Estimation multivariable complexe

Complex multivariable estimation

P. COMON
CEPHAG-ENSIEG, BP n° 46, 38402 SAINT-MARTIN-D'HÈRES

TABLE DES MATIÈRES

Introduction

En traitement du signal, on a parfois affaire à des problèmes d'optimisation dans $\mathbb{C}^p$, notamment lorsque les données sont délivrées dans le domaine spectral. Le problème d'optimisation dans $\mathbb{C}^p$ est très différent d'une optimisation dans $\mathbb{R}^{2p}$, dans les situations réelles auxquelles on est confronté où les données de $\mathbb{C}^p$ sont obtenues par transformation de données de $\mathbb{R}^p$. Poser le problème dans $\mathbb{R}^{2p}$ peut entraîner un alourdissement considérable des notations et même conduire à une formulation intractable; pour s'en rendre compte, nous invitons le lecteur à résoudre notre exemple 2 dans $\mathbb{R}^{2p}$. N'ayant trouvé dans la littérature que très peu de résultats concernant la théorie de l'estimation complexe [4, 8], nous espérons que l'exposé succint qui va suivre sera de quelque utilité. Le lecteur ne manquera pas de remarquer que certaines démonstrations et certains résultats sont différents de ceux établis dans $\mathbb{R}^p$ [7], bien que leur obtention exposée ci-après ne pose aucun problème. Deux exemples très simples permettent d'illustrer notre propos dans un environnement gaussien.

1. Notations

Les vecteurs et les matrices seront distingués des grandeurs scalaires par une impression grasse.

La densité de probabilité (ddp) d'une variable complexe $\mathbf{z}$ est définie comme étant la ddp conjointe de ses parties réelle et imaginaire. Les intégrales intervenant dans cet exposé devront être interprétées dans ce sens.

On notera h_n^2 l'espace hilbertien des vecteurs aléatoires du second ordre de dimension n. Le produit scalaire de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ de h_n^2 est défini par la moyenne statistique $E\{\mathbf{x}^+\mathbf{y}\}$, où $\mathbf{x}^+$ désigne le vecteur transposé et conjugué de $\mathbf{x}$. En outre, le vecteur $\mathbf{x}^T$ désignera le vecteur transposé de $\mathbf{x}$.

$L(\mathbf{x}, \boldsymbol{\theta})$: ddp de l'observable $\mathbf{x} \in h_n^2$, conditionnellement à $\boldsymbol{\theta}$, vecteur de paramètres déterministes inconnus à estimer; $L \in \mathbb{R}$ et $\boldsymbol{\theta} \in \mathbb{C}^p$. $L(\mathbf{x}, \boldsymbol{\theta})$ est donc la fonction de vraisemblance de $\mathbf{x}$ associée à $\boldsymbol{\theta}$.

$$\mathbf{S}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\mathbf{x}, \boldsymbol{\theta})$$

(*Voir* définition dans l'Annexe).

$\mathbf{S}$ est un vecteur de h_p^2, appelé parfois « score function ».

Posons $\mathbf{F}(\boldsymbol{\theta}) = E(\mathbf{S}\mathbf{S}^+)$. Cette matrice correspond à l'information de Fisher. Elle est déterministe et hermitienne de dimension $p \times p$

2. Préliminaire

2.1. On utilise la notation de dérivation d'une fonctionnelle réelle $\mathbf{h}(\boldsymbol{\theta})$ définie en annexe; $\mathbf{h}(\boldsymbol{\theta})$ étant à valeurs dans $\mathbb{R}^m$ et définie sur $\mathbb{C}^p$. Rappelons que:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{h}^T(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\alpha}} \mathbf{h}^T(\boldsymbol{\theta}) + j \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{h}^T(\boldsymbol{\theta}); \text{si } \boldsymbol{\theta} = (\boldsymbol{\alpha} + j\boldsymbol{\beta}), \boldsymbol{\theta} \in \mathbb{C}^p.$$

Cette notation établit une bijection entre $\mathbb{C}^p$ et $\mathbb{R}^{2p}$. Relevons en particulier les propriétés suivantes.

