

OBSERVATOIRE ROYAL
DE
BELGIQUE

KONINKLIJKE STERRENWACHT
VAN
BELGIE

GPS Report N° 13

INTRODUCTION AUX PROBLEMES INVERSES

O. FRANCIS

Dec. 1991

Av. Circulaire 3 - 1180 Bruxelles
Ringlaan 3 - 1180 Brussel

Avant-propos

Lors de la rédaction de cette introduction générale à la résolution des problèmes inverses, nous nous sommes largement inspirés des articles et du livre de Tarantola. Les grandes qualités de ses travaux sont la *simplicité* et la *généralité* du formalisme. En effet, à partir de considérations simples de la théorie des probabilités et de la théorie de l'information, une solution générale des problèmes inverses linéaires et non-linéaires est obtenue. Cette solution contient toutes les solutions par moindres carrés publiées précédemment.

Une grande partie de ce rapport est théorique. Cependant, afin de l'illustrer et de démystifier les équations, un problème inverse simple d'interpolation est traité en détails.

1. Introduction.

Tarantola et Valette (1982a) ont montré que le problème inverse peut être formulé comme un problème de combinaison d'informations : une information de type expérimentale sur les données, une information a priori sur les paramètres, et une information théorique. Utilisant leur approche, la solution générale du problème inverse non linéaire est unique et consistante (c'est-à-dire que résoudre le même problème avec les mêmes données, mais avec un système différent de paramètres, ne change pas la solution).

Les contraintes minimales nécessaires à la formulation d'un problème inverse doivent :

1. inclure aussi bien le problème linéaire que le problème fortement non linéaire.
2. inclure les problèmes sur-déterminés et sous-déterminés.
3. être consistantes vis-à-vis d'un changement de variables.
4. permettre n'importe quelle distribution générale des erreurs des données.
5. permettre l'incorporation formelle de n'importe quelle hypothèse a priori.
6. permettre l'incorporation d'erreurs théoriques d'une manière naturelle. Ces erreurs théoriques proviennent de l'existence de simplifications théoriques et de mauvaises ou fausses paramétrisations du problème.

Il est démontré que toutes ces contraintes peuvent être respectées si le problème inverse est formulé en utilisant une extension simple de la théorie des probabilités et de la théorie de l'information.

La présente approche se limite à l'étude des systèmes qui peuvent être décrits à l'aide d'un *ensemble fini de paramètres*. Cette limitation entraîne que l'on ne sera capable que de traiter des caractéristiques *quantitatives* des systèmes et que, pour les décrire, on utilisera des fonctions plutôt que des paramètres discrets.

Attirons l'attention sur le problème de l'échantillonnage qui n'est certes pas trivial. En effet, si l'intervalle d'échantillonnage est plus grand que la longueur de décorrélation du bruit, les erreurs sur les données peuvent être considérées comme indépendantes; ceci peut ne plus être vrai lorsque l'on densifie l'échantillonnage. On formule alors explicitement l'hypothèse que, dans la suite, la discrétisation a été faite avec soin de telle sorte que diminuer le pas d'échantillonnage n'altère pas les résultats.

2. Paramètres et information.

On définit ζ un système physique au sens large, c.-à-d. un système physique lui-même, la famille d'instruments de mesure et leurs mesures. Ce système, qui est discret ou bien a été discrétisé, peut être décrit par un ensemble fini de paramètres $X = \{X_1, \dots, X_m\}$ dont $x = \{x_1, \dots, x_m\}$ est un ensemble de valeurs particulières de l'ensemble X ($x \in X$). Enfin E^m est l'espace de dimension m où les paramètres X prennent leurs valeurs, cet espace

est appelé l'espace modèle ou espace des paramètres.

Définition : un système physique est dit paramétrisable lorsqu'il peut être décrit par un ensemble X de paramètres. La paramétrisation d'un système physique n'est pas unique: deux paramétrisations sont équivalentes si elles sont liées entre-elles par une bijection.

Le degré de connaissance que l'on peut avoir sur les valeurs des paramètres d'un système peut varier de la connaissance totale à l'ignorance complète. Ceci nous conduit à un premier postulat :

Tout état de connaissance des valeurs X peut être décrit par une fonction de densité mesurable $f(x)$ réelle, positive, intégrable localement au sens de Lebesgue telle que sa mesure positive définie par :

$$m(A) = \int_A f(x) dx \quad (A \subset E^m) \quad (1)$$

est absolument continue pour la mesure de Lebesgue définie sur E^m .

Définition : si $m(E^m)$ est finie alors $f(x)$ est normalisée de sorte que $m(E^m) = 1$. Dans ce cas, $f(x)$ est appelée densité de probabilité et $m(A)$ notée $p(A)$ est la probabilité de A .

Définition : on définit une densité particulière $\mu(x)$ associée à l'état d'ignorance complète. Souvent, à cet état d'ignorance complète correspond une fonction uniforme $\mu(x) = \text{const.}$, parfois pas. On fait l'hypothèse que $\mu(x) \neq 0$ partout dans E^m .

Définition : si, à l'état d'ignorance complète correspond une densité de probabilité $\mu(x)$ alors le contenu d'information de n'importe quelle densité de probabilité $f(x)$ est défini par :

$$I(f;\mu) = \int f(x) \text{Log} \frac{f(x)}{\mu(x)} dx \quad (2)$$

Cette définition a les propriétés suivantes qui se vérifient facilement :

a) I est invariant pour tout changement de variables

$$I(f;\mu) = I(f';\mu')$$

b) l'information ne peut être négative

$$I(f;\mu) \geq 0$$

c) l'information de l'état d'ignorance complète est nulle

$$I(\mu;\mu) = 0$$

la réciproque est vraie aussi

$$I(f;\mu) = 0 \Rightarrow f = \mu$$

Nous dirons que toute densité de probabilité (ou, par extension, fonction de densité) $f_i(x)$ est équivalente à un état d'information qui sera notée s_i .

Définition : l'opération qui combine deux fonctions de densité f_i et f_j (représentatifs de deux états d'information s_i et s_j) pour obtenir f (représentatif de la conjonction $s=s_i \wedge s_j$) notée :

$$f = f_i \wedge f_j \quad (3)$$

doit satisfaire les conditions suivantes :

- a) $f_i \wedge f_j$ doit être une fonction de densité. En particulier, le contenu d'information de $f_i \wedge f_j$ doit être invariant vis-à-vis de tout changement de paramètres.
- b) l'opération doit être commutative.
- c) l'opération doit être associative.
- d) la conjonction de tout état d'information f_i avec l'information nulle μ doit donner f_i , il ne peut y avoir de perte d'information : en d'autres termes, μ est l'élément neutre de l'opération.
- e) \forall ensemble A mesurable $\subset E^m$:

$$\left. \begin{array}{l} \int_A f_i dx = 0 \\ \int_A f_j dx = 0 \end{array} \right\} \Leftrightarrow \int_A (f_i \wedge f_j) dx = 0 \quad (4)$$

ce qui signifie qu'une condition nécessaire et suffisante pour que $f_i \wedge f_j$ donne une probabilité nulle à un sous-ensemble A est que soit f_i ou bien soit f_j donne une probabilité nulle pour A.

