delta v_0 + \frac{\mathrm{d}p}{\mathrm{d}t}(0)\,m\,\delta x_0 + \int_0^T \frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\,m\,\delta x(t)\,\mathrm{d}t \\ &+ \int_0^T 2\,m\,\omega\,p(t)\,x(t)\,\delta\omega\,\mathrm{d}t + \int_0^T p(t)\,k\,\delta x(t)\,\mathrm{d}t\end{aligned}$$

et après regroupement

$$\begin{aligned}\delta\mathcal{L} = &\frac{\mathrm{d}p}{\mathrm{d}t}(0)\,m\,\delta x_0 - p(0)\,m\,\delta v_0 + \int_0^T 2\,m\,\omega\,p(t)\,x(t)\,\delta\omega\,\mathrm{d}t + \\ &+ \int_0^T \left(\frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\,m + p(t)\,k + \big(x(t) - \widetilde{x}(t)\big)\right)\delta x(t)\,\mathrm{d}t\end{aligned}$$

si $x(t)$ est un oscillateur harmonique.

On **choisit** le multiplicateur de Lagrange $p(t)$
de sorte d'annuler les termes qui ne nous plaisent pas.

Il **suffit** donc de poser

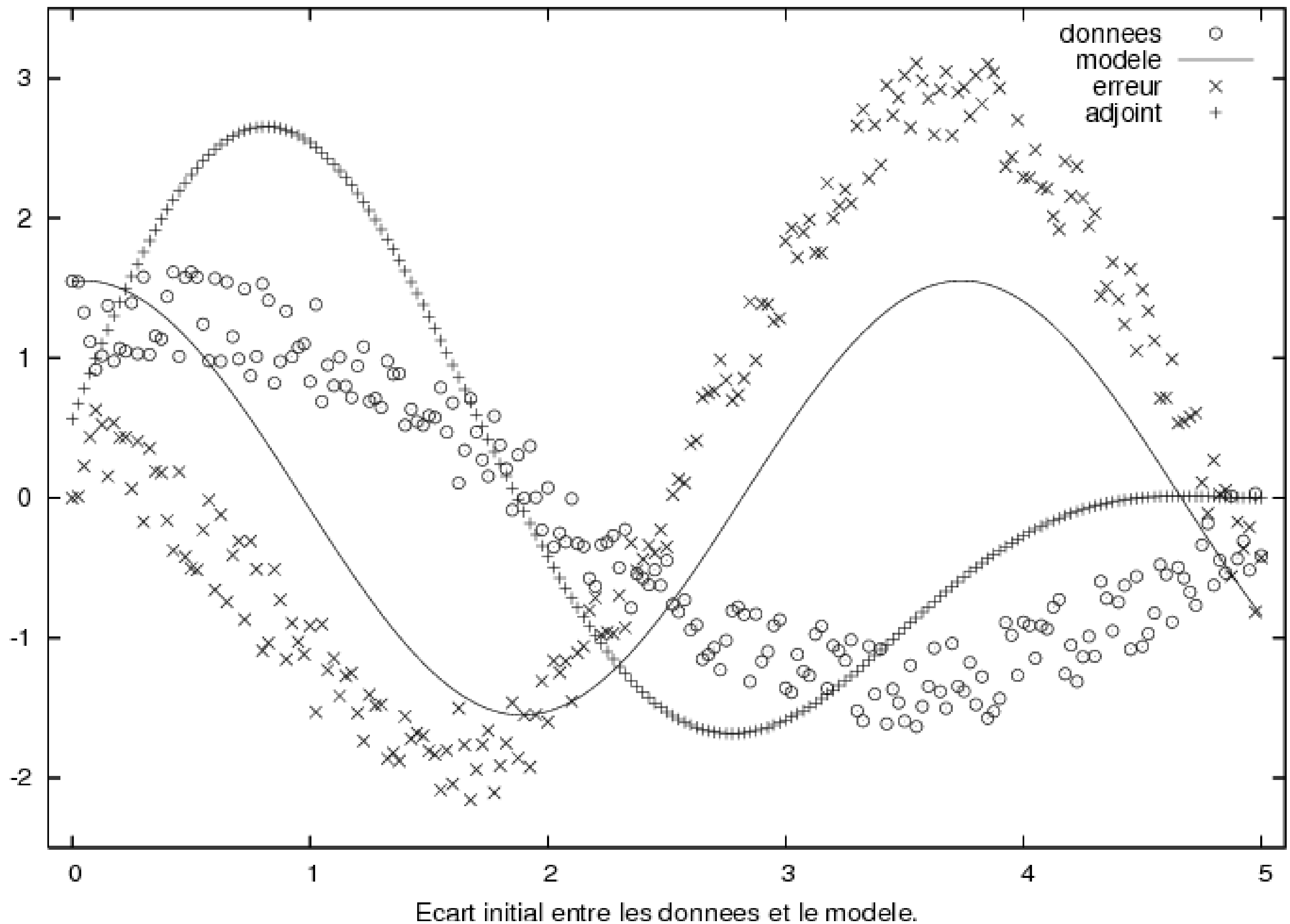
$$\frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\, m + p(t)\, k + \big(x(t) - \widetilde{x}(t)\big) = 0\,, \qquad 0 < t < T$$

