ANNEXE 1

GRADIENTS MATRICIELS

1 - DERIVATION D'UNE FONCTION REELLE 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 $\forall z$:

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

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 :

$$d(x,y) \triangleq (\partial F/\partial x)(x,y) + j\,(\partial F/\partial y)(x,y)$$

On convient de poser :

$$(\partial f/\partial z)(z) \triangleq \; d(\mathrm{Re}(z), \mathrm{Im}(z))$$

Cette définition est la seule qui nous semble rigoureuse, et a été utilisée dans la littérature par MONZINGO-MILLER 1980 (p.85) et HUDSON 1981 (p. 245).

Exemples :

$$f(z) = \mathrm{Re}(z) \rightarrow f'(z) = 1 \qquad \text{A1.1}$$

$$f(z) = \mathrm{Im}(z) \rightarrow f'(z) = j \qquad \text{A1.2}$$

$$f(z) = (z)^2 \rightarrow f'(z) = 2z \qquad \text{A1.3}$$

$$f(z) = \ln(z) \rightarrow f'(z) = 1/z^* \qquad \text{A1.4}$$

$$f(z) = \mathrm{Re}(az) \rightarrow f'(z) = a^* \qquad \text{A1.5}$$

$$f(z) = \mathrm{Im}(az) \rightarrow f'(z) = ja^* \qquad \text{A1.6}$$

$$(\partial f/\partial z^*) = (\partial f/\partial z)^* \qquad \text{A1.7}$$

$f(z) = |z|^2 \rightarrow f'(z) = 2z$

D'autres définitions peuvent être rencontrées mais ne semblent pas avoir de fondements théoriques. Nous les considérons plutôt comme moyens mnémothechniques (CF COMON 1984b).

2 - GRADIENTS MATRICIELS.

On peut définir la dérivée d'une quantité matricielle (pq) **A** par rapport à la variable matricielle (st) **B** par la matrice (sp tq) notée **∂A/∂B** comme suit :

A

$$\frac{\partial \mathbf{A}}{\partial \mathbf{B}} \triangleq \begin{bmatrix} \frac{\partial \mathbf{A}}{\partial \mathbf{B}_{11}} & \cdots & \frac{\partial \mathbf{B}}{\partial \mathbf{B}_{1t}} \\ \vdots & & \\ \frac{\partial \mathbf{A}}{\partial \mathbf{B}_{s1}} & \cdots & \frac{\partial \mathbf{A}}{\partial \mathbf{B}_{st}} \end{bmatrix} \qquad \text{A1.8}$$

Cette matrice est structurée par blocs de dimension (p q) que l'on peut noter

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

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

$$\partial \mathbf{A}/\partial \mathbf{B}^t = [\, \partial \mathbf{A}^t/\partial \mathbf{B} \,]^t \qquad \text{A1.9}$$

Et plus généralement, si **B** est à valeurs dans **C**, d'après A1.7 :

$$\partial \mathbf{A}/\partial \mathbf{B}^+ = [\, \partial \mathbf{A}^t/\partial \mathbf{B} \,]^+ \qquad \text{A1.10}$$

Cas particulier :

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

$$\nabla_B A = \partial A/\partial B$$

Un certain nombre de propriétés concernant les gradients matriciels peut être consulté, mais toutefois dans le cas réel uniquement, dans VETTER 1970, MARDIA 1979, BARD 1974. Malgré sa souplesse indiscutable, nous n'avons pas trouvé jusqu'à présent dans la littérature le formalisme de dérivation d'une fonctionnelle de $\mathbf{C}^n$.