Cette dernière condition entraîne que la mesure engendrée par $f_i \wedge f_j$ est absolument continue vis-à-vis de la mesure engendrée par f_i et f_j respectivement. Par le théorème de

Nykodym, on peut montrer que $f_i \wedge f_j$ doit nécessairement s'écrire sous la forme :

$$f_i \wedge f_j = f_i \cdot f_j \cdot \Phi(f_i, f_j) \quad (5)$$

où $\Phi(f_i, f_j)$ est localement intégrable au sens de Lebesgue et est une fonction positive.

Le choix le plus simple pour $\Phi(f_i, f_j)$ qui satisfasse les conditions b) et c) est de la choisir indépendante de f_i ou de f_j . La condition d) impose alors:

$$\Phi(f_i, f_j) = 1/\mu \quad (6)$$

et la condition a) est automatiquement vérifiée.

Tout ceci conduit à la définition suivante :

Soient s_i et s_j deux états d'information représentés par les fonctions de densité f_i et f_j et μ la fonction de densité de l'état d'ignorance complète. Par définition, la conjonction de s_i et s_j notée $s = s_i \wedge s_j$ est un état d'information représenté par la fonction de densité $f(x)$ donnée par :

$$f(x) = \frac{f_i(x) \cdot f_j(x)}{\mu(x)} \quad (7)$$

Remarques :

1. La fonction de densité $f(x)$ n'est pas nécessairement normalisable, mais à part certains cas, dans la plupart des problèmes réels lorsque l'une des deux fonctions de densité $f_i(x)$ ou $f_j(x)$ est normalisable, la fonction de densité $f(x)$ l'est aussi et est de ce fait une densité de probabilité;
2. Cette définition (7) permettra d'obtenir une solution simple du problème inverse généralisé.
3. La densité de probabilité conditionnelle peut être définie comme un cas particulier de la conjonction d'états d'information.
4. La conjonction définie ci-dessus ne peut être utilisée que pour combiner deux états d'information que si et seulement si ces états d'informations ont été obtenu indépendamment l'un de l'autre.
5. Les fonctions de densité sont interprétées en terme de *connaissance humaine* plutôt qu'en terme de propriétés statistiques.

3. Données et information a priori.

Parmi les paramètres X décrivant un système ζ , les paramètres décrivant les sorties ('outputs') des instruments de mesure sont appelées données $D=(D_1, \dots, D_r)$. Le reste des paramètres sont les paramètres à proprement parler et noté $P=(P_1, \dots, P_s)$. Partitionnant X en $X=(D,P)$, toute fonction de densité sur X peut s'écrire :

$$f(x) = f(\mathbf{d}, \mathbf{p}) \quad (8)$$

On fait l'hypothèse de l'existence d'une fonction de densité

$$\rho(x) = \rho(\mathbf{d}, \mathbf{p}) \quad (9)$$

appelée fonction de densité a priori, représentant à la fois les résultats des mesures et toute information a priori sur les paramètres.

4. Relations théoriques

Habituellement, une relation théorique est une fonctionnelle agissant sur les valeurs des paramètres :

$$F(x) = F(\mathbf{d}, \mathbf{p}) = 0 \quad (10)$$

cette écriture peut généralement se simplifier sous la forme:

$$\mathbf{d} = G(\mathbf{p}) \quad (11)$$

Cette relation est trop restrictive et ne correspond pas à la réalité. Dans la plupart des cas même si la valeur de \mathbf{p} est donnée, on ne peut pas calculer exactement la valeur correspondante de \mathbf{d} , car la théorie est incomplète (paramétrisation trop grossière, ...). Dans de tels cas, on doit plutôt parler de densité de probabilité de \mathbf{d} connaissant \mathbf{p} , c'est-à-dire de densité de probabilité conditionnelle :

$$\vartheta(\mathbf{d}|\mathbf{p}) \quad (12)$$

En toute généralité, on fera l'hypothèse que toute relation théorique peut être représentée par une fonction de densité :

$$\vartheta(x) = \vartheta(\mathbf{d}, \mathbf{p}) \quad (13)$$

or par définition de la probabilité conditionnelle, on a

$$\vartheta(d,p) = \vartheta(d|p) \vartheta_p(p) \quad (14)$$

où $\vartheta_p(p)$ est la fonction de densité marginale de p . On peut réécrire l'équation sous la forme :

$$\vartheta(d,p) = \vartheta(d|p) \mu_p(p) \quad (15)$$

où $\mu_p(p)$ est la fonction de densité de l'information nulle. Dans le cas particulier d'une théorie exacte, on aurait $\vartheta(d|p) = \delta(d-Gp)$ où δ est la fonction de Dirac :

$$\vartheta(d,p) = \delta(d-Gp) \mu_p(p) \quad (16)$$

Dans le cas où le calcul rigoureux de $\vartheta(d|p)$ ne peut pas être fait et si toutefois on a une idée sur les bornes d'erreurs sur la validité de la théorie σ_T , on choisit pour $\vartheta(d,p)$ quelque chose du type :

$$\vartheta(d|p) = \text{const.} \exp\left\{-\frac{1}{2} \frac{\|d-G(p)\|^2}{\sigma_T^2}\right\} \quad (17)$$

qui peut être une bonne approximation de l'erreur théorique.

5. Solution du problème inverse.

Soient ζ un système physique

X une paramétrisation de ζ

$\mu(x)$ l'état d'information nulle sur le système ζ

$\rho(x)$ toute information a priori sur le système (en particulier, les résultats des mesures et les contraintes a priori sur les paramètres)

$\vartheta(x)$ représente les relations théoriques entre les données et les paramètres.

La conjonction de $\rho(x)$ et $\vartheta(x)$ donne un nouvel état d'information appelé état d'information a posteriori. En utilisant l'équation (7), on obtient la fonction de densité correspondante :

$$\sigma(x) = \frac{\rho(x) \cdot \vartheta(x)}{\mu(x)}$$

(18)

On sépare les paramètres X en $X=(D,P)$ avec les données d'un côté et les paramètres de l'autre. L'équation (18) devient alors :

$$\sigma(d,p) = \frac{\rho(d,p) \cdot \vartheta(d,p)}{\mu(d,p)} \quad (19)$$

Expression à partir de laquelle, on calcule les fonctions de densité a posteriori

$$\sigma_d(d) = \int \frac{\rho(d,p) \cdot \vartheta(d,p)}{\mu(d,p)} dp \quad (20)$$

$$\sigma_p(p) = \int \frac{\rho(d,p) \cdot \vartheta(d,p)}{\mu(d,p)} dd \quad (21)$$

Cette dernière équation permet de transférer aux paramètres via les relations théoriques l'information contenue dans les données. C'est par définition la solution d'un problème inverse. L'équation précédente résout quant à elle le problème direct.