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{Re}(\boldsymbol{\theta}^T) = \mathbf{I}$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{Im}(\boldsymbol{\theta}^T) = j\mathbf{I}$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}^+ \mathbf{H} \boldsymbol{\theta} = 2\mathbf{H}\boldsymbol{\theta},$$

où $\mathbf{H}$ désigne une matrice hermitienne $p \times p$.

2.2. **Lemme**: *Soit* **a** *un vecteur de* $\mathbb{C}^n$ *et* **B** *une matrice hermitienne régulière. Alors*:

$$\underset{x \in \mathbb{C}^n}{\mathrm{Max}}\left[\frac{|\mathbf{a}^+\mathbf{x}|^2}{\mathbf{x}^+\mathbf{B}\mathbf{x}}\right]$$

est obtenue lorsque **x** *est colinéaire à* $\mathbf{B}^{-1}\mathbf{a}$. *Le maximum vaut alors* $\mathbf{a}^+\mathbf{B}^{-1}\mathbf{a}$.

Ce lemme résulte de l'inégalité de Schwarz dans h_1^2.

3. Moyenne statistique de la fonction S (x, θ)

Les calculs présentés ci-après sont légitimes si les fonctions considérées satisfont les conditions nécessaires autorisant l'interversion des opérateurs de dérivation et d'intégration réelles.

Théorème 1: $\mathrm{E}(\mathbf{S})=\mathbf{0}$.

Démonstration:

$$\int \dots \int \mathrm{L}(\mathbf{X}, \boldsymbol{\theta})\, dx_1 \dots dx_n = 1$$

Donc, comme $\mathrm{L}(\mathbf{X}, \boldsymbol{\theta}) \in \mathbb{R}$, il vient:

$$\int \dots \int \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{L}(\mathbf{X}, \boldsymbol{\theta})\, dx_1 \dots dx_n = \mathbf{0}$$

or

$$\mathrm{E}(\mathbf{S}) = \int \dots \int \frac{\partial \ln \mathrm{L}(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{L}(\mathbf{X}, \boldsymbol{\theta})\, dx_1 \dots dx_n$$

d'où

$$\mathrm{E}(\mathbf{S}) = \int \dots \int \frac{\partial \mathrm{L}(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\, dx_1 \dots dx_n = \mathbf{0}$$

4. Variance des estimateurs de θ: borne inférieure

Théorème 2: *Si* $\mathbf{h}(\mathbf{X})$ *est une fonction à valeur dans* $\mathbb{R}^p$, *c'est-à-dire de même dimension que* $\boldsymbol{\theta}$, *alors*:

$$\mathrm{E}\{\mathbf{S}.\mathbf{h}^{\mathrm{T}}\} = \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{E}\{\mathbf{h}^{\mathrm{T}}\}$$

Démonstration:

$$\mathrm{E}\{\mathbf{h}^{\mathrm{T}}\} = \int \dots \int \mathbf{h}^{\mathrm{T}}(\mathbf{x})\, \mathrm{L}(\mathbf{x}, \boldsymbol{\theta})\, dx_1 \dots dx_n$$

$$\Rightarrow \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{E}\{\mathbf{h}^{\mathrm{T}}\} = \int \dots \int \frac{\partial \mathrm{L}}{\partial \boldsymbol{\theta}} \mathbf{h}^{\mathrm{T}}\, dx_1 \dots dx_n$$

$$= \int \dots \int \frac{\partial \ln \mathrm{L}}{\partial \boldsymbol{\theta}} \mathbf{h}^{\mathrm{T}}.\mathrm{L}\, dx_1 \dots dx_n$$

$$= \int \dots \int \mathbf{S}.\mathbf{h}^{\mathrm{T}} \mathrm{L}\, dx_1 \dots dx_n$$

Théorème 3: *Si* $\mathbf{t}(\mathbf{x})$ *est un estimateur non biaisé de* $\boldsymbol{\theta}$, *paramètre de* $\mathbb{C}^p$, *alors*:

$$\mathrm{E}\{\mathbf{S}.\mathrm{Re}(\mathbf{t}^{\mathrm{T}})\} = \mathbf{I}$$
$$\mathrm{E}\{\mathbf{S}.\mathrm{Im}(\mathbf{t}^{\mathrm{T}})\} = j\mathbf{I}$$