3 - IDENTITES USUELLES DANS R^n

R^n	A, M, B : Matrices réelles quelconques (à éléments indépendants). x, y, a, b : Vecteurs réels.	
Si U scalaire :	$\partial U/\partial A^t$	$(\partial U/\partial A)^t$
LINEAIRE	$(\partial/\partial A)\,(x^t\,A\,y)$	$x\,y^t$
	$(\partial/\partial M)\ \mathrm{trace}\,(A\,M\,B)$	$A^t\,B^t$
	$(\partial/\partial x)\,(x^t\,A\,y)$	$A\,y$
A $n \times m$ ... $n \times m$	$(\partial/\partial A)\begin{cases}\mathrm{trace}\,(A\,B)\\ \mathrm{trace}\,(B^t\,A^t)\\ \mathrm{trace}\,(B\,A)\end{cases}$	B^t
il n'y a pas de transposée.	$(\partial/\partial A)\begin{cases}\mathrm{trace}\,(A^t\,B)\\ \mathrm{trace}\,(\,A\,B^t)\end{cases}$	B
QUADRATIQUE $x^tAMA^ty = \mathrm{trace}(AMA^tyx^t)$ $\frac{\partial}{\partial A}(\mathrm{trace}(AMA^tyx^t))$ $\frac{\partial}{\partial A}\mathrm{trace}(yx^tAM^tA^t)$	$(\partial/\partial A^t)\,(x^t\,AMA^t\,y)$	$xy^t\,AM^t + yx^t\,AM$
	$(\partial/\partial A^t)\,((Ax)^t\,.\,Ax))$	$2xx^t\,.\,A^t$
	$(\partial/\partial x)\,(x^t\ \ Ax)$	$Ax + A^t\,x$
	$(\partial/\partial x)\,[(Ax)^t\,.\,(Ax)]$	$2\,A^t\,A\,.\,x$
	$(\partial/\partial x)\,[\,(b^tx)\,.\,(x^t\,A)\,]$	$(b^tx)\,.\,A + b\,.\,(x^t\,A)$
	$(\partial/\partial M)\ \mathrm{trace}\,(A\,M\,B\,M)$	$A^t\,M^t\,B^t + B^t\,M^t\,A^t$
	$(\partial/\partial M)\ \mathrm{trace}\,(A\,M\,B\,M^t)$	$A^t\,M\,B^t + A\,M\,B$
AUTRE	$(\partial/\partial A)\ \mathrm{trace}\,(A^n)$	$n\,A^{(n-1)t}$
(Avec A régulière)	$(\partial/\partial A)\ \mathrm{Det}\,A$	$A^{-t}\ \det A$
	$(\partial/\partial A)\ \mathrm{Ln}\,\lvert \mathrm{Det}\,A \rvert$	A^{-t}
	$(\partial/\partial A)\,(x^t\,A^{-1}\,y)$	$-A^{-t}\,x\,y^t\,A^{-t}$
	$(\partial/\partial A^t)\,(x^t\,A^{-1}\,y)$	$-A^{-1}\,y\,x^t\,A^{-1}$
	$(\partial/\partial A^t)\ \mathrm{trace}\,(A^{-1}\,B)$	$-A^{-1}\,B\,A^{-1}$

TABLEAU A1.1

Les résultats précédents ne sont valables que si tous les éléments de la variable matricelle sont indépendants. Si cette dernière est assujettie à une contrainte, le calcul doit évidemment être repris.

Une contrainte couramment rencontrée est celle de la symétrie. Dans ce cas, les termes diagonaux restent inchangés par rapport aux résultats présentés dans le tableau précédent, tandis que les termes extradiagonaux subissent l'addition de leur symétrique. La matrice dérivée est alors symétrique (propriété A1.9). Le tableau ci dessous résume quelques identités remarquables :

R^n	A,B : Matrices quelconques. S : Matrice symétrique réelle. x,y : Vecteurs réels.	
LINEAIRE	$(\partial/\partial S)\,(x^t S\, y)$	$xy^t + yx^t - \mathrm{Diag}\,(x^t y)$
	$(\partial/\partial S)\,\mathrm{trace}\,(A\,S\,B)$	$A^t B^t + BA - \mathrm{Diag}\,(BA)$
QUADRATIQUE	$(\partial/\partial S)\,[\,(Sx)^t (Sx)\,]$	$2xx^t S + 2Sxx^t - \mathrm{Diag}\,(\,xx^t S + Sxx^t\,)$
	$(\partial/\partial S)\,\mathrm{trace}\,(A\,S\,B\,S)$	$ASB+BSA+B^t SA^t + A^t SB^t - \mathrm{Diag}\,(ASB+BSA)$
AUTRE	$(\partial/\partial S)\,\mathrm{Det}\,S$	$(2\,S^{-1} - \mathrm{Diag}\,S^{-1})\,.\,\det S$
	$(\partial/\partial S)\,\mathrm{Ln}\,\vert\mathrm{Det}\,S\vert$	$2\,S^{-1} - \mathrm{Diag}\,S^{-1}$
	$(\partial/\partial S)\,(x^t S^{-1}\,y)$	$-S^{-1}\,xy^t S^{-1} - S^{-1} yx^t S^{-1} + \mathrm{Diag}\,(S^{-1} xy^t S^{-1})$

TABLEAU A1.2

4 - IDENTITES DANS 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 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 R^{2n} un problème d'optimisation posé dans C^n peut alourdir considérablement les expressions matricielles. De plus, il est souvent nécessaire de retourner l'expression du résultat dans C^n, ce qui s'est avéré parfois très ardu sur certains exemples, voire inextricable (notamment dérivation de Ln | det H| par rapport à H).