Généralement, l'information a priori sur D est indépendante de l'information a priori sur P :

$$\rho(d,p) = \rho_d(d) \cdot \rho_p(p) \quad (22)$$

et la fonction de densité théorique est obtenue sous la forme d'une fonction de densité conditionnelle

$$\vartheta(d,p) = \vartheta(d|p) \cdot \mu_p(p) \quad (23)$$

L'équation (21) se simplifie en :

$$\sigma_p(p) = \rho_p(p) \int \frac{\rho_d(d) \cdot \vartheta(d|p)}{\mu_d(d)} dd \quad (24)$$

Si la relation théorique est exacte $d=G(p)$ et $\vartheta(d|p)=\delta[d-G(p)]$ et alors (23) devient :

$$\sigma_{\mathbf{x}}(\mathbf{p}) = \rho_{\mathbf{x}}(\mathbf{p}) \frac{\rho_{\mathbf{x}}(\mathbf{G}(\mathbf{p}))}{\mu_{\mathbf{x}}(\mathbf{G}(\mathbf{p}))} \quad (25)$$

qui est la solution inverse pour une théorie exacte non-linéaire, avec des contraintes a priori sur les paramètres ($\rho_{\mathbf{p}}$) et une distribution de probabilité arbitraire sur les données ($\rho_{\mathbf{d}}$).

Revenons à l'équation (21), si nous sommes intéressés à un paramètre particulier, disons P_1 , toute l'information sur P_1 est contenue dans la fonction de densité marginale :

$$\sigma_1(p_1) = \int \sigma_{\mathbf{p}}(\mathbf{p}) \cdot dp_2 \cdot dp_3 \cdot \dots \cdot dp_s \quad (26)$$

A partir des $\sigma_1(p_1)$, on peut extraire par exemple: la valeur moyenne, la valeur médiane, la valeur la plus probable, la déviation standard, ou tout autre estimateur dont on a besoin concernant le paramètre P_1 . La valeur moyenne a posteriori ainsi que la covariance a posteriori sont calculées par :

$$E(\mathbf{P}) = \int \mathbf{p} \cdot \sigma_{\mathbf{p}}(\mathbf{p}) \, d\mathbf{p} \quad (27)$$

$$C = E\{(\mathbf{P}-E(\mathbf{P})) \cdot (\mathbf{P}-E(\mathbf{P}))^T\} \\ = \int (\mathbf{p}-E(\mathbf{P})) \cdot (\mathbf{p}-E(\mathbf{P}))^T \cdot \sigma_{\mathbf{p}}(\mathbf{p}) \, d\mathbf{p} \\ = \int \mathbf{p} \cdot \mathbf{p}^T \cdot \sigma_{\mathbf{p}}(\mathbf{p}) \, d\mathbf{p} - E(\mathbf{P}) \cdot E(\mathbf{P})^T \quad (28)$$

Toutes ces expressions sont valables pour les problèmes inverses non-linéaires aussi.

6. Existence, unicité, consistance, robustesse, résolution.

L'*existence* de la solution est liée à l'existence de la densité de probabilité $\sigma(\mathbf{x})$ (équ. 18) qui est trivial puisque par hypothèse $\rho(\mathbf{x})$ et $\vartheta(\mathbf{x})$ existent. On peut obtenir des solutions $\sigma(\mathbf{x})$ avec des pathologies (non normalisables, variance infinie, valeur la plus probable non unique) mais la solution est la densité de probabilité avec toutes les pathologies qu'elle peut présenter.

La consistance est garantie car l'équation (18) est consistante.

La robustesse peut être obtenue si l'on suspecte des erreurs dans les données en étalant la fonction de densité $\rho_d(\mathbf{d})$. L'expérience montre que la solution $\sigma_p(\mathbf{p})$ devient alors insensible à une erreur.

La résolution peut être considérée sous deux points de vue différents : jusqu'à quel point un paramètre donné a-t-il été résolu par les données ou bien quelle est la résolution spatiale atteinte avec les données ? L'information sur le premier type de résolution nous est donnée par la comparaison entre les fonctions de densité a priori $\rho_i(p_i)$ et a posteriori $\sigma_i(p_i)$ pour chaque paramètre p_i . La résolution spatiale est obtenue en étudiant la corrélation a posteriori entre les paramètres :

$$C_{ij} = \int p_i \cdot p_j \cdot \sigma(p_i, p_j) \cdot dp_i \cdot dp_j - E(p_i) \cdot E(p_j) \quad (29)$$

La courbe C_{ij} en fonction de $i-j$ donne l'information sur la variance a posteriori du paramètre C_{ij} et sur la résolution spatiale (la longueur de corrélation).

Remarque :

L'information a priori sur les valeurs des paramètres contenue dans $\rho(p)$ ne doit pas être stipulée de manière à obtenir la solution **désirée** mais elle doit être choisie afin de correspondre au mieux à l'information a priori réellement disponible.

7. Cas Gaussien.

Le problème gaussien regroupe les problèmes dont la fonction de densité de tous les paramètres est une gaussienne :

$$\rho(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \cdot \mathbf{C}_0^{-1} \cdot (\mathbf{x}-\mathbf{x}_0)\right\} \quad (30)$$

où \mathbf{x}_0 est la valeur moyenne a priori et \mathbf{C}_0 est la matrice de covariance a priori. Prenons le problème linéaire c'est-à-dire celui pour lequel on considère la théorie comme exacte; la relation théorique entre les paramètres est linéaire :

$$\mathbf{F} \cdot \mathbf{x} = \mathbf{0} \quad (31)$$

la fonction de densité des erreurs théoriques est aussi choisie gaussienne :

$$\vartheta(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{F} \cdot \mathbf{x})^T \cdot \mathbf{C}_T^{-1} \cdot (\mathbf{F} \cdot \mathbf{x})\right\} \quad (32)$$

où \mathbf{C}_T est la matrice de covariance décrivant les erreurs théoriques. Enfin, l'information nulle est une fonction constante :

$$\mu(\mathbf{x}) = \text{const.} \quad (33)$$

La fonction a posteriori est donnée par :

$$\sigma(\mathbf{x}) = \rho(\mathbf{x}) \cdot \vartheta(\mathbf{x}) \quad (34)$$

$$\sigma(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \cdot \mathbf{C}_0^{-1} \cdot (\mathbf{x} - \mathbf{x}_0) + \mathbf{x}^T \cdot \mathbf{F}^T \cdot \mathbf{C}_T^{-1} \cdot \mathbf{F} \cdot \mathbf{x}\right\} \quad (35)$$

Après quelques manipulations matricielles (Tarantola & Valette, 1982a), on obtient :

$$\sigma(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}_*)^T \cdot \mathbf{C}_*^{-1} \cdot (\mathbf{x} - \mathbf{x}_*)\right\} \quad (36)$$

avec

$$\mathbf{x}_* = \mathbf{P} \cdot \mathbf{x}_0 \quad (37)$$

$$\mathbf{C}_* = \mathbf{P} \cdot \mathbf{C}_0 \quad (38)$$

$$\mathbf{P} = \mathbf{I} - \mathbf{Q} \quad (39)$$

$$\mathbf{Q} = \mathbf{C}_0 \cdot \mathbf{F}^T \cdot (\mathbf{F} \cdot \mathbf{C}_0 \cdot \mathbf{F}^T + \mathbf{C}_T)^{-1} \cdot \mathbf{F} \quad (40)$$

La fonction de densité a posteriori est donc aussi une gaussienne centrée en \mathbf{x}_* et de matrice de covariance \mathbf{C}_* .