Démonstration: On applique le théorème 2 à:

$$\mathrm{Re}(\mathbf{t}^{\mathrm{T}}) = \frac{1}{2}(\mathbf{t}^{\mathrm{T}} + \mathbf{t}^+) \quad \text{et} \quad \mathrm{Im}(\mathbf{t}^{\mathrm{T}}) = \frac{1}{2j}(\mathbf{t}^{\mathrm{T}} - \mathbf{t}^+)$$

Il suffit alors d'utiliser les propriétés rappelées dans le préliminaire.

Corollaire 3: *Soit* $\mathbf{t}(\mathbf{x})$ *un estimateur non biaisé de* $\boldsymbol{\theta}$, *et* $\mathbf{t}_c(\mathbf{x}) = \mathbf{t}(\mathbf{x}) - \boldsymbol{\theta}$, *alors d'après le théorème* 1:

$$\mathrm{E}\{\mathbf{S}.\mathbf{t}_c^{\mathrm{T}}\} = \mathbf{0} \quad \textit{et} \quad \mathrm{E}\{\mathbf{S}.\mathbf{t}_c^+\} = 2\mathbf{I}.$$

Théorème 4: *Si* $\mathbf{t}(\mathbf{x})$ *est un estimateur non biaisé de* $\boldsymbol{\theta}$, *alors*:

$$\mathbf{V}(\mathbf{t}) \geqq 4\,\mathbf{F}^{-1}(\boldsymbol{\theta})$$

où:

$$\mathbf{V}(\mathbf{t}) = \mathrm{E}(\mathbf{t}_c \mathbf{t}_c^+) \quad \textit{et} \quad \mathbf{t}_c = \mathbf{t} - \boldsymbol{\theta}$$

Autrement dit, la matrice $\mathbf{V} - 4\,\mathbf{F}^{-1}$ *est non négative.*

Démonstration: L'inégalité de Schwarz dans h_1^2 s'écrit:

$$|\mathrm{E}\{u^+ v\}|^2 \leqq \mathrm{E}\{u^+ u\}.\mathrm{E}\{v^+ v\}$$

Soient **a** et **c** deux vecteurs déterministes quelconques de $\mathbb{C}^p$. Pour $u = \mathbf{a}^+\mathbf{t}_c$ et $v = \mathbf{c}^+\mathbf{s}$, l'inégalité devient:

$$|\mathbf{a}^+\mathrm{E}(\mathbf{t}_c\mathbf{s}^+)\mathbf{c}|^2 \leqq \mathbf{a}^+\mathrm{E}(\mathbf{t}_c\mathbf{t}_c^+)\mathbf{a}.\mathbf{c}^+\mathrm{E}(\mathbf{s}\mathbf{s}^+)\mathbf{c}$$

et d'après le corollaire 3:

$$|2\mathbf{a}^+\mathbf{c}|^2 \leqq \mathbf{a}^+\mathbf{V}(\mathbf{t})\mathbf{a}.\mathbf{c}^+\mathbf{F}\mathbf{c}$$

Cette inégalité reste vraie en particulier lorsque la quantité

$$\frac{\mathbf{c}^+\mathbf{a}\mathbf{a}^+\mathbf{c}}{\mathbf{c}^+\mathbf{F}\mathbf{c}}$$

est maximale, soit, d'après le lemme, lorsque:

$$\mathbf{c} = \mathbf{F}^{-1}\mathbf{a}$$

Il vient donc que, $\forall\, \mathbf{a} \in \mathbb{C}^p$, l'inégalité suivante reste vérifiée:

$$4\,\mathbf{a}^+\mathbf{F}^{-1}\mathbf{a} \leqq \mathbf{a}^+\mathbf{V}(\mathbf{t})\mathbf{a}$$

La Borne inférieure de Cramer-Rao (BICR) vaut ici par conséquent $4\,\mathbf{F}^{-1}$ contre $\mathbf{F}^{-1}$ dans le cas réel.