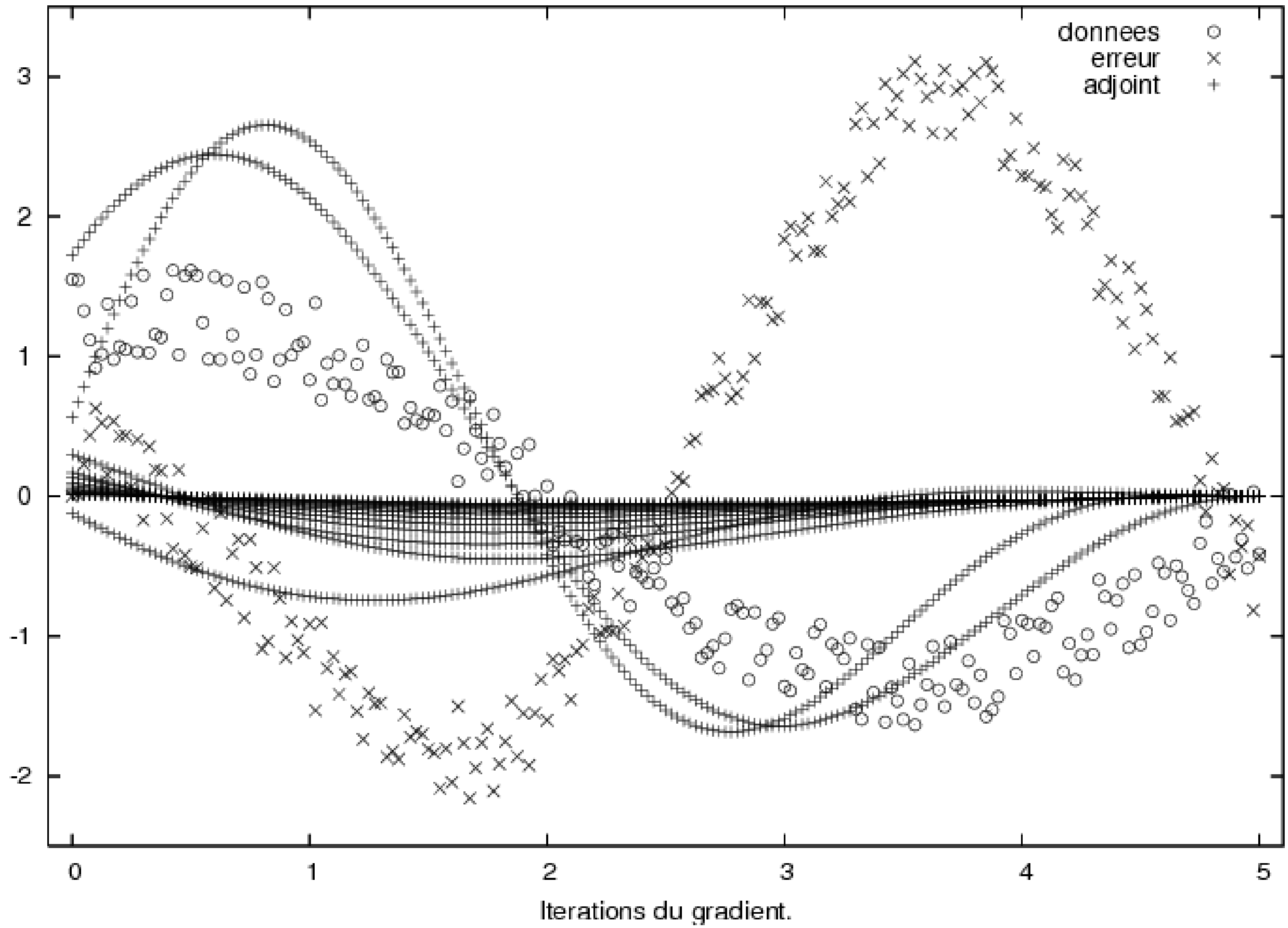
$$p(T) = 0 \quad \text{et} \quad \frac{\mathrm{d}p}{\mathrm{d}t}(T) = 0$$