Pour obtenir la solution classique des moindres carrés, il suffit de partitionner les paramètres \mathbf{X} en \mathbf{D} et \mathbf{P} :

$$\mathbf{x} = \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} \quad \mathbf{x}_0 = \begin{bmatrix} \mathbf{d}_0 \\ \mathbf{p}_0 \end{bmatrix} \quad \mathbf{x}_* = \begin{bmatrix} \mathbf{d}_* \\ \mathbf{p}_* \end{bmatrix} \quad \mathbf{C}_0 = \begin{bmatrix} \mathbf{C}_{dd} & \mathbf{C}_{dp} \\ \mathbf{C}_{pd} & \mathbf{C}_{pp} \end{bmatrix} \quad (41)$$

avec l'hypothèse :

$$\mathbf{F} \cdot \mathbf{x} = [\mathbf{I} \quad -\mathbf{G}] \cdot \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{d} - \mathbf{G} \cdot \mathbf{p} = 0 \quad (42)$$

réécrivant les équations 37 et 38, on obtient les équations 58 à 60 du paragraphe suivant.

8. Solution des problèmes inverses par moindres carrés.

La solution des problèmes inverses par moindres carrés est un cas particulier de la solution générale dans lequel les états d'information sont décrits par des densité de probabilité gaussiennes.

Toute information a priori même peu contraignante stabilise et rend la solution inverse unique.

8.1 Définition des moindres carrés.

Soient x_0 vecteurs des valeurs a priori

C_0 matrice de covariance a priori :

$$C_0 = \begin{pmatrix} C_{\alpha\alpha} & C_{\alpha\beta} \\ C_{\beta\alpha} & C_{\beta\beta} \end{pmatrix} \quad \text{où } C_{\alpha\beta} = C_{\beta\alpha}^T = 0 \quad \text{dans la plupart des applications}$$

On fait l'hypothèse que l'information a priori est une gaussienne. Alors x_0 et C_0 définissent dans l'espace des paramètres E^m une densité de probabilité gaussienne :

$$\rho(x) = \text{const.} \exp\left\{-\frac{1}{2} (x-x_0)^T \cdot C_0^{-1} \cdot (x-x_0)\right\} \quad (43)$$

l'équation théorique non-linéaire est :

$$f(x) = 0 \quad (44)$$

définissant dans E^m un sous-espace théorique non-linéaire \mathfrak{S} . La fonction de densité a priori $\rho(x)$ induit dans le sous-espace théorique une densité de probabilité qui est la fonction de densité a posteriori :

si \mathfrak{S} est un sous-espace linéaire alors la fonction de densité a posteriori est gaussienne.

si \mathfrak{S} est un sous-espace non-linéaire alors la fonction de densité a posteriori n'est généralement pas gaussienne.

On définit le problème des moindres carrés comme étant la recherche du point \hat{x} du sous-espace théorique pour lequel la densité de probabilité a posteriori est minimum, c'est-à-dire que \hat{x} vérifie :

$$\begin{aligned} f(\hat{x}) &= 0 \\ s(\hat{x}) &= (\hat{x}-x_0)^T \cdot C_0^{-1} \cdot (\hat{x}-x_0) \text{ minimum sur } \mathfrak{S} \end{aligned} \quad (45)$$

Si les distributions de probabilité ne sont pas gaussiennes alors les moindres carrés donnent des solutions inacceptables.

Le système d'équations est équivalent à l'équation implicite :

$$\hat{x} = x_0 + C_0 \cdot F^T \cdot (F \cdot C_0 \cdot F^T)^{-1} \cdot \{F \cdot (\hat{x} - x_0) - f(\hat{x})\} \quad (46)$$

où F est la matrice des dérivées partielles :

$$F^{ik} = \partial f^i / \partial x^k \quad (47)$$

prises au point \hat{x}_k .

L'équation implicite (46) est résolue par méthode itérative. Sous l'hypothèse que les éléments F^{ik} de la matrice F sont des fonctions continues de x, la plus simple des procédures de résolution est celle du point fixe :

$$\hat{x}_{k+1} = x_0 + C_0 \cdot F_k^T \cdot (F_k \cdot C_0 \cdot F_k^T)^{-1} \cdot \{F_k \cdot (\hat{x}_k - x_0) - f(\hat{x}_k)\} \quad (48)$$

où les dérivées sont prises au point \hat{x}_k . Cet algorithme converge si les non-linéarités ne sont pas trop importantes. Si il n'y a qu'un point stationnaire sur \mathfrak{J} , on peut démontrer que s est minimum et que l'algorithme convergera toujours vers \hat{x} qui minimise s, qu'elle que soit la valeur initiale x_0 . Si plusieurs point de convergence existent, ils seront obtenus en choisissant différentes valeurs initiales de x_0 .

8.2 Le problème linéaire

L'équation générale du problème linéaire $f(x) = 0$ est de la forme :

$$f(x) = F \cdot x = 0 \quad (49)$$

où F est une matrice dont on connaît parfaitement tous les éléments (s'ils ne le sont pas, ils doivent être considérés comme paramètres et le problème est alors non-linéaire). La solution (équ. 46) devient l'expression explicite suivante :

$$\hat{x} = x_0 - C_0 \cdot F^T \cdot (F \cdot C_0 \cdot F^T)^{-1} \cdot F \cdot x_0 \quad (50)$$

La matrice de covariance a posteriori peut aussi être obtenue et est donnée par :

$$C = C_0 - C_0 \cdot F^T \cdot (F \cdot C_0 \cdot F^T)^{-1} \cdot F \cdot C_0 \quad (51)$$

si on définit les opérateurs linéaires :

$$Q = C_0 \cdot F^T \cdot (F \cdot C_0 \cdot F^T)^{-1} \cdot F \quad (52)$$

$$P = I - Q \quad (53)$$

On montre facilement que ce sont des projecteurs orthogonaux à C_0^{-1} . Les équations (50) et (51) prennent alors la forme compacte :

$$\hat{x} = P \cdot x_0 \quad (54)$$

$$C = P \cdot C_0 \quad (55)$$

Ces équations montrent que c'est le même projecteur qui agit sur les valeurs a priori des paramètres et sur la valeur a priori de la matrice de covariance. Ces équations sont exactement les mêmes que celles obtenues dans le cas gaussien sans emploi du critère des moindres carrés.