Corollaire 4: *Du théorème précédent, il découle immédiatement les propriétés particulières*:

(i) $\mathbf{V}(t_i) \geqq 4\,\mathbf{F}_{ii}^{-1}, \forall\, 1 \leqq i \leqq p$.

(ii) $\det \mathbf{V}(\mathbf{t}).\det \mathbf{F}(\boldsymbol{\theta}) \geqq 4$.

5. Efficacité

Un estimateur $\mathbf{t}$ est usuellement dit « efficace » lorsque sa variance est égale à la BICR. Il faut remarquer qu'il n'est pas toujours possible d'atteindre cette borne. Ainsi, dans le souci de relier l'efficacité des estimateurs avec leur variance, il peut être intéressant d'évaluer la variance minimale atteignable plutôt que la BICR [10; p. 346]. Nous nous contenterons dans cette section de présenter un résultat fondamental concernant l'efficacité.

Théorème 5: *Soit* $\mathbf{t}(\mathbf{x})$ *un estimateur non biaisé de* $\boldsymbol{\theta}$. *Alors les propriétés suivantes sont équivalentes* :

(i) *La BICR est atteinte* ($\mathbf{t}$ *efficace*).

(ii) $\mathbf{t}(\mathbf{x})-\boldsymbol{\theta} \underset{\text{p.s.}}{=} 2\mathbf{F}^{-1}(\boldsymbol{\theta}).\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$.

(iii) *Il existe une matrice* $p\times p$ $\mathbf{K}(\boldsymbol{\theta})$ *telle que*

$$\mathbf{t}(\mathbf{x})-\boldsymbol{\theta} \underset{\text{p.s.}}{=} \mathbf{K}(\boldsymbol{\theta}).\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$$

Démonstration: Si (i) est satisfaite, alors $\mathbf{V}(\mathbf{t})-4\,\mathbf{F}^{-1}(\boldsymbol{\theta})=\mathbf{0}$.
Ceci implique d'après le corollaire 3 que:

$$E\{(\mathbf{t}_c(\mathbf{x})-2\mathbf{F}^{-1}\mathbf{S}(\mathbf{x}, \boldsymbol{\theta}))(\mathbf{t}_c(\mathbf{x})-2\mathbf{F}^{-1}\mathbf{S}(\mathbf{x}, \boldsymbol{\theta}))^+\}=\mathbf{0}$$

Ceci entraîne que $\mathbf{t}_c(\mathbf{x}) \underset{\text{p.s.}}{=} 2\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$, donc (ii) puis (iii).
Inversement, si (iii) est satisfaite, il vient en postmultipliant les deux membres par $\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})^+$ et en prenant l'espérance mathématique:

$$E\{\mathbf{t}_c(\mathbf{x})\,\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})^+\}=\mathbf{K}(\boldsymbol{\theta}).E\{\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})\,\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})^+\}.$$

Soit d'après le corollaire 3:

$$2\mathbf{I}=\mathbf{K}(\boldsymbol{\theta}).\mathbf{F}(\boldsymbol{\theta}).$$

en reportant dans (iii), nous avons (ii).
Il nous reste à montrer (i). Pour ce faire, il suffit de postmultiplier (ii) par $\mathbf{t}_c^+$ puis de prendre l'espérance mathématique des deux membres:

$$E\{\mathbf{t}_c(\mathbf{x})\,\mathbf{t}_c(\mathbf{x})^+\}=2\mathbf{F}^{-1}(\boldsymbol{\theta}).E\{\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})\,\mathbf{t}_c(\mathbf{x})^+\}.$$

Le résultat escompté est obtenu en utilisant une nouvelle fois le corollaire 3.

Remarque: Les résultats de ce genre ont été généralisés au cas d'estimateurs biaisés lorsque $\boldsymbol{\theta}$ est réel [10; p. 327, 1; p. 41 et 313]. Parallèlement, généraliser les résultats exposés ici au cas biaisé ne présente pas de problèmes; la seule différence avec la formulation dans $\mathbb{R}^n$ réside dans la covariance croisée $E(\mathbf{t}\mathbf{S}^+)\neq\mathbf{I}$.