Le tableau A1.3 ci-après rassemble de manière redondante quelques identités de base. Ces dernières peuvent être démontrées pour la plupart à partir des résultats présentés dans les tableaux A1.1 et A1.2, mais le plus simple peut consister parfois à remonter aux définitions des paragraphes 1 et 2.

$\mathbf{C}^n$	$\mathbf{A}, \mathbf{B}, \mathbf{C}$: Matrices complexes quelconques. $\mathbf{H}$: Matrice hermitienne ($\mathbf{H}^+ = \mathbf{H}$). x, y : Vecteurs complexes.	
U : Réelle scalaire	$\partial U/\partial \mathbf{A}^+$	$(\partial U/\partial \mathbf{A})^+$
LINEAIRE	$(\partial/\partial x)\,(x^+ \mathbf{A}\, y + y^+ \mathbf{A}^+ x)$	$2\, \mathbf{A}\, y$
	$(\partial/\partial \mathbf{H}^t)\,(x^+ \mathbf{H} x)$	$2\, x\, x^+ - \mathrm{Diag}\,(x\, x^+)$
	$(\partial/\partial \mathbf{A})\,(x^+ \mathbf{A}\, y + y^+ \mathbf{A}^+ x)$	$2\, x\, y^+$
	$(\partial/\partial \mathbf{A})\ \mathrm{trace}\,(\mathbf{AB} + \mathbf{B}^+\mathbf{A}^+)$	$2\, \mathbf{B}^+$
QUADRATIQUE	$(\partial/\partial x)\,(x^+ \mathbf{H}\, x)$	$2\, \mathbf{H}\, x$
	$(\partial/\partial \mathbf{A})\ \mathrm{trace}\,(\mathbf{A}^+ \mathbf{H}\, \mathbf{A})$	$2\, \mathbf{H}\, \mathbf{A}$
	$(\partial/\partial \mathbf{A})\ \mathrm{trace}\,(\mathbf{A}\, \mathbf{H}\, \mathbf{A}^+)$	$2\, \mathbf{A}\, \mathbf{H}$
	$(\partial/\partial \mathbf{A})\,(x^+ \mathbf{A}^+ \mathbf{H}\, \mathbf{A}\, x)$	$2\, \mathbf{H}\, \mathbf{A}\, x\, x^+$
	$(\partial/\partial \mathbf{A})\,(x^+ \mathbf{A}\, \mathbf{H}\, \mathbf{A}^+ x)$	$2\, x\, x^+ \mathbf{A}\, \mathbf{H}$
	$(\partial/\partial \mathbf{A})\ \mathrm{trace}\,(\mathbf{H}_1 \mathbf{A} \mathbf{H}_2\, \mathbf{A}^+)$	$2\, \mathbf{H}_1\, \mathbf{A}\, \mathbf{H}_2$
	$(\partial/\partial \mathbf{A})\ \mathrm{trace}[(\mathbf{BA}+\mathbf{C})^+ \mathbf{H}(\mathbf{BA}+\mathbf{C})]$	$2\, \mathbf{B}^+ \mathbf{H}(\mathbf{BA}+\mathbf{C})$
AUTRE	$(\partial/\partial \mathbf{H}^t)\,(x^+ \mathbf{H}^{-1}\, x)$	$-2\, \mathbf{H}^{-1} x\, x^+ \mathbf{H}^{-1} + \mathrm{Diag}\,(\mathbf{H}^{-1} x x^+ \mathbf{H}^{-1})$
	$(\partial/\partial\, \mathbf{H}^t)\ \mathrm{Ln}\,\lvert \det \mathbf{H} \rvert$	$2\, \mathbf{H}^{-1} - \mathrm{Diag}\, \mathbf{H}^{-1}$

TABLEAU A1.3