Séparons les paramètres X en l'ensemble des données D et l'ensemble des paramètres P. l'équation théorique (49) devient :

$$F \cdot x = \begin{bmatrix} I & -G \end{bmatrix} \cdot \begin{bmatrix} d \\ p \end{bmatrix} = d - G \cdot p = 0 \quad (56)$$

c'est-à-dire

$$d = G \cdot p \quad (57)$$

où G est une matrice dont les éléments sont parfaitement connus. On obtient finalement les expressions suivantes :

$$\hat{p} = p_0 + (C_{p_0 p_0} \cdot G^T - C_{p_0 d}) \cdot (C_{d d} - C_{d p_0} \cdot G^T - G \cdot C_{p_0 d} + G \cdot C_{p_0 p_0} \cdot G^T)^{-1} \cdot (d_0 - G \cdot p_0) \quad (58)$$

avec

$$\hat{d} = G \cdot \hat{p} \quad (59)$$

ce qui signifie que la solution par moindres carrés vérifient exactement la relation théorique.

La covariance a posteriori est

$$C_{pp}^{-1} = C_{p_0 p_0} - (C_{p_0 p_0} \cdot G^T - C_{p_0 d}) \cdot (C_{d d} - C_{d p_0} \cdot G^T - G \cdot C_{p_0 d} + G \cdot C_{p_0 p_0} \cdot G^T)^{-1} \cdot (G \cdot C_{p_0 p_0} - C_{d p_0}) \quad (60)$$

Remarques :

1. si $p_0 = 0$, on retrouve la solution proposée par Franklin (1970).
2. si $C_{d p_0} = (C_{p_0 d})^T = 0$, c'est-à-dire si les erreurs des données sont non-corrélées aux incertitudes sur les paramètres, on retrouve la solution de Jackson (1979).
3. si $C_{p_0 p_0} = \sigma^2$ avec $\sigma^2 \rightarrow \infty$, c'est-à-dire que l'incertitude sur les paramètres est complète, on obtient la solution classique par moindres carrés, dite solution normale.
4. En général, la (variance a posteriori) \leq (variance a priori). Si un paramètre est non résolu, c'est-à-dire que les données n'apportent aucune information sur ce paramètre, alors (variance a posteriori) = (variance a priori).
5. La résolution des problèmes non-linéaire se fera de manière classique par linéarisation des équations (pour plus de détails voir le papier de Tarantola et Valette, 1982b).

8.3 Le problème continu

Les résultats précédents sont étendus au cas où les données et/ou les paramètres sont fonctions de variables continues :

$$d(s) = \int g(s,r) \cdot p(r) dr \quad (61)$$

que l'on écrit sous la forme compacte

$$d = G \cdot p \quad (62)$$

où l'opérateur G est linéaire et la fonction $g(s,r)$ est le noyau de G .

Soient un système décrit par un ensemble X de vecteurs continus et/ou discrets, x_0 la valeur moyenne a priori de x , C_0 l'opérateur de covariance correspondant et une relation théorique non-linéaire de la forme :

$$f(x) = 0 \quad (63)$$

On définit le problème des moindres carrés comme étant la recherche du point \hat{x} minimisant:

$$s(\mathbf{x}) = [(\hat{\mathbf{x}} - \mathbf{x}_0), C_0(\hat{\mathbf{x}} - \mathbf{x}_0)] \quad (64)$$

parmi les points vérifiant l'équation (63) où $[\cdot , \cdot]$ est le produit scalaire. C'est le même problème que celui rencontré dans la section 8.1. La solution vérifie :

$$\hat{\mathbf{x}} = \mathbf{x}_0 + C_0 \cdot F^* \cdot (F \cdot C_0 \cdot F^*)^{-1} \cdot \{F \cdot (\hat{\mathbf{x}} - \mathbf{x}_0) - f(\hat{\mathbf{x}})\} \quad (65)$$

où l'opérateur linéaire F est la dérivée de l'opérateur non-linéaire $f(\mathbf{x})$ et F^* est l'adjoint de F . La solution est obtenue comme précédemment par la méthode du point fixe :

$$\boxed{\hat{\mathbf{x}}_{k+1} = \mathbf{x}_0 + C_0 \cdot F_k^* \cdot (F_k \cdot C_0 \cdot F_k^*)^{-1} \cdot \{F_k \cdot (\hat{\mathbf{x}}_k - \mathbf{x}_0) - f(\hat{\mathbf{x}}_k)\}} \quad (66)$$

9. Illustration numérique.

Examinons le problème de l'inversion d'un ensemble discret de données d^i , lorsque l'inconnue est une fonction $p(r)$. On fait l'hypothèse que la relation théorique non-linéaire peut s'écrire sous la forme

$$\mathbf{f}(\mathbf{x}) = \mathbf{d} - \mathbf{g}(\mathbf{p}) = \mathbf{0} \quad (67)$$

où \mathbf{g} est un vecteur de fonctionnelles non-linéaires :

$$d^i = g^i(p(r)) \quad (68)$$

Les données consistent en un vecteur de valeur moyenne d_0 et d'une matrice de covariance des erreurs des données $C_{d_0 d_0}$. La connaissance a priori de la fonction $p(r)$ est une valeur moyenne $p_0(r)$ et la fonction de covariance associée $C_{p_0 p_0}(r, r')$; on considère qu'il n'y a pas de corrélation entre d_0 et p_0 , et donc l'opérateur de covariance est donné par l'expression suivante :

$$C_0 = \begin{bmatrix} C_{d_0 d_0} & 0 \\ 0 & C_{p_0 p_0} \end{bmatrix} \quad (69)$$

Sous toutes ces hypothèses, l'algorithme général (66) devient :

$$\mathbf{p}_{k+1} = \mathbf{p}_0 + C_{p_0 p_0} \cdot G_k^* \cdot (C_{d_0 d_0} + G_k \cdot C_{p_0 p_0} \cdot G_k^*)^{-1} \cdot \{d_0 - \mathbf{g}(\hat{\mathbf{p}}_k) + G_k \cdot (\hat{\mathbf{p}}_k - \mathbf{p}_0)\} \quad (70)$$

Explicitement,

$$\hat{p}_{k+1}(r) = p(r) + \int dr' \sum_i \sum_j C_{popo}(r, r') \cdot G_i^j(r') \cdot (S_i^j)^{ij} \cdot \left\{ d_0^i - g^i(\hat{p}_k) + \int dr'' \cdot G_i^j(r'') \cdot [p(r) - p(r)] \right\} \quad (71)$$

où

$$S_k^{ij} = (C_{ooco})^{ij} + \int dr' \int dr'' G_i^j(r') \cdot C_{popo}(r', r'') \cdot G_i^j(r'') \quad (72)$$

et $G_i^j(r)$ est la dérivée de la fonctionnelle non linéaire $g^i(p)$ prise au point $P = \hat{p}_k$.