6. Exemple 1

Considérons le problème de l'estimation d'une moyenne sous contrainte linéaire. Le modèle d'observation adopté est le suivant:

$$\mathbf{x}^\mu=\boldsymbol{\Phi}\,\boldsymbol{\theta}+\mathbf{e}^\mu;$$

les $\mathbf{x}^\mu$ sont indépendamment $N_n^c(\boldsymbol{\Phi\theta}, \boldsymbol{\Gamma}_e)$, où:

— $\mathbf{x}$ désigne l'observable, vecteur aléatoire de dimension n, que l'on suppose normal complexe; l'exposant μ distingue les M réalisations indépendantes: $1\leqq\mu\leqq M$;
— $\mathbf{e}$ désigne la partie centrée de $\mathbf{x}$;
— $\boldsymbol{\Phi}$ est une matrice donnée de dimension $n\times p$, de rang $p\leqq n$;
— $\boldsymbol{\theta}$ est le paramètre inconnu à estimer, de dimension p;
— $\boldsymbol{\Gamma}_e$ est la matrice de covariance de $\mathbf{e}$, de taille $n\times n$:

$$\boldsymbol{\Gamma}_e=E\{\mathbf{e}\mathbf{e}^+\}; \qquad \boldsymbol{\Gamma}_x=E\{\mathbf{x}\mathbf{x}^+\}=\boldsymbol{\Gamma}_e+\boldsymbol{\Phi\theta\theta}^+\boldsymbol{\Phi}^+$$

Ce type de modèle est largement utilisé pour la réception des ondes, et de manière plus générale pour l'analyse sélective de puissance [6, 9, 11].

Le logarithme de la fonction de vraisemblance s'écrit:

$$\ln L(\mathbf{x}, \boldsymbol{\theta})=-n\,M\ln\pi-M\ln|\det\boldsymbol{\Gamma}_e| - \sum_{\mu=1}^{M}(\mathbf{x}^\mu-\boldsymbol{\Phi\theta})^+\boldsymbol{\Gamma}_e^{-1}(\mathbf{x}^\mu-\boldsymbol{\Phi\theta}).$$

Une condition nécessaire pour accéder à un maximum de cette fonction est l'annulation de la dérivée partielle par rapport à $\boldsymbol{\theta}$, qui s'écrit d'après le tableau d'identités de notre annexe:

$$\partial\ln L/\partial\boldsymbol{\theta}=2\,M\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}(\mathbf{X}-\boldsymbol{\Phi\theta});$$

en notant $\mathbf{X}=\sum\mathbf{x}^\mu/M$.

Il vient donc l'expression, bien connue dans le cas réel, de l'estimateur optimal:

$$\boldsymbol{\theta}^0=[\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}\boldsymbol{\Phi}]^{-1}\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}\mathbf{X}$$

Cette expression peut également s'écrire sous une autre forme:

$$\boldsymbol{\theta}^0-\boldsymbol{\theta}=[\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}\boldsymbol{\Phi}]^{-1}\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}(\mathbf{X}-\boldsymbol{\Phi\theta})$$

L'estimateur $\boldsymbol{\theta}^0$ est donc non biaisé et sa variance vaut:

$$\mathbf{V}(\boldsymbol{\theta}^0)=E\{(\boldsymbol{\theta}^0-\boldsymbol{\theta})(\boldsymbol{\theta}^0-\boldsymbol{\theta})^+\}=[M\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}\boldsymbol{\Phi}]^{-1}.$$

Ce résultat est parfois dénommé dans la littérature « formule de Capon » lorsque $\boldsymbol{\theta}$ est un réel scalaire. A l'aide du théorème 5, on peut évaluer l'efficacité de cet estimateur.

$$\mathbf{F}(\boldsymbol{\theta})=E\{\partial\ln L/\partial\boldsymbol{\theta}\,\partial\ln L/\partial\boldsymbol{\theta}^+\}=4\,M\boldsymbol{\Phi}^+\boldsymbol{\Gamma}_e^{-1}\boldsymbol{\Phi}.$$