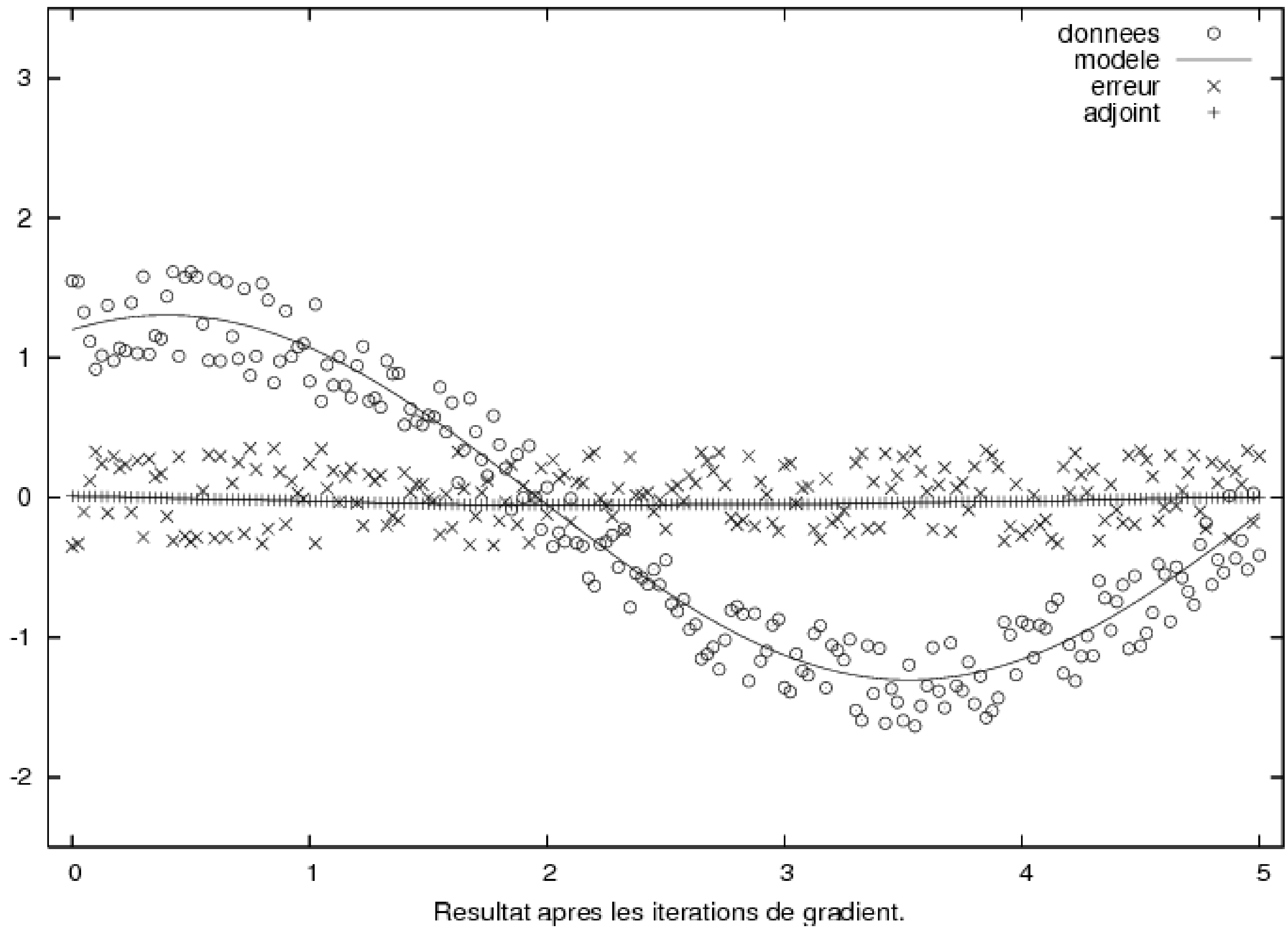
Alors on a facilement

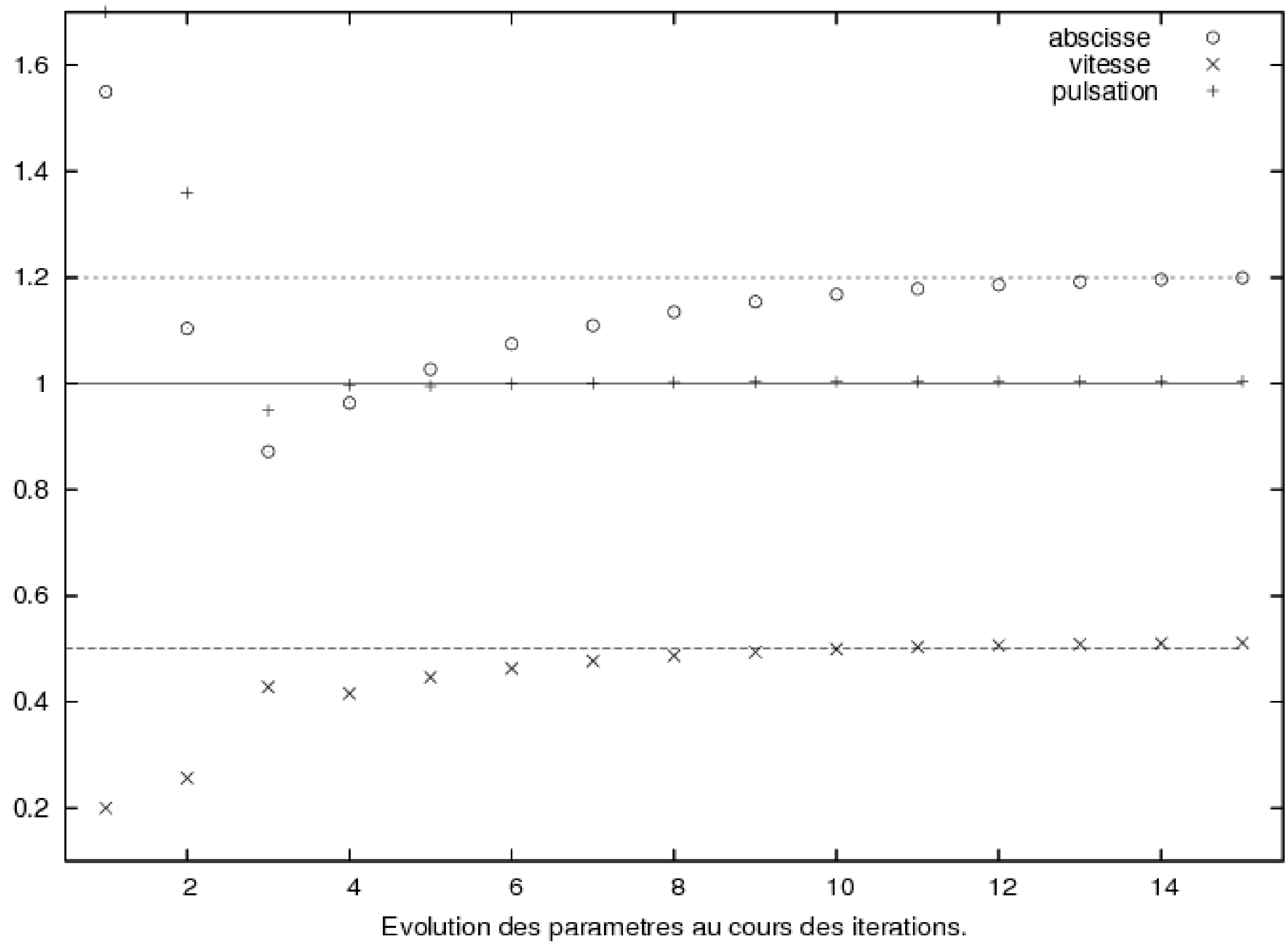
$$p(t)\, m = -\int_T^t \frac{\sin(\theta - t)}{\omega}\, \big(x(\theta) - \widetilde{x}(\theta)\big)\, \mathrm{d}\theta$$