Dans le cas d'un problème linéaire, l'équation (68) devient :

$$d^i = \int dr \cdot G^i(r) \cdot p(r) \quad (73)$$

L'équation (71) fournit alors la solution explicite :

$$\hat{p} = p_0 + C_{popo} \cdot G \cdot (C_{ooco} + G \cdot C_{popo} \cdot G)^{-1} \cdot \{d_0 + G \cdot p_0\} \quad (74)$$

explicitement,

$$\hat{p}(r) = p_0(r) + \int dr' \sum_i \sum_j C_{popo}(r, r') \cdot G_i^j(r') \cdot (S_i^j)^{ij} \cdot \left\{ d_0^i - \int dr'' \cdot G_i^j(r'') \cdot p_0(r'') \right\} \quad (75)$$

où la matrice S est

$$S^{ij} = (C_{ooco})^{ij} + \int dr' \int dr'' G_i^j(r') \cdot C_{popo}(r', r'') \cdot G_i^j(r'') \quad (76)$$

comme le problème est linéaire, la fonction de covariance a posteriori de $\hat{p}(r)$ peut être calculée exactement :

$$C_{pp}^{**} = C_{popo} - C_{popo} \cdot G \cdot (C_{ooco} + G \cdot C_{popo} \cdot G)^{-1} \cdot G \cdot C_{popo} \quad (77)$$

explicitement,

$$C_{pp}(r,r') = C_{popo}(r,r') - \int dr'' \sum_i \sum_j \int dr''' \cdot C_{popo}(r,r'') \cdot G(r'') \cdot (S^{-1})^{ij} \cdot G(r''') \cdot C_{popo}(r''',r') \quad (78)$$

Application:

Nous allons traiter le problème de l'interpolation d'une fonction dont on connaît la valeur en un certain nombre de points, $d(r^i)$. Dans ce cas particulier, l'équation théorique ou d'observation est simplement :

$$d^i = p(r^i) \quad (79)$$

de cette équation, on en déduit que le problème est linéaire et que le noyau est :

$$G^i(r) = \delta(r^i - r) \quad (80)$$

Les informations a priori disponibles sont : les valeurs observées d^i et leurs écarts types σ^i (voir table 1 et figure 1), une solution a priori $p_0(r)$ ($= 0$, dans cet exemple) ainsi que σ ($= 2$) l'incertitude sur cette solution a priori. Cela signifie que la dispersion de la solution a posteriori $p(r)$ autour de $p_0(r)$ est de l'ordre de σ . Enfin, on veut que la solution $p(r)$ soit lisse : pour ce faire, on impose que les solutions en deux points proches soient similaires au travers d'une fonction de covariance a priori sur les paramètres qui est choisie gaussienne :

$$C_{popo}(r,r) = \sigma^2 \exp\left\{-\frac{1}{2} \frac{(r-r')^2}{\Delta^2}\right\} \quad (81)$$

où σ^2 est la variance au point r et Δ est la longueur de corrélation. La solution de ce problème fortement sous-déterminé est obtenue en introduisant dans l'équation (75) les expressions des différents termes qui la composent. Après intégration, on obtient l'expression compacte :

$$\hat{p}(r) = p_0(r) + \sum_i \sum_j C_{\text{popo}}(r, r^i) (S^{-1})^{ij} [d_0^i - p_0(r^i)] \quad (82)$$

où

$$S^{ij} = (C_{\text{aodo}})^{ij} + C_{\text{popo}}(r^i, r^j) \quad (83)$$

et la covariance a priori (78) se simplifie en

$$C_{\text{pp}}(r, r') = C_{\text{popo}}(r, r') - \sum_i \sum_j C_{\text{popo}}(r, r^i) (S^{-1})^{ij} C_{\text{popo}}(r^j, r') \quad (84)$$

Si $r=r'$, on a la variance a posteriori au point r .

r^i	d_0^i	σ_0^i
5.	1.	0.2
5.5	1.5	0.4
7.5	1.25	0.4
8.25	0.85	0.1
10.	0.25	0.2
11.	0.5	0.2
12.5	1.35	0.3
13.	0.4	0.25
15.	-0.5	0.3

Table 1: positions, valeurs et écarts types des données de la figure 1.

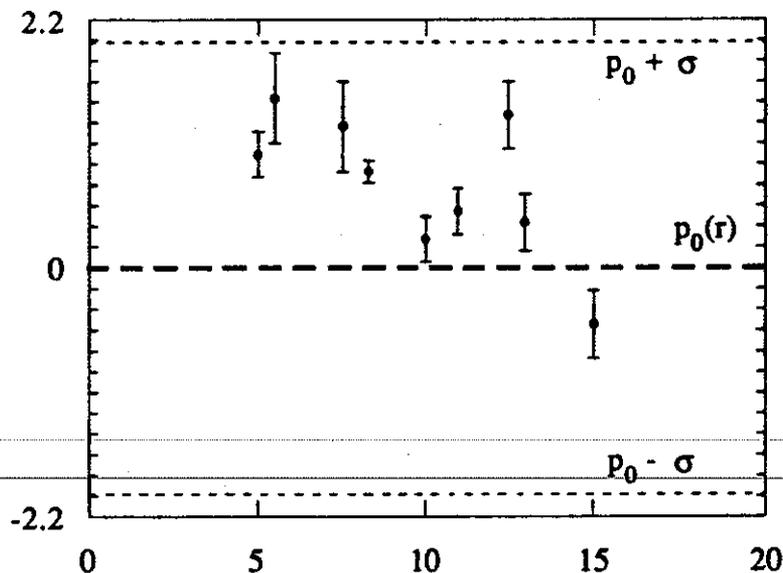


Figure 1: Données et leurs écarts types utilisées dans l'interpolation. La solution a priori $p_0(r)$ ainsi que l'incertitude a priori σ ($= 2$) de cette solution sont reproduits.

Ces formules (82) et (84) sont la base du code Fortran d'interpolation reproduit en annexe qui a été utilisé pour calculer les solutions des figures 2 à 4. Ces quelques exemples montrent qu'au plus on s'éloigne des données, au plus la solution et l'écart type ressemble à l'a priori. Les différentes solutions présentées correspondent à des choix différents de la longueur de corrélation Δ de la fonction de covariance a priori des paramètres. Au plus la longueur de corrélation est petite (fig. 2), au plus la solution passe par les données et l'écart type de la solution se rapproche de l'écart type de l'erreur des données. En effet, si la longueur de corrélation est petite, l'information contenue dans les données ne peut se propager d'une données à l'autre.