On vérifie ici, avec notre formalisme complexe, que l'estimateur linéaire du maximum de vraisemblance de $\boldsymbol{\theta}$ est efficace car:

$$\mathbf{V}(\boldsymbol{\theta}^0)=4\,\mathbf{F}(\boldsymbol{\theta})^{-1}.$$

7. Exemple 2

Il est également très illustratif de présenter le problème de l'estimation d'une matrice de covariance par le maximum de vraisemblance. C'est sur ce type d'exemple que l'on apprécie le mieux les outils de dérivation dans $\mathbb{C}^p$ que nous avons donnés en annexe (par ailleurs, ces outils peuvent être avantageusement utilisés pour l'étude de problèmes plus compliqués [2, 3]). Considérons simplement le modèle d'observation suivant :

$$\mathbf{x}^\mu = \boldsymbol{\theta} + \mathbf{e}^\mu;$$

les $\mathbf{x}^\mu$ étant indépendamment $N_p^c(\boldsymbol{\theta}, \boldsymbol{\Gamma}_e)$.
Le logarithme de la fonction de vraisemblance est le même que dans l'exemple précédent avec $\boldsymbol{\Phi} = \mathbf{I}$. Il vient donc, en tenant compte des identités présentées en annexe :

$$\frac{\partial \ln L}{\partial \boldsymbol{\Gamma}_e} = -2\,M\,\boldsymbol{\Gamma}_e^{-1} + 2\,\boldsymbol{\Gamma}_e^{-1} \sum_{\mu=1}^{M} (\mathbf{x}^\mu - \boldsymbol{\theta})(\mathbf{x}^\mu - \boldsymbol{\theta})^+ \boldsymbol{\Gamma}_e^{-1} + \operatorname{Diag}\left[M\,\boldsymbol{\Gamma}_e^{-1} - \boldsymbol{\Gamma}_e^{-1} \sum_{\mu=1}^{M} (\mathbf{x}^\mu - \boldsymbol{\theta})(\mathbf{x}^\mu - \boldsymbol{\theta})^+ \boldsymbol{\Gamma}_e^{-1} \right].$$

En annulant toutes les coordonnées de ce gradient, on peut remarquer que les termes diagonaux et extra-diagonaux sont alors régis par une équation unique :

$$\boldsymbol{\Gamma}_e^0 = \frac{1}{M} \sum_{\mu=1}^{M} (\mathbf{x}^\mu - \boldsymbol{\theta})(\mathbf{x}^\mu - \boldsymbol{\theta})^+.$$

On vérifierait facilement que pour toute matrice hermitienne $\mathbf{H}$, $\ln L(\boldsymbol{\theta}, \mathbf{H})$ est inférieure à $\ln L(\boldsymbol{\theta}, \boldsymbol{\Gamma}_e^0)$, ce qui assure bien un maximum; la méthode suivie serait tout à fait similaire à celle menée dans le cas réel [5, 10].

Conclusion

Ce bref exposé fait le point sur les outils nécessaires à l'estimation et l'optimisation dans $\mathbb{C}^p$. Une telle introduction synthétique et rigoureuse devrait être utile en raison des lacunes encore évidentes dans la littérature sur ce sujet.

Manuscrit reçu le 28 *mars* 1986

Annexe

1. Dérivation d'une fonction réelle de la variable complexe

Une fonction réelle de la variable complexe $f(z)$, non réduite à une constante n'est jamais holomorphe. En conséquence, il est impossible de définir la dérivée à partir du taux d'accroissement :

$$g(z, h) = [f(z+h) - f(z)]/h$$

On est donc amené très simplement à poser $\forall z$.

$$z = x + jy, \qquad (x, y) \in \mathbb{R}^2.$$

A toute fonction $f(z)$, on peut associer une fonction $F(x, y) = f(z)$ de manière unique. Lorsque le gradient de F est défini, il est possible de construire la quantité

$$\mathbf{d}(x, y) \stackrel{\triangle}{=} (\partial F/\partial x)(x, y) + j(\partial F/\partial y)(x, y)$$

On convient de poser

$$(\partial f/\partial z)(z) \stackrel{\triangle}{=} \mathbf{d}(\operatorname{Re}(z), \operatorname{Im}(z)).$$

Cette définition est la seule qui nous semble rigoureuse, et a été utilisée dans la littérature par Monzingo et Miller [8; p. 85] et Hudson [4; p. 245].