On peut visualiser l'état adjoint
pour l'oscillateur harmonique étudié précédemment


donnees
modele
erreur
adjoint
3
2
1
0
-1
-2
0
1
2
3
4
5
Ecart initial entre les donnees et le modele.


donnees
erreur
adjoint
3
2
1
0
-1
-2
0
1
2
3
4
5
Iterations du gradient.


donnees
modele
erreur
adjoint
3
2
1
0
-1
-2
0
1
2
3
4
5
Resultat apres les iterations de gradient.


abscisse
vitesse
pulsation
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
2
4
6
8
10
12
14
Evolution des parametres au cours des iterations.

Système dynamique où le vecteur d'état $y(t\,;\,v(\bullet))$

dépend du temps t

est commandé par un ensemble de variables $v(t)$

grâce à une équation différentielle ordinaire :

$$\frac{\mathrm{d}y}{\mathrm{d}t} \;=\; f(y(t),\, v(t),\, t)$$

à laquelle on joint une condition initiale

$$y(t_0\,;\,v(\bullet)) = x\,.$$

On cherche une solution optimale

associée au contrôle optimal $t \longmapsto u(t)$

de façon à minimiser la fonction coût J suivante :

$$J(v(\bullet)) \;\equiv\; \lambda(y(T)) \;+\; \int_{t_0}^{T} L(y(t),\, v(t),\, t)\,\mathrm{d}t\;,$$

où $L(\bullet,\,\bullet,\,\bullet)$ et $\lambda(\bullet)$ sont des fonctions réelles fixées.

Le système des équations adjointes s'écrit sous la forme

$$\frac{\mathrm{d}p}{\mathrm{d}t} + p\,\frac{\partial f}{\partial y} + \frac{\partial L}{\partial y} \;=\; 0$$

avec une condition finale

$$p(T) \;=\; \frac{\partial \lambda}{\partial y}(y(T))\,.$$

Considérer l'équation différentielle qui décrit l'évolution de l'état $y(t)$ comme une **contrainte** que doit satisfaire la variable liée y et de la traiter comme telle dans un Lagrangien.

On introduit donc un multiplicateur de Lagrange p associé à cette contrainte.

Ce dernier est un vecteur ligne fonction du temps : $p = p(t)$.

On pose

$$\mathcal{L}(y,\, v,\, p) \,\equiv\, \lambda(y(T)) \,+\, \int_{t_0}^{T} L(y,\, v,\, t)\,\mathrm{d}t \,-\, \int_{t_0}^{T} p\,\Big(\frac{\mathrm{d}y}{\mathrm{d}t} - f(y,\, v,\, t)\Big)\,\mathrm{d}t\,.$$

L'équation satisfaite par le multiplicateur est obtenue
en écrivant la différentielle de $\mathcal{L}(y,\, v,\, p)$
et en éliminant les termes
dus aux variations des dérivées temporelles de y
par intégration par parties.

On note δy, δv et δp les variations de y, v, et p respectivement.