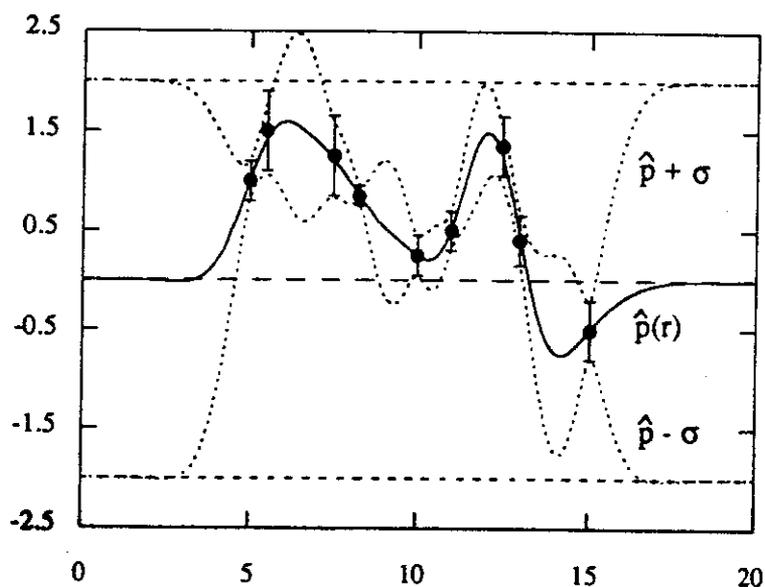


Figure 2: Solution obtenue (courbe continue) et l'écart type en chaque point comparée à la solution a priori pour une longueur de corrélation $\Delta = 1$.

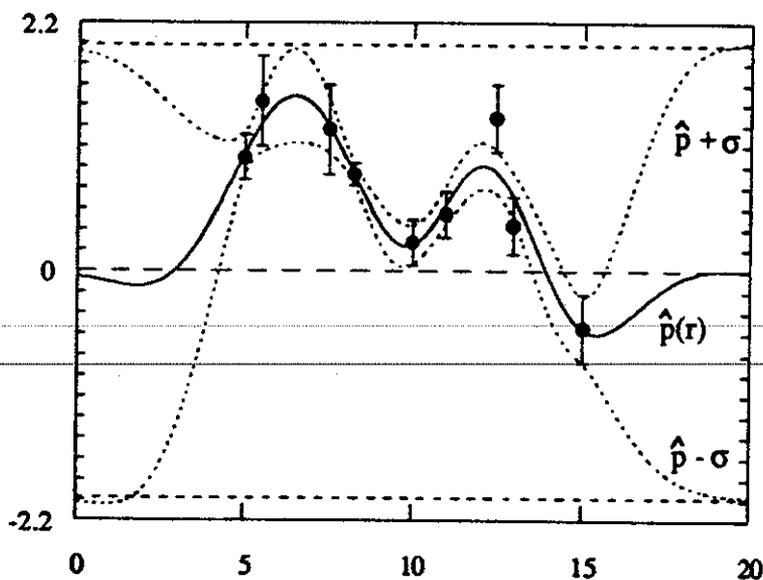


Figure 3: Solution obtenue (courbe continue) et l'écart type en chaque point comparée à la solution a priori pour une longueur de corrélation $\Delta = 2$.

Par contre, si la longueur de corrélation est très grande (fig. 4), la solution sera lisse et l'écart type a posteriori tend à s'uniformiser autour des données. Dans ce cas, l'information contenue dans une donnée très précise peut se propager vers une donnée proche moins précise et rendre l'écart type de la solution au point de moindre précision inférieur à l'écart type de l'erreur a priori de la donnée. La figure 3 reproduit une solution intermédiaire entre ces deux cas extrêmes.

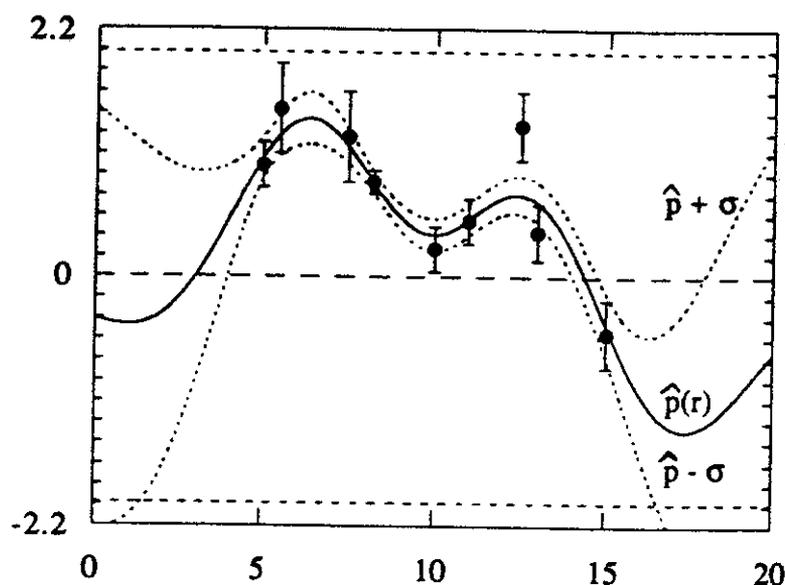


Figure 4: Solution obtenue (courbe continue) et l'écart type en chaque point comparée à la solution a priori pour une longueur de corrélation $\Delta = 3$.

10. Conclusion.

Tarantola et Valette (1982a et b) ont développé un formalisme qui permet d'obtenir une solution générale des problèmes inverses en termes de densités de probabilité. si on fait l'hypothèse que ces densités de probabilité sont gaussiennes, on obtient une solution générale des problèmes inverses par moindres carrés. Les expressions mathématiques de ces différentes solutions sont démontrées et décrites en détails. Enfin, un exemple numérique simple est proposé afin d'appliquer à un cas concret les formules.

Références

- Franklin, J.N., 1970. Well-posed stochastic extensions of ill-posed linear problems, *J. Math. Anal. Appl.*, 31, 682-716.
- Jackson, D.D., 1979. The use of a priori data to resolve non-uniqueness in linear inversion, *Geophys. J. R. Astron. Soc.*, 57, 137-157.
- Tarantola, A., 1987. *Inverse Problem Theory : Methods for Data Fitting and Model Parameter Estimation*, Elsevier, 613 pp.
- Tarantola, A. & Valette, B., 1982a. Inverse Problems = Quest for information, *J. Geophys.*, 50, 159-170.
- Tarantola, A. & Valette, B., 1982b. Generalized nonlinear inverse problems solved by using the least squares criterion, *Rev. Geophys. Space Phys.*, 20(2), 219-232.