Exemples :

$$\begin{aligned}
f(z) &= \operatorname{Re}(z) \to f'(z) = 1 \\
f(z) &= \operatorname{Im}(z) \to f'(z) = j \\
f(z) &= |z|^2 \to f'(z) = 2z \\
f(z) &= \ln|z| \to f'(z) = 1/z^* \\
f(z) &= \operatorname{Re}(az) \to f'(z) = a^* \\
f(z) &= \operatorname{Im}(az) \to f'(z) = ja^* \\
(\partial f/\partial z^*) &= (\partial f/\partial z)^*
\end{aligned}$$

D'autres définitions peuvent être rencontrées mais ne semblent pas avoir de fondements théoriques. Nous les considérons plutôt comme moyens mnémothechniques.

2. Gradients matriciels

On peut définir la dérivée d'une quantité matricielle $p \times q$ $\mathbf{A}$ par rapport à la variable matricielle $s \times t$ $\mathbf{B}$ par la matrice $sp \times tq$ notée $\partial \mathbf{A}/\partial \mathbf{B}$ comme suit :

$$\frac{\partial \mathbf{A}}{\partial \mathbf{B}} \stackrel{\triangle}{=} \begin{vmatrix} \dfrac{\partial \mathbf{A}}{\partial \mathbf{B}_{11}} & \cdots & \dfrac{\partial \mathbf{B}}{\partial \mathbf{B}_{tt}} \\ \vdots & & \\ \dfrac{\partial \mathbf{A}}{\partial \mathbf{B}_{st}} & \cdots & \dfrac{\partial \mathbf{A}}{\partial \mathbf{B}_{st}} \end{vmatrix}$$

Cette matrice est structurée par blocs de dimension $p \times q$ que l'on peut noter

$$\partial \mathbf{A}/\partial \mathbf{B}_{ij}; \qquad 1 \leqq i \leqq p, \quad 1 \leqq j \leqq q$$

A l'issue de cette définition, il est intéressant de noter que :

$$\partial \mathbf{A}/\partial \mathbf{B}^T = [\partial \mathbf{A}^T/\partial \mathbf{B}]^T$$

$\mathbf{C}^n$.
A, B, C: matrices complexes quelconques.
H: matrice hermitienne ($\mathrm{H}^+ = \mathrm{H}$).
x, y: vecteurs complexes.

U : réelle scalaire	$\partial \mathrm{U}/\partial \mathbf{A}^+$	$(\partial \mathrm{U}/\partial \mathbf{A})^+$
Linéaire	$(\partial/\partial x)\,(x^+\,\mathbf{A}\,y + y^+\,\mathbf{A}^+\,x)$	$2\,\mathbf{A}\,y$
	$(\partial/\partial\,\mathbf{H}^t)\,(x^+\,\mathbf{H}\,x)$	$2\,xx^+ - \mathrm{Diag}\,(xx^+)$
	$(\partial/\partial \mathbf{A})\,(x^+\,\mathbf{A}\,y + y^+\,\mathbf{A}^+\,x)$	$2\,xy^+$
	$(\partial/\partial \mathbf{A})$ trace $(\mathbf{AB} + \mathbf{B}^+\,\mathbf{A}^+)$	$2\,\mathbf{B}^+$
Quadratique	$(\partial/\partial x)\,(x^+\,\mathbf{H}\,x)$	$2\,\mathbf{H}\,x$
	$(\partial/\partial \mathbf{A})$ trace $(\mathbf{H}_1\,\mathbf{A}\,\mathbf{H}_2\,\mathbf{A}^+)$	$2\,\mathbf{H}_1\,\mathbf{A}\mathbf{H}_2$
	$(\partial/\partial \mathbf{A})$ trace $[(\mathbf{BA} + \mathbf{C})^+ \mathbf{H}\,(\mathbf{BA} + \mathbf{C})]$	$2\,\mathbf{B}^+\,\mathbf{H}\,(\mathbf{BA} + \mathbf{C})$
Autre	$(\partial/\partial\,\mathbf{H}^t)\,(x^+\,\mathbf{H}^{-1}\,x)$	$-2\,\mathbf{H}^{-1}\,xx^+\,\mathbf{H}^{-1} + \mathrm{Diag}\,(\mathbf{H}^{-1}\,xx^+\,\mathbf{H}^{-1})$
	$(\partial/\partial\,\mathbf{H}^t)\,\ln\,\lvert \det \mathbf{H} \rvert$	$2\,\mathbf{H}^{-1} - \mathrm{Diag}\,\mathbf{H}^{-1}$