Alors $$\delta \mathcal{L} \;=\; \frac{\partial \lambda}{\partial y}(y(T))\,\delta y(T) \;+\; \int_{t_0}^{T} \Big[\, \frac{\partial L}{\partial y}\,\delta y + \frac{\partial L}{\partial v}\,\delta v \,\Big]\, \mathrm{d}t$$

$$-\; \int_{t_0}^{T} p \,\Big(\frac{\mathrm{d}\delta y}{\mathrm{d}t} - \frac{\partial f}{\partial y}\delta y - \frac{\partial f}{\partial v}\delta v \Big)\, \mathrm{d}t \;-\; \int_{t_0}^{T} \delta p \,\Big(\frac{\mathrm{d}y}{\mathrm{d}t} - f(y,\, v,\, t) \Big)\, \mathrm{d}t$$

$$=\; \frac{\partial \lambda}{\partial y}(y(T))\,\delta y(T) \;+\; \int_{t_0}^{T} \Big[\, \frac{\partial L}{\partial y} + p\,\frac{\partial f}{\partial y} \,\Big]\, \delta y \, \mathrm{d}t$$

$$+\; \int_{t_0}^{T} \Big[\, \frac{\partial L}{\partial v} + p\,\frac{\partial f}{\partial v} \,\Big]\, \delta v \, \mathrm{d}t \;-\; \Big[\, p\,\delta y \,\Big]_{t_0}^{T} \;+\; \int_{t_0}^{T} \frac{\mathrm{d}p}{\mathrm{d}t}\; \delta y \, \mathrm{d}t\,.$$

$$\delta \mathcal{L} \;=\; \Big(\frac{\partial \lambda}{\partial y}(y(T)) - p(T) \Big)\, \delta y(T) \;+\; \int_{t_0}^{T} \Big[\, \frac{\mathrm{d}p}{\mathrm{d}t} + p\,\frac{\partial f}{\partial y} + \frac{\partial L}{\partial y} \,\Big]\, \delta y \;\mathrm{d}t$$

$$+\; \int_{t_0}^{T} \Big[\, \frac{\partial L}{\partial v} + p\,\frac{\partial f}{\partial v} \,\Big]\, \delta v \;\mathrm{d}t \;+\; p(0)\,\delta x\,.$$

car $\delta y(t_0) = \delta x$ compte tenu de la condition initiale.

$$\delta \mathcal{L} \;=\; \Big(\frac{\partial \lambda}{\partial y}(y(T)) - p(T) \Big)\, \delta y(T) \;+\; \int_{t_0}^{T} \Big[\frac{\mathrm{d}p}{\mathrm{d}t} + p\, \frac{\partial f}{\partial y} + \frac{\partial L}{\partial y} \Big]\, \delta y \;\mathrm{d}t$$
$$+ \int_{t_0}^{T} \Big[\frac{\partial L}{\partial v} + p\, \frac{\partial f}{\partial v} \Big]\, \delta v \;\mathrm{d}t \;+\; p(0)\, \delta x \,.$$

On annule les deux premiers termes du membre de droite
et on trouve l'équation adjointe

$$\frac{\mathrm{d}p}{\mathrm{d}t} + p\, \frac{\partial f}{\partial y} + \frac{\partial L}{\partial y} \;=\; 0$$

qui donne l'évolution du multiplicateur de Lagrange
ainsi que la condition finale

$$p(T) \;=\; \frac{\partial \lambda}{\partial y}(y(T)) \,.$$

Le troisième terme permet de calculer la variation
de la fonctionnelle $J(\bullet)$ dans une variation δv de la commande.

Dans de nombreuses situations, le calcul de l'état adjoint
facilite grandement l'évaluation du gradient de la fonctionnelle.

Difficultés

Non linéarités

Discrétisation des équations différentielles :

faut-il discrétiser l'équation adjointe
ou chercher l'adjoint discret du schéma numérique ?

Passage des modèles discrets ("équations différentielles ordinaires")
aux modèles continus ("équations aux dérivées partielles")

Peut-on utiliser des algorithmes meilleurs que
la descente le long de la ligne de plus grande pente ?