PROGRAM INTER

 INTERPOLATION (1D) PAR METHODE INVERSE

LA FONCTION DE COVARIANCE A PRIORI $C_{pp}(r, r')$ EST UNE GAUSSIENNE

REFERENCE : TARANTOLA, A. ET VALETTE, B., GENERALIZED NONLINEAR INVERSE
 ----- PROBLEMS SOLVED USING THE LEAST SQUARES CRITERION, REV.
 GEOPHYS. SPACE PHYS., 20(2), 219-232, 1982.

INPUTS :

 ND = NOMBRE DE DONNEES
 NP = NOMBRE DE POINTS D'INTERPOLATION - 1
 RMIN = COORDONNEE MINIMALE D'INTERPOLATION
 RMAX = COORDONNEE MAXIMALE D'INTERPOLATION
 R = VECTEUR COORDONNEE DES DONNEES
 D0 = VECTEUR DES DONNEES
 SIGMA0 = VECTEUR DES DEVIATIONS STANDARDS DES DONNEES
 P0 = SOLUTION A PRIORI
 SIGMA = INCERTITUDE SUR LA SOLUTION A PRIORI

OUTPUTS :

 SOL = SOLUTION
 COV = DEVIATION STANDARD DE LA SOLUTION

SOUS-ROUTINES :

 SSPCO = FACTORISATION D UNE MATRICE REEL SYMETRIQUE DEFINIE POSITIVE
 SPPSL = RESOLUTION D UN SYSTEME REEL SYMETRIQUE DEFINI POSITIF

 PARAMETER (ND=9, NDD=ND*(ND+1)/2, NP=200)

DIMENSION R(ND), D0(ND), SIGMA20(ND)
 DIMENSION AW(ND), AWS(ND), S(NDD)
 DIMENSION ZCOND(ND)

----- FICHER ENTREES/SORTIES -----

OPEN(1, FILE=' :MFH.FRANCIS.INVERSE.ARCHIVES.INPUTS_INT')
 OPEN(2, FILE=' :MFH.FRANCIS.INVERSE.ARCHIVES.OUTPUTS_INT')

----- LECTURE DES DONNEES -----

CALL LEC(1, RMIN, RMAX, P0, SIGMA2, DELTA, R, D0, SIGMA20, ND)

----- CALCUL ET FACTORISATION DE LA MATRICE S -----

CALL SCPPCDD(ND, NDD, R, SIGMA20, SIGMA2, DELTA, ZCOND, S)

----- INTERPOLATION -----

DELTAR=(RMAX-RMIN)/REAL(NP)

DO 4 L=1, NP+1
 RI=XMIN+REAL(L-1)*DELTAR

```

*
*   CALL INT(RI,R,D0,SIGMA2,DELTA,AW,AWS,S,ND,NDD,SOL,COV)
C
*   SOL=P0+SOL
*
*   --- IMPRESSION DES RESULTATS ---
*
*   CALL SORTIE(2,RI,P0,SIGMA2,SOL,COV)
*
4  *   CONTINUE
*
*   CLOSE(2)
*
*   STOP
*   END
*
*   -----
*
*   SUBROUTINE LEC(NR,RMIN,RMAX,P0,SIGMA2,DELTA,R,D0,SIGMA20,ND)
*
*   LECTURE DES INPUTS
*
*   -----
*
*   DIMENSION R(ND),D0(ND),SIGMA20(ND)
*
*   READ(NR,*) RMIN,RMAX
*   READ(NR,*) P0,SIGMA,DELTA
*
*   SIGMA2=SIGMA*SIGMA
*
*   READ(NR,*)
*
*   DO 100 I=1,ND
*     READ(NR,*) R(I),D0(I),SIGMA0
*     SIGMA20(I)=SIGMA0*SIGMA0
100 *   CONTINUE
*
*   CLOSE(NR)
*
*   RETURN
*   END
*
*   -----
*
*   SUBROUTINE SCPPCDD(ND,NDD,R,SIGMA20,SIGMA2,DELTA,ZCOND,S)
*
*   CONSTRUCTION ET FACTORISATION DE LA MATRICE S = Cpp + Cdd
*
*   -----
*
*   DIMENSION R(ND),ZCOND(ND),S(NDD),SIGMA20(ND)
*
*   K=0
*
*   DO 3 I=1,ND
*     RI=R(I)
*     DO 31 J=1,I
*       RJ=R(J)
*
*       K=K+1
*       S(K)=COVA(RI,RJ,SIGMA2,DELTA)
*
31 *   CONTINUE
*   S(K)=S(K)+SIGMA20(I)

```

```

3  CONTINUE
*
*  ----- FACTORISATION -----
*
CALL SPPCO(S,ND,RCOND,ZCOND,INFO)
*
PRINT *, 'RCOND=', RCOND
*
IF (INFO.NE.0) THEN
  PRINT 1002
  PRINT 1003
  PRINT *, 'INFO=', INFO
  STOP
ENDIF
*
PRINT *, 'INVERSION OK'
*
1002 FORMAT(5X, 'MATRICE S SINGULIERE', /)
1003 FORMAT(5X, 'BUG --> ARRET INVERSION')
*
RETURN
END
*
*  -----
*
SUBROUTINE INT(RI,R,D0,SIGMA2,DELTA,AW,AWS,S,ND,NDD,SOL,COV)
*
*  INTERPOLATION AU POINT RI : SOLUTION ET COVARIANCE A POSTERIORI
*
*  -----
*
DIMENSION R(ND),D0(ND)
DIMENSION AW(ND),AWS(ND),S(NDD)
*
DO 5 I=1,ND
  AW(I)=COVA(RI,R(I),SIGMA2,DELTA)
  AWS(I)=AW(I)
5  CONTINUE
*
CALL SPPSL(S,ND,AW)
*
SOL=0.
COV=0.
*
DO 6 I=1,ND
  SOL=SOL+AW(I)*D0(I)
  COV=COV+AW(I)*AWS(I)
6  CONTINUE
*
COV=SIGMA2-COV
*
RETURN
END
*
*  -----
*
FUNCTION COVA(R,R1,SIGMA2,DELTA)
*
*  FONCTION DE COVARIANCE GAUSSIENNE
*
*  -----
*
COVA=SIGMA2*EXP((-1)*0.5*(R-R1)*(R-R1)/DELTA/DELTA)
*

```


Table des matières

Avant-propos	1
1. Introduction	2
2. Paramètres et information	2
3. Données et information a priori	6
4. Relations théoriques	6
5. Solution du problème inverse	7
6. Existence, unicité, consistance, robustesse et résolution	9
7. Cas gaussien	10
8. Solution des problèmes inverses par moindres carrés	12
8.1 Définition des moindres carrés	12
8.2 Le problème linéaire	13
8.3 Le problème continu	15
9. Illustration numérique	16
10. Conclusion	21
Références	21
Annexe	22