Et plus généralement, si **B** est à valeurs dans **C**:

$$\partial \mathbf{A}/\partial \mathbf{B}^+ = [\partial \mathbf{A}^{\mathrm{T}}/\partial \mathbf{B}]^+.$$

CAS PARTICULIER

Si A est une fonction scalaire et **B** un vecteur $s \times 1$, on retrouve la définition du gradient usuel:

$$\nabla_{\mathbf{B}}\,\mathrm{A} = \partial \mathrm{A}/\partial \mathbf{B}.$$

Un certain nombre de propriétés concernant les gradients matriciels peut être consulté, mais toutefois dans le cas réel uniquement, dans [1], [7] et [12]. Malgré sa souplesse indiscutable, nous n'avons pas trouvé jusqu'à présent dans la littérature le formalisme de dérivation d'une fonctionnelle de $\mathbf{C}^n$.

4. Identités dans $\mathbf{C}^n$

Cette annexe perdrait son plus grand intérêt si nous ne donnions pas une liste presque exhaustive des identités de base de dérivation dans $\mathbf{C}^n$. En effet, nous n'avons trouvé pratiquement aucune information dans la littérature en ce qui concerne cette forme de dérivation. Pourtant, réexprimer dans $\mathbf{R}^{2n}$ un problème d'optimisation posé dans $\mathbf{C}^n$ peut alourdir considérablement les expressions matricielles. De plus, il est souvent nécessaire de retourner l'expression du résultat dans $\mathbf{C}^n$, ce qui s'est avéré parfois très ardu sur certains exemples, voire inextricable (notamment dérivation de $\ln \lvert \det \mathbf{H} \rvert$ par rapport à **H**).

Le tableau ci-dessus rassemble de manière redondante quelques identités de base. Ces dernières peuvent être démontrées pour la plupart à partir des résultats obtenus dans le cas réel, mais le plus simple peut consister parfois à remonter aux définitions des paragraphes 1 et 2.

BIBLIOGRAPHIE

[1] Y. BARD, *Non Linear Parameter Estimation*, Academic Press, 1974.

[2] P. COMON, Traitement de signaux magnétiques multivariables, *Thèse de Doctorat*, INPG, 9 décembre 1985, Grenoble.

[3] P. COMON et J. L. LACOUME, A Robust Adaptive Filter for Noise Reduction Problems, *Proceedings of ICASSP*, 7-11 avril 1986, Tokyo.

[4] J. E. HUDSON, *Adaptive Array Principles*, Peregrinus Ltd, 1981.

[5] A. M. KSHIRSAGAR, *Multivariate Analysis*, Dekker, 1972.

[6] T. T. LACOSS, Data Adaptive Spectral Analysis Method, *Geophysics*, 36, n° 4, 1971, p. 661-675.

[7] K. V. MARDIA, J. T. KENT et J. M. BIBBY, *Multivariate Analysis*, Academic Press, 1979.

[8] R. A. MONZINGO et T. N. MILLER, *Introduction to Adaptive Arrays*, Wiley, 1980.

[9] J. MUNIER, De l'analyse spatiale continue à l'analyse paramétrique, *X^e colloque GRETSI*, Nice, 1985, p. 307-312.

[10] C. R. RAO, *Linear Statistical Inference and its applications*, Wiley, 1965.

[11] J. C. SAMSON, Pure States, Polarized Waves and Principal Components in the Spectra of Multiple Geophysical Time Series, *Geophys J. R. Astr. Soc.*, 72, n° 3, 1983, p. 647-664.

[12] W. J. VETTER, Derivative Operations on Matrices, *IEEE Trans Automatic Control*, 15, n° 2, 1970, p. 241-244.